MORTALITY CASE NOTE REVIEW USE FOR HOSPITAL CARE QUALITY IMPROVEMENT: A METHODOLOGICAL, PSYCHOLOGICAL AND QUALITATIVE INVESTIGATION

By
AN PHUNG TE

A thesis submitted to the
University of Birmingham
for the degree of
DOCTOR OF PHILOSOPHY

Institute of Applied Health Research
College of Medical and Dental Sciences
University of Birmingham
June 2021

Thesis Abstract

Evaluating hospital care quality using case-note reviews is mandated in the United Kingdom and is endorsed by many high-income countries. This thesis separately addresses both the validity of case-note reviews and the use of case-note reviews for care quality improvement. On case-note reviewing validity, there are moderate-to-high levels of disagreement (*variability*) between multiple clinician case-note reviewers when evaluating the overall care quality of the same case-note. The sources of this disagreement (*variability*) are unknown. On case-note review use, the potential factors which affect case-note reviewing in hospitals has not been well-studied in relation to their contribution to hospital care quality improvement. This thesis presents the findings of three original studies and seeks to both identify the sources for this reviewer variability and the organizational factors which influence case-note review's likely contribution to hospital quality improvement. The introduction discusses the policy context and offers a critique of hospital mortality statistics with the prospective use of case-note reviews as an alternative approach for detecting care quality issues. Chapter 1 involved a systematic review of preventable mortality rates and a characterization of their measurement properties for evaluating care quality and subsequent hospital ranking. Findings concluded that a limitation of studies not accounting for variation between different hospitals, assuming equal variance, in the ranking process. Case-note reviews are presented as a workable alternative, to which this thesis is devoted to investigating. Chapter 2 presents the findings of an original systematic review which identified cognitive biases and heuristics related to case-note review care quality judgements. Cognitive biases and heuristics, sourced from two systematic reviews, are investigated with their plausible influence upon case-note reviewer care quality judgments using clinical scenarios derived using a systematic literature search and

informed by a panel consensus. Findings indicate the plausible influence of cognitive biases and heuristics. Chapter 3 investigates the influence of reviewer attitudes; their demographics and patient case-note review characteristics upon case-note reviewer care quality judgements. Selected attitudes did not significantly influence care quality judgements and a significant proportion of care quality judgement variability is unexplained by the included independent variables. Chapter 4 describes case study fieldwork in an acute NHS Trust which explored the organizational processes around case-note review including its embedding, information flow and its perceived quality improvement contribution. We found that case-note reviews were well-embedded, with there being limited information flow from ward-to-board. Chapter 5 is a critical reflection of the research process and the assumptions made in this thesis. Chapter 6 summarizes the thesis, discusses practical implications, and identifies opportunities for future research for quality improvement from case-note reviews.

Dedication

I would like to dedicate this work to my family who have supported me for all these years.

Acknowledgements

I would like to express my gratitude to the following for their support and guidance during the writing of this thesis:

My family for their steadfast love and forbearance.

Dr Semira Manaseki-Holland and Dr Kelly Schmidtke for their indefatigable support and encouragement throughout the PhD process.

Professor Russell Mannion for his steadfast support, his expertise on qualitative research methods and health systems, and his careful management of the manuscript submission process.

Professor Ivo Vlaev for his support on psychological theory and critical feedback towards the thesis.

Professor Timothy P Hofer for his consistent support on all things statistical and quality measurement.

Professor Richard Lilford for his steadfast support and contributions his knowledge on health systems and keen eye for manuscript details.

Professor Julian Bion for inviting me to be part of the HiSLAC study and the generous comments given on draft manuscripts.

Erica Ferris for her close support throughout the case-study work and encouragement throughout the PhD journey.

John George Arthur and Rachel Hu for all those conversations that have helped to keep me on the right PhD track.

Catherine Taylor and Melita Harris for their organisational ability and her joyful in helping me to bring this thesis to completion.

Dr Laura Jones and Professor Jayne Parry for their consistent support and guidance at key stages of the PhD.

My friends and close acquaintances whom I have not mentioned, you know who you are.

Authorship Statement

In this thesis, the research presented was undertaken by the author (AT) with guidance and support from the following academic supervisory team: Dr Semira Manaseki-Holland (SMH), Professor Richard Lilford (RL), Professor Russell Mannion (RM), Professor Ivo Vlaev (IV) and Dr Kelly Schmidtke (KS).

This thesis incorporates the quantitative analysis of clinical case-note reviewer and non-identifiable patient data from the High Intensity Specialist-led Acute Care Study (HISLAC) which was collected as a prospective non-funded investigation with ad-hoc ethics obtained. The research project was conceived by SMH, and the study materials were conceived and designed by RL, AT and RM and the following researchers: Professor Julian Bion (JB), Professor Timothy Hofer (TH), Mr. Gavin Rudge (GR), Ms. Jianxia Sun (JS)
AT developed the psychological and attitudinal direction and recruited IV first, closely followed by KS for support on the requisite behavioural science for this part of the thesis.

The following researchers were responsible for study supervision during the data collection period (HiSLAC and hospital case study (HCS)): SMH, TH, IV, KS, RM, GR, and JS. JB, JS, and GR were the lead researchers in charge of the HISLAC elements. For the HCS, AT was solely responsible for collecting data.

Following data collection, all data were coordinated and sent to AT for cross-checking and cleaning. AT consulted with TH in person and through email to design an appropriate analysis for the cleaned dataset, resolve queries and correct any errors in judgement.

AT was entirely responsible for selecting the HISLAC independent variables to be analyzed within the thesis, completing the analysis and visualizations. TH was involved in selecting the analysis method, developed STATA command-line code from which AT derived the final data analysis in the research manuscripts in Chapter 3. AT understood all elements of the code, and himself modified the code according to his purpose.

Specific chapter contributions are as follows:

Introduction - Introduction & Methodology: AT drafted the chapter, SMH, RM, RL, KS reviewed and edited the chapter for its content.

Chapter 1 – 'Ranking' paper: AT contributed to the data collection, data curation, formal analysis, visualization, validation, and original drafting of the published Milbank article.

Chapter 2 – Systematic review: The conceptualization, data curation and selection criteria were developed by AT with data collection from AT and Dr Lauren Quinn (LQ) with support from SMH, IV, KS and RL. AT refined the protocol and selection criteria and developed a comprehensive search strategy with support from IV, KS and Mrs. Susan Bayliss (SB).

Consensus panel: this project was conceived by SMH and designed by AT with input from the following researchers: TP, IV, KS, Chris Beet (CB), LQ, JB and RL. AT drafted the chapter, SMH, RL, JB, IV and KS reviewed the chapter for its content and focus.

Chapter 3 – Attitudinal Measures: The study was introduced by RL and SMH as pertinent to AT's research thesis. AT conceived of this research line with JB, RL, SMH, IV, KS and TPH.

Chapter 4 – Hospital Case Study: The study was conceived by SMH and designed by AT, RM and SMH. AT solely undertook the data collection, data curation, investigation, member-checking, and the writing of the original draft. EF, SMH, RM reviewed and edited for content and focus.

Chapter 5 – Study Reflections: AT drafted the chapter with all supervisors inputting from SMH, RM, KS, IV and RL, reviewed and edited the chapter for its focus and relevance to the field.

Chapter 6 – Discussion: AT drafted the chapter with all supervisors inputting from SMH, RM, KS, and RL, reviewed and edited the chapter for its focus and relevance to the field.

CONTENTS

List of Tables[1]

[1] excluding tables in the appendix

List of Figures

List of Papers

Papers under review directly from the thesis

A. P. Te, K. A. Schmidtke, C. Beet, L. Quinn, J. Bion, T. P. Hofer, et al. Using behavioural theory to characterize cognitive biases and heuristics in case record reviews: a conceptual study, Qualitative Psychology, 2021

An. P. Te, Erica Ferris, Semira-Manaseki-Holland, and R. Mannion, Using retrospective mortality case-note review for Quality Improvement (QI): evidence from a UK hospital case study, BMJ Open Quality 2021

An P Te, Kelly-Ann Schmidtkhe, Semira Manaseki-Holland, Richard Lilford, Julian Bion, Ivo Vlaev, et al. The influence of physician reviewer characteristics and attitudes on care quality judgements during case note reviews: a multilevel analysis, Health services research 2021

Papers under review based on the work included in this thesis

HOFER, T. P. 2019. Ranking Hospitals Based on Preventable Hospital Death Rates: A Systematic Review With Implications for Both Direct Measurement and Indirect Measurement Through Standardized Mortality Rates. *Milbank Quarterly,* 97**,** 228-284.

Abbreviations

AAA - abdominal aortic aneurysm

ADTU – Anxiety due to uncertainty

AMU – Acute Medical Unit

AOGG – Audit and Organisational Governance Group

AT - An Te (Researcher)

CCM – Constant comparison method

CGC – Clinical Governance Committee

CHKSA – N/A (provider of healthcare intelligence and quality improvement services)

CMA – Case mix adjustment

CGC – Clinical Governance Committee

CNR – Case Note Review

CP – Cognitive participations (NPT construct)

CQC – Care Quality Commission

CQUIN - Commissioning for Quality and Innovation

CPD - Continued professional development

CR – Critical Realism

EF – Erica Ferris (researcher)

EPHPP - Effective Public Health Practise Project

HISLAC – High Impact Specialist Led Acute Care Study

HISB - Healthcare Safety Investigation Branch

HMIC - Health Management Information Consortium

HSCIC - Health and Social Care Information Centre

HSMR – Hospital Standardised Mortality Ratios

ICC – Intra-class correlation

ICU – intensive care unit

ID – Individual Differences

IHI - Institute of Healthcare Improvement

LEDR - Learning Disabilities Mortality Review

MCNR – Mortality Case Note Review

MD – Medical Director

MIDSS - Measurement Instrument Database for the Social Sciences

NCEPOD - National Confidential Enquiry into Patient Outcome and Death

NMCRR – National Mortality Case Note Record Review (and methodology)

NELA - National Emergency Laparotomy Audit

NEWS - National Early Warning Score

NFC – Need for Cognition

NHS – National Health Service, UK

NICE – National Institute for Health and Care Excellence

NPT – Normalisation Process Theory

NQB – National Quality Board

PALS - Patient Advice and Liaison Service

PNS – Personal need for structure

PROSPERO - International prospective register of systematic reviews

PSN - Patient Safety Newsletter

ONS – Office of National Statistics

QI – Quality Improvement

RAMI - Risk-adjusted mortality index

RAND – "Research and development" Corporation

RCP – The Royal College of Physicians

SHMI - Summary hospital-level mortality indicator

SOP – Standard Operating Procedure

SJR – Structured judgement review

SSC - Surviving Sepsis Campaign

SSI – Semi-structured Interview

SWFT – South Warwickshire NHS Foundation Care Trust

YHASN – Yorkshire and Humber Academic Health Science Network

VA – Veteran Administration

INTRODUCTION: HOSPITAL CARE QUALITY IMPROVEMENT IN THE UNITED KINGDOM

"Remember that all models are wrong; the practical question is how wrong do they have to be to not be useful."

Quote by George E. P. Box, statistical methodologist

Abstract

The measurement of hospital care quality is important because it presents a way to ensure that the care quality is at least to the same standard the public and the patient expect. This then helps to establish a yardstick for future interventions to improve and build upon the existing care quality. This introductory chapter will first set out what care quality is, the purpose of quality improvement, the definition of quality improvement (QI) and care quality, the QI activity, the recent progress in the care and the overall strengths, weaknesses, opportunities, and challenges for care quality improvement. I then present the legacy of policy developments from the turn of the 21st century to the present day and discuss the transition of QI from a summative to a formative measure. Thirdly, I present some knowledge and evidence of hospital quality improvement and performance measurement and its practical use. Next, I discuss, in detail, the use case of hospital mortality statistics (HMS) with their strengths and weaknesses. I then discuss the potential of case-note reviews to overcome HMS weaknesses connecting these reviews to hospital quality improvement. Lastly, conceptual frameworks used in this thesis are individually described, and in relation to one another, to help us have an appropriate and efficacious methodology to understand whether case-note reviews can and do lead to hospital quality improvement.

Problems with defining care quality

What is care quality?

Care quality, or quality of care, is the central term of interest in this thesis. We would do well to discuss what care quality is and what factors have been shown to shape how we perceive or measure it. One group has identified two principal dimensions of quality of care for individual patients through *access* and *effectiveness*. It must be noted that these definitions concern the interaction between clinician and patient and do not consider in any great detail the community, administrative or economic aspects of care quality control. For focus and brevity, this thesis focusses on the caring element between clinicians and patients, and not on less salient factors that indirectly influence care quality. In chronological order, here are several approaches and definitions to care quality, as acknowledged in the literature:

The Institute of Medicine (IOM) definition from the USA(Richardson, 2001) espoused the Care Quality Standard as:

1. safe – avoiding injuries to patients from care that is intended to help them
2. effective – avoiding overuse and misuse of care
3. patient-centered – providing care that is unique to a patient's need
4. timely – reducing wait times and harmful delays for patients and providers
5. efficient – avoiding waste of equipment, supplies, ideas, and energy
6. equitable – providing care that does not vary across intrinsic personal characteristics

Also, one of the early definitions and descriptions of the components of care quality was promulgated by Avedis Donabedian(Donabedian, 2002). This is outlined in Table 1.

Table 1. Components of Quality: Definitions

1. **Efficacy** – the ability of the science and the technology of health care to bring about improvements in health when used under the most favourable circumstances.

2. **Effectiveness** – the degree to which attainable improvements in health care are, in fact, attained.

3. **Efficiency** – the ability to lower the cost of care without diminishing attainable improvements in health

4. **Optimality** – the balance of improvements in health against the costs of such improvements

5. **Acceptability** – conformity to the wishes, desires and expectations of patients and their families

6. **Legitimacy** – conformity to social preferences as expressed in ethical principles, values, norms, mores, laws, and regulations

7. **Equity** - conformity to a principle that determines what is just and fair in the distribution of health care and it benefits among members of the population.

Donabedian stresses the extent to which these quality-of-care components contribute to care quality across different situations.

Since Donabedian's care quality definitions, other Quality of Care definitions have since come to the fore. The essence of Donabedian's definitions have been preserved in current definitions of care quality.

Following the United Kingdom's healthcare reforms of the Health and Social Care Act 2012, the definition of quality of care is: "effective, safe and provides as positive an experience as possible." The patient's viewpoint is implicitly considered.(Department of Health, 2013) These three definitions share common components. For concision and common understanding for this thesis, the care quality will have the components of being **safe**, **effective**, **patient-centred, and appropriate**.

1. **Safe** i.e., patients are not intentionally harmed
2. **Effective** i.e., if science demonstrates its use as being effective, use it as directed. If not or it is in doubt, do not use it.
3. **Patient-centred** – Patients should be in control of their own care i.e., 'nothing about me without me.'
4. **Appropriate** – in combining the concepts of timeliness, efficiency and patient centredness, the further term "equitable" is summarised as being appropriate or not. Were the decisions appropriate given the situation's timing, resource use and consideration of other patients?

At the international level, the World Health Organization (WHO) have set standards for care quality through its own Care Quality Standard which is defined as the: "the extent to which health care services provided to individuals and patient populations improve desired health outcomes. To achieve this, health care must be safe, effective, timely, efficient, equitable and people-centred."(Organization, 2018) All the care quality definitions, thus mentioned, involve some element of safety, efficacy and people-centredness. Having underscored some of the dimensions common to all care quality constructs[2] or measures, where a construct refers to any measurement of any aspect of care quality, it will then be important to discuss how we can know how these dimensions are captured to then determine whether this care quality has indeed improved.

Defining patient safety and quality care

For this subject matter, there are two terms which are closely related but distinct, these being "patient safety" and "care quality." Many patient safety experts and researchers view quality of care as the overarching umbrella under which patient safety resides. In 2003, Harteloh reviewed multiple conceptions of quality and concluded with this abstract definition that: "*Quality [is] an optimal balance between possibilities realised and a framework of norms and values*." This conceptual definition reflects the fact that quality is an abstraction and does not exist as a discrete entity but as an interaction among relevant

[2] The term "construct" is used to indicate that there is a subjective understanding from the clinician, or any person for that matter, which lies in the mind of the beholder who thinks about care quality

actors who agree about standards (the norms and values) and components (the actualities from the possibilities).(Harteloh, 2003)

Work groups such as those in the IOM attempted to define quality of health care in terms of professional standards. Initially, the IOM defined quality as the "*the degree to which health services for individuals and populations increase the likelihood of desired health outcomes and are consistent with current professional knowledge*."(Lohr, 1990) This led to a definition of quality that employed quality indicators, as signature detectors of the standards. These standards are not necessarily in terms of the possibilities or conceptual clusters for these indicators. Further, most quality indicators were and often continue to be comprised of the 5Ds—death, disease, disability, discomfort, and dissatisfaction—rather than more positive components of quality.

The most recent work to identify the components of quality care for the 21st century is centered predominantly on the conceptual components of quality. Following a thematic analysis after a systematic literature search of articles published between 2006 and 2012, there were 4 defining attributes identified: (1) effective, (2) safe, (3) culture of excellence, and (4) desired outcomes. Based on these attributes, the literature's definition of healthcare quality is *the assessment and provision of effective and safe care, reflected in a culture of excellence, resulting in the attainment of optimal or desired health (from both patients and clinician's perspective)*.

Now turning to patient safety terms. The term patient safety is itself closely related to care quality. However, patient safety is usually defined as "the prevention of harm to patients"

with the emphasis placed upon the prevention of errors, learning from errors, and building a good safety culture. The idea of safety suggests that the activity of patient safety is to prevent potential harm to patients. The idea is not constructive for patients and clinicians to build on and it must be stated that one must be careful to have a learning culture, not a punitive one. Subsequently, this patient safety implies an ontology and epistemology of error; in other words, what types of errors are there and what is the nature of these errors? This must first be understood to combat these errors in any effective or systematic way. This is a section which I shall delve into next by addressing the usual normative approach to patient safety with the alternative suggestion that safety should not be seen as regular but as dynamic.

Reflecting on the definitions of care quality

Firstly, care quality must be loosely defined because care involves multiple components revolving around the patient, the care-provider and the health system around which care is received and provided. At the care-provider clinician level, there are established processes and standards, either formally or informally, for the medically qualified through their respective associations (e.g., Royal College of Physicians UK, American Medical Association). At the system level, which is really the policy-making level, governance and management play more of an integral role for care quality because resources need to be responsibly used, especially in the case of state-funded healthcare systems. And lastly, the patients have their own perception of the care journey. In summary, care quality is a multi-component construct comprised of personal (the patient), medical (the clinicians and care providers) and economic (health system) elements. For the express purpose of this thesis, I

will focus predominantly on the clinician side as it is the single group which is most able to competently survey all the critical steps in the delivery of good care quality.

Secondly, for the health system, measuring care quality is usually seen as a quality assurance activity. As the health care system and policymakers are more removed from the clinical bedside of patients and are concerned for the better interests of all their patients, we can find that these policymakers can endorse and apply a mixture of process and outcome measures in gauging the care quality of their services. Early on this century, quality improvement at the national level in the United Kingdom was separated into individual, group/team, organization and larger system/environment categories with a strict focus on outcomes rather than processes (Ferlie and Shortell, 2001) The last two decades have seen the proliferation of more process-of-care measures to assess care quality. These have become common for internal quality assessment and improvement activities. For instance, these can include cost (Hussey et al., 2013) and pay-for-performance process outcomes (Mendelson et al., 2017) that have been all used to gauge the care quality. The general trend has been a move from outcome measures to process indicators of care quality.

Thirdly, at the national level in the United States, for instance, there was a move away from dealing with individual clinical errors to helping providers emphasise the goodness of the care. Whilst at the international level, the Organization for Economic Co-operation and Development attempted in 2006 to set out a conceptual framework for the OECD's Health Care Quality Indicator (HCQI) which sought to identify the concepts and dimensions of quality of health care which ought to be measured and how, in practise, they should be measured. Questions posed included "what is being measured, what is care quality, what is

its construct?" As a corollary, care quality is directly connected to the approach used to measure (i.e., tools, methods, people involved and their interactions) and the nature of care quality. And why measure care quality at all, why is care quality important? It is important because care quality concerns must have a good reason why one would wish to pursue it, the "what for?" arises – which is teleology, in other words, purpose. In this OECD report, the care quality is traced back through the academic literature to a health determinants model which includes social and economic factors, in addition to the physical factors considered by clinicians.(Arah et al., 2006)

This broader outlook on care quality is more comprehensive because we, being more than physical beings, have complex social interactions with each other. As I am not a clinician, I cannot consider the clinical content of the care quality for instance whether that action was deemed medical negligence, or those events were all tell-tale signs of a myocardial infarct. But what I do consider is the broader approach used to survey care quality in the hospital, not just as a physical process but a socio-political and cultural process.

Given the complexity of addressing these many aspects of the care quality standard, it was considered beyond the scope of this PhD to address all these care quality dimensions before, during and after all care. Strictly, care can be extended to all domains of life, in other words, from cradle to grave. There is the care we give to those we care about and the care we give to ourselves, which is termed "self-care". However, this concept of care used is strictly about the services the patient has received in the hospital setting. This is what is meant by "hospital care quality." Care quality is defined for a reason, and this reason is for the improvement of the care quality. It is this that we shall address next.

What is quality improvement?

Quality improvement is a broad set of approaches to enhance the care quality and the hospital service delivery for patients now and in the future. The UK NHS has a precedent of putting much emphasis on the active improvement of care:

> *"From 1 April 2019, NHS England and NHS Improvement are working together as a new single organisation to better support the NHS to deliver improved care for patients."(NHS Improvement, 2020b) Quality improvement more broadly defined is "the combined and unceasing efforts of everyone—healthcare professionals, patients and their families, researchers, payers, planners and educators—to make the changes that will lead to better patient outcomes (health), better system performance (care) and better professional development. This definition arises from our conviction that healthcare will not realise its full potential unless change making becomes an intrinsic part of everyone's job, every day, in all parts of the system."(Batalden and Davidoff, 2007)*

For our intents and purposes, quality improvement is the responsibility of everybody who works in the NHS, and especially those healthcare professionals[3] who, directly or indirectly, deliver this care to the patient.

---

[3] Administrative staff are indirectly involved as they do oversee the medical records/health records which the medical professionals must make use of to determine the most

The aim of care quality improvement is to improve the care quality from a lower level of care quality to a higher level of care quality, which is represented in Figure 1. The arrow indicates the direction of the goal of quality improvement.

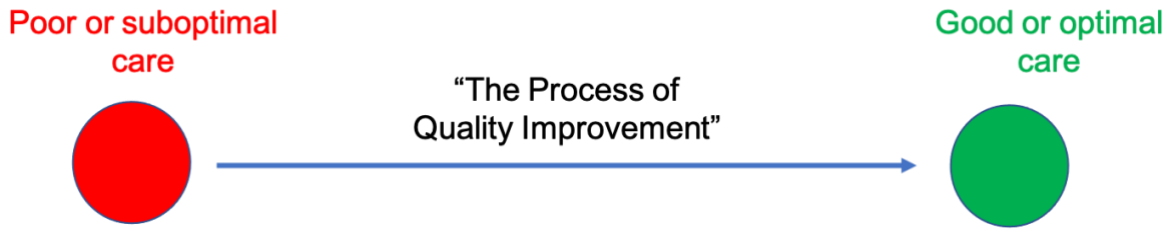


Figure 1. The process of care quality improvement

To know whether an improvement has been made, we need first to have established a baseline, a standard or rule. For instance, say we have someone who has hypertension and is moving in the right direction to resolve their hypertension. We first need to establish the various blood pressure ranges[4] for classifying someone as having hypertension.

---

appropriate next steps for medical intervention. For instance, if the notes were not assembled correctly, this would make the true state of the patient difficult for the clinician to gauge and thus put at risk the patient under question, either due to lost time or any (in)action. This would then raise the likelihood poor care is being given. We know it is the duty of medical professionals to care and promote the overall (physical and mental) wellbeing of their patients through the transparent documentation and reporting procedures. Thus, administrative staff have their role to play in delivery of care.

[4] ideal *blood pressure is between 90/60mmHg - 120/80mmHg

Furthermore, these pressure ranges must themselves have a standard, in other words a unit of measurement, which is typically millimetres of mercury (mmHb). In this same way, to know if the care quality is improving, we need to be able to define first what good care or bad care is to first establish then where would like the care to progress to next. These standards are implicit in all comparative evaluations of care quality (or anything else for that matter). There then needs to be an accurate and reliable way to measure the quality of care. It is this accuracy and reliability which concerns chapter 2 and 3 of the thesis, which concerns the "what" of care quality, the substance and quantity of care quality. In the next section, however, we shall first look at the "why" of care quality. Why should we even care about care quality improvement? And what is care quality for exactly?

What is the purpose of care quality improvement?

I will discuss some popular and recent theoretical frameworks on how to approach and study care quality improvement. This should help us to identify a systematic way in which to understand the subject better. Care quality improvement can simply be understood as "the improvement of the care which have been delivered to patients." Unpacked a little more, the purpose is to establish when, how and why the care quality intervention works, and to unpick the complex relationship between four key features. These are context (the situation, setting or organization in which the intervention is deployed), content (the nature or characteristics of the intervention itself), application (the process through which the

---

high blood pressure is considered to be 140/90mmHg or higher

low blood pressure is considered to be 90/60mmHg or lower

intervention is delivered) and outcomes (the results of the intervention). Each of these factors are subject to low variance, which is homogeneity, and high variance, which is heterogeneity. Given these features are active in various ways in different healthcare settings, any credible approach to quality improvement must account for the complex social interventions that can only be evaluated correctly if the context, content, application and outcomes are understood well together. (Walshe, 2007)

Another approach by Kaplan et al. (2012) went about obtaining opinions, initially sourced from systematic reviews from QI experts which identified that certain contextual factors in the QI team (i.e. leadership, culture supportive of QI, capability for improvement, motivation to change) have a greater bearing upon the outcome of QI interventions than factors within the organisation and the external environment (i.e. external motivators, project motivators, physician payment structure, data infrastructure). This approach places greater influence on QI upon the culture of the clinical team in question over both the organisation's culture and the external socio-political environment.(Kaplan et al., 2012)

And lastly, another theory espoused the need to consider the importance of middle-level theory (e.g., what the components of care quality are) and theories of change and behaviour, which is how those undertaking QI must interact well with key objects in their context. The activity of quality improvement has a programme, and for all intents and purposes, and these are relatively well-defined. The evaluation of quality improvement programmes has had extensive discussion in academic circles. Dixon-woods admits that the definition of quality improvement is elusive and has had little consensus across practitioners and researchers in this field. And that perhaps what is needed is not more

theory and discussion but brute evidence of QI's effectiveness.(Dixon-Woods, 2019) And this is why Kleinman and Dougherty put to task the "structure, process and outcome" bins requisite for health care quality measurement - first dubbed by Avedis Donabedian, a renowned health care researcher (Donabedian, 2002) – by respecting both the health care operations and health care contexts as a indivisible, hybrid pair. (Kleinman and Dougherty, 2013)

To summarise, the purpose and function of these quality improvement approaches is to consider the QI team in their context, which is the wider political and cultural environment and its context and the empirical evidence which demonstrates the effectiveness (or ineffectiveness) of QI within both the inner team and the extra-hospital team culture. This can be achieved using a variety of methods from quantitative to qualitative research methods. This thesis seeks to employ some of these tools to measure QI. Having discussed the importance of the objects, mechanisms, and context for QI, we turn now to how care quality improvement is defined.

How is care quality improvement defined?

As alluded to a little earlier, quality improvement assumes both 'direction' and 'magnitude' of improvement. It assumes that care quality is an improvement upon existing quality of care. Thus, for there to be any meaningful connection between propositions (i.e., collected data) and the external world (actual impact of case-note reviews) is necessary for these research purposes, there must be normative standards of care in place. If normative standards of care were not in place, we would expect one person's understanding of quality improvement to be also another's idea of harm. I appeal to normative standards to ground

the discussion of quality improvement with this issue discussed later in this thesis. I shall discuss a variety of methods previously used to define care quality.

Ought Patients Help Define Care Quality?

It is sensible for any care provider to consider the patient's perspective on care. But doing so must necessitate that patient values are prioritised over individual physician opinion. This entrenched paternalism requires culture change. In a conceptual framework conceived by Kamel et al (2018), first, patients will need to be engaged and active decision makers in their care. Second, at the physician-patient level, there is a need for processes to elicit patient preferences that are reproducible, quick, and feasible to implement. Elicited preferences could also generate patient-specific decision aids to facilitate decision making. Third, health systems and payers need to value patient-centered care and the process of eliciting preferences over the decision itself because informed patients may make decisions that are misaligned with their own broad definitions of quality or cost. Quality measure development would focus on ensuring preference elicitation and patient-centered care. Value would subsequently be both patient-centered, increasing its value, and leading to decreases in the overall cost of care. By refocusing on the patient through collaborative decision-making and preference elicitation, the patient's values can be optimized, improving care both from the perspective of the individual patient and from that of the health system, whose role it is to serve its patients. (Kamal et al., 2018)

Patient perceptions of the quality of health services

Studies have elicited a wide range of specific definitions offered by patients themselves. The definition categories were patient-centered care; access; communication and

information; courtesy and emotional support; technical quality; efficiency of care/organization; and structure and facilities. Within the hospital setting, it was the case that the total number of patient-reported problems with care was the strongest predictor of overall patient evaluation of care in a 1992 US survey of 6455 adults.(Sofaer and Firminger, 2005) In the UK, the Consumer Assessment of Health- care Providers and Systems (CAHPS) is a national, standardized, publicly reported survey of patients' perspectives of hospital care. From these surveys, the nurse's effective communication, the responsiveness of hospital staff, pain management and doctor-patient communication were the four most strongly correlated constructs with overall patient satisfaction. Measuring patients' satisfaction alongside their experiences of care and determining the relationship between these two concepts would provide an objective evaluation of whether care was delivered in a manner consistent with the principles of patient centeredness. This work emphasises the potential significance of relational aspects of care.(Kumah, 2017)

Having provided a concise overview of the approaches and methodologies (i.e., researcher, patient safety vs. care quality, safety II – a dynamic approach to understanding processes in healthcare, clinical, nursing, and patient indicators) concerning care quality, I will next discuss how quality, in practise, is assessed.

Factors connected to care quality

The section will be discussed through the three groupings of patient, care provider and health system. The first widely adopted formal definition of the multi-determinant model of healthcare was advanced in the Lalonde "White Paper" in Canada.(Lalonde, 1974) From

the literature[5] regarding patient perceptions of care quality, the evidence is sparse and weak. This is because care is delivered by the clinician or the care provider to the patients and almost always the patient does not have the complete picture on the decisions. In other words, the clinician has the clinical expertise to know much more about the care quality given to the patient, than the patient knows or acknowledges. This is not to say that the patient is not involved or not consenting, but simply that clinicians can understand the exact medical reason, through their medical education, and justify why one decision was made over another. Often this decision is steeped in medical terminology which cannot be understood well by patients; the overall arch of care is presented before the patient. But when we want to understand what care quality is, the exact details of the care are critical for evaluating how the care process went for a patient. Even the use of patient satisfaction surveys cannot completely capture all medical care elements which can only be seen to the greatest extent by the clinician or care provider. I address this in the next section.

Measuring quality of care using patient satisfaction surveys

Quality of care has been correlated with patient satisfaction outcomes. However, there needs to be consideration around how the validity of the instrument is relatable for both patients and healthcare practitioners. For instance, a low-quality judgement that expectations of the patient are too high or that there is poor quality of care, which is then

[5] I cannot be comprehensive as the extent of studies last search were over 5,900 studies relating care quality to some other factor or construct. This discussion forms only a small selection of factor known to correlate with care quality in a meaningful way.

passed on to quality improvement. One key issue of designing a valid instrument is that it must be understood in the same or similar way for all patients. to overcome this obstacle, focus groups, concept mapping and other qualitative research activities can be undertaken to reduce the conceptual distance between researchers and patients. But these activities are quite time and cost-consuming.(van Campen et al., 1995) There has been progress on the development of theoretical models which can capture the relevant information from an instrument.(Sixma et al., 1998) The surveys may help triangulate care concerns and thus lead to greater awareness of opportunities for improvement especially when used in conjunction with other care quality instruments. Overall, it can be difficult to measure the quality of care due to the variability of standards from patients and care providers.(Al-Abri and Al-Balushi, 2014) But patient satisfaction survey can still be an developmental tool for care providers looking to identify opportunities to improve quality of care. For instance, in one study there were quality improvement activities in 13 tertiary care hospitals in response to public reports of patient satisfaction. The results showed that each of these hospitals introduced a range of quality improvement initiatives that addressed different aspects of care.(Barr et al., 2006) Given the limitations of patient satisfaction surveys and their incomplete understanding of clinical care, this thesis will stress the relevance of the physician-perspective over the patient's on care quality matters. Having set the scene on the factors which are connected to care quality, there will next be a discussion on what hospitals services are provided in the UK to introduce the context.

The Donabedian approach to quality of care

There have been attempts to define quality of care, but by far the most comprehensive and long-lasting approach was developed by Avedis Donabedian. His approach was to first

define what "quality of care" means. Many problems are present at this fundamental level, and being no different, quality of care is a remarkably difficult concept to define. In a classic paper, Lee and Jones offer "articles of faith," which are attributes or properties of the process of care and others as goals or objectives of that process.(Lee, 1933) These "articles" clearly convey the impression that the criteria of quality are value judgments that are applied to several dimensions of a process called medical care. As such, the definition of quality may be almost anything anyone wants it to be. It reflects values and goals current in the medical care system and in the larger society of which it is a part. Few empirical studies delve into what the relevant dimensions and values are at any given time in each setting. But this really is what is required. One classic paper found that 24 "administrative officials," had between them gave 80 criteria for evaluating "patient care."(Klein et al., 1961) They conclude that patient care, like morale, cannot be considered as a unitary concept and ". . . and so it appears that there will never be a single comprehensive criterion by which to measure the quality of patient care."

Yet is it an indubitable fact that the criteria selected to define quality will have profound influence on the approaches and methods used in the assessment of medical care. But often, the assumptions of this term “quality” is implicit or no addressed at all in studies. And so, any future endeavour to explore care quality improvement must make explicit its own methods and assumptions, and indeed, the level of warrant for their assumptions.

Furthermore, Donabedian elicited a tri-partite approach to care quality using structure, process and outcome.(Donabedian, 1966) For example, the structure describes the raw, physical materials required for the delivery of care and these are typically the hospital buildings, staff, human resourcing, and medical supplies. The process denotes the

transactions had been patients and providers throughout the delivery of healthcare which include the diagnosis, treatment and all the other medical efforts to help enhance the patient's health state. These also include errors in care. And outcomes are the final effects of healthcare on the patients, which include the changes to health status, behaviour, or knowledge as well as patient satisfaction and health-related quality of life. Despite its age as a concept, it is a robust heuristic for those in health service research especially when it comes to evaluating quality of healthcare.(McDonald et al., 2007) This is generally accepted by all stakeholders in the care quality circle, albeit implicitly. Next, I shall look at some of the issues of measuring care quality.

Issues with the measurement of quality

Having addressed the important elements which go into a care quality definition and its construct, we now need to see how care quality is currently being measured.

How Quality of Care can be assessed

Much concerning the problem of how to measure something begins with the definition of that certain something. Much of it depends on what people can measure, and then need to compromise with what practically they can measure given the circumstances. This can be around the healthcare practitioners, the contributions of patients or the broader health system. Furthermore, the definition of health, otherwise termed the ontology "being" of health needs to be settled as well as who the groups are who are responsibility for the upkeep of this health. Any such characterisation of quality must consider individual and social preferences and so there is likely a different optimum for those of different groups. It is only when who is involved, who is responsible, consensus around what health is and

practical measurement considerations are made can there then be a feasible programme of care quality assessment.(Donabedian, 1988) Otherwise there is likely to be no real measurement of the same construct and people will most likely be talking past each other with different versions of the programme, which vary in any one of these aforementioned components. I next turn to how other approaches from various stakeholders have been used to capture care quality.

A perspective from US researchers

Researchers from the United States indicate that the definition of care quality is a fluid concept which lies in the eye of the clinician or the patient. One of the ways they suggest that help to accommodate for this fluid meaning for care quality is to abstract information from the medical records and have the data available for physicians, administrators, and patient groups to use. And that critically, the clinician must always consider the needs and the wants of their patients in any decisions they do make.(Brook et al., 2000) Before further clarifying care quality, it is necessary to distinguish between two often conflated terms; care quality and patient safety. Clinical indicators, which are indicators of the quality of the patient care delivered, have been employed for quality improvement. This will be discussed next.

Clinical indicators for quality improvement

Clinical indicators have been used to measure the extent to which set targets are achieved. They are expressed as numbers, rates, or averages and provide a basis for clinicians, organizations, and planners with the aim of improving care and the processes by which care is provided. They can be measured via structure, process, and outcome, either as generic

measures relevant for all diseases, or disease-specific measures describing the quality of patient care relating to a specific diagnosis.(Mainz, 2003)

In general, indicator data have been of interest to patients, purchasers, and providers and outcome data may be of major interest to consumers and providers of care. Providers who are receiving data for quality improvement purposes need detailed data about the process of care to sufficiently inform the actions to be taken.(Mainz, 2003)

Clinical indicators should be both valid and sensitive to the events and changes they are designed to detect. Furthermore, clinical indicators should be well defined to avoid the measurement of changes in the patient's status arising from unmeasured external factors. Only evidence-based clinical indicators predict patient outcomes and are true measures of quality, although indicators based on professional consensus without evidence may be all that is feasible for certain conditions, treatments or patient populations.(Mainz, 2003) And so these evidence-based indictors should be sought after. Patient health outcomes are determined by many other factors besides the quality of health care itself, and this needs to be kept in mind when using clinical indicators as proxies for health care quality.(Manary et al., 2013, Stewart, 1995, Cervero and Gaines, 2015) Risk adjustment therefore can play an important role in comparison to using outcomes data, for the adjustment of any confounding factors.(Iezzoni, 2003)

The surveillance of health care quality is greatly aided by using relevant quantitative indicators, supplementing other approaches that may include qualitative analyses of specific events or processes.(Mainz, 2003) And so this thesis itself will adopt a

combination of quantitative and qualitative approaches to broach both the conceptual and contextual elements of this work. Next, I look at both the relevance and stability of using patient perceptions about care quality.

Patient perceptions of the quality of health services

It is sensible and ethical to include, wherever possible, the input of the patient and there is no subject matter more personal than the care that has been given to the patient. However, there are several obstacles that should be considered around patient perceptions on this matter. One of the major limitations for using patient perceived quality of care scores is that it is rarely possible to determine if such systematic differences should be attributed to differences in patient expectations, perceptions, or the actual care received. More work is needed to assess whether patient reports of outcomes, particularly functional status outcomes, are less reliable than are those of clinical judgement and opinion.

What then about patient reports of technical processes, rather than outcomes? Can we depend on patients to self-report whether specific actions were taken that are consistent with the evidence-based care? What actions can they reliably report? And which can they not? Additional research is needed to grapple with the uncertainty of whether to believe a medical record or the patient's report of his/her experience.(Sofaer and Firminger, 2005) Uneasy answers to these suggest that clinicians, if we had to choose one group, are the most suited to make a care quality decision.

How does hospital mortality rate relate to care quality?

How does standardized hospital mortality rate relate to care quality?

Standardized hospital mortality rates (HSMR) have been used to try and measure care quality by controlling for a host of factors in the hospital setting. Standardized mortality rates are relativized on an arbitrary scale or central point (i.e., for HSMR, the central point is 100 which indicates the mortality rate one taking into consideration key factors within that Trust). Hospital mortality rates are raw, crude, and absolute estimates of mortality over a given period and cannot be compared across Trusts given the demographic and hospital differences. In one example from the Netherlands, it was found that the overall hospital HSMRs and mortality at individual diagnostic group level can be monitored using statistical process control charts to give an early warning score of possible problems with quality of care.(Jarman et al., 2010) And a UK study concluded that, for three common procedures, risk prediction with discriminatory ability comparable with that obtained from clinical databases is possible using routinely collected administrative data.(Aylin et al., 2007) The HSMR for the Netherlands is statistically robust model that can be used as an indicator for hospital deaths to help Dutch hospitals improve quality of care. The statistical model is robust enough to include all hospitals with more than about 100 deaths per year, an average case mix and good quality data, varying in size and function, into one analysis. However, random variation and data quality issues do need to be considered when interpreting the results of these statistical models.(Manaseki-Holland et al., 2019a) HSMRs have been used to highlight hospitals that have significantly higher mortality and it has been claimed that the impact of interventions on mortality reduction can be monitored using this measure. HSMRs have also been calculated for several other countries. But

before we go to discuss the UK approaches, we need to have a summary of the actual hospital services found in hospitals to understand the business context.

Hospitals Services in the United Kingdom

UK hospitals are institutions providing medical and surgical and nursing care for sick and injured people. A wide range of services are provided which include(Webster, 2002):

Surgical operations, Specialist medical treatments, Accident and emergency, Consultations, Diagnostics, Maternity and neonatal, Pathology, Termination of pregnancy services. For specialist treatment, referrals from primacy care providers are required and are only accessible on a need's basis.(McKee and Healy, 2002, Wunsch et al., 2008) In other high-income country, the access condition based on need, but some other criteria, may vary but the health service provision are typically uniform across countries. However, in low-income countries, hospital service provision can vary widely.(Hensher et al., 1999, Kobusingye et al., 2005) Specialist hospital services such as:

- Head and neck oncology
- Perinatology (high-risk pregnancies)
- Neonatology (high-risk new-born care)
- PET scans
- Organ transplantation
- Trauma surgery
- High-dose chemotherapy for cancer cases
- Growth and puberty disorders
- Neurology and neurosurgery

and are provided in high-income countries, but less so in low-income countries. From this point on, the terms “care quality” or “quality of care” will refer hospital care quality. Having outlined what services are typically offered to patients are constitutive of hospital care quality, I will next discuss what care quality improvement is.

Variations in the care quality standard held in the clinician’s mind, may influence how and what quality improvement is achieved from what is maximally possible. I consider that the care quality standard (or construct in the mind of the clinician), may serve a role along factors such as judgement biases, attitudes, and organisational-cultural factors. The care quality standard has yet to be formally measured or studied in terms of its effects upon care quality judgements. This should be a driver to pursue further research to understand the importance of the care quality criteria employed in the clinician or patient’s mind. Thus, there must be an attempt to elicit its component parts. Despite not explicitly knowing the actual care quality standard employed, it would be sensible to establish what is first good care and subsequently what is not good care.

Broad frameworks for quality improvement in the UK

Healthcare quality improvement and safety management are important for the safe, effective care and treatment of patients. Preventable deaths have been used as a proxy of care quality.(Manaseki-Holland et al., 2019b) Preventable hospital deaths are known to occur in UK NHS hospitals.(Hogan, 2016) Preventable death in the hospital setting is defined by the Office of National Statistics (ONS) as:

'if, in the light of medical knowledge and technology at the time of death, all or most deaths from that cause (subject to age limits if appropriate) could be avoided through good quality healthcare.'(Statistics, 2013)

Importantly, preventable hospital deaths are a concern for hospital staff, management, and the public. They should not be happening. If they are happening, they must be acted upon to prevent their occurrence. Quality improvement was allowing the adaptation of the latest knowledge and methods to meet the needs of current patients, healthcare professionals and policymakers.

Measuring care quality using "preventable deaths" as proxy: how do we identify a preventable death?

Since 2001, the main approach has been to use hospital data repositories and produce summary statistics from them.(Jarman et al., 2005) Hospital mortality statistics have since been used in England, Wales and Scotland as indicators of care quality. In 2009, this was developed by the Dr Foster statistical team, a collaborative unit of public health professionals, statisticians, epidemiologists and informaticians at Imperial College, London who created the hospital standardised mortality ratio (HSMR).(Aylin et al.) The HSMR considers only the most life-threatening conditions of the in-hospital deaths. Dr Foster assumed that adjusted mortality statistical models were more accurate and reliable than non-adjusted ones.(Aylin et al.) Their use received criticism and there was no reasonable and robust evidence, as it stands, to suggest a clear association between high mortality indices and poor hospital care.(Shahian et al., 2010, Mohammed et al., 2009a) Across four acute NHS hospitals, the differences in both coding and coding depth explained the two interaction effects observed for the Charlson co-morbidity index, emergency

admission,[6,](Dixon et al., 1998) admission practices across the hospital, specifically the changeable hospital-specific admissions policy that altered the proportion of non-emergency-to-emergency admissions.(Mohammed et al., 2009b) Shahian et al. used four common methods to calculate hospital mortality-wide statistics with each leading to different outcomes.(Shahian et al., 2010) This is evidence that challenges the integrity of HSMRs.

Since October 2011, the Health and Social Care Information Centre (HSCIC) have produced the summary hospital-level mortality indicator (SHMI) on a monthly basis.(NHS Digital, 2020) The SHMI is the ratio of the observed number of deaths to the expected number of deaths. The numerator uses 'deaths report of patients admitted to non-specialist acute NHS Trusts who die while in hospital or within 30 days of discharge' (excluding specialist trusts, mental health trusts and community trusts). The denominator is the prediction that adjusts for the patient's diagnostic condition, age, sex, co-morbidities (Charlson index), and the method of admission to hospital. The SHMI adjusts for patient case-mix, which is the mix of patients coupled with the patient population context.(Campbell et al., 2012)

There are differences between the HSMR and the SHMI. The former uses 99.8% and the latter 95% control limits as thresholds for the 'above-expected' and 'below-expected' mortality-rate levels. Also, the SHMI considers all deaths within 30 days after discharge

[6] Depth of coding refers to the extent to which the provider populates primary, secondary, and other subsidiary diagnostic information pertaining to the patient record.

whilst the HSMR places no such limit. An alternative hospital mortality statistic is the risk-adjusted mortality index (RAMI), that was used in parts of Wales prior to the Palmer review. RAMI adjusts for the underlying health of patients and the procedures undertaken. However, this is insensitive to external factors such as deprivation levels, patient lifestyle, community health service provision(Pitches, 2014) and the specific statistical model adopted.(The Welsh Government, 2014)

The use of hospital mortality statistics to measure care quality

With over 15 years of quality and patient safety improvement, the field has advanced approaches from the routine use of statistical charts and adjusted hospital mortality statistics i.e. Summary Hospital Mortality Index (SHMI), the risk adjusted mortality index (RAMI)(DesHarnais et al., 1988) and hospital standardised mortality ratios (HSMR),(Aylin et al.) to incorporate more culture-motivated approaches utilised by such organisations as the United States' Institute of Health Improvement (IHI).(Berwick et al., 2008)

Weaknesses of hospital mortality statistics

There are several weaknesses for hospital mortality statistics. Dr Foster noted that hospitals recording an average of 2.5 codes per patient had a HSMR of approximately 15-20 points higher those that recorded 6 codes per patient.(Jarman et al., 2005) Ergo, the HSMR is sensitive to coding depth. With one unintended consequence, the presence of perverse motivations coupled with financial incentives led to an increased prevalence of coding depth, a phenomenon dubbed as '*over-coding*', has been extensively practised in the US.(Simborg, 1981) If over-coding is systematically applied, the hospital mortality statistic is more likely to misrepresent the true care quality. One must cumulatively consider *coding*

accuracy, depth, and patient case-mix to truly represent the actual quality of care. Next, I set out a minimal set of questions any hospital care quality measurement tool must answer:

- Is care quality accurately measured using hospital mortality indices? Different care quality measurement methodologies each lead to markedly different hospital mortality statistics.(Shahian et al., 2010) The statistic is very sensitive to the initial modelling assumptions.
- Are there reasonable criteria to direct our methodology in any given hospital context? As it stands, there are none found in the literature. And furthermore, care quality in one context may use a hospital mortality statistic differently than in another context. In one study, the second level examinations, those data which are used to support quality of care measures, of mortality data were important, however, the actions taken with this information were not reasonably defensible. For instance, in United States Veteran Medical Centers, '*hospital acquired infection rates, rates of hypoglycaemia and hyperglycaemia, throughput, and deep venous thrombosis prophylaxes* were chosen by most clinical leads as proxies for care quality. However, there are no evidence-based reasons for their selection.(Render et al., 2011) There are no clear clinical indicators strongly correlated with care quality.

Statistical models were incorporated to help identify the important signal of "care quality" from the noise. '*Case-mix adjustment*'(CMA) is a statistical approach used to adjust for patient severity and other hospital-specific features using hospital mortality statistics. This adjustment could reduce or even introduce more bias into the statistic. CMA assumes the risk relations hold indefinitely between measured indicators. This condition is called the

ceteris paribus condition, which is a common assumption in empirical inquiry where investigators assume that all other variables except those under immediate consideration are held constant to permit the illustration of concepts pertaining to a specific area of inquiry.(Reutlinger et al., 2011b) However, the invocation of *ceteris paribus* are unjustified and it can lead to misleading care quality measurements. This phenomenon is referred to as the '*constant risk fallacy.'*(Nicholl, 2007) Here I offer an example of CMA in action. Assuming perfect coding obtains, one must consider that the risk of a patient dying from, for example, chronic heart failure in a region is not the same for another region. There are several factors for why this might not obtain. Smoking and alcohol lifestyle factors are not captured in the Hospital Episode Statistics (HES) data and are certainly likely to influence risk of death. The assumed CMA state of *ceteris paribus*, a term meaning '*all other things remaining constant,*' is too stringent and idealized to be found in the complex healthcare setting.(Cartwright and Cartwright, 1983, Reutlinger et al., 2011a, Cartwright and Bradburn, 2011) Such *ceteris paribus* relations tend to apply in well-characterized physical and logically consistent systems.(Schiffer, 1991) CMA in care quality measurement egregiously assume a range of factors (i.e. environmental, psychological, political, social and cultural factors) are not influencing the care quality measurement. Without sufficient warrant, the CMA assumes that only salient factors for care quality measurement have been considered with non-salient factors excluded. However, healthcare systems are complex and are themselves embedded in broader health systems, policy contexts and cultures which each could modify the adjusted HSMR statistic.(Mannion and Davies, 2018b, Mannion and Braithwaite, 2012) Consequently, if the mortality statistics is possibly misleading and incorrect, any subsequent decisions based on these statistics are misinformed and possibly detrimental. This is using CMA to generate hospital mortality statistics is methodologically

biased and practically harmful, if not misleading. An alternative to hospital mortality statistics should be found. It is quite possible that CMA leads to unidentified harm to hospital patients.(van Gestel et al., 2012) Case-mix has also been indicated to influence the indirect standardization of these HSMRs, and should be carefully considered before evaluating HSMR as a reliable and accurate care quality proxy.(Pouw et al., 2013)

Strengths of hospital mortality statistics

To improve hospital care quality, our efforts must accurately and reliably measure care quality. Our work must appraise whether the hospital mortality statistics are an accurate and reliable measure of care quality. Good coding accuracy and depth are essential for ensuring the precise adjustments can be made. For instance, good coding accuracy ensures one is measuring what seeks to measure; good coding depth ensures that the coding is to a sufficient granularity to be representative of the true quality of care. A further strength is the rapidity with which these statistics can be generated and interpreted to a lay audience over comparable methods such as case-note reviews, which are laborious and time-consuming to undertake. There is the low-cost to hospital mortality statistics, which cannot be stated for case-note reviews, which take up valuable clinician time and administrative staff resourcing and management.

To summarise, hospital mortality statistics for quality improvement has several insuperable drawbacks. Despite being easy to generate, interpret, manipulate, and produce propensity scores on key variables for, the weaknesses abound for these statistics. Firstly, the use of a metric entails that the information-richness from the medical context will be lost. Secondly, there are assumptions made in the model which generates the final statistic, with no

methodologist nor clinician entirely clear why one model was chosen over another. Furthermore, there are often no good reasons for choosing one method over the other. Thirdly, there is the loss of contextual information by siphoning the patient data and medical record through a series of variables to compute the statistic; and important contextual information lost is irretrievable. Fourthly, there are likely unobserved confounders "lurking" variables especially when there is no causal model to guide the identity and mechanism of the phenomena, which here is how certain selected variables (should) relate to the hospital mortality statistic model.(Williamson et al., 2014) And to re-emphasise the fact that there are a number of these models and each one produces slightly (and in some instances markedly) different outputs mortality ratios.(Mohammed et al., 2009a) Despite bearing all these methodological weaknesses in mind, hospital mortality statistics still has a place in the quality improvement toolkit as a first-level smoke screen for poor care. However, which data should be presented and the decisions which need to be made around hospital mortality statistics remains less clear as discussed by Render and colleagues.(Render et al., 2011) Next, the focus is on another process to measure the process of care through errors and the outcome of adverse events.

The field of human error in patient safety was highlighted in a ground-breaking report published by the Institute of Medicine, *To Err is Human: Building a Safer Health System*.(Kohn et al., 2000). Since this publication, the UK Chief Medical Officer and his panel of health experts have since published a white paper[7], '*An Organisation with a Memory.*' In this report, the learning culture emphasises that the local hospital culture is a

[7] White papers are government reports giving information or proposals on issues.

likely source of patient safety concerns.(Department of Health Expert Group (Chairman, 2000) To situate our current patient safety landscape, I discuss the influence of key white papers on the patient safety landscape.

UK 'Patient Safety' White Papers

The *Francis Report*, written by QC Robert Francis, was a national inquiry into the care quality concerns arising from the Mid-Staffordshire NHS Foundation Trust (2005-2009). This report showed us how we know what the care quality was like then. The report identified a variety of problems between commissioning, supervisory and regulatory organisations and the Trust culture and system. At the Trust board level, the report found a negative culture, the presence of professional disengagement, patients not being heard, poor governance and a greater concern for hospital finances over patient quality of care. The report found a culture that lacked an openness to criticism with there being a wanton disregard for patients placed in the Trust's care. The Francis Report was a landmark report that helped raise the scrutiny of hospital quality of care process. Out the report came extensive recommendations(Francis, 2010) which, one year on from the report, received a government response which accepted the need for NHS vigilance and reform in the instance where care was found to be poor.

Three years later, and in a similar vein, the *Berwick Report* highlighted the importance of hospitals embracing an ethic of learning.(Berwick, 2013) The *Keogh Review* scrutinised the quality of care and treatments provided by 14 hospital trusts in England. The review identified key domains with which hospitals were assessed for healthcare quality improvement. These include concerns around patient experience, workforce, clinical and

operational effectiveness, governance and leadership and safety.(Keogh, 2013) Pertinent to quality improvement, the review found that root cause analyses, which was a method for identifying root causes of faults or problems(Wilson, 1993), were poorly conducted with limited dissemination of learning. Such conditions would have promoted little tangible quality improvement from any pertinent lessons available.

These documents all continued to emphasise the importance of learning from mistakes. This emphasis of promoting a learning culture was emphasised in the report entitled, '*Learning, Candour and Accountability*' by the Care Quality Commission (CQC). The first report recommended that a widespread learning culture was necessary to improve the care process. To achieve this learning culture, indicators were measured which include family and carer involvement, identification and reporting of deaths, steps to review suspicious deaths and the establishment of strong, good governance and learning cultures. Its singular purpose was to learn from patient deaths to improve hospital care quality.(Care Quality Commission, 2016)

In 2017, the National Quality Board report entitled 'A Framework for NHS Trusts and NHS Foundation Trusts on Identifying, Reporting, Investigating and Learning from Deaths in Hospitals' recognized that quality of care was a function of leadership and system-wide factors.(National Quality Board UK, 2017) The purpose of the report inaugurated in the UK a national initiative to review and investigate deaths to prevent their recurrence. The Royal College of Physicians (RCP) and the Yorkshire and Humber Academic Health Science Network (YHAHSN) were commissioned by NHS England to deliver the Structured Judgement Review (SJR) national case-note review training. This training was cascaded

between 2016 and 2018 to identify learning points from deceased patients that had died in hospital.(The Royal College of Physicians, 2016d, The Royal College of Physicians, 2016b, The Royal College of Physicians, 2016c) The report provided guidance on how Trusts could deal with certain types of deaths (i.e. learning disability, mental health needs, an infant or child death and a stillbirth or maternal death) and how deaths were reviewed using the SJR methodology which was adapted for local Trust requirements. There was also the need for each Trust to provide a public release of a mortality review policy communicating the measures taken by NHS Trusts to learn from their deaths by September 2017. The case-note review training was conducted by RCP-YHAHSN with facilitation by Health Education England (HEE) and the Healthcare Safety Investigation Branch (HSIB). I consider the policy history behind the different case-note review formats in section.

At the time, preventable deaths were deemed to be a burdensome problem, however due to the evocative and everchanging definition of 'preventable death' and case-note reviews then used to reduce these preventable deaths. Preventable deaths implied that some deaths should not have occurred, which is hard to determine with even all the care evidenced and discussed by clinicians and coroners. Rather than leading to positive learning culture, it has often fomented a blame and finger-pointing culture. Due to the difficulties around this definition, the RCP replaced it with 'learning from deaths' instead. This was not to suggest that preventable deaths are not present or no longer being detected but more so that the narrative around mortality reviews stresses 'learning from deaths' as opposed to the blame that is typically associated with events following the identification of a preventable death. Finding a 'preventable death,' as it were, would imply lessons to which the 'preventable' events are not to be repeated.(Stewart et al., 2016) Since 2019, United Kingdom's (UK)

National Health Service (NHS) improvement, the quality improvement branch of the UK's NHS, have been rolling-out a medical examiner system across England and Wales to deliver on greater scrutiny of deaths.(The Royal College of Pathologists, 2016)

The role of medical examiner was developed out of the recommendations in the 2003 Home Office Fundamental Review of Death Certification and Investigation(Services and Luce, 2003) and in response to concerns raised by Janet Smith in her third report into the murders committed by English general practitioner Harold Shipman.(Smith, 2004) This recommendation was endorsed by Robert Francis in his investigation into avoidable deaths at the Mid-Staffordshire NHS Foundation Trust(Inquiry, 2013) and Bill Kirkup in his review of deaths and patient safety at Morecombe Bay Hospitals.(Kirkup and Investigation, 2015) A medical examiner is an independent senior doctor who is accountable to the national medical examiner.(The Royal College of Pathologists, 2016) Their role is to manage three matters related to the cause of death, handling relevant medical documentation, taking the views of bereaved relatives into consideration. First, a medical certificate of cause of death (MCCD) is completed by the medical examiner and they confirm that the content is as accurate as possible. Second, where a case needs to be notified to a coroner, the medical examiner will make sure it is undertaken as swiftly and accurately as possible. Finally, the medical examiner will help detect and report clinical governance concerns as swiftly as possible.

Independent Inquiries of Hospital Deaths

Since the mid-20th century, the UK initiated a series of independent enquiries into hospital-related mortality involving case-note mortality review. During the years 1952-1954, the first confidential enquiry into mortality reviews in England and Wales was conducted

around maternal deaths and it has since developed to cover the four countries in the UK.(Ministry of Health, 1957) Fast forwarding 35 years, '*The report of a confidential enquiry into perioperative deaths'* was published with a focus on the mortality and outcomes of anaesthesia and surgical patients in the UK. This publication led to the creation of the *National Confidential Enquiry into Patient Outcome and Death* (NCEPOD) whose purpose was to '*to assist in maintaining and improving standards of care for adults and children for the benefit of the public by reviewing the management of patients...*' and remains in place to this day.(Raja and Thomas, 2019) In 1987, the second National Inquiry into Perioperative Deaths identified that most surgical procedures were performed on the frail and the elderly with 0.7% of these surgical procedures resulting in death. And overall, there were recommendations for trainees to be overseen by experienced doctors.(NCEPOD, 2018) In 1996, '*Safer Services*', the National Confidential Inquiry into Suicide and Homicide provided an extensive analysis on suicide, homicide and sudden unexplained death by patients across the UK.(Amos et al., 1997) In 2015, the *Learning Disabilities Mortality Review* (LeDeR) programme aimed at making improvements to the lives of people with learning disabilities.(Disabilities;, 2016) And most recently, the *National Mortality Case Record Review* (NMCRR) *Programme* aimed to improve understanding and learning around the problems in healthcare associated with mortality to help share best practice.(The Royal College of Physicians, 2016b) This has set the scene to turn to the measurement aspects around hospital mortality. In the next subsection, I outline an approach which has the potential to change alter the science of patient safety and care quality improvement, the Safety-II approach.

Re-thinking the approach to care quality: a Safety-II approach

Most conceptions of safety are seen as the absence of accidents and incidents (or as an acceptable level of risk). With this perspective, this is called ”Safety I”, and is defined as a cultural stance which minimises errors.(Hollnagel, 2018) A Safety I approach holds that things go wrong because of identifiable failures or malfunctions of specific components (i.e. technology, procedures, the human workers and the organisations in which they are embedded). Human agents are therefore viewed predominantly as a liability or hazard, principally because they are a key source of variability.

The purpose of accident investigation using this approach is to identify the causes and contributory factors of adverse outcomes, while risk assessments aim to determine their likelihood. The safety management principle is about responding when something problematic occurs or is considered an unacceptable risk, usually eliciting the response of eliminating causes or working around these risks. This view of safety became widespread in the safety critical industries (nuclear, aviation, etc.) between the 1960s and 1980s. At that time performance demands were significantly lower than today and systems simpler and less interdependent.

It was tacitly assumed then that healthcare systems could be decomposed and that the components of the system functioned in a bimodal manner either worked correctly or incorrectly. These assumptions led to detailed and stable system descriptions that enabled a search for causes and fixes for malfunctions. But these assumptions do not fit today’s world, neither in industries nor in health care. In health care, systems such as an intensive

care or emergency setting cannot be decomposed neatly, neither in detail nor for as a whole system. On the contrary, clinical work is and must be variable and flexible.

The Safety I view does not consider why human performance practically always goes right. Procedures in healthcare do not go correctly because people behave as they are supposed to, but because people can and do adjust what they do to match the conditions of work. As systems continue to develop and introduce more complexity, these adaptive adjustments become increasingly important to maintain acceptable performance. The challenge for safety improvement is therefore to understand these adaptive adjustments in other words, to understand how performance usually go well despite the uncertainties, ambiguities, and goal conflicts that pervade complex work situations. Despite the obvious importance of things going right, traditional safety management has paid little attention to this.
And so, the choice between safety I or a safety II will influence the understanding of what quality is. In the safety I perspective, care quality definitions and approaches will assume that there are minimal errors and that good care is present when there are no errors in care. In the safety II perspective, the concept of care quality consider how the care delivered adapts to the patient and healthcare context.(Hollnagel et al.) We next consider what needs to be done to monitor this.

Monitoring system

Using adverse events and errors to measure quality

One of the ways to measure care quality using Donabedian's outcome is to capture adverse event as an outcome measure. Adverse event monitoring is the systematic detection, investigation and analysis of events which indicate that a patient received poor quality care.

First developed in the USA in the mid-1970s to study the potential levels of medical negligence. Adverse event monitoring provides a systematic form of clinical audit, which can be implemented across most specialties or an entire hospital relatively easily. Because it is, by nature, a multidisciplinary, multispecialty approach, it can identify issues which cross specialty and professional boundaries and helps to ensure that a minimum level of audit takes place in every area.(Walshe et al., 1995) Adverse events during hospital admission have been shown to affect nearly one out of 10 patients and a substantial proportion of these events are considered by clinician observers to be preventable.(de Vries et al., 2008)

Adverse event monitoring does not, in and of itself, provide a full picture of the quality of care. It focuses on the process of in-patient care, and takes no direct account of final health outcomes, nor of out-patient or community-based care services, except indirectly when they result in or affect in-patient hospital care. It does highlight poor quality care but does not of itself help to define what constitutes good quality care. Moreover, adverse event monitoring is fundamentally dependent on the quality of the medical records – and so poor, sketchy or badly organized medical records are not a good condition for reliable records-based audits. To address information stream upstream of these adverse events, the focus needs to be on an alternate process to detect care quality; case-note reviews. We turn briefly now to the measurement of errors as a proxy for care quality.

Errors are different to adverse events. Errors, as defined by James Reason (a human factors expert), do not necessarily harm patients, whereas the term adverse event does imply harm.(Reason, 1990) And it is shown also that an individual error or adverse event is

typically the result of numerous latent errors in addition to the active error committed by a clinical practitioner. There are a range of error detection methods which are beset by several limitations ranging from hindsight bias, reporting bias, high expense, its potential reliance on incomplete and inaccurate data and their poor ability to detect upstream latent errors, which set off the error cascade.(Thomas and Petersen, 2003) Due to these weaknesses, we turn now to case-note reviews as another potential solution to accurately and reliably detect the signal of care quality.

The use of case-note reviews as a care quality improvement tool

Why use case-note reviews for quality improvement?

Case-note reviews offer the most granular form of information of all systematic forms of medical information. These reviews have been shown to reliably detect adverse events.(Vincent et al., 2001, Brennan et al., 1991, Wilson McL et al., 1995, Thomas and Petersen, 2003) It was shown that a routine incident reporting system may not provide an accurate picture of the extent and severity of patient safety incidents, particularly those resulting in harm to patients. And the study furthermore found that healthcare organisations should consider using routine structured case-note review on samples of medical records as part of quality improvement.(Sari et al., 2007)

Case-note reviews: its history, methods, and use

Case-notes are a complete record of care, at least, for a particular hospital episode. [8] An episode is defined as the period between admission to discharge or death. Case-note reviews are the evaluation of the care quality of these compiled case-notes. It can take one of these two forms. In this thesis, two case-note review forms are mentioned. The first is the general 'case-note review' format which is voluntarily undertaken in routine practise, audits, investigations, formal inquiries and in accordance with established topic-specific case-note reviewing programmes. The second is 'mortality case-note review', which is mandated by England's National Quality Board, for all hospitals to use case-note reviews to learn from in-hospital deaths.(National Quality Board UK, 2017)

The methodology of case-note reviews

Now I offer some historical context on the development of the method for case-note reviews. The first formal methodology for case-note review was developed by RAND, an organisation based in North America, whose organisational aim is to improve policy and decision-making through research and analysis. This organisation developed the structured-implicit form of case-note review that recommended that the case-note reviewer have their clinical judgement guided by a semi-structured proforma.(Rubenstein et al., 1991) Case-note reviews are also historically undertaken for medico-legal purposes. Our review context of interest is tied to routine screening and the examination of in-hospital care quality which

[8] In certain case-notes, some previous episodes are collated into the case-note.

is distinct from the judicial processes of medical negligence cases.(Thomas, 2009) This form of review is the most widely utilised today and so I refer to this specific case-note review methodology from now on.

A signal-to-noise problem

Case-note reviews have been considered to offer more granular information compared with hospital mortality statistics in measuring the construct of preventability (or avoidability) following hospital care. Thus, there has been a drive over the last few years that has sought to identify these preventable death 'signals' using hospital mortality statistics.(Intelligence, 2009, NHS Digital, 2020, Aylin et al.) The signal-to-noise ratio is a term referring to the level of desirable signal (true care quality construct) to the level of background noise (unconsidered factors modifying care quality construct). However, this has been too little or no avail as the small 'signal' of preventable deaths is drowned out by the large 'noise' from deaths from non-preventable causes. This is because non-preventable deaths have too much noise around the signal whilst preventable deaths have less noise but more measurement error. (Lilford and Pronovost, 2010a) This measurement error is most likely attributable to the subjective case-note reviewer judgement over the care quality.

In general, the case-note review process can be described in Figure 2. The case-note reviewer makes observations from the case-note review content, which include the current perceived standard of care and the circumstantial nature of the case-notes. The documentation, the case-note, is reviewed by the case-note reviewer after having read all, if not in the very least, the pertinent parts of the case-note. The reviewer will invoke a care quality standard, a measurement scale, with which they are rating the care quality against.

The reviewer will deliberate on the overall care quality given to this patient from the case-note. This is typically given on a 5- or 6-point Likert scale, but other scales have been used.

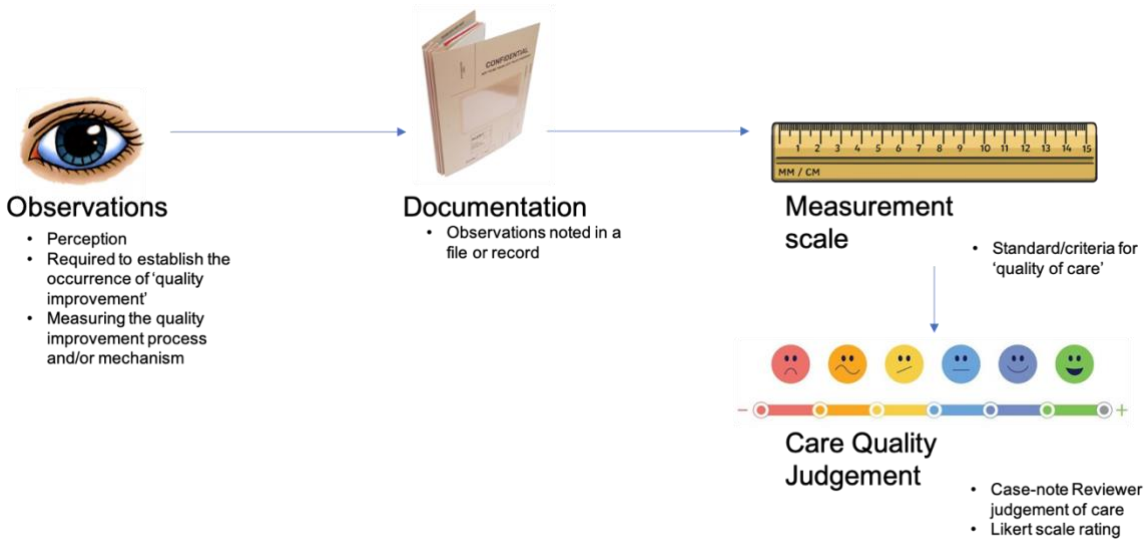


Figure 2. The process of reviewing the care quality of case-notes

What about the signal-to-noise ratio?

If we desire to measure hospital care quality accurately and reliably, then the signal must be measured well according to these standards and the noise must be mitigated or removed all together. The identifation of this noise can be done in several ways:

- First, there could be the calibration of reviews by obtaining the central tendency with its variation across care quality judgements from a set number of reviews. However, this would not establish whether the care quality judgements were accurate and truthful, but only that they were reliable across case-note reviewers.(Manaseki-Holland et al., 2016, Smith et al., 1997b, Hofer et al., 2004) For instance, there could be details in the case-notes which unintentionally lead to inaccurate care quality judgements, which arise from cognitive biases and heuristics or some other source of bias. It must be stated though, that it is the least bad of

methods, as generally speaking, reviewers would be able to broadly distinguish between very good and very poor care quality. The issue lies in the borderline cases between good and bad care.[9,10]

- Second, training can be provided to improve the awareness of good reviewing protocol and ways in which uncertainty can be managed during case-note reviews.
- Thirdly, one can seek to obtain the detailed rationale behind the reviewer's care quality judgements. This can be achieved through either the review proforma, clinical meetings or medical re-validation, which will likely raise the transparency of case-note review decision-making to help reduce the noise. This has been proposed by a group of researchers (Hutchinson et al., 2010e) and implemented nationally for reviewers across the UK.(The Royal College of Physicians, 2016c) The final recommended approach requires that each reviewer present reasons for their care quality judgements during the case-note review.[11] The only concern with this is that sharing reasons for a reviewer's judgement are time-consuming and lengthy and would very likely conflict with their clinical duties. Proformas and IT

---

[9] Another factor which could reduce the validity of calibration is if the standard of care quality, in the mind of the reviewer, varies markedly in the judgement of case-notes. Disagreements may well arise from the employment of different care quality standards used during case-note review.

[10] This element of 'good faith' on the reviewer's behalf will be scrutinized as ill-placed faith when reviewers systematically err.

[11] This is a method to engage the reviewer's rationale for judgement and thus make more explicit the reviewer's quality of care standard used to judge the case or the decision.

systems can be streamlined to work more seamlessly alongside their clinical work, to free up more time for this important scrutiny and reasoning behind their care quality judgements. If implemented widely, we should find the more precise line of reasoning reviewers have taken to form the overall care quality judgement about a case-note. Transparency is essential for clinicians and patients to understand how a decision was made.

- Fourth, an effective and systematic way to organise hospital data and its containment in the records could reduce the total noise in the case-notes. This will likely reduce the noise of care quality judgements over these case-notes as demonstrated with other types of medical documentation.(Donoghue et al., 2011) It will have to be explored at the hospital level (or higher) what the most appropriate form of data organisation will be taking into consideration the existing data infrastructure, time available, money available and stakeholder willpower.

The above have outlined several the possible approaches to isolate this small signal-to-noise ratio.

The different types of case-note review methodologies

In the UK, researchers have investigated the advantages and disadvantages of the criterion-based approach compared with the structured-implicit approach in capturing the care quality and determining the death's preventability. The care quality and the preventability of the death are overlapping terms, as an actual preventable death could not possibly indicate good care. Care quality is a broader and more inclusive term as good care can still be demonstrated in a preventable death case.

The criterion-based (focused-explicit) case-note review has a higher level of inter-rater agreement score than the holistic (structured-implicit) approach.(Ashton et al., 1999) A drawback of the criterion-based approach has been its low granularity and sensitivity(Camacho and Rubin, 1998) for detecting weak or noisy care quality signals.(Mohammed et al., 2005) The two have been compared and assessed for their measurement properties in rating quality of care.(Hutchinson et al., 2010a) The '*holistic*' component provides clinicians the opportunity to exercise their expert clinical judgement. This could help identify other non-listed factors that are also potential influencers of care quality.

The '*structured*' criterion-based approach has a shortcoming. A UK care quality review of four '*top-performing'* hospitals and four '*struggling*' hospitals for stroke revealed that 'organisation of care' and '*do not resuscitate*' orders were influencers of care quality with this being a surprise finding for the researchers.(Mohammed et al., 2006) Thus, the criterion-based approach failed to identify factors influencing care quality. The upshot is that it is better to identify '*implicit*' influencers using a more holistic situational judgement approach. However, a full '*implicit*' review has been shown to be less sensitive in measuring the hospital quality of care construct as it tends to decrease the '*signal-to-noise*' ratio(Gibbs et al., 2001) and also increase the time and effort to review cases.(Hutchinson et al., 2010b, Iezzoni, 1996) This has also been formally discussed by Girling et al, who partitioned the mortality arising from optimal care (U) and suboptimal care (V) and giving consideration to the variability of key quantities such as the preventable hospital mortality rates, the overall hospital mortality rates and proportion of the variance in total mortality rates as explained by the model.(Girling et al., 2012a)

$$M = U + V\ (Mortality\ partition\ equtation.)$$

The structured-implicit case-note review is the *via media* between explicit 'structured' and full implicit 'holistic' review forms.(Pearson et al., 2000) It strikes the balance between fidelity and practicability. Future research could identify the source(s) of variance (and lack of agreement) from the structure-implicit review format. Case-note reviewers armed with sources of variation are better placed to mitigate any unwanted influence during care quality decisions. This has the potential to enhance the reliability of care quality judgements/decisions.

The types of scales used to judge case-note care quality

Regarding the scale used for care quality improvement, a Likert scale is non-inferior to a continuous scale (0—100). There is a high correspondence between Likert and continuous scales for measuring case-note review's preventability of death. The inter-rater reliability of both remain low.(Manaseki-Holland et al., 2016) Thus, it must be carefully calibrated for its use as a care quality measure. Otherwise, it is recommended as an ordinal measure (e.g., the distance from 0 to 1 is not the same as option 3 to 4 on the Likert scale) and not an interval measure (e.g., the Fahrenheit scale).(Trochim, 2016) Training is essential in establishing 'exemplar' cases to set preventability standards, most likely through a common criterion. It would be advisable to use the Likert scale as it is a more facile tool to use over the continuous scale and there is little difference in the granularity of the care quality judgements between these two scales.(Norman, 2010)

The threshold used by doctors to assess the preventability of a hospital death varies and was contingent on several factors (e.g., healthcare culture, patient-mix, clinician experience, adequacy of the medical records, existing formal processes for audit and learning from mortality and morbidity meetings, precedents, personal experience etc.). Thus, there are a panoply of factors that could influence the case-note review care quality judgement. This was the case within and between reviewers.(Goldman, 1994, Goldman, 1992, Armstrong et al., 2007, Hutchinson et al., 2010c) These case-note review studies are presented in Table 2 with the reviewer agreement statistics. It must be noted that Fleiss kappa, a statistic that accounts for chance agreement, was used to compute the reliability between different observers. The term "kappa" denotes the Fleiss kappa, unless otherwise stated. And if relevant, implicit, or explicit labels are given to denote the review type. This table will be revisited in the final chapter (Chapter 6Chapter 6) to see the extent to which our practical knowledge has improved to solve some the disagreement issue between reviewers.

Table 2. Case-note review studies on reviewer agreement levels

Journal Article	Key Findings	Implications for Case-note review
The reliability of peer assessments of quality of care R. L. Goldman Journal of American Medical Association 1992 Vol. 267 Issue 7 Pages 958-960	The degree of agreement beyond chance was compared with accepted standards in 12 remaining studies. Most of these studies found agreement corrected for chance to be in the range regarded as poor, indicating that physician agreement regarding quality of care is only slightly better than the level expected by chance.	Given the wealth of the resources devoted to quality assurance and the centrality of peer assessment, there is a need for a re-examination of this peer review process. Several proposals appear to have potential for improving the peer review process including the more objective assessment procedures, multiple reviewers, higher standards for reviewers, elimination of systematic reviewer bias, use of outcome judgments, and adoption of practice guidelines.
Inter-rater reliability of preventable death judgments. The Preventable Death Study Group E. J. MacKenzie, D. M. Steinwachs, L. R. Bone, D. J. Floccare and A. I. Ramzy	The review of non-central nervous system (CNS) deaths agreed in only 36% of the cases (kappa = 0.21). Agreement among panels reviewing CNS deaths was higher at 56% (kappa = 0.40). Most of the	There is low agreement between reviewers around non-CNS deaths compared to CNS deaths. Agreement levels differ by patient age, and for patients with less severe injuries.

J Trauma 1992 Vol. 33 Issue 2 Pages 292-3 discussion 302-3	disagreements, however, were in judging whether deaths were "not preventable" or "possibly preventable." Agreement was higher for early deaths and less severely injured patients. For non-CNS deaths, the agreement was higher for younger patients than older patients.	
The Reliability of Peer Assessments: A Meta-Analysis	*Implicit Review*	Implicit review has poor inter-rater agreement levels.

R. L. Goldman Evaluation & the health professions 1994 Vo Issue 1 Pages 3-21	The weighted mean kappa of the 21 findings from the 13 studies was 0.31 (SD=0.10).	
Inter-rater reliability of case-note audit: A systematic review R. Lilford, A. Edwards, A. Girling, T. Hofer, G. L. Di Tanna, J. Petty, et al. Journal of Health Services Research and Policy 2007 Vol. 12 Issue 3 Pages 173-180 DOI: 10.1258/135581907781543012	Mean kappa values ranged from 0.32 to 0.70. *Explicit vs. Implicit* Measured reliabilities were found to be higher for case-note reviews based on explicit, as opposed to implicit, criteria *Focus (adverse events, causality, process error)*	Explicit reviews have higher agreement levels than implicit reviews. The different focus of the review affects the agreement levels.

	Reviews that focused on outcomes (including adverse events) had higher reliabilities compared with process errors.	
Comparison of case note review methods for evaluating quality and safety in health care A. Hutchinson, E. Coster J, L. Cooper K, A. McIntosh, and J. Walters S	*Implicit (holistic) review* Across heart failure and COPD patients, the weighted mean kappa[12] between **doctors** is (95% CI) is 0.51(0.40 to 0.61). The kappa score is lower and more uncertain for **nurse/clinical** vs.	Implicit reviews have low inter-rater reliabilities. Doctors have the most reliable inter-rater reliabilities. Further research needs to explore the reason for the lower and more dispersed kappas for clinicians and non-clinical audit staff.

[12] A weighted mean kappa ascribes different weights to decisions made with ordinal scales such as here with reviewer care quality judgements. On a 5-point Likert scale for instance, two reviewers have given a score of 1 and 2 for the first case-note, respectively, and 2 and 5 for a second case-note respectively. It is evident the difference in weights is markedly larger for the second compared to the first case-note. This method is used to account for these ordinal level differences.

Health Technology Assessment 2010 Vol. 14 Issue 10 Pages 170 DOI: 10.3310/hta14100	nurse clinical staff case-note reviews with a weighted mean kappa 0.22 (0.08 to 0.36) **Non-clinical audit staff** vs non-clinical audit staff have a low (not the lowest) and most dispersed confidence intervals with a weighted mean kappa of 0.30 (–0.01to 0.61)	
Large scale organisational intervention to improve patient safety in four UK hospitals: Mixed method evaluation	*Implicit review*	Explicit reviews have higher agreement levels than implicit reviews.

A. Benning, M. Ghalib, A. Soukis, M. Dixon-Woods, J. Dawson, N. Barber, et al. BMJ 2011 Vol. 342 Issue 7793 Pages 369 DOI: 10.1136/bmj.d195	Agreement across two observers, with a κ of 0.71 and 0.70 across epochs 1 and 2 *Holistic review* Reviewers assessed 91 sets of case notes with low reliability between them (κ=0.15, SE 0.08)	
Manaseki-Holland, Semira, Lilford, Richard J., Te, An P., Chen, Yen-Fu, Gupta, Keshav K., Chilton, Peter J., Hofer, Timothy P. **Milbank Quarterly, Vol. 97, Issue 1, pp.228-284**	Twenty-three studies were included from 1985 to 2017. Recent larger studies suggest consistently low rates of preventable deaths (interquartile range of 3.0%-6.0% since 2008). Reliability of a single review for	The interquartile range of reliability reported for the ability of a single review to distinguish between cases with respect to whether death was preventable was 0.27 to 0.45. At a representative level of reliability of 0.35 for a single review, we can estimate that an average of 8 reviews per case note would be required to achieve a

	distinguishing between individual cases regarding the preventability of death had a Kappa statistic of 0.10-0.50 for deaths and 0.21-0.76 for adverse events. A Kappa of 0.35 would require an average of 8 to 17 reviews of a single case to be precise enough to have confidence in high stakes decisions to change care procedures or impose sanctions within a hospital as a result. No study estimated the variation in preventable deaths across hospitals, although we were able to reanalyse one study to obtain an estimate.	reliability of 0.8 when distinguishing between cases. Seventeen reviews per case would be required to achieve a reliability of 0.9, a level often recommended for testing with high-stakes consequences. If the reliability of a single review were as high as 0.5, then only 4 or 9 reviews per case note would be needed for a reliability of 0.8 or 0.9, respectively. However, any given operational program would have to determine the reliability of its measurement procedure in its population to figure out the number of cases needed to review.

	Based on this estimate, 200 to 300 total case note reviews per hospital could be required to reliably distinguish between hospitals. The studies displayed considerable heterogeneity: 13/23 studies defined preventable death with a threshold of greater than or equal to four in a six-category Likert scale and 11/24 involved a two-stage screening process with nurses at the first stage and physicians at the second. Fifteen studies provided expert clinical review support for reviewer disagreements, advice, and quality	

	control. A “generalist/internist” was the modal physician specialty for reviewers, and they received one to three days of generic tools orientation and case note review practice. Methods did not consider the influence of human or environmental factors.	

How are case-note reviews connected to quality improvement?

Case note review are a common means of retrospectively assessing quality of care,(Thomas et al., 2002, Wilson et al., 1995, Baker et al., 2004, Zegers et al., 2011, Hogan et al., 2012b, Hogan et al., 2015) despite the widely-acknowledged methodological and practical challenges of this review method.(Hutchinson et al., 2010c, Lilford et al., 2007, Hayward and Hofer, 2001) We must first be able to measure quality before we can improve it. Two main review methods are used: explicit criterion-based methods and implicit (sometimes called holistic) methods which focus on unstructured clinical judgement.

First, criterion-based methods, usually using frameworks of pre-determined criteria to identify elements of care which are either met or not met, are useful for large-scale audits of care or for screening case notes using criterion-based trigger tools.(Hutchinson et al., 2010d) Second, implicit review methods are based on clinical judgement, and are likely more effective for identifying and recording in more detail the nuances of care.(Mohammed et al., 2006) Thus, implicit review methods are likely more appropriate for in-depth exploration of the care for people who die in hospital. However, unstructured implicit review formats have been criticised for their low inter-rater reliability (high variability) and for potential reviewer bias, (Hutchinson et al., 2010c, Lilford et al., 2007, Hayward and Hofer, 2001, Hutchinson et al., 2010e) whereas structured implicit reviews limit this variability by creating specific frameworks to help reviewers to justify and organise statements on the care.(Kahn et al., 1989)

There have been recent studies which have found quality improvement contributions from case-note reviews.(Mendu et al., 2020, Kobewka et al., 2017) However, there are no systematic approaches which have been adopted across institutions that could form a body of

recommendations for how case-note reviews should be systematically used in practise. Next, there will be a review, in chronological order, of the two types of stances the UK government has taken for the use of case-note reviews

The summative use of case-note reviews: for and against

The initial argument to use case-note reviews to detect preventable deaths is because it could detect more accurately and reliably the extent of preventable deaths as the instrument was more sensitive to changing contextual factors of the medical records and the health system itself. The case-note reviews potential ability to detect the signal of concern.(Lilford and Pronovost, 2010b) However, it was then shown that the reliability of the reviewers was subject to variability (Goldman, 1992, MacKenzie et al., 1992, Goldman, 1994, Lilford et al., 2007, Hutchinson et al., 2010c, Benning et al., 2011), and also that any preventability of death estimates were subject to a wide range, unless a large number of case-notes were undertaken to establish estimate parameters (i.e. calibration review sets to obtain an accurate and reliable variance around the preventability judgement in any given healthcare context (e.g. hospital or Trust)). So, for case-note reviews to be fairly represented, there is the need to understand the degree of variation in the judgement of care quality to factor this into summative evaluations about the care in a hospital or Trust. Without this, the preventability estimates are not comparable and will certainly lead to the erroneous care quality classification of under-performing and hyper-performing institutions.

The formative use of case-note reviews: for and against

When healthcare institutions and policymakers became aware of the variability between different hospital preventable death estimates and the underlying variability found within the

care quality judgement process between reviewers and their own local context, the decision to adopt case-note reviews as a formative tool became much more appealing. This decision was made by the Department of Health, UK, (through the Health Quality Improvement Partnership (HQIP))(Health Quality Improvement Partnership (HQIP), 2015a) the National Quality Board[13](National Quality Board UK, 2017) and also the Care Quality Commission(Care Quality Commission, 2016) in separate policy releases and white papers which endorsed the use of case-note reviews for obtaining learning from in-hospital deaths. The joint drive has helped to embed the awareness that despite it being a mandatory process for each hospital Trust and the requirement to publicly publish the number of case-note reviews undertaken per fiscal quarter, there is no supra-hospital league table of any information obtained from these case-note reviews. The desire alone is for the Trust to develop its own learning and strategy in line with its own goals as an organisation given the Trust status and to be semi-autonomous from Department of Health input. It is the hospital and/or its specialties which will learn from these reviews. However, one of the arguments against the within-hospital use of case-note reviews is that the process may differ significantly within different hospitals at the same Trust. To counter this claim, training was provided with oversight from the Royal College of Physicians (RCP)(The Royal College of Physicians, 2016d, The Royal College of Physicians, 2016a) and the intellectual content delivered through the Yorkshire and Humber Academic Health Science Network (YHAHSN) to make better use of the structured-implicit review tool. One of the reasons against the formative use of case-note reviews is that the reviewing process is not actively monitored at a

[13] an independent panel of senior clinical experts, professional leaders and patient and public representatives who legislate for the championing quality improvement in healthcare

supra-hospital level (that is, the regional or national level) but the hospital level. This makes case-note reviewing an activity susceptible to the hospital work culture and other contextual factors (i.e., environment, work processes, board priorities) within the hospital; it is, as yet, unexplored how the case-note review process is affected by such factors and warrants further exploration in this thesis.

However, it must be noted that the structured implicit case-note review method was originally designed to measure the quality of care, not preventable death, and has much literature describing its use for such a purposes.(Kahn et al., 1989) Being perhaps informed by the limitations of inter-hospital use of case-note reviews, the UK's national policy on case-note reviewing has shifted from its emphasis on summative work to formative local hospital processes over the half decade or so. Case-note reviews are best used as a formative tool, that is it is there to inform and instruct local change rather than be used formally as a benchmarking tool. The UK has officially adopted this stance towards case-note review through the policy for the mandatory use of retrospective case-note reviews in hospitals for each Trust. This mandate requires that each Trust publicly reports the number of case-note reviews undertaken per fiscal quarter.(National Quality Board UK, 2017) This policy has cemented the use of case-note reviews for formative purposes across all UK hospitals.

The gradual shift from summative to formative assessment for case-note reviews

As case-note reviews became more popular in hospital administrations, they were used for summative purposes, which, in other words, was the formal use for benchmarking purposes over the recent decades. This benchmark use involved the measurement of preventable hospital deaths using structured-implicit case-note reviews. This is showcased in the

Department of Health tendering a bid for system for "*retrospective record review to estimate the proportion of hospital deaths due to problems in care in English acute NHS trusts*."(Health Quality Improvement Partnership (HQIP), 2015b) It is now known that the relatively low estimates for preventable death rates using case-note reviews entail that ranking hospitals using HSMR or SHMI would lead to a high-risk of misattribution and the mislabelling of outliers.(Manaseki-Holland et al., 2019a) This is the reason why mortality rates are no longer used as a proxy for preventable deaths due to the high variability around these estimates. It must be indicated that there was a strong desire from policymakers in the United Kingdom to employ case-note reviews as a formal, summative tool for comparative purposes, specifically to identify the true proportion of preventable deaths in hospitals. This is admirable, but the question remains as to whether this is feasible given the data and methods employed around it. This desire to use case-note reviews to measure preventable deaths in the UK is first found in a white paper *An Organisation with a Memory*, which estimated that 60,000–255,000 NHS patients suffer serious disability or death due to healthcare interventions.(Donaldson, 2002) These UK estimates themselves were derived from case note review studies conducted in the USA decades before.(Kohn et al., 2000, Leape et al., 1991, Thomas et al., 2000) It is important to note that these methods were not designed to measure the proportion of deaths that were preventable when it was first introduced.[14](Wilson et al., 1995, Davis et al., 2003, Vincent et al., 2001)

[14] These studies are later addressed in the systematic review in Chapter 1

The potential for case-note reviews as a care quality improvement tool

Case-note reviews are superior to hospital mortality statistics because they are not beset by issues of statistical mal-adjustment. I have discussed the potential of hospital mortality statistics and mortality case-note reviews as an accurate and reliable indicator for hospital care quality. There are clear weaknesses with hospital mortality statistics, and it is worthwhile to explore the extent to which case-note reviews can be used as a hospital care quality indicator. But as has been identified, there are some drawbacks with the reliability of case-note review and their judgement across qualified peers in Table 2. Thus, the rest of this thesis will investigate whether psychological and attitudinal factors contribute to the inter-reviewer care quality judgement variability in Chapter 2 and Chapter 3 . The penultimate chapter will identify and discuss the organisational dynamic of case-note review use for quality improvement in a UK hospital Trust (Chapter 4). Having presented the policy context for mortality reviews, I will consider how the thesis content relates to the existing literature on this topic. This will not neglect the assumptions of conceptual frameworks for intellectual trends evidently shape what is practicable also. I shall spend time to discuss the warrant for these conceptual frameworks in this thesis.

To recap, we shall compare the general strengths and weaknesses of hospital mortality statistics and case-note reviews. Comparing the strengths and weaknesses of these two methods can help highlight the most appropriate use of each method. Hospital mortality statistics are easily generated as it is routinely collected and is relatively easy to interpret if there is enough complete data. The data can be used for a variety of purposes for healthcare staff, reporting to central information repositories such as NHS Digital or to board members and other stakeholders (Table 3).

Table 3. The strengths and weaknesses of using hospital mortality statistics and case-note reviews for evaluating hospital care quality

	Hospital Mortality Statistics	**Case-note reviews**
Strengths	Easily generated	Detailed insights into a specific case
	Ease of interpretation	
	Easy manipulation	
	Account for propensity by collecting data on many key patient variables	
Weaknesses	Loss of important signal	Time-consuming
	Inaccurate statistics due to abundance of assumptions concerning the case-note context	Labour-intensive
	Loss of contextual features	Difficult to generalise learning
	Incorrect propensity scores due to unconsidered local noise from "lurking variables"	Dependent on good administrative and IT skills
		Requires localised calibration before use

Summary

Quality improvement in hospitals has a legacy in the United Kingdom. Over its two decades of employment, the hospital mortality statistic has been used to measure the care quality in hospitals. However, there are known methodological issues around this type of statistical modelling approach with its limitations likely undermining its usefulness for accurate and reliable care quality assessments requisite for hospital care quality improvement. In response to these insuperable issues of mortality statistics, case-note reviews have been expounded as an alternative approach to measure the quality of care in hospitals. Yet, case-note reviewers typically have significant levels of disagreement for the same case-note. To explore the potential of case-note reviews, it is useful and informative to better understand some of these sources of variability for case-note review care quality judgements. It is also not known how case-note reviews are used for quality improvement; this will be explored also. This thesis explores this in further depth.

Thesis format

This thesis is presented in accordance with the University of Birmingham Alternative Thesis Guidelines, which allow chapters formatted for submission to peer-reviewed journals and published articles to be included. Some of the work in this thesis has been prepared for submission to peer-reviewed journals (Chapter 2 , Chapter 3 , Chapter 4). In addition, part of the work presented in this thesis has been published or accepted for forthcoming publication in peer-reviewed journals (section 1.4.).

Please note that the page numbers of the publications will not be included in the pagination sequence of the thesis and that the referencing and numbering of tables and figures will be

self-contained within each chapter. Please also note that the inclusion of publication and publication-style chapters will result in duplication as each chapter will have self-contained components that overlap with parts of other thesis sections.

The publication chapter appendices (chapter 2-4) are found at the end of each chapter.

The full regulations for the University of Birmingham Alternative Thesis Guidelines can be found at:
http://www.birmingham.ac.uk/Documents/university/legal/regulations-part7.pdf
https://intranet.birmingham.ac.uk/as/studentservices/graduateschool/documents/public/rsa/alternative-format-thesis-guidelines.pdf

Thesis aims

This research aims to inform the selection of measurement methods for hospital care quality in the UK and high-income countries with self-sustaining healthcare systems. There is the specific aim of separately exploring the validity and use of mortality case-note reviews for hospital care quality improvement. This research has identified some limitations in existing approaches for measuring the quality of care in hospitals using only hospital statistics. Alternatives need to be explored and one of these is retrospective mortality case-note review with the intention of reliably measuring individual patient and hospital-wide care quality. The research aim was carried out through quantitative, qualitative, and psychological research methods via the research objectives. Each analytical chapter examined a separate research objective.

Research Objectives

1. Explore the measurement characteristics of preventable hospital deaths in high-income countries (Chapter 1)
2. Examine the plausible influence of cognitive biases and heuristics on case-note reviewer care quality judgements (Chapter 2)
3. Identify the influence of case-note reviewer demographics and select attitudinal measures and case-note features upon case-note reviewer care quality judgements and their care quality confidence judgements (Chapter 3)
4. Examine the potential quality improvement potential and any mechanisms from the use of case-note reviews in hospitals (Chapter 4)
5. Discuss the research process and limitations (Chapter 5)
6. Implications for stakeholder groups (Chapter 6)

The conceptual frameworks, theories, analytical and statistical methods employed in this thesis

In this section I discuss the assumptions of conceptual frameworks which are likely to help us understand how mortality case-note reviews can *and* do contribute to hospital care quality improvement. I will expand on what each framework is and how each framework contributes to the overall thesis. Note, that these concepts will be discussed in greater detail in their own respective chapters, but for now I offer a cursory introduction for their relevance for this thesis. Hereon in, I employ the following conceptual frameworks in this thesis for this purpose:

Critical realism

Critical realism, as an epistemology and ontological framework, holds that the social phenomena is not directly observable.(Archer et al., 2013) Thus, any social interactions around case-note review and its quality improvement process within a hospital and Trust setting will likely remain hidden. And so, an appropriate investigative method needs to be chosen which respects the "hidden" non-trivial nature of this social and physical reality of case-note reviews in their use for hospital care quality improvement.(Fletcher, 2017) The epistemological and ontological frameworks help the researcher to develop a more in-depth understanding of the reviewing phenomenon through multiple methods and approaches. Critical realism does not assume there is one approach to understanding the constructed reality, but that reality is constructed and can be taken from different angles and different levels.[15] This brings the benefit of ensuring that we come to understand the phenomenon at different levels so that the research findings may help deliver a higher likelihood of practical benefit. This is the main benefit of tackling this topic from multiple perspectives.

---

[15] Levels refer to whether it is concerned with the cognitive or the organisation cultural. On the whole, these are the two different levels considered in this thesis which accords to the most informative levels of abstraction. But this is not necessarily the only levels which can be considered for this research. I consider the neural - that is within the individual -, national and supra-national levels may have some bearing upon case-note reviews but is not directly addressed in this thesis for obvious reasons of being beyond the thesis scope and research granularity.

Case study methodology

Case study methodology is an investigative framework which respects the context of the investigation and preserves the analytic generalizability to other settings through analytic, generalizable theory.(Yin, 2010, Yin, 1981, Yin, 2017) It is the appropriate methodology for a critical realist stance because it respects the multiple layers of social and physical phenomena within a complex organization, such as a hospital. The case study method preserves well the context of study and can investigate the hidden interactions assumed to occur through the critical realist stance. The main benefit of conducting a case study is that the quality improvement context is preserved so that the actual barriers and facilitators in play can be highlighted and identified.(Baker, 2011a) It concerns the social and cultural dynamics of one context and seeks to develop generalizable theory for both similar and dissimilar settings.(Yin, 1981)

Dual process theory

Dual process theory is a theory of cognition which espouses that there are two broad and universal human cognitive systems which are contrastingly typified by one through slow, effortful and deliberate thought and the other which is rapid and intuitive cognition.(Tversky and Kahneman, 1975, Kahneman, 2011, Evans, 2008) This theory is relevant because reviewers judging the case-note review's care quality must come to a cognitive judgement about the quality of care. It is plausibly anticipated that these cognitive and biases have a role to play in the reviewer's way of thinking. The flawed human mind, which includes the reviewer's mind, is at the heart of any care quality decision and judgement about the case-notes. All kinds of decisions and considerations must be deliberated on by the case-note

reviewer, and it is expected that the reviewer must employ heuristics. In just thinking about the case-notes, they are likely to commit an error in their judgement. This theory is explored further in Chapter 2 where I assess in further detail its plausible influence upon case-note reviewer care quality judgements. This dual process theory has the benefit of helping researchers develop a more nuanced understanding of the clinical reviewer's decision-making during case-note reviewer, which can help highlight influencers and sticking-points for those who are affected by these reviews. It would be a novel view to not only study the review method and tool, but also the reviewer themselves. The focus on the reviewer will be sure to shed some more insights into the psychology of case-note reviewing.

Having identified that there are concerns around the high variability between case-note reviewer judgements, one began to explore the academic literature. The cognitive bias readings ranged from seminal literature from Daniel Kahneman and Amos Tverksy(Kahneman et al., 1982, Kahneman and Tversky, 1984, Kahneman, 2011) to the latest reference works (Koehler and Harvey, 2008) and popular science releases, which have garnered a wide readership worldwide and brought the concept to worldwide acclaim. (Kahneman and Klein, 2009, Marewski and Gigerenzer, 2012a, Marewski and Gigerenzer, 2012b, Gilovich, 2008, LeGault, 2006, LeDoux, 1998, Casti, 2010, Gigerenzer, 2008b, Groopman and Prichard, 2007) This list is not exhaustive but illustrative of the type of literature I engaged with. The first stage of the review process involved the systematic search, screening and appraisal of articles sourced from online databases, which were informed by the search terms employed by two existing systematic reviews exploring the influence of cognitive biases and heuristics in clinical and/or medical decision-making.(Saposnik et al., 2016, Blumenthal-Barby and Krieger, 2015a) One limitation of these two systematic reviews was that the search terms may not have completely captured the relevant studies available at

the time. However, this is unlikely given the major cognitive biases and heuristics were mentioned and cited across these two reviews as judged by two authoritative lists.(Wikipedia, 2020, Benson, 2016)

Normalization Process theory

As case-note review is a complex intervention and needs to be embedded in a setting for it to make any impact in its context. A tool cannot expect to be used unless it has found a place among the constellation of quality improvement interventions. And so, normalization process theory studies the embedding of complex interventions. As embedding of an intervention, such as case-note review, is the first requirement for their acceptance and its subsequent use, this theory shaped which interview questions were used and how the interviews were conducted. Case-note review embedding is a *sine qua non* for information flow and, our outcome of interest, which is whether there is any quality improvement obtained from case-note reviews.(May and Finch, 2009) The Normalization Process Theory provides the benefit of seeing the case-note review process as an intervention which must be first embedded before it can be used. This fundamental assumption of the theory helps me to present a more realistic account of the pressures faced when implementing the case-note review programme in one Trust, over simply assuming it has been implemented to some degree or other. The theory's realism is its strength as it helps build a cogent and more representative body of evidence from the interviews and documents collected from the case-study and how the nature of reality is directly coupled with the social world of reviewers and hospital managers.

Framework method

Framework method is a common approach to manage the analysis of qualitative data in health research. It comprises multiple steps in this sequence: transcription, familiarization with the interview, coding of interview scripts, development of a working analytical framework, application of the analytical framework, charting data into the framework matrix and interpreting the data.(Gale et al., 2013) This is a common and useful method proven to identify key themes and manage these through an analytical framework. The benefit of this method is that it ensures that the main themes will be identified.

Multi-level modelling

This type of model was used to respect the nesting within the dataset. Nesting is also called clustering, and it is an undesirable analytical feature because it leads to the underestimation of the standard errors. A multi-level model can account for these effects in a statistical model which better represent these dependencies with the hierarchical structure in the data. The difference between a single level regression model and a multi-level model is that the latter separates the variance components at the individual and the group level. The benefit of this approach is that it will more accurately represent the true effect size of any nesting effects, thus rendering the calculation of standard errors and subsequently the 95% confidence interval more accurately.

Documentary analysis

I employ documentary analysis because I had documents to collect from the case study. Documents can play a significant role in providing information from a case study context and

so it was important to have an analytical approach to collecting and making sense of them. I specifically considered using triangulation approaches for interview and document data.(Farmer et al., 2006) The documentary analysis benefits this thesis by adding more breadth to the data to consider all documents at all levels of the organization pertinent to case-note reviews. This helps to ensure that the widest range of evidence is considered leading us to the least biased and most representative findings from the given context.

Data science

Data science is the practice of analyzing large volumes of often unstructured data, with the goal of producing meaningful insights. It is an applied branch of statistics to help make real-world impact from the hidden patterns latent in large and noisy datasets. Akin to the statistics behind multi-level modelling, data science is a set of wide-ranging approaches, algorithms, and tools that can be applied to most business problems. This includes applying it to the problem of learning from case-notes and the reviewing process. As a case-note is essentially a data repository about the patient, information can be extracted and mined for insights. Considering the problem in this way gives data science approaches the possibility of developing practical solutions. Though this thesis does not directly employ data science methods, the discussion section will look to engage with this in greater detail.

Narrative mining

Much of the data available in mortality case-note are unstructured, making it challenging for clinicians, healthcare analysts, coders and informaticians to apply their tools in a systematic and unbiased manner to this information. As a sub-branch of data science, narrative mining methods can manipulate, summarize, and visualize characteristics of the text in case-notes with the potential to help case-note reviewers make better judgements in the future, through

the aid of automatic data extraction from the scanned documents. These are explored in the discussion section of this thesis in Chapter 6, section 6.4.

How has the mixed methods approach contributed to the thesis?

The approach to undertake a mixed methods thesis was had the purpose of illuminating the reviewer-specific factors and the organisational factors which were likely to contribute to the low-to-moderate inter-rater reliability between case-note reviewers and the QI from mortality case-note reviewing, respectively.

First, there is the individual case-note reviewer to consider. In part one of looking at the variability within the reviewer, I looked to the field of cognitive biases and heuristics, which found a plausible list of fifteen systematically searched for in the literature and identified by our consensus panel. Though we do not present any empirical findings on any single bias/heuristic, the sweeping identification of this list and the mechanisms, as illustrated in the clinical scenarios, will help a wider audience understand the potential relevance and context of our research. The cognitive bias and heuristic literature bring together studies from the medical literature and generates new connections by linking them with case-note review care quality judgements through a consensus panel. From this there has been a conceptual contribution in understanding the individual reviewer. In psychological terms, there has also been a deeper understanding of the case-note reviewer.

As part two of the reviewer-specific contribution to the care quality judgement, our attitudinal measures are not strongly correlated with changing care quality judgements scores; however, we did identify that about one fifth of the variability in reviewer care quality judgements are attributable to the reviewer. And compellingly, this attributable fifth is not accounted for by any of our independent variables. In sum, the reviewer is suggested to be

influenced by key biases and heuristics in certain ways, and that attitudes, which are unrelated to these biases, were another way which influenced the reviewer. However, this was shown to not obtain for the three attitudinal measures we considered. This thesis has not identified a positive finding but has found independently that biases and heuristics may be responsible for care quality judgement variability, and the need for cognition, personal need for structure and anxiety due to uncertainty are not shown to significantly influence care quality judgements. It is the case perhaps that attitudes serve a lesser role in their influence of care quality judgements, or it could be that other less obvious attitudinal measures are stronger influencers of care quality score. There is now the possibility of developing further empirical work across the priority cognitive biases and heuristics from the list of 15 to help explain this reviewer-specific variability in the care quality judgement. In sum, the systematic-review informed conceptual study coupled with the multi-level model identifying an absence of attitudinal effects but a reviewer-specific variability portion both helps to lay down new knowledge about the reviewer contribution to care quality judgements.

Second, qualitative research persuades through rich description and strategic comparison between and/or across cases, thereby overcoming the "abstraction inherent in quantitative studies" and permitting generalization to theory.(Yin, 2017) Qualitative research usually answers research questions that address "how" and "why" whereas quantitative research typically addresses "how often" and "how many". These suggestions indicate that a mix of quantitative and qualitative methods can be fruitful for obtaining profoundly new empirical insights, especially of complex phenomena. As quantitative methods need valid conceptual grounding, qualitative methods are probably always a necessity to understand social phenomena.(Malina et al., 2011)

Thirdly, the advent of data science and its methods has gained significant proliferation in all fields and industries today. Given the glut of data in today's modern healthcare system

and the need for data-driven decision-making from healthcare executives and patients, there is a need to employ the methods of this modern age to discover the patterns behind the data. Thus, there is a warrant for taking case-notes into the modern era by scanning and mining this scanned information. This is explored and expanded on in the final discussion chapter in section 6.4. Though data science methods have not been employed in this thesis, it was the thesis remit to not mine the information but investigate the rigour and actual use of case-note reviews in greater depth and with greater temporal mapping. The use of data science may open doors to explore both the rigour and use of case-notes themselves for more fluid and accessible data from the case-notes.

It is hoped that this methodologically flexible approach will foster scholarly interaction on this topic. By contextualising the research better, mixed method approaches mirror more the way case-note reviewers interact in their usual hospital setting. Specifically, the reviewer's mind is very important considering our focus is to understand the care quality judgement better and factors affecting it; this was coupled with qualitative data to provide a more complete picture of its use for quality improvement. Naturally, reviewer-specific components are important and QI, if it is instantiated, must arise at the organisational level, A qualitative case study could capture well any QI better than a quantitative study could. The individual reviewer is nested within the organisation. But there is a need to have the first to elaborate more deeply the role of the reviewer, before then capturing the reviewer's wider context within the organisation and its QI from this method. These three-prongs, the conceptual, reviewer-specific, and qualitative, when considered together has provided more impact and insights than would have been possible through drilling down into only one of these methods alone.

Thesis content and structure

This first chapter introduces measures of care quality used for hospital quality improvement purposes and information around their strengths and their limitations. This includes the moderate-to-high inter-reviewer reliability across the same case-note review. As the case-note is a document, its interpretation is reliant upon several case-note and reviewer features. These include the quality of the case-note, the administration process, inadvertent clinician biases, and the reviewer's attitudes. Two chapters, Chapter 2 (conceptual review and study) and Chapter 3 (attitudinal measures), discuss and test these features to see if they could explain this moderate-to-high inter-reviewer reliability. I investigate how case-note review is used in a UK NHS hospital Trust and explore whether any quality improvement obtained from the process. In Chapter 4 , I investigate through a case-study design the facilitators and barriers for any quality improvement from case-note reviews with the embedding and information flow as secondary research questions. In Chapter 5, I reflect on the research process. In Chapter 6 , I summarise the findings and discuss the implications for future case-note review use as a hospital care quality improvement tool, and where reasonable, I suggest areas of future research to clarify gaps in knowledge.

THESIS OVERVIEW

This thesis is organised as follows, there are six chapters – an introduction to the policy and methodological context, three analytical chapters, a summary of studies, a reflections chapter, and a concluding discussion chapter. Broadly speaking, the thesis addresses two parts of case-note reviews for care quality improvement. Firstly, Chapter 2 and Chapter 3 address the validity of case-note review care quality judgments. Secondly, Chapter 4 addresses the use of case-note reviews in hospitals for care quality improvement. Each analytical chapter is

presented in a publication format with appropriate bookends. A listing of the chapters is included in this section as:

Introduction

The measurement of hospital care quality has taken several forms over the last few decades. Since the year 2000, hospital mortality statistics have been used for this purpose. Historically, case-note reviews have also been used for this purpose but now competes with the ease-of-implementation and superior efficiency of hospital mortality statistics. Case-note reviews are used in hospitals across specialties, within hospital groups and external reviewing groups to investigate post-mortality case-note of patients and review these for the quality of the care received. As a potential quality improvement tool, mortality case-note reviews have been mired in some contention concerning the meaningfulness of its contribution to quality improvement. Mortality case-note reviews, specifically referring to the United Kingdom's national mandate for exploratory case-note reviews of in-hospital patient deaths, have more recently been adopted to obtain lessons for quality improvement. This chapter presents information which find that hospital mortality statistics are susceptible to important methodological weaknesses. On the other hand, case-note reviews have a better chance of identifying the true care quality "signal." This chapter lays the policy context for forthcoming chapters. **In the introduction, I discuss how care quality is measured and possible implications for its use for care quality improvement in the United Kingdom.**

Chapter 1: The measurement of preventable hospital deaths, care quality and their associated problems

The systematic review found that mortality rates in hospitals are commonly used in high-income countries to measure preventable death rates. There are limitations to the use of

hospital mortality rates as a proxy for preventable deaths. This systematic review was published during the PhD research. The systematic review gives cautionary wind to the measurement characteristics involved in the measurement of the true quality of care using hospital mortality statistics and case-note reviews. The rest of this chapter has been included in the form of a published peer-reviewed systematic review article. **In Chapter 1, a systematic review is presented on the measurement of care quality using preventable deaths and the measurement requirements for any reliable and accurate measure of care quality and the potential of case-note review for this purpose.**

Chapter 2: A conceptual review and expert panel study of cognitive biases and heuristics plausibly influencing global care quality judgements during mortality case-note review (on the validity of case-note reviews)

As with any human judgement, they are susceptible to bias. Care quality judgements of case-note reviewer are no different. This chapter explores the plausible influence of cognitive biases and heuristics, which are irrational errors in human decision-making and simple strategies and personal knowledge biases on these care quality judgements. For example, the reviewer knows that the consultant gastroenterologist who oversaw the care for Mr Smith's case-note is famous for his delicate bedside manner and compassionate care. Thus, case-note reviews care quality judgements are susceptible to such cognitive biases and heuristics, which have been demonstrated to influence humans across a variety of settings. **In chapter 2, I discuss the plausible influence of cognitive biases and heuristics on case-note reviewer global care quality judgements.**

Chapter 3 : A multi-level model investigating the influence of clinician case-note reviewer attitudes, their demographics, and the patient case-note characteristics upon global care

quality judgements (on the validity of case-note reviews). A variety of reviewer-specific characteristics, attitudes and demographics plausibly influence the global care quality judgements of reviewer global care quality judgements. It is with the High-intensity Specialist-led Acute Care (HiSLAC), where I investigate the influence of attitudes, specifically using three attitudinal measures (Need for Cognition, Personal need for Structure, Anxiety due to Uncertainty), demographics and case-note review (patient) characteristics (collectively termed 'individual differences') upon *global care quality judgements* and *confidence in global care quality judgements* using a multi-level model. **In chapter 3, I investigate the influence of reviewer attitudes, their demographics and the patient case-note characteristics upon case-note reviewer care quality judgements and their confidence in these judgements.**

Chapter 4 : Using case-note review for quality improvement: an NHS Hospital Trust case study (on the use of case-note reviews)

In a case-study, I explore whether mortality case-note reviews contribute to hospital quality improvement. For this, a qualitative case-study was undertaken which involved more than twenty interviews and the collection of on-site documentary materials. As case-note reviews have been mandated in the UK for learning purposes, it is important to know the extent to which case-note reviews are embedded in the hospital, as any intervention needs to be embedded before it can be usefully employed. I also identify barriers and facilitators to case-note review derived information flow. I also investigate the perceived case-note review derived hospital quality improvement. **In Chapter 4, I describe the barriers and facilitators for quality improvement from a mortality case-note review in a case study of an NHS case study in addition to the information flow and the embedding of these reviews.**

Chapter 5: Study Reflections

In this reflective chapter, I present my impressions of the research process and the reasonableness of some methodological assumptions. I reflect on all aspects of this process and the methods involved. **In Chapter 5, I discuss the research process and the methodological assumptions and limitations of the thesis.**

Chapter 6 : Discussion

In this final chapter, I summarise the thesis content thus far. I relay recommendations to stakeholders based on the thesis content. With any enquiry, there likely remain gaps in knowledge which present opportunities for further research around the digitisation and narrative mining of the data from mortality case-note reviews. **In Chapter 6, I summarise the thesis, give stakeholder recommendations and discuss future opportunities for research.**

CHAPTER 1: THE MEASUREMENT OF PREVENTABLE HOSPITAL DEATHS, CARE QUALITY AND THEIR ASSOCIATED PROBLEMS

1.1. Chapter Preface

Whilst undertaking this PhD, I was involved in a systematic review to identify studies which used hospital mortality statistics, such as standardized mortality ratios, to measure hospitals preventable death rates and their use to inform the ranking of hospitals. This systematic review seeks to identify the true level of hospital preventable mortality rates having considered the variation between hospitals in each of the studies. This review seeks to identify whether such studies account for inter-hospital variation in estimations of the preventable hospital death rates. This systematic review is important for this PhD as both are intent on identifying accurate and reliable methods for the measurement of hospital care quality. The use of mortality statistics is truly only as good as the data behind them. If the data are imprecise, which is the case much of the time due to the variability and noise in data collection methods across time and inability to control for all confounding factors, the use of mortality statistics for ranking and assessment of hospitals is very likely over-optimistic which could lead to over-interpretation of the data with further unintended consequences for the patients and their care providers. However, the position of hospital mortality statistics have recently been recommended as a smokescreen tool for measuring the quality of care across healthcare organisations.(Manaseki-Holland et al., 2019b) This published manuscript is enclosed in Chapter 1.

My specific contributions to the review included updating the search strategy from the initial systematic review search and collection of data. From the new terms obtained through dialogue with our team and the authors of the studies from the review's previously studies, I updated the search strategy to reflect any shift in the terminology since the initial search. I ran

the search and pooled the additional studies by applying the inclusion criteria and exclusion criteria; if there were any difficulties, I consulted with a senior academic on the team for their input. I extracted the data from studies into a tabular form. I corresponded with study authors via email when there was missing information or clarifications needed. I contributed to the early drafts of the paper and was on hand to provide clarifications for senior academics if additional information was needed.

This Chapter presents the findings of the systematic review paper that has been published in the Milbank Quarterly (2-year impact factor: 4.195, 5-year impact factor: 7.290) as:

Manaseki-holland, S., Lilford, R.J., Te, A.P., Chen, Y.F., Gupta, K.K., Chilton, P.J. and Hofer, T.P., 2019. Ranking hospitals based on preventable Hospital death rates: a systematic review with implications for both direct measurement and indirect measurement through standardized mortality rates. The Milbank Quarterly, 97(1), pp.228-284. https://doi.org/10.1111/1468-0009.12375

1.2. Ranking Hospitals Based on Preventable Hospital Death Rates: A Systematic Review With Implications for Both Direct Measurement and Indirect Measurement Through Standardized Mortality Rates

THE MILBANK QUARTERLY
A MULTIDISCIPLINARY JOURNAL OF POPULATION HEALTH AND HEALTH POLICY

Original Scholarship

Ranking Hospitals Based on Preventable Hospital Death Rates: A Systematic Review With Implications for Both Direct Measurement and Indirect Measurement Through Standardized Mortality Rates

SEMIRA MANASEKI-HOLLAND,[*] RICHARD J. LILFORD,[†] AN P. TE,[*] YEN-FU CHEN,[†] KESHAV K. GUPTA,[‡] PETER J. CHILTON,[§] and TIMOTHY P. HOFER[‖]

[*]*University of Birmingham;* [†]*Warwick Medical School, University of Warwick;* [‡]*Sandwell and West Birmingham Hospitals, NHS Trust;* [§]*Warwick Business School, University of Warwick;* [‖]*Institute for Healthcare Policy & Innovation, University of Michigan*

Policy Points:

- The use of standardized mortality rates (SMRs) to profile hospitals presumes differences in preventable deaths, and at least one health system has suggested measuring preventable death rates of hospitals for comparison across time or in league tables. The influence of reliability on the optimal review number per case note or hospital for such a program has not been explored.
- Estimates for preventable death rates using implicit case note reviews by clinicians are quite low, suggesting that SMRs will not work well to rank hospitals, and any misspecification of the risk-adjustment models will produce a high risk of mislabelling outliers.

The Milbank Quarterly, Vol. 97, No. 1, 2019 (pp. 228-284)

- Most studies achieve only fair to moderate reliability of the direct assessment of whether a death is preventable, and thus it is likely that substantial numbers of reviews of deaths would be required to distinguish preventable from nonpreventable deaths as part of learning from individual cases, or for profiling hospitals.
- Furthermore, population- and hospital system–specific data on the variation in preventable deaths or adverse events across the hospitals and providers to be compared are required in order to design a measurement procedure and the number of reviews needed to distinguish between the patients or hospitals.

Context: There is interest in monitoring avoidable or preventable deaths measured directly or indirectly through standardized mortality rates (SMRs). While there have been numerous studies in recent years on adverse events, including preventable deaths, using implicit case note reviews by clinicians, no systematic reviews have aimed to summarize the estimates or the variations in methodologies used to derive these estimates. We reviewed studies that use implicit case note reviews to estimate the range of preventable death rates observed, the measurement characteristics of those estimates, and the measurement procedures used to generate them. We comment on the implications for monitoring SMRs and illustrate a way to calculate the number of reviews needed to establish a reliable estimate of the preventability of one death or the hospital preventable death rate.

Methods: We conducted a systematic review of the literature supplemented by a reanalysis of authors' previously published and unpublished data and measurement design calculations. We conducted initial searches in PubMed, MEDLINE (OvidSP), and ISI Web of Knowledge in June 2010 and updated them in June 2012 and December 2017. Eligibility criteria included studies of hospital-wide admissions from general and acute medical wards where preventable death rates are provided or can be estimated and that can provide interobserver variations.

Findings: Twenty-three studies were included from 1985 to 2017. Recent larger studies suggest consistently low rates of preventable deaths (interquartile range of 3.0%-6.0% since 2008). Reliability of a single review for distinguishing between individual cases with regard to the preventability of death had a Kappa statistic of 0.10-0.50 for deaths and 0.21-0.76 for adverse events. A Kappa of 0.35 would require an average of 8 to 17 reviews of a single case to be precise enough to have confidence in high-stakes decisions to change care procedures or impose sanctions within a hospital as a result. No study estimated the variation in preventable deaths across hospitals, although we were able to reanalyze one study to obtain an estimate. Based on this estimate, 200 to 300

total case note reviews per hospital could be required to reliably distinguish between hospitals. The studies displayed considerable heterogeneity: 13/23 studies defined preventable death with a threshold of greater than or equal to four in a six-category Likert scale and 11/24 involved a two-stage screening process with nurses at the first stage and physicians at the second. Fifteen studies provided expert clinical review support for reviewer disagreements, advice, and quality control. A "generalist/internist" was the modal physician specialty for reviewers and they received one to three days of generic tools orientation and case note review practice. Methods did not consider the influence of human or environmental factors.

Conclusions: The literature provides limited information about the measurement characteristics of preventable deaths, suggesting that substantial numbers of reviews may be needed to create reliable estimates of preventable deaths at the individual or hospital level. Any operational program would require population-specific estimates of reliability. Preventable death rates are low, which is likely to make it difficult to use SMRs based on all deaths to validly profile hospitals. The literature provides little information to guide improvements in the measurement procedures.

Keywords: avoidable, preventable, hospital deaths, hospital mortality, systematic review, variation.

Standardized mortality rates (SMRs) for hospitals are currently used as an indicator of institutional quality and to compare hospitals in order to identify outliers.[1] The rationale for their use is that they are a proxy for excess or preventable deaths, but there are compelling arguments that any signal (preventable death) will be obscured by the noise (all other unavoidable deaths).[2,3] Some policymakers are considering using direct measurements of preventable mortality, rather than trying to infer it indirectly from SMRs, as with the summary hospital-level mortality indicator (SHMI) used in the NHS in England.[4-7] For example, the NHS in England has instituted a system of mandatory physician retrospective case record review (RCRR) of deaths in hospitals in order to establish (and publish) the number of preventable deaths for local trust use and to learn from mistakes.[8,9] A direct measurement of preventable death is also an obvious way to validate the widespread use of SMRs to measure the quality of care delivered to people prior to their death.

However, preventable death, as well as preventable adverse events (AEs) more broadly, can only be directly measured by the judgment of expert clinical observers who retrospectively review case notes.

Although no systematic review has been done for preventable deaths, such judgment-based assessments have generally reported low reliability, meaning that they lack consistency across repeated reviews. Thus, current and future policy and research agendas that propose measuring any preventable AEs, and specifically preventable mortality, should push us to define, and if possible, improve the measurement characteristics of those estimates. Only then can we use case note review measurements in research to validate SMRs, to design operational systems for learning from AEs within hospitals, and to compare preventable deaths between hospitals, possibly augmenting or even replacing comparisons by means of SMRs.

To this end, we conducted a systematic review firstly to summarize data from existing studies reporting avoidable deaths and the measurement characteristics of those estimates and applied these in order to determine the number of reviews needed to establish a reliable preventable death estimate at the individual or hospital level. Secondly, we summarize the heterogeneity between the measurement procedures used in these studies, including reviewer characteristics, selection, and training factors, to assess whether there are potential opportunities to improve the reliability of the measurement procedure. This is the first review of methods to measure preventable mortality rates.

Methods

Literature Search

We conducted an initial search in PubMed and ISI Web of Knowledge in 2010. We updated and supplemented this in June 2012 and December 2017 with a broader search in MEDLINE (OvidSP), incorporating a wider range of terms covering preventability and errors, deaths and AEs, hospitals, and case note reviews (Online Appendix 1). After our last search and before finalizing this manuscript, we were made aware of two studies that met our inclusion criteria.[10-12] These studies are included in our review to ensure that our findings remain up-to-date. Reference lists of included studies were also hand searched to find additional articles.

Study Selection

The inclusion criteria were studies that (a) evaluated the preventability of hospital deaths (deaths primarily from general and acute medical

wards) or preventable AEs contributing to death from a hospital-wide sample or primarily from general and acute medical wards; (b) provided a quantitative estimate of preventability of death or allowed this to be calculated; and (c) incorporated retrospective case record review that elicits the reviewer's own expert judgment in reaching the conclusion about preventability. Only articles published in English were considered. Two reviewers (Gupta, Chilton, or Te) independently examined titles and abstracts retrieved from literature searches and selected studies for inclusion. Disagreements were resolved by consensus after retrieval of full-text articles and further discussions with a third reviewer (Chen). The review protocol was not submitted to PROSPERO as the review process was initiated before the establishment of PROSPERO.

Data Extraction and Synthesis of Evidence

Two reviewers (Gupta, Chen, Chilton, or Te) extracted data from the selected studies, including all data tabulated in Tables 1-3. The characteristics and findings of included studies were tabulated and summarized in a narrative form. We did not plan to pool results across studies given the underlying differences in settings and methods between the studies. Where data were missing, we wrote to the study authors and obtained details.

Number of Reviewers Required for a Reliable Measurement

Reliability describes the consistency of measurement and can be used to quantify the ability to distinguish between the objects of measurement. Reliability ranges from zero to one and increases with a measurement procedure that makes multiple independent measurements and averages them. Most reports of the reliability of case note review give a number that describes the ability of a single review of any one case note to distinguish between a preventable and a nonpreventable death. In Online Appendix 2, we describe one method that makes use of equations that allow you to calculate how reliability improves as the number of measurements is increased.

Table 1. Characteristics of Included Studies and Methods Used for Assessing the Preventability of Deaths or Adverse Events (AEs)

Author	Location; Date of Study	Target Group/ Type of Hospital	Grading of Preventability[a]	Threshold for Defining a Preventable Case	Kappa (ICC) for Preventability	Interhospital Variance/ ICC	Comments
Dubois et al, 1987; 1988[26-28]	United States; 1985	12 private hospitals	1-4[b]	≥3[b] Death as "probably preventable"	$\kappa = 0.4$, 0.3 and 0.2[c] preventability of **death** (182 charts, each reviewed by three physicians)	Not reported	• Hospital-wide medical wards with conditions specific to cerebrovascular accident, pneumonia and myocardial infarction • Acute care hospitals that were considered outliers with higher and lower than expected mortality • Preventable mortality estimated from data • 14% of deaths (of all deaths) were preventable

Continued

Table 1. *Continued*

Author	Location; Date of Study	Target Group/ Type of Hospital	Grading of Preventability[a]	Threshold for Defining a Preventable Case	Kappa (ICC) for Preventability	Interhospital Variance/ ICC	Comments
Brennan et al, 1991[22]	New York, United States; 1984	51 private and nonfederal acute care hospitals	1-6	≥ 4 negligence is more likely than not	$\kappa = 0.24$ / preventability of **AE** (based on duplicated review of 318 cases (2/51 hospitals)	Not reported	• Hospital-wide, excluding psychiatric patients • Nonfederal, acute care hospitals • Preventable mortality estimated from data • Weighted figures based on events discovered during index hospitalization only • 13.6% of patients with AEs died

Continued

Table 1. *Continued*

Author	Location; Date of Study	Target Group/ Type of Hospital	Grading of Preventability[a]	Threshold for Defining a Preventable Case	Kappa (ICC) for Preventability	Interhospital Variance/ ICC	Comments
Hayward et al, 1993[10]	United States; 1988-1990	1 teaching hospital	1-6	≥ 5 better quality care could have prevented the death	$\kappa = 0.5$ Death preventable by better quality of care (based on dual reviews of 79 deaths)	N/A (Insufficient denominator)	• Hospital-wide medical wards with no single diagnostic-related group contributing ≥ 5% of patient admissions • Acute care university teaching hospital • 9% of patient deaths preventable

Continued

Table 1. *Continued*

Author	Location; Date of Study	Target Group/ Type of Hospital	Grading of Preventability[a]	Threshold for Defining a Preventable Case	Kappa (ICC) for Preventability	Interhospital Variance/ ICC	Comments
Best and Cowper, 1994[21]	United States; 1986	16 Veterans Affairs Medical Centers	1-4	≥ 3 Somewhat likely that better management in the hospital might have prevented patient's death	$\kappa = 0.33$ "agreement = ≤ 2 positions on 9-point scale" (111 match-pairs from high and low mortality risk Veterans Affairs Medical Centers)	Not reported	• Veterans Affairs Medical Centers (small, med/large and psychiatric/long-term types) • 21.6% of patients with better care management might have prevented death (or near the time of death)

Continued

Table 1. *Continued*

Author	Location; Date of Study	Target Group/ Type of Hospital	Grading of Preventability[a]	Threshold for Defining a Preventable Case	Kappa (ICC) for Preventability	Interhospital Variance/ ICC	Comments
Wilson et al, 1995[43]	New South Wales and South Australia; 1992	28 private and public acute care hospitals	1-6	≥ 4 "Preventability more likely than not, more than 50/50 but close call"	$\kappa = 0.33$ preventability of **AE** (based on duplicated review of 6,200 cases [all cases positive for screening criteria])	Not reported	• Hospital-wide excluding day-only admissions and admissions to psychiatric wards • Preventable AEs and preventable mortality estimated from data • 4.9% of patients with AEs died

Continued

Table 1. *Continued*

Author	Location; Date of Study	Target Group/ Type of Hospital	Grading of Preventability[a]	Threshold for Defining a Preventable Case	Kappa (ICC) for Preventability	Interhospital Variance/ ICC	Comments
Thomas et al, 1999; 2000a; 2000b; 2002[39-42]	Utah and Colorado, United States; 1992	28 private and public hospitals	1-6	≥ 4 "More likely than not, > 50:50 but close call"	$\kappa = 0.19$ to 0.23 (95% CI, 0.05 to 0.37) preventability of **AE** (based on 3 independent reviews of 500 records)	Not reported	• Hospital-wide (13 in Utah and 15 in Colorado), excluding psychiatric and veterans hospitals and patients < 16 • Number of patients with AEs not specified, only total number of AEs • Based on events discovered during index hospitalization only • 6.6% of patients with AEs died

Continued

Table 1. *Continued*

Author	Location; Date of Study	Target Group/ Type of Hospital	Grading of Preventability[a]	Threshold for Defining a Preventable Case	Kappa (ICC) for Preventability	Interhospital Variance/ ICC	Comments
Hayward and Hofer, 2001[31]	United States; 1994-1995	7 Veterans Affairs hospitals	1-5[d]	≥ 4[d] "probably" – the death was preventable by optimal care	ICC = 0.34 preventability of **death** (based on 383 reviews of 111 cases)	N/A (Insufficient denominator)	• Hospital-wide, excluding data of patients receiving comfort care and nonveterans • Public hospitals • Patients with hospital-acquired laboratory abnormality over-sampled • Reviewed deceased patients only

Continued

Table 1. *Continued*

Author	Location; Date of Study	Target Group/ Type of Hospital	Grading of Preventability[a]	Threshold for Defining a Preventable Case	Kappa (ICC) for Preventability	Interhospital Variance/ ICC	Comments
Davis et al, 2001; 2003[24,25] Briant et al, 2006[23]	New Zealand; 1998	13 public acute care hospitals	1-6	≥ 4 "Close call, > 50:50"	Not reported	Not reported	• Hospital-wide excluding specialist institutions • Public hospitals • Over all hospitals there were: 850 AEs; 315 avoidable AEs ≥ 4; 531 ≥ 2 • 4.5% of patients with AEs died • 6.1% of avoidable AEs; unclear concerning disability/death status

Continued

Table 1. *Continued*

Author	Location; Date of Study	Target Group/ Type of Hospital	Grading of Preventability[a]	Threshold for Defining a Preventable Case	Kappa (ICC) for Preventability	Interhospital Variance/ ICC	Comments
Baker et al, 2004[20]	Canada; 2000	20 public acute care hospitals	1-6	≥ 4 "Preventability more than likely (more than 50/50, but close call)"	$\kappa = 0.69$, (95% CI, 0.55-0.83) / preventability of **AE** (based on duplicated review of a random sample of 10% of cases)	Not reported (Hospital size groupings preclude *de novo* calculation)	• Hospital-wide, excluding psychiatric and obstetric hospitals, day-only admission and patients < 18 • Acute care hospitals • Weighted percentages to account for total charts per hospital and hospitals per type per province • 15.7% of patients with AEs died

Continued

Table 1. *Continued*

Author	Location; Date of Study	Target Group/ Type of Hospital	Grading of Preventability[a]	Threshold for Defining a Preventable Case	Kappa (ICC) for Preventability	Interhospital Variance/ ICC	Comments
Michel et al, 2007[35]	France; 2004	71 private and public hospitals	1-6	≥ 4 "more likely than not"	preventability of AE $\kappa = 0.31$ (95% CI, 0.05-0.57) /(based on 58 cases judged to have AE by both reviewers)	Not reported	• Hospital-wide, excluding obstetric hospitals • Retrospective case note review and 7-day observation with data collection across 294 wards • Patients with (preventable) AEs not noted • 8.2% of patients with AEs died

Continued

Table 1. *Continued*

Author	Location; Date of Study	Target Group/ Type of Hospital	Grading of Preventability[a]	Threshold for Defining a Preventable Case	Kappa (ICC) for Preventability	Interhospital Variance/ ICC	Comments
Soop et al, 2009[37]	Sweden; 2003-2004	28 public acute care hospitals	1-6	≥ 4 "more than 50% likelihood"	$\kappa = 0.76$ / preventability of **AE** (based on duplicated review of 642 cases [all cases positive for screening criteria])	Not reported	• Hospital-wide, excluding psychiatric, rehabilitation, and palliative hospitals and day-only admission • Acute care hospitals with high proportion of elderly patients; all deaths occurred in elderly/ critically ill patients • Preventable mortality estimated from data • 4.1% of patients with AEs died

Continued

Table 1. *Continued*

Author	Location; Date of Study	Target Group/ Type of Hospital	Grading of Prevent-ability[a]	Threshold for Defining a Preventable Case	Kappa (ICC) for Preventability	Interhospital Variance/ ICC	Comments
Aranaz-Andrés et al, 2008; 2009[16,17]	Spain; 2005	24 public hospitals	1-6	≥ 4 "positive" – not defined	Not reported	Not reported	• Hospital-wide • Retrospective cohort study • Patients had 655 AEs; 278 preventable AEs (with at least moderate evidence) • Patients with preventable AEs estimated based on 42.6% of AEs being preventable • 4.4% of patients with AEs died; Kappa was reported only for the identification of AEs between reviewers and "gold standards"

Continued

Table 1. *Continued*

Author	Location; Date of Study	Target Group/ Type of Hospital	Grading of Preventability[a]	Threshold for Defining a Preventable Case	Kappa (ICC) for Preventability	Interhospital Variance/ ICC	Comments
Aranaz-Andrés et al, 2011[15]	Argentina, Colombia, Costa Rica, Mexico and Peru; 2005	58 public hospitals	1-6	≥ 4 "positive" – not defined	κ ranged from 0.27 to 0.74 between countries / preventability of AE (sample size not stated)	Not reported	• Hospital-wide • Retrospective case note review and prospective data collection • Preventable mortality estimated from data • 5.8% of patients with AEs died
Martins et al, 2011[34]	Brazil; 2003	3 teaching hospitals	1-6	≥ 4 (wording not described)	Not reported	Not reported	• Hospital-wide, including obstetric wards • 38% of patients with AEs died

Continued

Table 1. *Continued*

Author	Location; Date of Study	Target Group/ Type of Hospital	Grading of Prevent-ability[a]	Threshold for Defining a Preventable Case	Kappa (ICC) for Pre-ventability	Interhospital Variance/ ICC	Comments
Hogan et al, 2012[32]	England; 2009	10 acute hospi-tals	1-6	≥ 4 "Probably pre-ventable, more than 50/50 but close call"	$\kappa = 0.49$ (95% CI, 0.2-0.8) / pre-ventability of **death**, based on duplicated review of 250 cases (25% of sample)	"There were no significant differences between proportions of preventable deaths found at each hospital."[32(p740)]	• Hospital-wide, excluding obstetric and psychiatric wards, pediatric patients, and palliative care • 100 cases randomly selected from each acute hospital • Reviewed deceased patients only

Continued

Table 1. *Continued*

Author	Location; Date of Study	Target Group/ Type of Hospital	Grading of Prevent-ability[a]	Threshold for Defining a Preventable Case	Kappa (ICC) for Preventability	Interhospital Variance/ ICC	Comments
Sorinola et al, 2012[38]	England; 2009	1 acute hospital	1-6	≥ 4 "Preventable death"	None given for preventability of death; Reported $\kappa = 0.75$ (from sample of 400 notes) only for "determination of a problem in care" (more equivalent to presence of an AE)	N/A (Insufficient denominator)	• Hospital-wide, excluding obstetric and psychiatric wards, pediatric patients, and palliative care • 400 death cases selected consecutively in 2009 • Preventable mortality estimated from data

Continued

Table 1. *Continued*

Author	Location; Date of Study	Target Group/ Type of Hospital	Grading of Preventability[a]	Threshold for Defining a Preventable Case	Kappa (ICC) for Preventability	Interhospital Variance/ ICC	Comments
Gupta et al, 2013[30]	United States; 2009-2012	1 acute hospital	1-5	≥ 4 "Possibly preventable"	$\kappa = 0.10$ Preventability of **death** agreement between provider classification and a mortality review committee (15 cases only)	N/A (Insufficient denominator)	• Hospital-wide • 2,483 patients died, 1,683 had surveys completed • Preventable mortality estimate provided

Continued

Table 1. *Continued*

Author	Location; Date of Study	Target Group/ Type of Hospital	Grading of Preventability[a]	Threshold for Defining a Preventable Case	Kappa (ICC) for Preventability	Interhospital Variance/ ICC	Comments
Baines et al, 2013; 2015[18,19] Zegers et al, 2007; 2009; 2011a; 2011b[14,44-46]	The Netherlands; 2004 and 2008	33 acute hospitals	1-6	≥ 4 AE was found to be preventable when the care did not comply with existing professional standards and/or due to shortcomings of a health care practitioner, management or system	$\kappa = 0.4$ for preventability of adverse events[46]	Preventable AEs ICC = 3.7% (hospital-level)	• Hospitals including palliative care and excluding psychiatric, obstetric, and pediatric patients • Hospitals were randomly selected on location • Reviewed patients discharged alive and deceased patients • Higher proportion of preventable AEs in deceased than patients discharged alive

Continued

Table 1. *Continued*

Author	Location; Date of Study	Target Group/ Type of Hospital	Grading of Preventability[a]	Threshold for Defining a Preventable Case	Kappa (ICC) for Preventability	Interhospital Variance/ ICC	Comments
Hogan et al, 2015[33,e]	England; 2012-2013	24 acute hospitals	1-6	≥ 4 Probably avoidable, more than 50/50	$\kappa = 0.45$ (95% CI, 0.24-0.66) / based on random sample of 486 avoidable **death** cases (grade 4-6)	Not reported	• Hospitals, excluding obstetric, psychiatric, and pediatric patients • 100 cases randomly selected from each acute hospital • Reviewed only deceased patients

Continued

Table 1. *Continued*

Author	Location; Date of Study	Target Group/ Type of Hospital	Grading of Preventability[a]	Threshold for Defining a Preventable Case	Kappa (ICC) for Preventability	Interhospital Variance/ ICC	Comments
Manaseki-Holland et al, 2016[13,d,f]	England and Wales; 2003-2009	22 hospitals	1-5	≥ 3 On the balance of probability (ie, > 50% chance)	$\kappa = 0.27$ (95% CI, 0.19-0.39) intra-class correlation across a single review	Not reported	• Hospitals with inclusion of only respiratory conditions from medical wards • 191 case notes for those admitted with respiratory complaints and those 65 years and over • Case notes randomly assigned to 2-7 reviewers (total of 653 reviews)

Continued

Table 1. *Continued*

Author	Location; Date of Study	Target Group/ Type of Hospital	Grading of Preventability[a]	Threshold for Defining a Preventable Case	Kappa (ICC) for Preventability	Interhospital Variance/ ICC	Comments
Flaatten et al, 2017[29]	Norway; 2011	3 acute hospitals	1-5	≥ 4 "Possibly preventable"	Not reported	Not reported	• All hospital deaths across 3 hospitals in 2011 (including emergency departments) • 1,185 death notes reviewed across one-year period • Case notes assigned to six consultant reviewers each from different specialties

Continued

Table 1. *Continued*

Author	Location; Date of Study	Target Group/ Type of Hospital	Grading of Preventability[a]	Threshold for Defining a Preventable Case	Kappa (ICC) for Preventability	Interhospital Variance/ ICC	Comments
Kobewka et al, 2017[11,12]	Canada; 2013	1 acute hospital	0-100	> 50 "Possibly preventable"	ICC = 0.14 (480 deaths each reviewed by 4 reviewers; reliability for average of four reviewers reported as 0.68)	N/A (Insufficient denominator)	• Hospital-wide, excluding pediatrics • 480 deceased case notes (structured case abstracts) produced across 3-month admission period • Case notes randomly assigned to 4 physician reviewers

Continued

Table 1. *Continued*

Author	Location; Date of Study	Target Group/ Type of Hospital	Grading of Preventability[a]	Threshold for Defining a Preventable Case	Kappa (ICC) for Preventability	Interhospital Variance/ ICC	Comments
Roberts et al, 2017[36]	United Kingdom; 2012-2015	4 Northeast England, UK acute care trusts (~23 hospitals)	1-6 (PRISM) 1-5 (NCEPOD)	≥ 4 (PRISM) ≥ 3 (NCEPOD)	κ = N/A Not reported for this study, authors cited a reliability estimate of $\kappa = 0.45$ from PRISM[32,33]	N/A	• All hospital deaths across 4 trusts • 7,370 medical records reviewed • Case notes reviewed predominantly by consultants, some by nurses

Abbreviations: CI, confidence interval; ICC, intraclass correlation coefficient; NCEPOD, National Confidential Enquiry into Patient Outcome and Death; PRISM, the Preventable Incidents Survival and Mortality Study.

[a]Scale of degree of preventability. This tends to range from "6, (virtually) certain evidence of preventability" to "1 (virtually) no evidence for preventability."

[b]We have reversed the scale to facilitate comparisons with other studies. The original scale ranged from 1, *(definitely) preventable death* to 4, *(definitely not) preventable death.* Cases with a grade of 2 or lower *(probably or definitely),* on the original scale, were considered as preventable.

[c]For cerebrovascular accident, myocardial infarction and pneumonia, respectively.

[d]We have reversed the scale to facilitate comparisons with other studies. The original scale ranged from 1, *(definitely) preventable death* to 5, *(definitely not) preventable death.* Cases with a grade of 2 or lower *(probably or definitely),* on the original scale, were considered as preventable.

[e]"In your judgment, is there some evidence that the patient's death was avoidable if the problem/s in health care had not occurred?"

[f]The England study has been extracted from the 2016 paper as the US data have been included in Hayward and Hofer.[31]

Table 2. Summary of Study Processes and Review Methods

Category		No.	References
Inclusion of a screening stage	No screening stage	4	32, 33, 36, 37
	Yes (16-18), criteria	15	10, 14-26, 31, 34, 35, 38-46
	Trigger tool	4	15, 26, 34, 38
Scale used for implicit judgment	Binary	0	
	4-point Likert	2	21, 26
	5-point Likert	3	13, 31, 36
	6-point Likert	16	10, 14-20, 22-25, 32-46
	Continuous	2	11-13
Reviewer screening stage 1	Physician	7	13, 14, 18, 19, 27-29, 32, 33, 36, 44-46
	Nurse	11	14-19, 21-25, 34, 35, 37-42, 44-46
	Pharmacist	1	38
Reviewer review stage 2	Physician expert advice available	15	14-25, 27, 28, 34-46
	Pharmacist support	0	
	Nurse support	0	
Duration of expert advice	Indefinite duration	3	10, 33, 36
	Temporary duration	3	16, 17, 21, 23-25
	No stated duration	2	13, 33
Reviewer affiliations	External to the institution being reviewed	20	10-26, 31-35, 37-46[a]
	Internal	2	21, 36[a]
Hospital anonymization	Undertaken	5	13, 23-25, 31-33

Continued

Table 2. *Continued*

Category		No.	References
	NOT undertaken	17	10-12, 14-22, 26-28, 34-46
Clinical experience of physicians	< 5 years	0	
	5-10 years	4	11, 12, 15-17, 20
	> 10 years	7	21, 32-34, 36, 37, 43
	Previous experience not mentioned	2	10, 39-42
	No mention of experience	5	22-28, 35
Speciality of physicians	General medicine/ internal medicine (alone)	13	10, 15-17, 20-25, 32, 34, 35, 37, 38, 43
	Internal medicine and specialists	9	11-14, 18, 19, 21, 26, 31, 33, 36, 39-42, 44-46
Review discrepancies and disagreements reconciled	Physicians	3	14, 18, 19, 36, 43-46
	Nurses	0	
	Medical health analysts /records analysts	1	22
	Executive board	2	16, 17, 37
	Information not available	6	20, 21, 23-28, 39-42
Physician reviewer training duration	≤ 1 day	7	14, 18, 19, 21, 23-25, 27, 28, 32, 33, 38, 44-46
	1-3 days	7	13, 20, 31, 34, 36, 39-43
	≥ 3 days	3	16, 17, 35, 37
	Not stated	4	10-12, 15, 26

Continued

Table 2. *Continued*

Category		No.	References
Training content	Case note exposure	12	10, 13, 14, 18-28, 31, 36, 37, 44-46
	Specialist advice provided	8	14, 16-19, 21, 23-25, 27, 28, 31, 32, 36, 44-46
	Absence of preventability definition	18	10, 13-20, 22-26, 31-35, 37-46
	Familiarity with study tools	14	10, 13, 14, 18-25, 27, 28, 33, 34, 36-42, 44-46

[a]Best and Cowper[21] was half external and half internal.

These commonly reported reliability estimates, which describe the ability to distinguish between case notes of patients who died, can quantify the confidence with which one can act on the presumption that a specific avoidable death had occurred, such as by investing in a root cause analysis to establish proximate causes, or possibly for establishing legal liability or determining compensation *for an individual case*. However, such reliability estimates tell you nothing about determining the performance of *different providers*, such as different hospitals. A key determinant of reliability in any measurement is the variation across the things you want to distinguish between; thus, to distinguish between hospitals requires an estimate of the variation of preventable death rates *across* hospitals.

No study was found to have published an estimate of this quantity despite its critical relevance to any policymaking with respect to preventable deaths. We were able to reanalyze data from one study of 22 hospitals to produce the variance estimates required to make a provisional "best available" calculation of the optimal number of reviews per case and per hospital required to produce a reliable estimate of the hospital preventable death rate (see Online Appendix 2).[13] Only one other study had quantified hospital variation for a more global measurement of preventable AEs that included deaths, and the study authors reported a hospital variance estimate similar in magnitude to the one we estimated.[14]

Table 3. Preventable Mortality and/or Adverse Events (AEs) Reported in the Included Studies

Author, Year (Country)	No. of Admitted Patient Case Notes Sampled for Review	No. of Deceased Patient Case Notes Reviewed	No. of Admission Case Notes Selected After Screening for Review by Physicians	Preventable AEs (% of admissions)	Preventable AEs (% of all AEs)	Preventable Mortality (% of admissions)	Preventable Mortality (% of deceased)	Threshold for Preventability & Comments[a,b]
Dubois et al, 1987; 1988 (United States)[26-28]	1,946	182	1,946	NR	NR	4.6% (weighted estimate, calculated n = [90]/1,946)	26.9% 49/182 14% 25/182	Preventability score ≥ 3 out of 4[c] (majority decision) Preventability score ≥ 3 out of 4[c] (unanimous decision)

Continued

Table 3. *Continued*

Author, Year (Country)	No. of Admitted Patient Case Notes Sampled for Review	No. of Deceased Patient Case Notes Reviewed	No. of Admission Case Notes Selected After Screening for Review by Physicians	Preventable AEs (% of admissions)	Preventable AEs (% of all AEs)	Preventable Mortality (% of admissions)	Preventable Mortality (% of deceased)	Threshold for Preventability & Comments[a,b]
Brennan et al, 1991 (United States)[22]	30,121	NR	7,743	306 (1.02% weighted)	3.96% 306/7,743	0.30% 89/30,121	NR	Causation score ≥ 1 on a 0-6 scale; preventability score ≥ 4 out of 6
Hayward et al, 1993 (United States)[10]	675	135 (calculated, reported as 20% of sample)	675	NR	NR	0.44% [3]/675 Weighted for over-sample of deaths	9% [12]/135 (n = 12 calculated from rate reported)	Preventability score ≥ 4 out of 6

Continued

Table 3. *Continued*

Author, Year (Country)	No. of Admitted Patient Case Notes Sampled for Review	No. of Deceased Patient Case Notes Reviewed	No. of Admission Case Notes Selected After Screening for Review by Physicians	Preventable AEs (% of admissions)	Preventable AEs (% of all AEs)	Preventable Mortality (% of admissions)	Preventable Mortality (% of deceased)	Threshold for Preventability & Comments[a,b]
Best and Cowper, 1994 (United States)[21]	NR	222[d]	NA	NR	NR	NR	21.6% median	Preventability score ≥ 3 out of 4
Wilson et al, 1995 (Australia)[43]	14,179	114	1,718	1,205 (8.50%)[e]	NR	0.55% 78/14,179	29.00%	Causation score ≥ 2 out of 6; preventability score ≥4 out of 6

Continued

Table 3. *Continued*

Author, Year (Country)	No. of Admitted Patient Case Notes Sampled for Review	No. of Deceased Patient Case Notes Reviewed	No. of Admission Case Notes Selected After Screening for Review by Physicians	Preventable AEs (% of admissions)	Preventable AEs (% of all AEs)	Preventable Mortality (% of admissions)	Preventable Mortality (% of deceased)	Threshold for Preventability & Comments[a,b]
Thomas et al, 1999; 2000a; 2000b; 2002 (United States)[39-42]	14,700	NR	448	3.00% 448/14,700	NR	0.265% 39/14,700	NR	Causation score ≥ 4 out of 6; preventability: "an adverse event was considered preventable if it was avoidable by any means currently available unless that means was not considered standard care." The implicit judgment methods are similar to those used in Brennan et al.[22]

Continued

Table 3. *Continued*

Author, Year (Country)	No. of Admitted Patient Case Notes Sampled for Review	No. of Deceased Patient Case Notes Reviewed	No. of Admission Case Notes Selected After Screening for Review by Physicians	Preventable AEs (% of admissions)	Preventable AEs (% of all AEs)	Preventable Mortality (% of admissions)	Preventable Mortality (% of deceased)	Threshold for Preventability & Comments[a,b]
Hayward and Hofer, 2001 (United States)[31]	NA	111	NA	NA	NR	0.23%-0.61% (at least possibly preventable) (95% CI)	22.7%; 6.0% (weighted for sampling design)	Preventability score ≥ 3 out of 5[f] Preventability score ≥ 4 out of 5[f]
Davis et al, 2001; 2003; Briant et al, 2006 (New Zealand)[23-25]	6,579	118	850	6.28% 413/6,579	48.6% 413/850	0.36% 24/6,579	19.8%-20.7%	Causation score ≥ 2 out of 6 Preventability score ≥ 2 out of 6

Continued

Table 3. *Continued*

Author, Year (Country)	No. of Admitted Patient Case Notes Sampled for Review	No. of Deceased Patient Case Notes Reviewed	No. of Admission Case Notes Selected After Screening for Review by Physicians	Preventable AEs (% of admissions)	Preventable AEs (% of all AEs)	Preventable Mortality (% of admissions)	Preventable Mortality (% of deceased)	Threshold for Preventability & Comments[a,b]
Baker et al, 2004 (Canada)[20]	3,692	236	1,512	2.8% (95% CI, 2.0% to 3.6%)[h]	7.01% 106/1,512[g]	0.66% (95% CI 0.37% -0.95%)[h] (death from preventable AE)	16.9% 40/236[f]	Causation score ≥4 out of 6 Preventability score ≥ 4 out of 6[f]
Michel et al, 2007 (France)[35]	8,754	NR	NR	1.08% 95/8,754	NR	0.09% 8/8,754	NA	Causation score ≥ 4 out of 6 Preventability score ≥ 4 out of 6
Soop et al, 2009 (Sweden)[37]	1,967	10	241	8.6% 169/1,967	70.1% 169/241	0.25% 5/1,967	NR	Causation score ≥ 4 out of 6

Continued

Table 3. *Continued*

Author, Year (Country)	No. of Admitted Patient Case Notes Sampled for Review	No. of Deceased Patient Case Notes Reviewed	No. of Admission Case Notes Selected After Screening for Review by Physicians	Preventable AEs (% of admissions)	Preventable AEs (% of all AEs)	Preventable Mortality (% of admissions)	Preventable Mortality (% of deceased)	Threshold for Preventability & Comments[a,b]
Aranaz-Andrés et al, 2008; 2009 (Spain)[16,17]	5,624	225	1,755	11.65% 655/5,624	37.3% 655/1,755	0.07% 5/5,624[i]	4.5%	Causation score ≥ 4 out of 6 Preventability score ≥ 4 out of 6
Aranaz-Andrés et al, 2011 (Argentina, Colombia, Costa Rica, Mexico, Peru)[15]	11,379	NR	1,754	10.47% 1,191/ 11,379	59% 674/1,144	NR	NR	Causation score ≥ 4 out of 6 Preventability score ≥ 4 out of 6
Martins et al, 2011 (Brazil)[34]	1,103	94	1,103	5.07% 56/1,103	5.07% 56/1,103	2.3% 25/1,103 (coexisting previous AE and death)	26.6% 25/94	Causation score ≥ 4 out of 6 Preventability score ≥ 4 out of 6

Continued

Table 3. *Continued*

Author, Year (Country)	No. of Admitted Patient Case Notes Sampled for Review	No. of Deceased Patient Case Notes Reviewed	No. of Admission Case Notes Selected After Screening for Review by Physicians	Preventable AEs (% of admissions)	Preventable AEs (% of all AEs)	Preventable Mortality (% of admissions)	Preventable Mortality (% of deceased)	Threshold for Preventability & Comments[a,b]
Hogan et al, 2012 (England)[32,j]	NR	1,000	NA	NR	NR	NR	5.2% 52/1,000	Preventability score ≥ 4 out of 6 (reporting 1 of 3)
Sorinola et al, 2012 (England)[38]	NR	400	NA	NR	NR	NR	3.5% 14/400	Preventability score ≥ 4 out of 6
Gupta et al, 2013 (United States)[30]	NR	1,683	NR	NR	NR	NR	2.50% 42/1,683	Preventability score ≥ 4 out of 5
Baines et al, 2013; 2015; Zegers et al 2007; 2009; 2011a; 2011b (The Netherlands)[14,18,19,44-46]	11,949	762	1,130	NR	NR	NR	4.5%	Preventability score ≥ 4 out of 6

Continued

Table 3. *Continued*

Author, Year (Country)	No. of Admitted Patient Case Notes Sampled for Review	No. of Deceased Patient Case Notes Reviewed	No. of Admission Case Notes Selected After Screening for Review by Physicians	Preventable AEs (% of admissions)	Preventable AEs (% of all AEs)	Preventable Mortality (% of admissions)	Preventable Mortality (% of deceased)	Threshold for Preventability & Comments[a,b]
Hogan et al, 2015 (England)[33,k]	NR	2,400	NA	NR	NR	NR	3% 101/2,400	Preventability score ≥ 4 out of 6
Manaseki-Holland et al, 2016 (England)[13]	NR	191[l]	NA	NR	NR	NR	10% (median) Q1 3% Q3 28%	Preventability score ≤ 2 out of 5
Flaatten et al, 2017 (Norway)[29]	59,605	1,167	NR	NR	NR	0.057% 34/59,605	2.91% 34/1,167	Preventability score ≥ 50 out of 100
Kobekwa et al, 2017 (Canada)[11,12]	14,267	480[j]	NR	NR	NR	0.22% 31/14,267	6.46% 31/480[m]	Preventability score ≥ 50 out of 100[m]

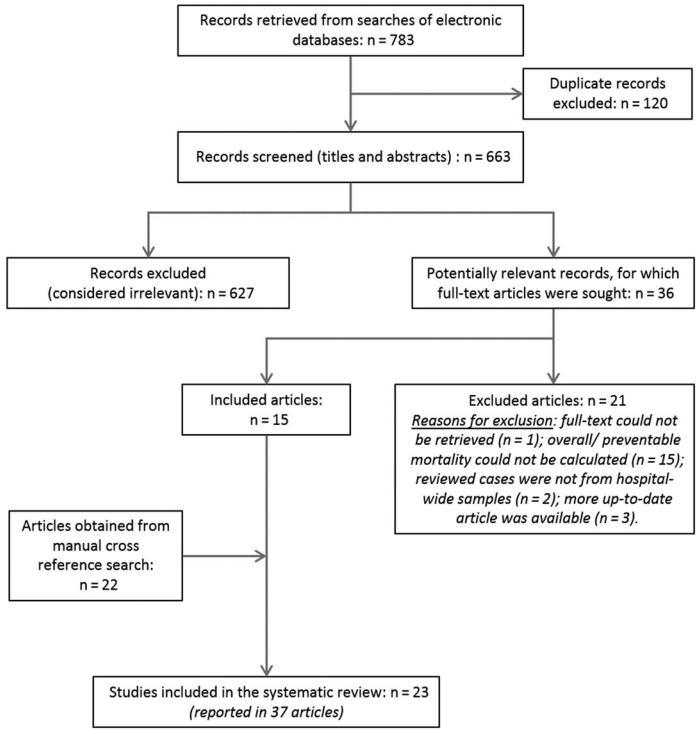
Continued

Table 3. *Continued*

Author, Year (Country)	No. of Admitted Patient Case Notes Sampled for Review	No. of Deceased Patient Case Notes Reviewed	No. of Admission Case Notes Selected After Screening for Review by Physicians	Preventable AEs (% of admissions)	Preventable AEs (% of all AEs)	Preventable Mortality (% of admissions)	Preventable Mortality (% of deceased)	Threshold for Preventability & Comments[a,b]
Roberts et al, 2017 (UK)[36]	NR	7,194	NR	NR	NR	NR	0.47% 34/7,194	Preventability score ≥ 50 out of 100

Abbreviations: NA, not assessed; NR, not reported.
[a]Causation score is the score given to the likelihood of the adverse event being caused by medical care/management. A causation score of ≥ 2 out of 6 corresponds to "at least slight to modest evidence of management causation"; a causation score of ≥ 4 out of 6 corresponds to "management causation more likely – more than 50/50.".
[b]A preventability score of ≥ 2 out of 6 corresponds to "at least slight to modest evidence of preventability"; a preventability score of ≥ 4 out of 6 corresponds to "preventability more than likely – more than 50/50."
[c]We have reversed the scale to facilitate comparisons with other studies. The original scale ranged from 1, *(definitely) preventable death* to 4, *(definitely not) preventable death.* Cases with a grade of 2 or lower *(probably or definitely),* on the original scale, were considered as preventable.
[d]Pairs were matched across high observed-to-expected mortality (OTEM) and low OTEM Veterans Affairs hospitals.
[e]This indicator is for deaths considered with a high level of preventability.
[f]Figures are taken from direct author response rather than published data.
[g]Of 255 patients with iatrogenic adverse events, 106 had > 50% probability of preventability.
[h]Adjusted for sampling frame.
[i]Associated with preventable AE.
[j]"Was the patient's death due to problems in the healthcare or did problems in healthcare contribute to the death?"
[k]"In your judgment, is there some evidence that the patient's death was avoidable if the problem/s in health care had not occurred?"
[l]Multiple reviews were undertaken with the case notes.
[m]> 50% probability of membership in the "possibly preventable" class.

Figure 1. Review Flow Diagram of Article Retrieval and Inclusion



Results

Article Retrieval and Inclusion

Our electronic searches yielded 663 records after duplicates were removed (Figure 1). A citation search of included studies identified 22 additional articles. In all, 37 articles (representing 23 studies) were included.[10-46] The characteristics of included studies are shown in Table 1. The study selection process and reasons for exclusion are summarized in Online Appendix 3. We were unable to find all the elements we required in the 37 published articles for any of the 23 studies. We

wrote to the authors of these studies for more detail and of these, 14 of the authors responded.

Twelve studies[10-13,21,26-33,36,38] had the reviewers focus on an assessment of whether a death was preventable. Eleven studies[14-20,22-25,34,35,37,39-46] aimed primarily to identify and evaluate whether AEs were preventable. These AEs could include or accompany the death of a patient. All but two studies were in high-income countries and conducted between 1984 and 2015. They involved a median of 20 hospitals (interquartile range = 23) and 230 deaths reviewed (range 10 to 7,194).

Methods for Assessing Preventable Deaths and Preventable Adverse Events Contributing to Deaths

The majority of the published studies did not present enough details to obtain the information required for this review, and unpublished data were obtained by author communications. Through writing to the authors, we obtained additional data on 14 of the 23 studies. These are summarized in Table 2 and Online Appendices 4 and 5.

Tools and Stages of Review. A plurality of the studies (9/23) followed the method of the Harvard Medical Practice Study,[22] which in turn was based on an approach called structured implicit physician review developed by the RAND Corporation in the 1980s.[47] This measurement procedure includes an initial screening of patient notes to identify cases in which it is more likely that an adverse event might have occurred. The other studies provided a varied amount of information on methodology, and therefore we wrote to the authors for details. These details are summarized in Table 2 and Online Appendices 4 and 5.

In structured implicit case note review, the structured component guides the reviewer systematically and more or less temporally through the hospital admission, asking him or her to focus and rate specific elements of the patient's care in sequence before making an overall judgment about the quality of care.[48] The "implicit" component is inherent in the summary judgments produced by the reviewer about the case, as well as the exercise of professional situational judgment in deciding whether deviations from ideal processes represent an error or are appropriate in the clinical context. This can be contrasted with

generating a score based on a checklist where the use of any judgment is much more restricted. A non-structured implicit review has been found to be less reliable in estimating hospital quality of care, presumably owing to the less standardized approach for navigating a record and building up to an overall rating.[49,50]

In our sample, most studies used a kind of structured implicit (or criterion-based implicit) review pro forma. Although the details of the structured component varied, in all cases adopting structured implicit review, the "structured" component required the reviewer to review and make quality judgments over phases of care (such as diagnostic or treatment phases). The reviewer was often asked to write explicit comments about areas of concern (as free-flow text) for each phase, and finally to score quality for each phase of care.

The decision on preventability was made on a scale applying implicit judgment of the physician reviewers. The majority (15/23) of the studies used a six-category grading system (Likert scale) to classify the preventability of deaths and/or AEs.[10,14-20,22-25,32-35,37-46] The categories were inevitably collapsed into a binary outcome. Deaths (and/or AEs) that were considered to have more than a 50/50 chance of being preventable were considered preventable in most studies. Three studies[11-13,31] used a continuous scale (0-100) probability of preventability, which was compared with the Likert scale; the 0-100 scale was found to have the same constructs and to impart comparable information to the Likert scale.[13]

Only five studies noted an attempt to anonymize the patient and hospital identifiers in case notes[13,23-25,31-33] to prevent bias during reviews. No study blinded the reviewers to the outcome in these samples selected on the basis of death as the outcome.

Reviewer Selection and Training. In all studies, reviewers were external to the institutions from which case notes were derived to reduce internal institutional bias. For reviewer selection, seven studies did not have a first-stage screening process and deployed only physicians for these reviews.[13,14,18,19,26-29,32,33,36,44-46] Fifteen studies used two stages; a screening process that involved mainly nurses at the first stage and exclusively physician reviewers at the second stage.[10,14-26,31,34,35,38-46] Seven studies used an experienced or supervisor reviewer physician: in six studies for settling disagreements between the physician reviewers[14,16-19,22-25,37,43-46] and in one study for quality control purposes (see Table 2 and Online Appendices 4 and 5).[39-42]

The required reviewer experience (where recorded) varied widely across the studies in both nurses and physicians. For physicians, regular handling of case notes, a lengthy period of clinical work (ie, more than five years of clinical/reviewing experience), postgraduate education, and independent accreditation were used as criteria. For example, in the US studies, reviewers were board-certified with a general preference for generalists/internists.[10,21,22,43] The UK studies used reviewers from specialties across general medicine and intensive care consultants.[13,32,33,38] Eight studies deployed general physicians,[11-19,22-25,37,43-46] and in seven of these a panel of specialists was available to advise individual reviewers when required.[11,12,14-19,22-25,37,43-46]

Various forms of reviewer training and support were provided. The training duration ranged from one to three days. Nurses and physicians had the same training in eight studies.[14-20,22-25,35,37,44-46] Eleven studies were explicit about the exposure to case notes during the training.[10,13,14,18-28,36,37,44-46] Six studies did not disclose reviewer training information. Where enough details were available, training did not define preventability, but rather offered clinicians an opportunity to understand the aims, merits, and some caveats (eg, hindsight bias[51,52]) of the case note review process, to familiarize them with the pro forma for data extraction, and to exchange views on approaches to difficult cases after practicing the review on one or more case notes.[13,16,17,19,20,22-24,26,31-34,46]

Estimated Preventable Mortality

The proportion of deaths judged to be preventable depends on the cut-off threshold used in the Likert scale. One study chose to estimate preventability at the lowest threshold, namely any probability that the death could have been prevented (eg, two or more out of six),[23,24] whereas most used a threshold of more than three out of six or three to four out of 5. Preventable mortality rates as a proportion of *all admissions* were estimated between 0.07% and 4.62% (Table 3). Most reports were below 0.7%; the 2.27% reported in Brazil[34] and the 4.6% in the Dubois study[26-28] were exceptionally high. Preventability rates as a proportion of *all deaths* were estimated between 0.47% and 29%.[10-13,16-19,20,21,26,28-34,36,38,43-46,53] The studies focusing more broadly on AEs varied in approach when estimating preventable deaths. Their approaches ranged from asking reviewers to rate whether the

identified AE contributed to death, to positing that a death is preventable if accompanied by a preventable AE, no matter how minor. The estimates are more direct and consistent when considering the larger studies specifically focused on preventable deaths from only more recent years (2008 to 2017). These have a median preventable death rate of 3% with an interquartile range of 3.0%-6.0% (range 0.47%-10%).

The studies that evaluated preventability of any AE as a proportion of *all admissions* reported generally higher but widely variable figures, ranging from 1.02%[22] to 11.65%,[16,17] and preventable AEs as a proportion of *all AEs* ranging from 3.96%[22] to 70.1%.[37]

Interrater Reliability (Kappa Statistic)

The reliability of a single review assessing preventability is reported for 17 of the 23 studies.[10-13,15,18-22,26,28,30-33,35,37,39-46] Fifteen are reported as Cohen's Kappa, a statistic that was developed to measure the agreement between raters taking into consideration the agreement that occurs by chance,[55] although for these ordinal measures the intraclass correlation (reported for the remaining two) is comparable and would probably be preferred.[56,57] The reliability for assessing the preventability of death is reported for nine studies with a median reliability of 0.33 and an interquartile range of 0.27-0.45 (range 0.10-0.50). If limited to the reported reliabilities from five larger studies done in the past 10 years (that included a median of 1,080 deaths), the reliability has a median of 0.27 (range 0.10-0.49). A further eight studies reported the reliability for preventing an AE with a median of 0.36 and an interquartile range of 0.29-0.58 (range 0.21-0.76). No data were found on the effects of reviewer selection, characteristics, or training on the reliability of the judgment of preventability by the reviewers.

Calculating the Optimal Number of Reviews and Reviewers per Case Note to Estimate Preventable Death per Case Note and per Hospital

The interquartile range of reliability reported for the ability of a single review to distinguish between cases with respect to whether death was preventable was 0.27 to 0.45. At a representative level of reliability of 0.35 for a single review, we can estimate that an average of 8 reviews per case note would be required to achieve a reliability of 0.8 when

distinguishing between cases. Seventeen reviews per case would be required to achieve a reliability of 0.9, a level often recommended for testing with high-stakes consequences. If the reliability of a single review were as high as 0.5, then only 4 or 9 reviews per case note would be needed for a reliability of 0.8 or 0.9, respectively. However, any given operational program would have to determine the reliability of its measurement procedure in its population to figure out the number of cases needed to review.

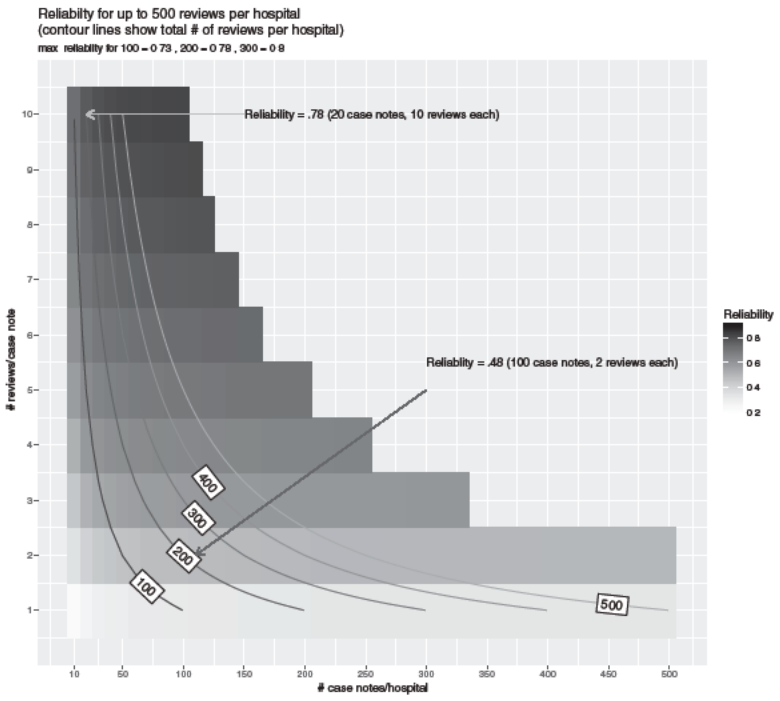
About 200 to 300 total reviews per hospital would be required to reach a reliability of 0.8 for distinguishing between hospitals, based on the limited evidence available about the between-hospital variance and other components of variance (see Online Appendix 2 for the estimates used and methods to project sample size). However, given 300 reviews in total, better reliability is achieved with more reviews per patient and fewer patients overall. Holding the total number of reviews constant, increasing the number of reviews per case increases reliability (eg, 10 reviews per case for 30 cases) more than selecting more cases per hospital (eg, 150 cases per hospital with two reviews per case). A strategy of only one review per case would provide at best fair reliability (0.20-0.40) no matter how many total reviews were done per hospital. Figure 2 illustrates how the reliability changes as the numbers of reviews and reviewers per hospital vary.

It is important to emphasize that more extensive and particularly population-specific data about the sources of variability in the review procedure could substantially change the projected number of reviews needed in either direction. In general, more heterogeneity across hospitals, more consistent reviewers, evaluating change over time within a hospital, and a focus on relative as opposed to absolute probability of preventable death would result in a more modest and feasible number of reviews needed to produce a reliable estimate.

Discussion

We set out to review the literature on measuring preventable deaths and to determine if it would allow us to project how many reviews and reviewers would be required for hospitals to learn lessons from reviewing preventable deaths and for a hospital system to profile hospitals based on their preventable death rates. Secondly, we looked at whether the

Figure 2. Reliability for Up to 500 Reviews per Hospital



literature contained any information on how the reliability of physician retrospective case record review to identify preventable deaths could be improved by refining the measurement procedure. To this end we conducted a review of studies of preventable hospital deaths published from 1984 to 2015.

The first important finding is that the preventability of death was consistently low in the reviewed studies and remarkably consistent across the more recent large studies. After our review was completed, one additional study from Norway of 1,000 deaths was published online ahead of print, reporting a preventable death rate of 4.2%, consistent with the interquartile range of 3%-6% from the larger studies of the past decade that we describe (reliability was not estimated).[71] While some

studies did vary the probability thresholds and Likert scale anchors for defining preventability as described earlier, most studies used a similar operational definition of more than a 50/50 chance on balance of probability for defining that a death was preventable. However, the difficulty of establishing how representative the deaths reviewed were for many studies, as well as the heterogeneity of the measurement procedures employed, made it impossible in our mind to develop a generalizable summary estimate.

Nevertheless, a low prevalence of preventable death should substantially heighten concern about using SMRs calculated from discharge data to profile hospitals. If 95% of deaths are nonpreventable, detection of outlier hospitals has an extremely low positive predictive value[3] and any misspecification of risk adjustment models will also necessarily introduce substantial bias in any judgment using SMRs about which hospitals have higher or lower rates of preventable deaths.

Another important finding is the lack of any published estimates in the literature of how much variation there is in preventable death rates across hospitals. Without this it is impossible to estimate the reliability for distinguishing between hospitals with respect to their preventable death rates or to design an operational program to do so. Using direct measurement, we estimated that as many as 300 or more total reviews could be required per hospital to distinguish between hospitals in a league table with high-stakes relegation and promotion consequences. Additionally, holding the total number of reviews per hospital constant, the optimal number of cases per hospital and reviews per case would require trade-offs to ensure the maximum generalizability and precision.

Furthermore, recall that the explicit purpose of comparing SMRs is to identify differences in preventable or avoidable death rates for which the SMR is just a proxy. The only study to look at this found little correlation between SMRs and preventable deaths across hospitals.[33] If it is found more broadly that the rates are not correlated, or that the variation in SMRs across hospitals is substantially larger than the variation in preventable death rates as directly measured, it would add substantial support to the concerns voiced by a number of critics that SMRs are measuring something else, most likely unmeasured case-mix differences. Yet, profiling hospitals based on SMRs remains ubiquitous and in the United States is tied to significant and increasing financial risk to hospitals in the absence of this critical piece of information that could further support or call into question the validity of SMRs.

The literature does provide more data about the reliability of a single measurement to distinguish individual cases with respect to whether a preventable death or preventable adverse event more generally occurred. This reliability estimate is relevant for quality reviews of sentinel cases by hospitals to learn from possible mistakes or for reviews by licensing boards or for cases subject to litigation. It is clear that high reliability is desirable before possible sanctions or major changes in work flows or procedures are contemplated on the basis of a judgment that a preventable death has occurred. For a typical reliability of 0.35 from the fairly wide range observed, between 8 and 17 reviewers could be required to reliably distinguish between patients with respect to whether a preventable death occurred. This number is far larger than is commonly used for credentialing, legal cases, and sentinel case and root cause analysis reviews.

However, providing these specific calculations as examples should not obscure the more important point that different measurement questions and different patient and hospital populations will each require their own estimates of reliability. These reliability estimates can then in turn be used to develop questions and population-specific calculations of the number of reviewers and reviews per record required so that an estimate with the required precision can be obtained. The numbers may vary substantially based on the setting and question.

We also summarize variation in the measurement procedures across studies (Online Appendix 5). We provide previously unpublished and summary data about many aspects of the procedures used as they were often not reported in the published papers. While the assessment methods had areas in common across the studies, on the whole they were quite heterogeneous. We found no empirical assessment of how single vs two-stage assessments, pro forma tools, reviewer selection or training, reviewer characteristics, and environmental influences affect consistency of measurement. Formal reliability or generalizability studies to evaluate different aspects of training and measurement procedures could be built into an operational program to facilitate improvements in the reliability of measurement. Details of these criteria and methodological issues as related to existing literature are discussed in Online Appendix 5.

Finally, it is worth reiterating that the structured implicit case note review method was originally designed to measure quality, not preventable death, and has a large literature describing its use for this purpose.[58] We should perhaps abandon attempts to measure the absolute proportion

of deaths that are preventable as an impossible quest.[13] Physicians are not good at estimating prognostic survival probabilities much less the even more challenging counterfactual probabilities such as "what is the probability of survival *if an event had not occurred*," which raises concern about the validity of such estimates.[59-61] Rather, structured implicit review could be used to directly measure the quality of care in the period before a patient's death, in keeping with how these methods were originally designed when developed 30 to 50 years ago.[47,62-64] This might be particularly useful if it were successfully demonstrated that quality problems were more common in those who eventually died during a hospitalization than in randomly selected cases.

The systematic review component of this study has several limitations. Because of practical reasons, we excluded studies not published in English. We found a large variation in the reported preventable mortality, but with only a limited number of studies we are unable to confirm the exact source of the observed heterogeneity. We have focused on overall hospital mortality and acute general medicine cases in this review.

Conclusions

Based on available information, preventable deaths comprise a relatively small fraction of all deaths, raising concerns about the feasibility of using SMRs as a proxy for preventable deaths. Structured implicit review is a challenging measurement task and it is likely that relatively large numbers of reviews are needed either to allow for learning from individual cases or to compare hospitals. Furthermore, there is a critical lack of any reported estimates of hospital variance in preventable death rates, which is required to design systems in a responsible way that profile hospitals based on preventable death rates, whether measured directly or indirectly. There is little evidence on factors that affect the consistency of case note reviews other than reviewer experience, and agreement between reviewers remains fair to moderate.

Any operational system assessing hospital quality around deaths will need to invest in a substantial ongoing effort to quantify the variation across hospitals and reviewers, although the cost of this would still be small relative to the cost of the operational system itself. It is also important to evaluate how the selection and training of the reviewers and measurement procedures can make the reliability more consistent

(see Online Appendix 5 for an expanded discussion).[65] Attempting to measure preventable deaths on an absolute scale would require engagement with the behavioral science and cognitive psychology literature, pertinent to human and system-wide errors[66] in health care,[67] that best locate the bounded rationality of human decision making,[68] and the biases that plague it.[69,70] However, whether measuring preventable deaths, or quality more generally as we would recommend, those who want to profile providers must recognize that no program can be designed to distinguish between providers without stable estimates of the amount of variation that exists across those providers.

References

1. *The Dr Foster Hospital Guide 2009: How Safe Is Your Hospital?* London, United Kingdom: Dr Foster Intelligence; 2009.
2. Lilford R, Pronovost P. Using hospital mortality rates to judge hospital performance: a bad idea that just won't go away. *BMJ*. 2010;340:c2016.
3. Girling AJ, Hofer TP, Wu J, et al. Case-mix adjusted hospital mortality is a poor proxy for preventable mortality: a modelling study. *BMJ Quality Saf*. 2012;21(12):1052-1056.
4. Scott IA, Brand CA, Phelps GE, Barker AL, Cameron PA. Using hospital standardised mortality ratios to assess quality of care—proceed with extreme caution. *Med J Aust*. 2011;194(12): 645-648.
5. Nicholl J. Case-mix adjustment in non-randomised observational evaluations: the constant risk fallacy. *J Epidemiol Comm Health*. 2007;61(11):1010-1013.
6. Lilford R, Edwards A, Girling A, et al. Inter-rater reliability of case-note audit: a systematic review. *J Health Serv Res Pol*. 2007;12(3):173-180.
7. Tender for Retrospective case record review programme - Lot 2 Tender Documentation for a National Programme to use retrospective record review to estimate the proportion of hospital deaths due to problems in care in English acute NHS trusts [press release]. 2015. [No longer available online but program description at https://www.rcplondon.ac.uk/projects/national-mortality-case-record-review-programme.]
8. National Quality Board, UK. *National Guidance on Learning from Deaths: A Framework for NHS Trusts and NHS Foundation Trusts on Identifying, Reporting, Investigating and Learning*

from Deaths in Care. March 2017. https://www.england.nhs.uk/wp-content/uploads/2017/03/nqb-national-guidance-learning-from-deaths.pdf. Accessed January 4, 2019.

9. The Royal College of Physicians. RCP hosts pioneering national Mortality Case Record Review Programme. February 15, 2016. https://www.rcplondon.ac.uk/news/rcp-hosts-pioneering-national-mortality-case-record-review-programme. Accessed January 4, 2019.
10. Hayward RA, McMahon LF Jr, Bernard AM. Evaluating the care of general medicine inpatients: how good is implicit review? *Ann Intern Med*. 1993;118(7):550-556.
11. Kobewka DM, Van Walraven C, Taljaard M, Ronksley P, Forster AJ. The prevalence of potentially preventable deaths in an acute care hospital: a retrospective cohort. *Medicine (Baltimore).* 2017;96(8):e6162.
12. Kobewka DM, Van Walraven C, Taljaard M, Ronksley P, Forster AJ. The prevalence of potentially preventable deaths in an acute care hospital: a retrospective cohort [Appendices]. *Medicine (Baltimore)*. 2017;96(8):e6162.
13. Manaseki-Holland S, Lilford RJ, Bishop JR, et al. Reviewing deaths in British and US hospitals: a study of two scales for assessing preventability. *BMJ Qual Saf*. 2017;26:408-416.
14. Zegers M, de Bruijne MC, Spreeuwenberg P, Wagner C, van der Wal G, Groenewegen PP. Variation in the rates of adverse events between hospitals and hospital departments. *I J Qual Health Care*. 2011;23(2):126-133.
15. Aranaz-Andrés JM, Aibar-Remón C, Limón-Ramírez R, et al. Prevalence of adverse events in the hospitals of five Latin American countries: results of the 'Iberoamerican study of adverse events' (IBEAS). *BMJ Qual Saf*. 2011;20(12):1043-1051.
16. Aranaz-Andrés JM, Aibar-Remón C, Vitaller-Burillo J, et al. Impact and preventability of adverse events in Spanish public hospitals: results of the Spanish National Study of Adverse Events (ENEAS). *I J Qual Health Care*. 2009;21(6):408-414.
17. Aranaz-Andrés JM, Aibar-Remón C, Vitaller-Murillo J, Ruiz-López P, Limón-Ramírez R, Terol-García E. Incidence of adverse events related to health care in Spain: results of the Spanish National Study of Adverse Events. *J Epidemiol Community Health*. 2008;62(12):1022-1029.
18. Baines RJ, Langelaan M, De Bruijne MC, et al. Changes in adverse event rates in hospitals over time: a longitudinal retrospective patient record review study. *BMJ Qual Saf*. 2013;22(4): 290-298.

19. Baines RJ, Langelaan M, De Bruijne MC, Wagner C. Is researching adverse events in hospital deaths a good way to describe patient safety in hospitals: a retrospective patient record review study. *BMJ Open*. 2015;5(7):e007380.
20. Baker GR, Norton PG, Flintoft V, et al. The Canadian Adverse Events Study: the incidence of adverse events among hospital patients in Canada. *Can Medl Assoc J*. 2004;170(11):1678-1686.
21. Best WR, Cowper DC. The ratio of observed-to-expected mortality as a quality of care indicator in non-surgical VA patients. *Medical Care*. 1994;32(4):390-400.
22. Brennan TA, Leape LL, Laird NM, et al. Incidence of adverse events and negligence in hospitalized patients: results of the Harvard medical practice study I. *N Engl J Medicine*. 1991;324(6):370-376.
23. Briant R, Buchanan J, Lay-Yee R, Davis P. Representative case series from New Zealand public hospital admissions in 1998 - III: adverse events and death. *N Z Med J*. 2006;119(1231).
24. Davis P, Lay-Yee R, Briant R, Ali W, Scott A, Schug S. Adverse events in New Zealand public hospitals II: preventability and clinical context. *N Z Med J*. 2003;116(1183):U624.
25. Davis P, Lay-Yee R, Briant R, et al. *Adverse Events in New Zealand Public Hospitals: Principal Findings From a National Survey*. Wellington, New Zealand: Ministry of Health; 2001.
26. Dubois RW, Brook RH. Preventable deaths: who, how often, and why? *Ann Intern Med*. 1988;109(7):582-589.
27. Dubois RW, Rogers WH, Draper D, Brook RH. Author's reply: does hospital mortality predict quality? *N Engl J Med*. 1988;318(24):1622.
28. Dubois RW, Rogers WH, Moxley JH III, Draper D, Brook RH. Hospital inpatient mortality. *N Engl J Med*. 1987;317(26):1674-1680.
29. Flaatten H, Brattebø G, Alme B, et al. Adverse events and in-hospital mortality: an analysis of all deaths in a Norwegian health trust during 2011. *BMC Health Serv Res*. 2017;17(1):465.
30. Gupta M, Fuchs B, Cutilli C, et al. Preventable mortality: does the perspective matter when determining preventability? *J Surg Res*. 2013;184(1):54-60.
31. Hayward RA, Hofer TP. Estimating hospital deaths due to medical errors: preventability is in the eye of the reviewer. *JAMA*. 2001;286(4):415-420.
32. Hogan H, Healey F, Neale G, Thomson R, Vincent C, Black N. Preventable deaths due to problems in care in English acute hospitals: a retrospective case record review study. *BMJ Qual Saf*. 2012;21(9):737-745.

33. Hogan H, Zipfel R, Neuburger J, Hutchings A, Darzi A, Black N. Avoidability of hospital deaths and association with hospital-wide mortality ratios: retrospective case record review and regression analysis. *BMJ*. 2015;351:h3239.
34. Martins M, Travassos C, Mendes W, Pavão ALB. Hospital deaths and adverse events in Brazil. *BMC Health Serv Res*. 2011;11.
35. Michel P, Quenon JL, Djihoud A, Tricaud-Vialle S, De Sarasqueta AM. French national survey of inpatient adverse events prospectively assessed with ward staff. *Qual Saf Health Care*. 2007; 16(5):369-377.
36. Roberts AP, Morrow G, Walkley M, et al. From research to practice: results of 7,300 mortality retrospective case record reviews in four acute hospitals in the North-East of England. *BMJ Open Qual*. 2017;6(2).
37. Soop M, Fryksmark U, Köster M, Haglund B. The incidence of adverse events in Swedish hospitals: a retrospective medical record review study. *Int J Qual Health Care*. 2009;21(4):285-291.
38. Sorinola OO, Weerasinghe C, Brown R. Preventable hospital mortality: learning from retrospective case record review. *JRSM Short Rep*. 2012;3(11):77.
39. Thomas EJ, Brennan TA. Incidence and types of preventable adverse events in elderly patients: population based review of medical records. *BMJ*. 2000;320(7237):741-744.
40. Thomas EJ, Studdert DM, Brennan TA. The reliability of medical record review for estimating adverse event rates. *Ann Intern Med*. 2002;136(11):812-816.
41. Thomas EJ, Studdert DM, Burstin HR, et al. Incidence and types of adverse events and negligent care in Utah and Colorado. *Medical Care*. 2000;38(3):261-271.
42. Thomas EJ, Studdert DM, Newhouse JP, et al. Costs of medical injuries in Utah and Colorado. *Inquiry*. 1999;36(3):255-264.
43. Wilson RM, Runciman WB, Gibberd RW, Harrison BT, Newby L, Hamilton JD. The quality in Australian health care study. *Med J Aust*. 1995;163(9):458-471.
44. Zegers M, Bruijne MC, Wagner C, Groenewegen PP, Waaijman R, Wal G. Design of a retrospective patient record study on the occurrence of adverse events among patients in Dutch hospitals. *BMC Health Serv Res*. 2007;7(1):1.
45. Zegers M, de Bruijne MC, de Keizer B, et al. The incidence, root-causes, and outcomes of adverse events in surgical units: implication for potential prevention strategies. *Patient Saf Surg*. 2011;5:13.
46. Zegers M, De Bruijne MC, Wagner C, et al. Adverse events and potentially preventable deaths in Dutch hospitals: results of a

retrospective patient record review study. *Qual Saf Health Care*. 2009;18(4):297-302.

47. Rubenstein LV, Kahn KL, Harrison E, Sherwood MJ, Rogers WH, Brook RH. *Structured Implicit Review of the Medical Record*. Santa Monica, CA: RAND Corporation; 1991.
48. Pearson ML, Lee JL, Chang BL, Elliott M, Kahn KL, Rubenstein LV. Structured implicit review: a new method for monitoring nursing care quality. *Medical Care*. 2000;38(11):1074-1091.
49. Gibbs J, Clark K, Khuri S, Henderson W, Hur K, Daley J. Validating risk-adjusted surgical outcomes: chart review of process of care. *Int J Qual Health Care*. 2001;13(3):187-196.
50. Rubin HR, Rogers WH, Kahn KL, Rubenstein LV, Brook RH. Watching the doctor-watchers. How well do peer review organization methods detect hospital care quality problems? *JAMA*. 1992;267(17):2349-2354.
51. Fischhoff B. Hindsight is not equal to foresight: the effect of outcome knowledge on judgment under uncertainty. *Journal of Experimental Psychology: Human Perception and Performance*. 1975; 1(3):288.
52. Fischhoff B. Hindsight ≠ foresight: the effect of outcome knowledge on judgment under uncertainty. *Qual Saf Health Care*. 2003;12(4):304-311.
53. Brett AS, Roberts M, Ruderman AJ, et al. Does hospital mortality predict quality? *N Engl J Med*. 1988;318(24):1622-1624.
54. Hanskamp-Sebregts M, Zegers M, Vincent C, van Gurp PJ, de Vet HCW, Wollersheim H. Measurement of patient safety: a systematic review of the reliability and validity of adverse event detection with record review. *BMJ Open*. 2016;6:e011078.
55. Cohen J. A coefficient of agreement for nominal scales. *Educational and Psychosocial Measurement*. 1960;20:37-46.
56. Maclure M, Willett WC. Misinterpretation and misuse of the kappa statistic. *Am J Epidemiol*. 1987;126(2):161-169.
57. Chmura Kraemer H, Periyakoil VS, Noda A. Kappa coefficients in medical research. *Statistics in Medicine*. 2002;21(14):2109-2129.
58. Kahn KL, Rubenstein LV, Sherwood MJ, Brook RH. *Structured Implicit Review for Physician Implicit Measurement of Quality of Care*. Santa Monica, CA: RAND Corporation; 1989. http://www.rand.org/pubs/notes/N3016/. Accessed January 9, 2018.
59. Chow E, Harth T, Hruby G, Finkelstein J, Wu J, Danjoux C. How accurate are physicians' clinical predictions of survival and the available prognostic tools in estimating survival times in terminally ill cancer patients? A systematic review. *Clin Oncol (R Coll Radiol)*. 2001;13(3):209-218.

60. Christakis NA, Lamont EB. Extent and determinants of error in doctors' prognoses in terminally ill patients: prospective cohort study. *BMJ*. 2000;320(7233):469-473.
61. Glare P, Virik K, Jones M, et al. A systematic review of physicians' survival predictions in terminally ill cancer patients. *BMJ*. 2003;327(7408):195.
62. Rochester region perinatal study. Medical review project, Empire State Medical, Scientific and Educational Foundation, Inc. *N Y State J Med*. 1967;67(9):1205-1210.
63. Richardson FM. Peer review of medical care. *Med Care*. 1972; 10(1):29-39.
64. Sheps MC. Approaches to the quality of hospital care. *Public Health Rep*. 1955;70(9):877-886.
65. Hogan H. The problem with preventable deaths. *BMJ Qual Saf*. 2016;25:320-323.
66. Reason J. *Human Error*. New York, NY: Cambridge University Press; 1990.
67. Reason J. Beyond the organisational accident: the need for "error wisdom" on the frontline. *Qual Saf Health Care*. 2004;13(suppl 2):ii28-ii33.
68. Kahneman D. A perspective on judgment and choice: mapping bounded rationality. *Am Psychol*. 2003;58(9):697.
69. Saposnik G, Redelmeier D, Ruff CC, Tobler PN. Cognitive biases associated with medical decisions: a systematic review. *BMC Med Inform Decis Mak*. 2016;16(1):138.
70. Blumenthal-Barby J, Krieger H. Cognitive biases and heuristics in medical decision making: a critical review using a systematic search strategy. *Med Decis Making*. 2015;35(4):539-557.
71. Rogne T, Nordseth T, Marhaug G, et al. Rate of avoidable deaths in a Norwegian hospital trust as judged by retrospective chart review. *BMJ Qual Saf*. 2018;pii:bmjqs-2018-008053. https://doi.org/10.1136/bmjqs-2018-008053 [Epub ahead of print].

Funding/Support: PJC and RJL were supported by the National Institute for Health Research (NIHR) Collaboration for Leadership in Applied Health Research and Care (CLAHRC) (grant #IS-CLA-0113-10018) West Midlands. This paper presents independent research and the views expressed are those of the

author(s) and not necessarily those of the NHS, the NIHR, or the Department of Health and Social Care.

Conflict of Interest Disclosures: All authors completed the ICMJE Form for Disclosure of Potential Conflicts of Interest. No conflicts were reported.

Acknowledgments: We would like to politely thank the study authors (14/23 of whom we contacted) who provided further information for this systematic review. This has helped to improve the systematic review content and quality.

Address correspondence to: Timothy P. Hofer, VA Center for Practice Management and Outcomes Research, PO Box 130170, Ann Arbor, MI 48113 (email: [redacted])

Supplementary Material

Additional supporting information may be found in the online version of this article at http://onlinelibrary.wiley.com/journal/10.1111/(ISSN)1468-0009:

Online Appendix 1. Search Strategies and Results
Online Appendix 2. Methods and Results for Estimating the Number of Reviews Needed
Online Appendix 3. List of Excluded Studies
Online Appendix 4. Details of Mortality Review Process by Study
Online Appendix 5. Discussion of the Findings of Methods Used for Reviewing Case Notes in Our Included Studies

1.3. Appendices

Online appendices can be accessed under “Supporting information” section through the following URL:

https://onlinelibrary.wiley.com/doi/full/10.1111/1468-0009.12375

Having considered issues around the fidelity of hospital mortality statistics, I conclude that these measures are flawed and ill-suited to the adequate capture of the care quality construct. Without the use of hospital mortality statistical measures, alternative methods must be considered for the purpose of measuring hospital care quality. I shall now consider whether case-note reviews can satisfy this role and explore whether psychological (cognitive) biases plausibly influence case-note reviewer care quality judgements.

CHAPTER 2: USING BEHAVIOURAL THEORY TO CHARACTERISE COGNITIVE BIASES AND HEURISTICS IN CASE RECORD REVIEWS: A CONCEPTUAL STUDY

2.1. Chapter Preface

Case-note reviewers reviewing the same case-note have considerable disagreement across their care quality judgements over the same set of case-notes. In Chapter 2, I will explore cognitive biases and heuristics and human factors as a plausible source of this disagreement. Clinical care of patients have been demonstrated to be cognitively and emotionally challenging.(Fox, 1989, Byrne, 2013) To orient this investigation, we expect cognitive biases and heuristics and human factors to influence care quality judgements during case-note reviews. Thus, we endeavour to explore the plausible contribution of cognitive biases and decision heuristics to case-note reviewer care quality judgements. This chapter underscores the need to direct attention to identify the plausible influence of cognitive biases and heuristics upon case-note reviewer care quality judgements.(Te et al., 2017)

2.2. Abstract

Objective: Identify which cognitive biases and heuristics influence case-note reviewer care quality judgements and how they influence these judgements.

Data sources: Systematic literature search pooled articles from PubMED, Health Management Information Consortium (HMIC), PsycINFO, Cochrane Library; recent systematic reviews identified longlist of cognitive biases and heuristics

Study design: Systematic literature search, hypothesis, pilot study

Data collection/extraction methods: During the systematic literature search, the Cochrane Data Extraction form for randomised control and non-randomised control trials was used to extract study information. Any quantitative data quality was assessed using the Effective Public Health Practise Project (EPHPP) Quality assessment tool. From two systematic reviews, the articles of the first studies demonstrating the biases' effect and the medical decision-making articles had the data extracted for salient features and patterns relating to the observable influence of each bias/heuristic by the lead author and subsequently screened by behavioural scientist (KAS). An expert panel exercise was conducted with nine panel members with relevant clinical and academic expertise. The heuristics/biases considered were identified from these two systematic reviews. Guided by dual process theory, an established theory which accounts for lapses in human decision-making, all panel members collaboratively developed clinical scenarios each demonstrating each biases' influence upon clinical reviewer care quality judgements. All members independently rated the likely influence of each bias on care quality judgements using a five-point Likert scale, then dichotomised to highlight the percentage of panel members identifying each bias as likely to influence case-note reviews. These percentage ratings were subsequently rank ordered.

Principal findings: One study, on the outcome bias, was found with a quality rating of 1, the highest rating on the EPHPP risk of bias assessment. The data extraction of salient features identified no studies and no clear way to develop a taxonomy to explore the judgment of case-note reviews. Of the 23 distinct heuristics/biases considered by the panel, 15 were identified with the potential to influence care quality judgements. One clinical scenario was developed for each bias. The panel members' independent ratings (ICC = 0.74) prioritised the following biases as having the most likely influence: ambiguity intolerance, outcome bias, availability bias, confirmation bias and omission bias. Using dual process theory, strategies are developed to show how these biases can be mitigated.

Conclusions: Several heuristics and biases have the potential to influence reviewer care quality judgements during case-note reviews. We propose potential mitigating strategies and opportunities for future research.

Keywords: cognitive biases, heuristics, decision-making, quality improvement methodologies, healthcare quality improvement, standards of care

2.2.1. What is known on this topic?

- Cognitive biases and heuristics are widely recognised to influence medical decision-making.
- There is considerable unexplained variation between different case-note reviewers in their assessments of the care quality of the same batch of case-notes.
- Whether cognitive biases and heuristics explain this variation in case-note reviewer care quality judgements remains unknown.

2.2.2. What this study adds

- Clinical scenarios describing how cognitive biases and heuristics might influence case-note review care quality judgements help to facilitate their evaluation.
- Key cognitive biases and heuristics (*ambiguity intolerance, outcome bias, availability bias, confirmation bias and omission bias*) are prioritised as influencers of case-note review care quality judgements.
- Behavioural science research offers a model for evaluating methods for identifying and mitigating these cognitive biases and heuristics during case-note review care quality assessments.

2.3. INTRODUCTION

Retrospective hospital case-note reviews are a common component of assessing care quality(Manaseki-Holland et al., 2019b), informing quality improvement initiatives(The Royal College of Physicians, 2016d), and medico-legal processes.(Rogers and Pleat, 2010) A case-note is alternatively known as or is derived from a medical record. During mortality case-note reviews, expert clinical reviewers look for evidence of the care patients received before their death and then provide a judgement of care quality.(Manaseki-Holland et al., 2019b) In the United States, these reviews are used in medico-legal processes(Sullivan and Anderson, 2010), for hospital accreditation(Makary et al., 2011) and adverse event research(Edwards, 2013). In the United Kingdom, National Health Service (NHS) hospital trusts are required to "*publish an updated policy…on how it responds to, and learns from, deaths of patients*" using mortality case-note reviews.(National Quality Board UK, 2017) Though the importance of mortality case-note reviews are widely acknowledged, there are

concerns about their reliability, as several studies have shown substantial levels of disagreement between expert clinical reviewers when undertaking independent examination of the same records(Goldman, 1992) (Hofer et al., 2000) The Royal College of Physicians has introduced structured judgement reviews in an attempt to harmonise the way reviewers analyse and interpret case records.(The Royal College of Physicians, 2016b) However, there is no 'diagnostic framework' for characterising the sources for these inter-reviewer disagreements, which leaves both methodologists and health practitioners uncertain of the methods useful for its mitigation.(Manaseki-Holland et al., 2019b) The ramification of this persisting uncertainty around inter-reviewer reliability extend into healthcare quality improvement(Edwards, 2013, Marcin et al., 2018, Forster et al., 2012) and into courts of law. (Gowensmith et al., 2012) For instance, medico-legal rulings tend to be intuitive decisions which are likely subject to these cognitive biases and heuristics.(Guthrie et al., 2007) Thus, we take the perspective in developing the case that cognitive biases and heuristics likely influence case-note reviewer care quality judgement. Cognitive biases and heuristics are widely regarded as active during all types of decisions.(Tversky and Kahneman, 1974, Kahneman, 2011) However, no evidence of their influence has been demonstrated during the clinician's care quality evaluation of these case-notes. And nor has any impact of this phenomena been studied in relation to these clinician judgements.

Ergo, we seek to develop a case for the plausible influence of these cognitive biases during case-note review care quality judgements and go on to suggest ways in which we could mitigate these cognitive biases and heuristics.

The focus on cognitive biases and heuristics were selected, because their influence upon medical decision-making has already been demonstrated across a range of clinical specialties and so plausibly extends to the care quality judgements of clinicians reviewing

case-notes. (Blumenthal-Barby and Krieger, 2015a) In addition, the case-note review process is cognitively demanding which likely leads reviewers to unintentionally employ heuristics that could bias decisions, as clinical reviewers experience cognitive limits in receiving, storing, retrieving, transmitting information.(Dobler et al., 2019) Heuristics are problem solving "rules of thumb" that work in various situations and are subject to their misapplication.(Tversky and Kahneman, 1974) One example is the "representativeness heuristic", e.g. a reviewer may overestimate the likelihood of a patient dying from a dramatic event such as a cancer or a high-speed accident and so fail to recognize inappropriate care quality. Cognitive biases are systematic patterns of deviation from rational norms in judgements. One example is the "outcome bias", e.g., a reviewer may unintentionally judge care quality more harshly when they have foreknowledge the patient had died compared to if they were unaware the patient was deceased.(Blumenthal-Barby and Krieger, 2015a) For simplicity, the current paper will refer to heuristics and cognitive biases from hereon in as "biases."

The aim of our research is to describe and prioritise biases which could influence reviewer judgments and demonstrate their proposed influence through clinical scenarios. We do this so that the biases can be made explicit in training and be targets for future interventions which can help to identify and minimise any negative influence upon clinician reviewer care quality judgements. To do this, we identified a list of biases by rationalising, via individual expert panel judgements, those biases most likely to influence case-note reviewers, and then employed a panel consensus procedure to prioritise biases based on clinical scenarios.

2.3.1. Dual process theory: a way to think about decision-making

One way to approach this problem of developing a diagnostic framework is to treat variability in reviewing and adjudicating case records as a form of modifiable behaviour. Traditionally, behaviour change policies and interventions have tended to focus on providing new *information*, which seek to change the way people think about their behaviour or different [financial or legal] *incentives* that change the consequences of behaviour.(Cecchini et al., 2010) These interventions can only get us so far, because changing reflective or rational intentions would account for less than one-third of the variance in behaviour.(Webb and Sheeran, 2006) In contrast to economic (and social science) models of rational choice suggesting that we respond to information and price signals, insights from across the behavioural sciences suggest that human behaviour is actually led by our very human, and often fallible brain, and influenced greatly by the environment within which decisions are taken.(Kahneman, 2011) The 'dual process' model has been proposed as a theoretical basis for understanding health behaviours.(Marteau et al., 2012, Sheeran et al., 2013) In particular, psychologists and neuroscientists have recently converged on a description of brain functioning that is based on two types of cognitive processes, also interpreted as two distinct systems [or sets of systems]: evolutionarily older 'System 1' processes described as automatic, uncontrolled, effortless, associative, fast, unconscious and affective, and more recent, characteristically human 'System 2' processes described as reflective, controlled, effortful, rule-based, slow, conscious and rational.(Evans, 2008) Neurobiological evidence of separate brain structures for automatic processes provides substantial support to this model.(Anderson et al., 2004)

Those two types of mental processes (the reviewers' conscious (planned) and automatic (heuristic) thought) are involved in reviewing a case record and determine the accuracy of judgements. Such judgments are affected by training and by attending to environmental factors (i.e., time of day, avoiding fatigue, freedom from interruptions etc.). The conscious aspects of reasoning are likely stimulated by the professional requirement to undertake case record reviews, or the desire to participate in research and quality improvement for example. But the automatic, heuristic aspects of thinking may be obscure to the reviewer as they operate mainly at a subconscious level and are subject to many potential biases of which the reviewer may be unaware. These biases and heuristics need to be mapped to a mitigation framework to make it easier to address them in a systematic manner (Dolan et al., 2012) (see Figure 3).

Policies that change the context or 'nudge' people in particular directions have captured the imagination of policymakers at the same time as the limitations of traditional approaches have become apparent.(Hofmann et al., 2008) Popularised in Richard Thaler and Cass Sunstein's book *Nudge*, the theory underpinning many of the policy suggestions are built on decades of research in the behavioural sciences, and particularly the growing field known as 'behavioural economics.'(Thaler and Sunstein, 2008) Behavioural economics is the science of human decision making, which combines insights from economics and psychology, provides new ways to think about the barriers and drivers to a range of behaviours.

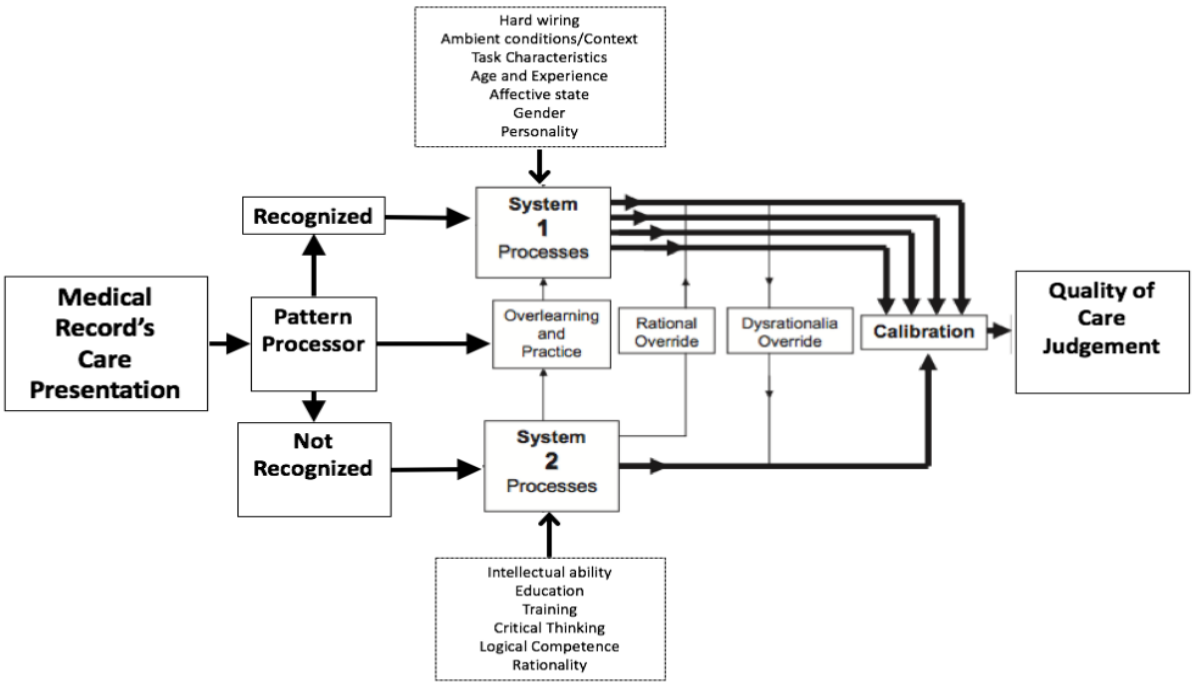


Figure 3. Model for diagnostic reasoning based on pattern recognition and dual process theory. This model has been adapted with permissions from Croskerry et al. 2009

2.4. METHODS

The methods sections present three main junctures which reflects the novelty and exploratory nature of this research. At each juncture, the next steps to take were carefully guided by the rationale to identify cognitive biases and heuristics which influence case-note review care quality judgements. Due to the exploratory nature of this chapter, the methods and the contents evolved to answer the research aim to identify any influence of cognitive biases and heuristics on the care quality judgements of the case-note reviewer.

2.4.1. Step 1: Systematic literature search[16]

In March 2017, a systematic review search was undertaken to identify in physicians and nurses, which cognitive biases, heuristics and human factors influence the reviewing and assessment of hospital medical records or "case-notes." (Appendix 1a: PROSPERO form and the Appendix 1b: PRISMA flow diagram)

2.4.2. Step two: Conceptual review of the literature

Due to the underwhelming number of cognitive biases and heuristics studies on case-note reviews and medical records, we decided to undertake a conceptual review of the literature for cognitive bias and heuristic studies which related to medical decision-making, a contiguous topic to our own. Having identified two systematic reviews, the review narrative and conclusions did not present clear indicators for which biases or heuristics would influence medical records or case-note review care quality judgements. Due to the lack of clear direction, AT independently extracted data from these two reviews and collected information relating to the following: origin, study characteristics (sample size, intervention, demographics) exploring medical decision-making and the potential extension to case-note reviewing care quality judgement context. Features noted include the types of study designs and processes plausibly relating to case-note review/diagnostic reasoning/clinical decision-making and the study limitations. The overall aim of this activity was to find indications any cognitive biases and heuristics extending to the case-note review setting. No studies were

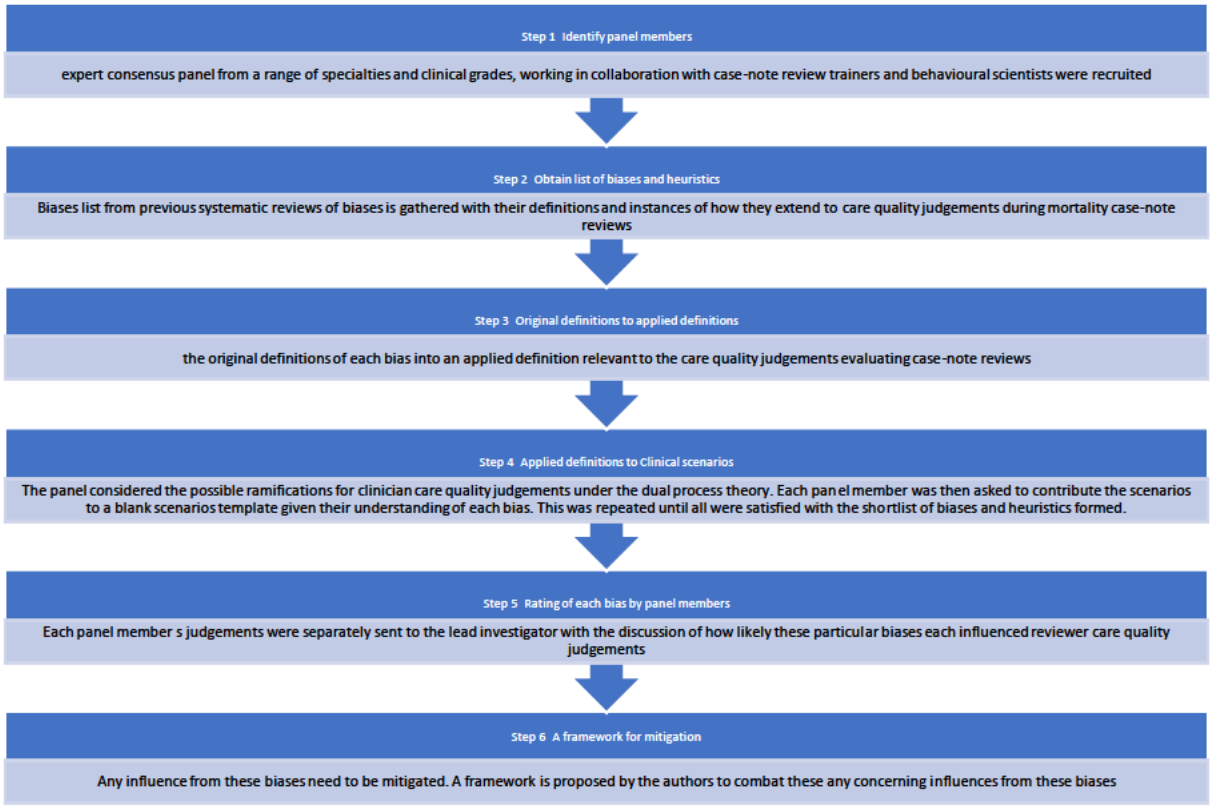
[16] This systematic is now out-of-date and was undertaken as a prelude to step 2. There was no need to update this search as it served its purpose as a heuristic to identify study sources (i.e., systematic reviews) on cognitive biases and heuristics.

found which were directly generalisable to the case-note review or medical records setting. Neither did a clear taxonomy or hierarchy of cognitive biases and heuristics present itself for this setting. Such a taxonomy could have helped to identify the cognitive processes which could similarly operate in our case-note review context. (See Flowchart 1)

2.4.3. Step three: Consensus panel

Due to these evidence gaps relating to the study of cognitive biases and heuristics in case-note reviews, we applied current behavioural theory and identified cognitive biases and heuristics from medical decision-making[17] literature to undertake a panel consensus process which explored the plausibility of these cognitive biases and heuristics (see Flowchart 1 below.)

[17] a contiguous field to decision-making with case-note reviews



Flowchart 1. The consensus panel process

Our panel process had six steps. This first step involved identifying members of the expert consensus panel from a range of specialties and clinical grades, working in collaboration with case-note review trainers and behavioural scientists. Note, the sample size of the consensus panel was not important as this was a qualitative exercise to capture the perceptions of relevant local experts. The concept was more important than the sample size. However, upscaling of this will involve more expert participants from different regions of the UK and countries.

The second step involved creating a list of biases from previous systematic reviews of biases shown to influence medical decision-making that, by their definitions, could extend to care quality judgements in mortality case-note reviews.

The third step involved transforming the original definitions of each bias into an applied definition relevant to the care quality judgements evaluating case-note reviews informed by the dual process theory.

The fourth step had members of the expert panel meet to transform the applied definitions into clinical scenarios to demonstrate how each bias may influence care quality judgements. Our panel considered the possible ramifications for clinician care quality judgements under the dual process theory. Each panel member was then asked to contribute their own scenarios to a blank scenarios template given their understanding of each bias. The first author of this manuscript (AT) revised all clinical scenarios through iterative and collaborative consultations with expert panel members until all members were satisfied with the resultant scenarios' representation of each bias, its definition and clinical scenario.

The fifth step involved expert panel members independently rating the likelihood of each bias influencing these care quality judgements. To do so, AT emailed each panel member the final list of biases, their applied definitions, and their clinical scenarios, along with a link to an anonymous online survey, which they submitted their responses through. In the survey, each panel member *indicated the extent to which each heuristic/bias is likely to influence a reviewer's global care quality judgement during a mortality case-note review on the quality of care received* from only one of the five Likert categories: "very unlikely", "unlikely", "not unlikely", "likely" or "very likely" (Appendix *2*).

The sixth step involved pooling the most pertinent biases, as assessed by the authors, from the survey. We then applied dual process theory as a framework to help us to describe the system 1-type effects of these biases, the mitigating actions required by a system 2-type intervention(s) and the predicted outcome following these system 2-type intervention(s) upon case-note reviewer care quality judgements.

Survey responses were analysed using SPSS (v. 26). The five response categories were collapsed to dichotomous categories, which more clearly establish ratings as either "*likely*" (composite of "likely" or "very likely") or "*not likely*" (composite of "very unlikely", "unlikely", "neither likely nor unlikely") to influence care quality judgements. The intraclass correlation (ICC) estimates and their 95% confident intervals for the dichotomised scores, were calculated based on a mean-rating, absolute-agreement, 2-way random-effects model. The resultant percentage for each bias was then rank ordered, in descending order from the bias most likely to the least likely to influence care quality judgements.

2.5. RESULTS

2.5.1. Systematic literature search

The research team initially conducted a systematic literature search that described cognitive biases, heuristics and human factors identified as influencing case-note reviews (see Appendix 3).(Te et al., 2017) The aims of this review were as follows:

- To identify within physicians and nurses, which cognitive biases, heuristics and human factors influence the reviewing and assessment of hospital medical records.
- To understand how these cognitive biases, heuristics and human factors influence the reviewing and interpretation of hospital medical records.

It was a team decision for the "human factors" term to be removed as this search term yielded too many diverse study designs, diversity of outcomes which were not amenable to a critical and focussed narrative discussion. The reason for this was due to the undefined relationship between "human factors" and "cognitive biases and heuristics." Cognitive biases and

heuristics were undefined, and yet unexplored in this topic area and so it was not pertinent to include another topic, human factors, to bulk up our systematic review search and its content. Despite the best intentions to explore both these topics areas, there is no clear extant framework, model or theory which broaches these two topics. Consequently, one of the topics had to be chosen over the other. Cognitive biases and heuristics were selected over human factors for two reasons. Firstly, cognitive biases and heuristics is more directly linked to reviewer care quality judgements, as cognitive biases and heuristics has been strongly evidenced to influence many aspects of judgements.(Heller and Saltzstein, 1992, Marewski and Gigerenzer, 2012b, Kahneman et al., 1982, Finucane et al., 2000, Weber and Stern, 2011, Weber, 2010, Kahneman, 2011, Kahneman et al., 2021) We considered it a more likely source of variation in the reviewer care quality judgement as the mechanisms of it are more distinctly related to decisions and judgements than human factors. There is also little to substantiate any mechanisms for how human factors come to influence clinical decision-making. All we had was that human factors might be implicated but there are an innumerable number of such objects, and the most likely of them was cognitive biases and heuristics. Secondly, the cognitive biases and heuristics is better defined and circumscribed than the human factors literature in that the biases and heuristics have paradigmatic studies which illustrate their effects. This is not the case with human factors as there are no set paradigmatic studies which also vary considerably from context-to-context. Without this, our team could not be certain that we had captured all if not most of the relevant literature for our purposes. Thus, it as a joint decision to not include the “human factors” search term on an undefined and potentially unbounded topic. Human factors are synonymous with the term “ergonomics.” I include a list of “human factors” and “ergonomics” studies obtained from the search, which is exhibited in Appendix 4 to indicate their sheer diversity. Our team

subsequently decided to focus on cognitive biases and heuristics, which was a better defined and focussed field for our research question.

As this search only located one study concerning "*outcome bias*", this was insufficient information to inform a meaningful discussion for future research directions. The characteristics of pertinent studies were extracted and are presented in Table

Table 4. Characteristics of studies included in the systematic literature search

Author	**Year**	**Country**	**Number of participants**	**Aim**	**Summary**	**Methods**	**Cognitive bias**	**Heuristics**	**Human Factors**
Caplan *et al.*	1991	USA	112 anaesthesiologists (of possible 115)	Explore the influence of temporary and permanent injury outcomes on the assessment of care quality	112 practicing anaesthesiologists judged the appropriateness of care in 21 cases involving adverse anaesthetic outcomes. Alternative outcomes were generated for cases where the care quality, in general, could have led to either a permanent or temporary injurious outcome. Independent ratings were given by each reviewer on each as appropriate, inappropriate, or	Case record review	Outcome bias	-	-

					impossible to judge. The proportion of ratings for appropriate care decreased by 31 percentage points when the outcome was changed from temporary to permanent and increased by 28 percentage points when the outcome was changed from permanent to temporary.				

Caplan et al. was included as it satisfied the study criteria of cognitive biases/heuristics, decision-making from clinicians and it being relatable to case-note reviews or medical records. The aim of the study was to determine whether a permanent injury was more likely to elicit a rating of inappropriate care than a temporary injury. 115 anaesthesiologists originally agreed to participate, with 112 (97%) completed their case reviews. Reviewers were in 36 states plus Washington, DC. All 30 US administrative districts of the American Society of Anaesthesiologists were represented. The average (SD) age of reviewers was 51 ± 8 years, and the average number of years in practice was 20 ± 8 years. The study had a retrospective case record review design with the random allocation of case outcomes (i.e., permanent, or temporary harm) to reviewing clinicians. There were 21 distinct cases which had their outcomes altered (i.e., temporary to permanent or vice-versa). Thus, there were 42 permutations from these 21 cases. The cases were randomly allocated to the anaesthesiologist case-note reviewers for their care quality assessment by these options “Appropriate care” or “Inappropriate care” or “Impossible to judge.”(Robert A. Caplan, 1991)

The study found that the proportion of ratings for appropriate care decreased by 31 percentage points when the outcome was changed from temporary to permanent. In the converse direction, the percentage points increased by 28 points when the outcome was changed from permanent to temporary. We conclude that knowledge of the severity of outcome can influence the reviewer's judgment of the appropriateness of care.

The quality assessment, using an adapted Cochrane Data Collection form for RCTs and non-RCTs(Cochrane Collaboration, 2019) with the Effective Public Health Practise Project (EPHPP) Quality assessment tool(Effective Public Health Practise Project (EPHPP), 2018), rated this study as good quality because of reviewer-matched cases receiving both outcome

forms of the same case (permanent-to-temporary and temporary-to-permanent outcomes), a range of clinical complications presented (eight in total), appropriate selection of statistical tool and the random allocation of cases (case permutations) to the reviewers (for full data extraction information see (Appendix 5: Cochrane Systematic Review Data extraction tool). There are limitations to this study. First, the alteration of outcomes may have not been equivalent across all the case-notes. For instance, let us consider two outcomes. The "*loss of hearing*" following the incident is a permanent outcome. However, the "*loss of hearing and subsequent employment*" is a worser permanent outcome than the former statement. Thus, the outcome bias effects could have been a consequence of imbalances during the construction of these contrived outcomes. Second, only anaesthesiologists were recruited into this study. There is reasonable chance that these results may not generalise to other clinician groups. Third, this study was undertaken with US-accredited clinicians. The exposure to the US culture and education may reduce or increase the effect from the outcome bias. Due to the time elapsed since the study, we cannot know the proportion of all clinicians who were US-accredited.

A single study with this design, with its associated flaws, was not sufficient material for us to develop a standalone article to elicit the significance of cognitive biases and heuristics. In the least, we would only be qualified to discuss the influence of the outcome bias, and not cognitive biases and heuristics in general, which was the aim of this systematic literature search. Thus, rather than base our findings on this single study, this paper sought to undertake a conceptual review of the literature, which are sourced from systematic reviews of cognitive biases and heuristics in medical decision-making.

2.5.2. Conceptual review of the literature

Second, the two reviews had the study information extracted by cognitive biases or heuristic. Interventions included qualitative (reflective writing, group discussion, interviews), vignettes, different presentation of messages, educational materials, emotion induction[18], survey and questionnaires. Several studies were shortlisted for further exploration. However, it remained unclear which biases we needed to select as the study method was adapted for its specific study question and did not directly concern the case-note review context. There were no framework or schemata in existence which could guide the identification of cognitive biases and heuristics pertinent to case-note review care quality judgements. The same issue applied to the article's study design which satisfied the requirements of their study question (i.e., specialty or health condition). This exercise did not identify specific cognitive biases and heuristics which influence case-note review care quality judgements. Neither did this activity identify study designs with which we could explore any bias influence upon case-note reviewer care quality judgements due to the sheer diversity of methodologies found. Rather than base our findings on results of the systematic literature search and the conceptual review, this article proposes that cognitive components, specifically cognitive biases, and heuristics operating through system 1 processes, do plausibly influence the care quality judgements of case-note reviewers during case-note review. This indicates a paucity of clinical studies concerned with cognitive biases and heuristics. There was a diversity of study designs, population. More details around this conceptual review search and its results can be found in Appendix 6.

[18] e.g., bursting balloon to induce anxiety

2.5.3. Consensus panel

A consensus panel was undertaken to rank the biases for their plausible significance to case-note review care quality judgements. In step 1, the consensus panel included nine individuals (all co-authors on the present manuscript) providing expertise in clinical practice (n=6), training case-note reviewers (n=5), research related to biases influencing medical decision-making (n=3), and a PhD student and a doctor in training[19] who reviewed the literature on case-note reviews and biases. Further details about each panel member and their contributions are present in Appendix 7.

In step 2, we identified two literature reviews describing biases that influence medical decision-making (see Appendix *2*). Blumenthal-Barby *et al.*'s review included papers on medical decisions made in a clinical setting by either patients or medical personnel(Blumenthal-Barby and Krieger, 2015a); Saposnik *et al.*'s review differed as they identified studies involving only physicians.(Saposnik et al., 2016) Nineteen biases were found in Blumenthal-Barby *et al.*'s review and 16 in Saposnik *et al.*'s. Thirteen of the biases were repeated across reviews, leaving 22 unique biases to consider. However, because Saposnik *et al.*'s bias list did not clearly distinguish between the "ambiguity aversion" and "ambiguity intolerance" biases, further literature was consulted to determine whether to treat them as one or two biases.(Tanaka et al., 2015) As it was determined they were two distinct constructs, one additional unique bias was considered. Next, of these 23 unique biases, 8 were excluded by logical application of their original definitions. For example, the "impact

[19] Who has since become a diabetic registrar

bias" describes a tendency to overestimate the length or the intensity of future emotional states, and as case-note reviewers are not asked to estimate the duration or intensity of future emotional states in their care quality judgements, therefore this bias cannot logically extend to case-note reviewer's care quality judgements. The shortlist of fifteen biases considered in the current review is displayed in the first column of Table 5 with full steps in Appendix 8: Shortlisting process for Biases. A full list of the biases was compiled, and any reasons for re-organisation or exclusion are provided in Appendix 9: Excluded biases.

In combining step 3 and 4, the developed case-note review applied definitions and clinical scenarios are found in second and third columns of Table 5. Each bias has one scenario. The scenarios varied across clinical areas and procedures, e.g., about medication, surgery, and staff clinical fatigue. The scenarios were between fourteen to thirty-one words in length. Medical terms were cross-checked for accuracy with both Black's (US) and Webster's (UK) medical dictionaries.

Table 5. Shortlist of system 1 biases, their case-note review applied definitions and "system 1" clinical scenarios applied to clinician case-note reviewer care quality judgements

System 1 Biases	Bias definition applied to case-note reviews	"System 1" clinical scenario[20]
Affect Heuristic	Reviewer care quality judgements may be influenced by incidental (case review unrelated) and integral (case review related) affective states	Clinicians in a state of negative affect can consider information more critically and with higher accuracy or other emotions that may modulate care quality judgements. For example, a heavily sleep deprived reviewer, may have her care quality rating more heavily influenced by affect from the case which may stimulate a higher (or lower) level of scrutiny above that of a routine case review. The care quality ratings may consequently be lower (or higher).
Ambiguity Aversion	Reviewer care quality judgements may be influenced by how the	Warfarin and Rivoroxaban are comparably effective anti-coagulants.(Cohen et al., 2016) However, the latter has a smaller evidence base. All things being equal, the medical team are accustomed to, and prefer using, warfarin. In the absence of clear

[20] For the demonstrative purposes of these scenarios, it is assumed higher ratings equate to better care, lower ratings poorer care. (i.e., a 5-point scale, 1 = very poor, 5 = very good)

	reviewer perceives therapeutic risks for evidence treatments compared to less or non-evidenced therapeutics	evidence, this judgement has been made because of ambiguity over the effectiveness of Rivoroxaban. The reviewer is also inclined to prefer warfarin because there is far more evidence for warfarin's effectiveness than Rivoroxaban. Thus, ambiguity aversion leads to higher (better) care quality judgements when warfarin is used compared to newer anti-coagulating agents.
Ambiguity Intolerance	Reviewer care quality judgements influenced by self-perceived uncertainty relating to case-note content	A reviewer examines a near complete set of case-notes and grades the care quality received for this case as exemplary. A different reviewer grades the same case-notes as merely "good" because of ambiguity aversion caused by the absence of information. The first reviewer is tolerant to missing information and gives a higher rating whilst the second reviewer is intolerant to missing information and gives a lower rating.
Anchoring Bias	Reviewer care quality judgements may be influenced by the past numerical value (i.e., through exposure to past numbers/figures which are	During clinic work, the reviewer gynaecologist routinely reviews blood test results. She then undertakes a retrospective case-note review of a male patient whose haemoglobin is 125 grams per litre (normal range for males = 138-172 g/L). She does not consider this a concern as she has been "anchored" to the normal female haemoglobin range of 121-151 g/L in their most recent exposures to clinical practice.

	quasi-fixed procedural standards)	
Availability Bias	Reviewer care quality judgements may be influenced by personal clinical experiences which then shape their perceived frequency of future events	A retrospective case-note reviewer who has been an expert witness in a coroner's court case because of a patient who died from cardiac arrest secondary to hyperkalaemia, will likely be more attentive to serum potassium levels, throughout a retrospective case-note review process. He may account harm to hyperkalaemia, even though there are other erroneous causes of serum hyperkalaemia e.g., clotted blood sample.
Bandwagon Effect	Reviewer care quality judgements may be influenced by the widespread, popular adoption of a medical procedure or treatment	A reviewer observes that an acute kidney injury (AKI) treatment protocol is not used. In the reviewer's hospital, the AKI protocol has gained little traction. The reviewer overlooks this omission of the AKI protocol and gives a higher care quality judgement score compared to another reviewer from a hospital which has fully adopted the protocol to his routine practice.

Commission Bias	Reviewer care quality judgements weigh more heavily active treatment over no treatment or *"watch & wait*."	Immediate antimicrobial treatment is not recommended for asymptomatic pyuria. However, the retrospective case-note reviewer has a general preference for active intervention rather than "watch & wait" strategies, and therefore gives a higher quality rating to cases in which an antibiotic has been prescribed.
Confirmation Bias	Reviewers' care quality judgements may be influenced by their tendency to find more support for their initial beliefs and to overlook information that opposes their initial beliefs	A retrospective case-note reviewer is reviewing a case in which the lead consultant for the patient was a close colleague and one whom the retrospective case-note reviewer respected highly. He therefore judges the quality of care as good because he considers this consultant very unlikely to make mistakes.
Loss/gain framing bias	Reviewer's care quality judgements may be influenced by concerns that their own low-quality	The reviewer refrains from escalating his findings from a retrospective case-note review because he fears that doing so will involve additional work, such as the need to undertake a root-cause analysis. He therefore reports no issues with the case and

	judgement may have adverse personal consequences	provide a higher care quality judgement than can be justified by the care quality content of this case-note.
Omission Bias	The converse of commission bias, where reviewer care quality judgements weigh more heavily no action or *'watch & wait*' over action	An elderly woman who is at high risk of surgery, had few viable options available than to undergo a salvage procedure for recurrence of cancer. The retrospective case-note reviewer gives a low rating for care quality judgement because she feels the surgery was inappropriate. Her preference would have been to watch and wait.
Order Effects: Primacy/recency bias	Reviewers' care quality judgements may be influenced by where different types of information are found in case-notes i.e., at the beginning, middle or end.	A reviewer identifies one error both at the very beginning and very end of a case-notes whilst the almost all parts with the central portion indicating good care content. These errors are more prominent in the reviewer's mind. He ascribes a lower care quality judgement than if the errors had been confined to the central portions (main body) of the record.
Optimism Bias	Reviewer care quality judgements tend to favour positive judgements of care	A renal physician reviews the case-note of a cardiology patient for its care quality. She rates an omission in care more favourably than a cardiologist reviewer familiar with the correct treatment pathways for this patient.

	quality for specialties or areas with which they are unfamiliar.	
Outcome Bias	Reviewer care quality judgements may be influenced by the case-note's presentation to the reviewer concerning therapeutic, diagnostic outcomes, or overall health outcome (e.g., death, adverse outcome, successful procedure, legal case, resolution with relatives	Having noticed the poor outcome of the case-note, the reviewer is primed to judge the processes of care more harshly. She thus incorrectly concludes that any minor deviations from recommended care were responsible for the outcome.

Representativeness heuristic	Reviewer care quality judgements may be influenced by the resemblance, in the reviewer's mind, between their current retrospective case-note review and previously reviewed case-notes, personal clinical experience or clinical knowledge (anecdotal/evidence-based)	A reviewer spends many hours trying to evaluate the case-note of a patient with a rare respiratory condition with which the reviewer is unfamiliar. She consequently feels compelled to find faults in care to justify her lack of knowledge concerning this case-note's specialty area and rates the care quality lower than a reviewer expert in this condition might have done.
Sunk-cost effect	Reviewer's care quality judgements may be influenced by the amount of money, effort, or time which	A gastroenterologist reviews a case-note of a patient treated for a rare respiratory condition. The gastroenterologist spends 20+ hours reviewing this case, which is in small part attributable a lack of familiarity with the clinical area. She consequently feels compelled to find faults in care to justify the time spent, and rates care quality lower than a competent respiratory physician reviewer might have done.

	the reviewer invested in a retrospective case-note review	

Note: This table demonstrates the cognitive biases and heuristics shortlisted by our expert panel. In column 1, there is the cognitive bias or heuristic. In column 2, there is a definition applied to case-note reviewer care quality judgements sourced from the original cognitive bias and heuristic definition in the literature. Column 3 contains a clinical scenario which demonstrates the influence of a cognitive bias or heuristic on the reviewer's care quality judgement.

In step 5, the percentage of expert panel members reporting each bias as likely or very likely to influence case-note reviewers is provided in Table 6. The top-ranked five biases were *ambiguity intolerance, outcome bias, availability bias, confirmation bias* and *omission bias* (all >75%). The >75% threshold was arbitrarily selected as a cut-off above which biases were considered more than "likely" to influence care quality judgements and were worth taking to the next stage. A moderate level of reliability was found between panel members ratings. The mean ICC measure was 0.74 with a 95% confidence interval from [0.48] to [0.90] ($F([14],[98])= [4.01]$, $p<0.001$]).(Koo and Li, 2016)

Table 6. Consensus panel ranking of the biases likely influencing case-note review care quality judgements

Biases	Percentage of Raters Reporting "likely" or "very likely" (%)
Ambiguity Intolerance	89
Outcome Bias	89
Availability Bias	78
Confirmation Bias	78
Omission Bias	78
Affect Heuristic	56
Bandwagon Effect	56
Commission Bias	56
Optimism Bias	44
Representativeness heuristic	44
Ambiguity Aversion	33
Anchoring Bias	22
Loss/gain framing bias	22

Order Effects: **Primacy/recency bias**	22
Sunk-cost effect	11

2.6. DISCUSSION

We have identified fifteen cognitive biases and heuristics which may influence case-note reviewer global care quality judgements. Using clinical scenarios to exemplify and contextualise these biases, panel members prioritised ambiguity intolerance and outcome bias as having the greatest potential to influence mortality case-note reviewer care quality judgements, followed by availability bias, confirmation bias and omission bias. We also demonstrate how we could mitigate these biases. To our knowledge, this is the first systematic approach to identify biases which could influence mortality case-note reviewer care quality judgements and characterise their relative influence by using clinical scenarios.

In 1991, Caplan et al. responded to a discussion article by Wood and Cooks(Cook and Woods, 1994), which addressed the complexity of human error, re-iterated the likely activity of the outcome bias upon clinicians in care quality judgement scenarios. This goes as follows:

> *"We wholeheartedly agree with Woods and Cook. Recognizing the presence and potential magnitude of outcome bias in physician judgments creates a healthy stimulus for improving the process of retrospective case review. Such improvements will benefit not only the patient and the physician, but also the much broader group of individuals and institutions who engage in the design, provision, and financing of health care. It is*

important—and exciting—to recognize that effective remedies may be close at hand. As Woods and Cook point out, some of the fundamental concepts and tools for managing outcome bias have already been explored in the literature of experimental psychology and human performance. The next logical step will be to determine whether this basic knowledge can be effectively adapted and extended to meet the specific needs of health care analysis." This means the reviews could have identified more studies. However, this was not the case.(Robert A. Caplan, 1991)

Since this recognition of the outcome bias was made in 1991, it begs the question why healthcare professionals are not more aware of the outcome bias that three decades have elapsed since its coverage in the Journal of the American Medical Association, which is a prestigious and highly cited medical journal. This lack of progress may be due to the intrinsic difficulty of studying this phenomenon or the difficulty for clinicians to appreciate the implications of outcome bias for their daily work. Worse still, it could simply be that they are implicitly resigned to the inevitability of the outcome bias in medical decision-making. For the first point, outcome bias has been studied to some extent since this articles publication mainly sourced around medical education and diagnostic reasoning studies.(Zimmermann et al., 2000, Wolfson et al., 2000, Siminoff and Fetting, 1989, Ritov and Baron, 1995, Mazzocco and Cherubini, 2010, Loewenstein and Prelec, 2000, Hemmerich et al., 2012, Graz et al., 2005, Chapman, 2003, Carling et al., 2010, Shi et al., 2011, Gupta et al., 2011) Thus, there is much literature on the subject.

Making care quality judgements requires human judgements. It is widely acknowledged that human judgements, across all domains, have been well documented to be influenced by cognitive biases and heuristics. This influence may be for better or ill but whilst this influence remains unknown it can affect care quality judgements in unpredictable ways that are not desirable for case-note reviewers, the measurement of the care quality and improvement of these judgements.

The specific contribution of using the perspective of cognitive biases and heuristics is to primarily identify some of the sources of variability around the reviewer judgements on care quality assessments. There are likely several sources[21] of this variability, but biases and heuristics are higher up the list given the complexity of these judgements and the shortcuts they may indeed be applying. It is known that the inter-rater reliability of case-note reviewers over the same batch of case-note reviews is moderately poor (0.3-0.5), and it was considered plausible that cognitive biases and heuristics could explain some of this variability.(Goldman, 1994, MacKenzie et al., 1992, Lilford et al., 2007, Hutchinson et al., 2010c, Benning et al., 2011, Goldman, 1992) Though I have not gone as far as to provide the empirical support for their influence, I rather deliberated on deducing the precise influence of any cognitive biases or heuristic, it was necessary to take little, but sure, steps to first characterise which cognitive biases and heuristics were most likely active during case-note review care quality judgements. I had pondered for some time and decided that it was more important to obtain a

[21] Another source is the variability from the environment (i.e., lighting, noise, etc.) and different instruments used in the case-note review process (i.e., administration to the format of the review proforma)

lay of the “conceptual” land. Attempting to elucidate an empirical effect would have assumed that biases, in principle, would have an effect, but it was unclear whether a certain subject or all ought to be studied closely. In choosing to survey the cognitive landscape through a conceptual-consensus exercise (Chapter 2), I sought to identify which cognitive biases and heuristics were plausibly going to influence these care quality judgements and provide tentative possibilities on how they were likely to influence these care quality judgements based on the extant conceptual and empirical literature. Cognitive biases and heuristics could explain some of this variability between reviewers. If this can be done, then there are methods which could be employed to reduce the extent of this reviewer disagreement. As things stand, one solution is to have 8-17 reviewers to review the same case-note content to raise the reliability to 80-90%.(Manaseki-Holland et al., 2019b) However, this is infeasible given the current clinical demands placed on most clinicians. In short, it takes far too much clinician time to undertake, and very few reviews would be undertaken. What identifying biases could help with is understand better some of the significant sources of variation which are attributable to cognitive biases and heuristics. With this understanding, the bias and/or heuristics can be mitigated by employing different cognitive and behavioural strategies, as will be outlined later in this chapter through the MINDSPACE framework. In sum, the contribution of cognitive biases and heuristics stand to reduce the variability between reviewer care quality judgement disagreements through understanding how this variability arises in practise. As a corollary, one can then seek to employ existing cognitive and psychological methods to raise awareness and actively mitigate for the undesirable action of these biases and heuristics. Understanding and mitigating the influence of these biases and heuristics is both far more cost-effective and less onerous on clinician’s time and workload. These two sure advantages justify the further exploration of any cognitive bias and heuristic activity in care quality judgements.

In specific regards to the mitigation of cognitive biases and heuristics, de-biasing, or the mitigation of their effects, according to Baruch Fischhoff, an early contributor to this field, necessitates four conditions. First, that they are aware of the bias. Second, they must understand the direction of the bias. Third, they must receive immediate feedback when falling prey to the bias. And fourth, they must be subsequently provided relevant training with regular feedback and coaching.(Fischhoff, 1977) These are optimistic conditions for any person and training specialist to fulfil, however, the healthcare system has one of the most cognitively able, motivated and capable technologists that can help to establish a culture and training system to fulfil these criteria.

Research indicates that bias awareness strategies work when a bias can or is attributed to others(Larrick, 2004), which is itself similar to stereotype reactance where inspired motivation is sourced from an awareness of the error which is followed by a wilful acceptance of this fact, which then leads to its resolution or release.(Miron and Brehm, 2006) These range from simple reflective writing and perspective-taking exercises which can reveal in different ways the emotionally charged nature of implicitly-held clinical narratives and so help to re-orient any underlying prejudiced beliefs, and help them to consider the weaknesses of their case/perspective.(Babcock and Loewenstein, 1997) Biases can be mitigated through changing the socially shared definitions of the group. For instance, if learning and collaboration for QI can be redefined or renormalized according to another framework than what is conventionally understood by these terms, groups and organisations stand to greatly benefit, as was the case with the non-governmental organisation renormalizing reconciliation efforts in Rwanda.(Paluck, 2009) If this can be done at the national level in Rwanda, there is every hope that it can be implemented at a smaller scale, the hospital level. These are proven

de-biasing strategies. Furthermore, a systematic review identified a significant number of studies favouring a meaningful de-biasing effect of interventions on health-related judgements.(Ludolph and Schulz, 2018)

2.6.1. Dual Process Theory: reflections on this theory

I adopt the term "dual process theory" but there are nuances within this theory which need elaboration. Firstly, there are several terms which refer to dual process like processes such as "dual processes" which assume that cognitive tasks involve two strict forms of processing. Dual types indicate the qualitative distinction between the two. Dual systems refer to system 1 and system 2, which have their origins in evolutionary origins. However, this term is avoided because it suggests that these systems are neurologically distinct, which they are not. This proliferation of dual-process theories has been more confusing with many of the distinctions unclear and hard to isolate through empirical methods.

For practical reasons, I needed to have a cogent starting point without consulting all the possible schools of psychology relating to my research question. I refer specifically to the ideas and work of Gary Klein and Gerd Gigerenzer. Gary Klein, a decision-making researcher, who pioneered the field of naturalistic decision-making (NDM) that studies human decision-making in real-world situations. Klein and other NDM researchers have provided evidence of preternatural decision-making speed and accuracy under high-stake situations.(Klein, 1999, Salas and Klein, 2001) Gerd Gigerenzer, another decision-making researcher, studies bounded rationality, which is when the ability to make rational decisions is hindered by intrinsic (and extrinsic) factors (i.e. human cognitive limitations, fatigue, time).(Simon, 1979) Kahneman and Gigerenzer both study bounded rationality, however, the

latter is more optimistic about the human ability to adapt to decision-making situations using "rules of thumb" or heuristics to the structure of an environment.(Gigerenzer and Brighton, 2009, Gigerenzer and Gaissmaier, 2011b) These two researchers depart from the dual process theory in significant ways. They were not included because the dual process theory has been widely accepted amongst most academics. And the grounds for this acceptance served the purpose of demonstrating the error-prone nature of decision-making, which could explain the observed care quality judgement variability between case-note reviewers. The dual process theory, as with any theory, must be subject to empirical validation before it can be accepted as scientific. If it is found sufficiently wanting in terms of its predictive and explanatory power, it will have to be abandoned in favour of other alternative theories, but for now, its assumptions are robust enough to help me illustrate its possible contribution to the variability of care quality judgements. It is not the only approach to understand human decision-making.

In particular, Gary Klein(Klein, 1999, Salas and Klein, 2001) is optimistic about accuracy and reliability of expert intuition whilst Gerd Gigerenzer(Gigerenzer, 2018, Gigerenzer and Gaissmaier, 2011b) consider heuristics, on the whole, as reliable and helpful cognitive aids rather than being much prone to error. These two people depart from the view of cognitive biases and heuristics as demonstrative of the error-prone nature of human decision-making. I, for one, hope these discussions can continue as these different perspectives are central to understanding any psychological phenomena. I felt, as did my supervisors, that these two systematic reviews were representative and comprehensive in their discussion of this field. But in the spirit of diversity and multi-paradigmatic science, one must, at some point, discover why the current paradigm has ascended to its current status.(Kuhn, 2012, Feyerabend, 1993) However, the ambition to detail a complete account of the supremacy of one account over another, comparing, contrasting and evaluating its strengths and weaknesses

exceeds the scope of this thesis. For instance, one discussion might revolve around whether mental states influence quality improvement, such as beliefs or desires. The foundations of thought and the nature of cognition per se need to be examined further. These foundations were not systematically expanded at the outset; however, I consider it beyond the domain of the health services research, which is invariably concerned with the practical application of theory and much less concerned with the conceptual and ontological foundations of thought. It cannot be neglected that understanding foundations and the assumptions we make can help one to better understand the world we live in.

In the third stage of the review, a panel consensus exercise helped to select cognitive bias and heuristic studies based on their influence on the judgement of the case-note reviews and/or medical records. Quality may not appear on first appearances to be a psychological affair. However, any improvement must itself be sufficiently resistant or immune to erroneous measurements and judgements. Otherwise, improvement will lapse into ad-hoc changes which may or may not improve the quality of care. In this event, I stressed the importance of their plausible influence and action.

Further, behavioural science practices can be used to devise de-biasing strategies. The MINDSPACE framework, which is a mnemonic for nine influencers – messenger, incentives, norms, defaults, salience, priming, affect, commitments and ego developed out of the popular, dual process theory (Dolan et al., 2012) This has been shown in peer-reviewed literature as effective ways to nudge or reinforce positive behaviour. (Dolan et al., 2012, Vlaev et al., 2016, Vlaev and Dolan, 2015) In Table 7, we present nudge interventions, otherwise known as “nudges”, which could be used for our five prioritised cognitive biases and heuristics.(Vlaev et al., 2016, Pettigrew and Mays, 2021) This table could pave the way

for further research around the influence of cognitive biases and heuristics upon case-note reviewer care quality judgements.

Table 7. Possible Nudges

Cognitive bias or heuristic	**MINDSPACE Mnemonic**	**Nudge (Dolan et al., 2012)**	**Case-note review proforma (location)**
Ambiguity Intolerance	Norms i.e., the case-note contains all the information necessary for evaluating its care quality	Encourage reviewer's information gather from sources other than the case-note, where possible (i.e., hospital records, primary care record)	"Do you feel you need more information to assess the care quality?" (middle) "What further information did you need to judge the care?" (end) "Is your care quality judgement on the information **only** contained in this case-note?" (end)
Outcome Bias	Default i.e., The death of the patient, especially if errors and slips in	*Blind the outcome*, if possible, through redaction or electronic reviews	The last few hours of case-note are masked, if possible

	care have occurred, was, to some extent, iatrogenic in origin.	If blinding is not possible, use the "consider-the-opposite" approach.	If a patient died, evaluate care quality as if the patient lived.
Availability Bias	Defaults i.e., what is pre-set in the reviewer's mind when reviewing specific case-note details? e.g., demographics, diseases etc.	Reviewers pay closer attention to defaults and first impressions	"Have you previously observed care like this before? If so, how is this pertinent to this case-note?" (middle) "Have you received a case-note like this before? If so, please justify similar features." (end)
Confirmation Bias	Salience i.e., the reviewer draws attention to novel case-note review aspects for personal reasons without clear justification.	Reviewers pay closer attention their own beliefs and values by active role-playing e.g. "*six thinking hats*."(De Bono, 2017) and clear justification and, where necessary, with supporting evidence.	"What would your professional colleague say about this care?" (middle) "Would your departmental colleagues agree with your overall care quality judgement?" (end)

Omission Bias	Defaults i.e., harm from the direct action is considered worse than the indirect action	Consider the significance all explicit and implicit care processes (i.e., evidence, reasoning)	"Think of things very likely done for the patient which are not documented in the case-note." (middle) "Evaluate the care quality from case-note information the case-note and any very likely justifiable actions not documented in the case-note." (end)

2.6.2. Contributions of the cognitive and heuristics approach

There are several advantages of employing cognitive biases and heuristics their influence on case-note reviewer care quality judgements. A list of the benefits with its corresponding result are given below (Table 8):

Table 8. Advantages of cognitive biases and heuristics and their examples

Advantages of cognitive biases and heuristics	**A concrete example sourced from the literature**
Helps to identify sources of biases within individuals in their organizational setting	There have been cognitive biases and heuristics which identify biases and heuristics within key decision-makers such as entrepreneurs and managers in large organisations.(Busenitz and Barney, 1994)
Can help individuals become more aware of these biases to thus help mitigate their effects	Greater contextualised clinical knowledge helps clinicians make better clinical decisions around “uncertain” diagnoses. This enhanced clinical knowledge helps reviewers be more immune to cognitive biases and heuristics.(Mamede et al., 2020)

Some strengths of the study are now discussed. First, the internal validity of these elements was preserved by having the literature search terms sourced from pre-existing psychological and cognitive science terms. The cognitive biases and heuristics in medical decision-making literature searched for concepts that could extend to case-note reviewer care quality judgements. The ranking exercise was methodologically deduced, and the protocol actioned.

Second, the external validity of the literature search was modelled on pre-existing systematic review search terms and the ranking exercise was led by the literature by methodologists familiar with consensus exercise procedures, behavioural science and case-note review styles (e.g., structured-implicit, explicit). Third, a reliability statistic known as the intraclass correlation coefficient was used to calculate the resemblance between individual member ratings. This helps to identify common ground and understanding across the panel, which accounts for chance agreement. Fourth, the objectivity was preserved by having the biases sourced from the literature. Medical decision-making biases were subsequently excluded for those that did not apply to case-note reviews with the remaining considered in alphabetical order by diverse panel members in the consensus exercise.

2.6.3 Limitations

Our study has several limitations. Our list of biases may have been incomplete. In this discussion, we first review several limitations of our study and then provide recommendations for future research and practice. Removing "human factors" as one of the topic search terms may have omitted several human factors which were likely to influence case-note review care quality judgements. In spite of the wealth of human factors research into complex decision-making, which include the military, civil and medical sectors, there are no clearly evidenced human factors which affect medical decision-making pertinent to case-note reviewer care quality judgements.(Cook et al., 2007) The psychological decision-making research area devoted to studying decision-making in naturalistic settings, such as in healthcare organisations, find plausible application to naturalistic settings like the emergency medical diagnosis.(Keller et al., 2010) Yet, there is little evidence for the human factors influencing case-note reviewer care quality decision-making. Firstly, that we could not

anticipate exploring in-depth human factors, nor a limited subset, of the topic search if the "human factor" search terms, once implemented into the searchers, were likely to not capture a large proportion of the studies pertaining to "human factors" and "medical decision-making." There was also no extant coherent theory "human factor" in the literature of its influence during medical decision-making at the time of the literature search (2017), nor was there a workable, cogent theory from the cognitive biases literature which overlapped with and encompassed "human factors" and could help us relate this to case-note reviewer care quality judgements. Notably, the cognitive biases and heuristics models were more coherent with respect to relevance of the content of their theory to case-note reviewer care quality judgements. It was simply unclear how human factors relate even to case-note review care quality judgements, as this was neither posited nor evidenced in any theory or empirical study in contiguous medical fields. We recommend "human factors" are explored separately when form (theory and research) comes to match its function, for its future exploration in relation to reviewer care quality judgements; without substantive theory on how a human factor influence medical decision-making or care quality judgements, the search would have been for exploratory purposes not founded on hypothesis or prior research. Contrastingly, the cognitive biases literature had both theories and some evidence supporting its relevance to case-note reviewer judgements. A standard conceptual framework for understanding cognitive bias in clinical decision-making does exist because it is overly difficult, perhaps impossible, to measure the real-time cognitive activity of a person (or group of people) in a complex naturalistic setting in which medical decisions are made. In theory, there ought to be one theory for how clinicians think and systematically err, however, it is too cumbersome for them to be explicit about their decisions and to control the external environment and reduce undesirable confounding from uncontrolled factors. Such controlled conditions do not exist, and it would be quite close to impossible to control for these conditions in the mind. And if

human factors are explored also in relation to case-note reviewer care quality judgements, then both the mind and the external environment need to be correctly understood and extraneous influencing factors controlled for. This is a most demanding task for the human factor's practitioner. As a result, human factors were unlikely to feature in how care quality decisions were made based on this scoping. However, human factors would be pertinent to understand the differences between using paper-based and electronic case-notes or medical records. If we did add this to the mix with cognitive biases and heuristics, the review question would have been far too specific (i.e., looking at case-notes with biases, heuristics, and human factors by electronic or paper-based records). Once more, cognitive biases were simply more pertinent to care quality judgements than human factors, and our review team settled on this rationale for human factors' exclusion for this Chapter 2. On a methodological point, the research question informing the systematic review research was overly refined and the two distinct fields of biases-heuristics and human factors do inhabit quite different theoretical and practical spheres[22] would have undermined the potential implications of this

[22] The theories are likely to present a different model of influence, with a different ontology and mechanism, which would very likely affect any effective reconciliation of one with the other for a consistent discussion and implications. Simply put, they were better sought separately, and cognitive biases-heuristics subject was chosen because it was more focussed and coherent compared with the human factors models, which had a unboundedness from the definition of human factors taken from the Health and Safety Executive (on 08/05/2017) as "Human factors refer to environmental, organisational and job factors, and human and individual characteristics, which influence behaviour at work in a way which can affect health and safety."

research. The likely reason cognitive biases and heuristics and human factors are not currently explored together in a single conceptual framework is because the human factors often do not consider the cognitive, mentalistic component of the behaviour. And likewise, the cognitive bias model does not explore much the human interaction with the work environment.(Carayon et al., 2014)

A possible exception to this was the work by Mica Endsley to develop a situational awareness model of human factors. Situational awareness is foundational to any naturalistic decision-making, of which medical decision-making is a sub-category of.(Endsley, 1995) Such social-psychological stressors include mental load, time pressure and the importance or difficulty of the task being undertaken. Juggling these factors may result in a reduction of available bandwidth, that is the brain's limited information processing ability to make decisions, to deal with the tasks at hand, as they too require cognitive input. One boon of Endsley's human factors model is that it considers the cognitive information processing systems. But it does not consider the mechanisms of how, in addition, cognitive biases happen, and the misuse of any decision-heuristic arise. Because of the lack of overlap between extant theories of human factors and cognitive biases, one had to be chosen. And to re-iterate, the more cogent models on cognitive bias and heuristics demonstrate any plausible influence on reviewer care quality judgements more convincingly than extant human factors models. As it stands to date from the writing of this thesis, there is no theoretical model which broaches human factors and cognitive biases and heuristics.

A reason for this could be that a human factors model must have many inputs, as many inputs as there are discrete entities or stimuli found in the work environment, that is it difficult to control for these factors. Given the complexity at this work environment level alone, the addition of cognitive biases and heuristics would raise the model's complexity at

least 100-fold given the number of these known.[23] It would be a conceptual model which would be very difficult for a practitioner to use effectively; there is simply too much information, assuming it is pertinent, for the practitioner to wade through and apply. In short, human factors have their sphere of concern and cognitive biases and heuristics has another.

Another reason is that it is not known exactly which work environments trigger certain biases in the real world. But the most-widely accepted convention in cognitive biases and heuristics research is to study one cognitive bias or heuristic at a time. The work environment changes and there is very likely not one bias operating, but several. A cognitive bias model extended into real-life practise may simply not be adaptable to dynamical systems such as in healthcare behaviour and decision-making. This use of the biases model is at odds with what the practitioners needs to know and would like to benefit from in real-time; it would simply be too difficult for practitioners to meaningfully implement and benefit from.

One of the possible limitations is that the search term, human factors, was left out of the review. It is obvious the human environment plays a role; however, its role is underplayed even in a medical professional. For instance, radiologists, who make interpretations from scanned images; only a single line of human factors research, namely fatigue, was found discussing anything remotely related to human factors research. The rest of the literature

[23] There are over 200 cognitive biases and heuristics on the Wikipedia entry (29/01/2022):

search[24] was underwhelming on human factor topics influencing any area of medical decision-making through the sifting abstracts stage.[25]

A larger list of biases could have been considered. However, there is acceptance that these biases are categorizable into different "bins" according to their underlying psychological constructs.(Benson, 2016) Over 100 biases have been described, but many of these are unlikely to influence medical decision-making, e.g., "*plant blindness.*"(Wikipedia, 2020, Blawatt Ken, 2016) Our focus was on identifying biases relevant to medical decision-making, not the general literature. Third, our panel members were recruited regionally, involving those working in or near the West Midlands, United Kingdom with relevant expertise; future panels may convene a larger panel of internationally recognised experts in case-note review and behavioural science. The geographical distribution of the research could have been wider, but we felt that engaging with overseas academics would have been difficult; the many-steps and the limits on AT's funding for the consensus procedure were prohibitive to a long-distance academic relationship with overseas collaborators, who needed to both travel (per-covid-19 pandemic) and would have expected to be reimbursed. Nevertheless,

---

[24] This involved a broad-brushed search of over 100 article abstracts from Sciencedirect.com and Google Scholar on 1 Nov 2021 failed to yield any pertinent human factor lines of research. The search terms used were: "human factors radiologists." The prominent lines of research for this search were artificial intelligence, deep learning methods, gestalt and learning theories as support decision aids were the most frequent lines of research for radiologists.

[25] On Sciencedirect.com and Google Scholar from 2017 onward with the search term "medical decision-making human factors" on 01/11/2021

engagement with global researchers could be encouraged in future if the research is important and has the potential for development.

Concerning the case-note review training status of panel members, this was limited to 4 of all the panel members, of whom in total have over 20 years of case-note reviewer training experience. However, our focus was not to recruit only those closely involved in case-note review training and its methodology but to consider having a diverse panel of researchers, clinicians, and behavioural scientists to contribute their knowledge and insights to burgeoning field of inquiry. The number of panel members with overlapping interests from existing case-note and "biases" work would have provided the best group of panel members; this is indeed the case with 6 of 9 straddling more than one of these disciplinary areas. Case-note reviewers have their own conceptual goggles as so do behavioural scientists, and so it is important to consider not only the direct experience of case-note reviewers but how the case-note review process can be conceptually conceived.

The choice of the biases examined were obtained from systematic review sources which spanned the formal conception of the field of "cognitive biases and heuristics" in the late 1960s to the present time of the reviews. The choice to include or exclude the bias/heuristic in step 3 was undertaken on the assumption that any bias/heuristic must logically apply to the case-note review context and the care quality judgement (or any judgement) on this special

document type. For instance, the Gambler's fallacy[26] was a bias/heuristic we excluded at this stage. It does not apply to case-note review judgements because the same set of case-note circumstances cannot be run again as a one-off event; the Gambler's fallacy works only in controlled games with fair calibration of instruments to which their probabilities can be assigned and critically run again.

The process was for the panellists to write the clinical scenarios and have them proceed to rank these cognitive biases may appear circular. To reply to this, those who did not write that scenario for that bias/heuristic outnumber this person 8 to 1 and thus would have had their biased voice diluted by others. Furthermore, being blinded to these clinical scenarios wasn't possible given we had no starting point for this field prior to this consensus exercise. This research commenced to develop a new field and if this perhaps has the pangs of subjective qualitative research, it is because there has been no empirical work on the intersection between the respective topics of case-note review judgements and cognitive biases and heuristics. It is virgin territory and oftentimes this requires exploratory, value-laden research and thought experiments. There is a certain level of internal and external validity to the thought experiments, which we consider are present given our methods and clear rationale for each step.(Wilson, 2016) And each of our examples presuppose two conditions required for the external validity of the case of any thought experiment, of which ours is like one because we have no empirical proof of the influence. First, there must share common normative

[26] The fallacy is defined as "*…when subjects act as if every segment of a random sequence/set of events must reflect the true proportion; if the sequence has strayed from the population proportion, a corrective bias in the other direction is expected…*"

context to actual medical practise and expertise and second, the presupposition of causal structures significantly like those found in the context of interest, that is, the case-note reviewer and their review . Great pains have been taken to ensure that this is indeed apparent at the outset of the study to its completion and so one is inclined to think there is relevance to these clinical scenarios. Embarking on the analysis, the process has its rigor preserved as the scenario generating stage was conducted with each panel independent from the other panellists. There was no peer pressure to conform to any standard other than the scenario which they personally felt elicited a salient effect of the bias/heuristic upon the case-note reviewer care quality judgements. And it was they themselves that provided supporting rationale for their scenario generation and choice. Also, the concern that the close connection the panel members had to this doctoral research has influenced the research is largely dispelled as each person had the opportunity to independently deliberate on their own decisions without conference with others at consensus stage 3, 4 and 5. This process was sufficiently far-removed from any personal agenda or peer pressure to significantly bias the consensus process. The small number of participants and the low-regional differences will not likely influence the generalisability of these concepts given that the numbers should not be stressed too strongly but the content of the justification for the influence of each bias through the applied definitions and the clinical scenarios. Yet, this type of research would benefit greatly from the involvement of individuals from contiguous fields for their perceptions on the validity of the research and its applicability to the case-note reviewer. This will likely allow for further conceptual and practical progress in this area.

Fourth, our hypothetical clinical scenarios are simple abstractions that demonstrate the potential impact of each bias on medical decision-making. They do not anticipate other factors that simultaneously influence decision-making, e.g., time constraints, and they might not fully reflect the complexity of examining original clinical case records. However, they allowed the reader to understand rapidly the potential relevance of each bias and stimulate reflection on how the bias could operate. To ensure scenarios were relevant, practitioners involved in case-note review training were involved at all stages of this review. Fifth, the frequency and severity of each bias's proposed influence on case-note reviews were not separately assessed: a future review might analyse each component independently. Sixth, deliberative thinking has been criticised for its inability to improve decision-making outcomes in academic (Gigerenzer and Gaissmaier, 2011a, Klein, 1999) and popular circles (Gladwell, 2005, De Becker, 1997) who situate intuitions as favourable evolutionarily adaptations. Intuitions in less-restrictive contexts do serve a role, however, intuitions may be less applicable to highly specialised, evidence-based areas of decision-making (e.g., medicine). Also, the recollection of positive instances of "gut feelings" are likely plagued by the hindsight bias. Thus, context is important when considering the best way for bias mitigation and this is endorsed by "nudge" approaches.

2.6.4. Implications for health services

As discussed in the introduction, retrospective case-note reviews are an established method of evaluating the quality of care delivered in hospitals. In the United States and in the United Kingdom, mortality case-note reviews are required in many hospitals. Retrospective case-note reviews require significant cognitive effort, monetary expenditure, and clinician time. The opportunity costs could be considerable if these efforts are not well utilised in the health

economy. A recently completed study of error and care quality in 4000 emergency admissions to hospital confirmed low inter-reviewer agreement at the case-note level but with very high agreement when case-notes were aggregated by hospital Trust.(Bion et al., 2020) This supports the possibility that training reviewers in bias awareness could improve patient-level assessments.

Where reviewers cannot agree on the judgement of care quality, the validity of the review process is in question. Attempts to improve consistency using consensus between reviewers produces a spurious, chance improvement in agreement without improving the accuracy of care judgement.(Hofer et al., 2000) Cognitive biases may act randomly in a systematic way, decreasing reliability and/or increasing inaccuracy. However, until biases in care quality judgements have been empirically demonstrated, these judgements and any subsequent actions arising from these judgements, will remain subject to doubt and speculation. To answer this speculation, our expert panel prioritised five biases as being of particular concern, thereby research and intervention should be first targeted at these.

2.6.5. Mitigation

There are ways to mitigate these cognitive biases and heuristics. Some research suggests that cognitive biases are resistant to intervention(Gigerenzer, 2008a, Epley and Gilovich, 2006), a broader behaviour modification approach is worth investigating. Aligning with the COM-B framework(Michie et al., 2011), capabilities could be improved with training or education sessions to raise reviewer awareness of biases(Sandars, 2009), improve clinical reasoning(Norman et al., 2017) or use physical exercise to improve general cognitive function.(Mandolesi et al., 2018) Environmental sources of bias can be mitigated by ensuring optimal reviewing environments (for example, light, noise or interruptions have been

implicated in work environs)(Banbury and Berry, 2005) and reviewing periods following meals and/or rest.(Danziger et al., 2011) Reviewers should also complete pre-review motivational checks to declare any professional or personal interests which conflict with their reviewing duties. Such motivational checks have a precedent in judicial review.(Wellington, 1982)

We use dual process theory as the framework to demonstrate ways to mitigate the proposed influence of these biases. We present examples of bias mitigation to demonstrate their action. To limit the influence of *ambiguity intolerance* during the selection of an antibiotic treatment, the case-note reviewer could apply critical thinking to override the emotive "system 1" urge to personally favour the sheer magnitude of evidence and re-evaluate the weight of evidence for warfarin and then for Rivoroxaban. (Mumma and Wilson, 1995) To limit the influence of any *availability bias* effects in evaluating the care of patients deceased by cardiac arrest with secondary hyperkalaemia, comprehensive clinical training around "hyperkalaemia" could help reinforce the consideration of all the likely differential diagnoses (i.e., alternative medical explanations). (Morewedge et al., 2015)

The de-biasing literature suggests two main avenues to combat biases, either by changing the person or changing the environment. Regarding changing the person, Morewedge *et al.* suggest training may decrease negative effects of biases.(Morewedge et al., 2015) Training can help re-entrain the use of system 2 processes. This generally involves raising people's awareness of the bias in ways that motivate intrinsic behaviour modification. Applying dual process theory, we can prompt reviewers to deliberately use the "*consider-the-opposite*" technique to diminish any anchoring bias during the differentials workup during the case-note review care quality assessments.(Mumma and Wilson, 1995) We can also recommend

reviewers use bespoke clinical training to their own specialty or topic area to raise awareness. This has been demonstrated to improve the clinician discrimination between similar disease-specific case presentations.(Mamede et al., 2020) Furthermore, the art of doing nothing has been shown to improve the quality of clinical decisions and should be considered as an method to employ in medical education.(Keijzers et al., 2018) Regarding the environment, Thaler *et al.* suggests that alterations to the environment, known as "*nudges*", can help reduce the likelihood of biased behaviour in favour of preferred responses.(Thaler and Sunstein, 2008) For example, contextual modification may be used to minimise biases, for example, by obscuring the patient's outcome before or during a case-note review. We recommend that the outcome bias is objectively explored for its effects upon reviewer care quality judgements. It is clear that Caplan's study was not designed to elicit the cognitive processes behind the effects from the outcome bias.(Robert A. Caplan, 1991) And it remains unclear whether the Caplan's review process was because of poor care or the bad outcomes.(Schroeder and Kabcenell, 1991) Thus, there is the need to scientifically understand the source of this effect. The effectiveness of nudges can be explained with the dual process theory as nudges involve the "*deliberate (system 2)*" crafting of "*automatic (system 1)*" choice architecture. As with any intervention, we envisage any nudges will need to respect the choice architecture of each specialty to maximise the possibility of mitigating any active biases. Furthermore, the study design needs to be calibrated for the measurement of such effects.

Our approach could be used by reviewers and policymakers to mitigate system 1 biases during the evaluation the care quality of case-note reviews. There is no reason these biases would not extend to other medical decision-making areas (e.g., diagnosis, examinations). For scientific rigor, we suggest reviewers and policymakers employ our methods to scrutinise this proposal. And the results of our extensive critical analysis of the two reviews suggest that

judicious use of existing study designs are adapted to develop new studies to empirically confirm this proposal.

The second point is real but difficult to evaluate as there have been no organisational-wide studies which have captured the perceptions of clinicians on the activity of cognitive biases and heuristics and its influence on their daily activity, let alone on case-note review care quality judgements. The last point concerns the necessary use of qualitative research methods to explore the level of awareness that case-note staff have concerning these cognitive biases and heuristics and, in addition, how mitigable they are in daily practise. There is currently no[27] exploration of this relating to medical records/case-note reviews in the healthcare setting, let alone in hospitals. Further exploring these three questions will help us to understand the extent to which these biases and heuristics could be mitigated.

2.7. Conclusion

Retrospective case-note reviews are a standard method for assessing care quality, but variability in reviewer judgements limit the utility of this approach. From our systematic literature search and conceptual review of systematic review literature, there was no evidence of an emerging taxonomy for these biases in this area. After undertaking the consensus panel exercise to rank order the influence of biases, the inter-reviewer variability may be driven by several biases, of which *ambiguity intolerance, outcome bias, availability bias, confirmation bias* and *omission bias* are likely the most important. Further research is needed to test methods for objectively identifying and mitigating the impact of these biases, to enhance the

[27] Or very little, which has escaped the notice of me and my research team

reliability of care quality judgements and the reviewer's self-confidence in their own care quality judgements, with consequential benefits for patient care, professional development, and medical regulation.

2.8. References

ANDERSON, J. R., BOTHELL, D., BYRNE, M. D., DOUGLASS, S., LEBIERE, C. & QIN, Y. 2004. An integrated theory of the mind. *Psychological review,* 111**,** 1036.

BABCOCK, L. & LOEWENSTEIN, G. 1997. Explaining bargaining impasse: The role of self-serving biases. *Journal of Economic perspectives,* 11**,** 109-126.

BANBURY, S. P. & BERRY, D. C. 2005. Office noise and employee concentration: Identifying causes of disruption and potential improvements. *Ergonomics,* 48**,** 25-37.

BENNING, A., GHALEB, M., SUOKAS, A., DIXON-WOODS, M., DAWSON, J., BARBER, N., FRANKLIN, B. D., GIRLING, A., HEMMING, K. & CARMALT, M. 2011. Large scale organisational intervention to improve patient safety in four UK hospitals: mixed method evaluation. *Bmj,* 342**,** d195.

BENSON, B. 2016. *Cognitive bias cheat sheet* [Online]. Available: https://betterhumans.pub/cognitive-bias-cheat-sheet-55a472476b18#.f6daic2o3 [Accessed 13/08/2018].

BION, J., ALDRIDGE, C., GIRLING, A. J., RUDGE, G., SUN, J., TARRANT, C., SUTTON, E., WILLARS, J., BEET, C. & BOYAL, A. 2020. Changes in weekend and weekday care quality of emergency medical admissions to 20 hospitals in England during implementation of the 7-day services national health policy. *BMJ Quality & Safety*.

BLAWATT KEN, R. 2016. Appendix A: List of Cognitive Biases. *Marconomics.* Emerald Group Publishing Limited.

BLUMENTHAL-BARBY, J. S. & KRIEGER, H. 2015. Cognitive biases and heuristics in medical decision making: A critical review using a systematic search strategy. *Medical Decision Making,* 35.

BUSENITZ, L. W. & BARNEY, J. B. Biases and heuristics in strategic decision making: Differences between entrepreneurs and managers in large organizations. Academy of management proceedings, 1994. Academy of Management Briarcliff Manor, NY 10510, 85-89.

BYRNE, A. 2013. Mental workload as a key factor in clinical decision making. *Advances in Health Sciences Education,* 18**,** 537-545.

CARAYON, P., WETTERNECK, T. B., RIVERA-RODRIGUEZ, A. J., HUNDT, A. S., HOONAKKER, P., HOLDEN, R. & GURSES, A. P. 2014. Human factors systems approach to healthcare quality and patient safety. *Applied ergonomics,* 45**,** 14-25.

CARLING, C. L. L., KRISTOFFERSEN, D. T., OXMAN, A. D., FLOTTORP, S., FRETHEIM, A., SCHÜNEMANN, H. J., AKL, E. A., HERRIN, J., MACKENZIE, T. D. & MONTORI, V. M. 2010. The effect of how outcomes are framed on decisions about whether to take antihypertensive medication: a randomized trial. *PLoS ONE,* 5.

CECCHINI, M., SASSI, F., LAUER, J. A., LEE, Y. Y., GUAJARDO-BARRON, V. & CHISHOLM, D. 2010. Tackling of unhealthy diets, physical inactivity, and obesity: health effects and cost-effectiveness. *The Lancet,* 376**,** 1775-1784.

CHAPMAN, G. B. 2003. Time discounting of health outcomes. *Time and Decision: Economic and Psychological Perspectives on Intertemporal Choice.* Russell Sage Foundation.

COCHRANE COLLABORATION 2019. Data Extraction and Assessment Form - Cochrane EPOC for Randomised and Non-randomised controlled trials.

COHEN, H., HUNT, B. J., EFTHYMIOU, M., ARACHCHILLAGE, D. R., MACKIE, I. J., CLAWSON, S., SYLVESTRE, Y., MACHIN, S. J., BERTOLACCINI, M. L. & RUIZ-CASTELLANO, M. 2016. Rivaroxaban versus warfarin to treat patients with thrombotic antiphospholipid syndrome, with or without systemic lupus erythematosus (RAPS): a randomised, controlled, open-label, phase 2/3, non-inferiority trial. *The Lancet Haematology,* 3**,** e426-e436.

COOK, M. J., NOYES, J. M. & MASAKOWSKI, Y. 2007. *Decision making in complex environments*, Ashgate Publishing, Ltd.

COOK, R. I. & WOODS, D. D. 1994. Operating at the sharp end: the complexity of human error. *Human error in medicine,* 13**,** 225-310.

DANZIGER, S., LEVAV, J. & AVNAIM-PESSO, L. 2011. Extraneous factors in judicial decisions. *Proceedings of the National Academy of Sciences,* 108**,** 6889-6892.

DE BECKER, G. 1997. *The Gift of Fear: Listening to the Intuition that Protects Us from Danger*, Little, Brown & Company.

DE BONO, E. 2017. *Six thinking hats*, Penguin uk.

DOBLER, C. C., MORROW, A. S. & KAMATH, C. C. 2019. Clinicians' cognitive biases: a potential barrier to implementation of evidence-based clinical practice. *BMJ Evidence-Based Medicine,* 24**,** 137.

DOLAN, P., HALLSWORTH, M., HALPERN, D., KING, D., METCALFE, R. & VLAEV, I. 2012. Influencing behaviour: The mindspace way. *Journal of Economic Psychology,* 33**,** 264-277.

EDWARDS, M. T. 2013. A longitudinal study of clinical peer review's impact on quality and safety in US hospitals. *Journal of Healthcare Management,* 58**,** 369-384.

EFFECTIVE PUBLIC HEALTH PRACTISE PROJECT (EPHPP) 2018. Quality assessment tool.

ENDSLEY, M. R. 1995. Toward a theory of situation awareness in dynamic systems. *Human factors,* 37**,** 32-64.

EPLEY, N. & GILOVICH, T. 2006. The anchoring-and-adjustment heuristic: Why the adjustments are insufficient. *Psychological science,* 17**,** 311-318.

EVANS, J. 2008. Dual-processing accounts of reasoning, judgment, and social cognition. *Annu. Rev. Psychol.,* 59**,** 255-278.

FEYERABEND, P. 1993. *Against method,* Verso.

FINUCANE, M. L., ALHAKAMI, A., SLOVIC, P. & JOHNSON, S. M. 2000. The affect heuristic in judgments of risks and benefits. *Journal of behavioral decision making,* 13**,** 1-17.

FISCHHOFF, B. 1977. Perceived informativeness of facts. *Journal of Experimental Psychology: Human Perception and Performance,* 3**,** 349.

FORSTER, A. J., TALJAARD, M., BENNETT, C. & VAN WALRAVEN, C. 2012. Reliability of the peer-review process for adverse event rating. *PloS one,* 7**,** e41239.

FOX, R. C. 1989. The sociology of medicine: a participant observer's view.

GIGERENZER, G. 2008. Bounded and rational. *Philosophie: Grundlagen und Anwendungen/Philosophy: Foundations and Applications.* mentis.

GIGERENZER, G. 2018. The Bias Bias in Behavioral Economics. *Review of Behavioral Economics,* 5**,** 303-336.

GIGERENZER, G. & BRIGHTON, H. 2009. Homo Heuristicus: Why Biased Minds Make Better Inferences. *Topics in Cognitive Science,* 1**,** 107-143.

GIGERENZER, G. & GAISSMAIER, W. 2011a. Heuristic decision making. *Annu Rev Psychol,* 62.

GIGERENZER, G. & GAISSMAIER, W. 2011b. Heuristic decision making. *Annu Rev Psychol,* 62**,** 451-82.

GLADWELL, M. 2005. Blink. London: Allen Lane. Penguin.

GOLDMAN, R. L. 1992. The reliability of peer assessments of quality of care. *Journal of American Medical Association,* 267**,** 958-960.

GOLDMAN, R. L. 1994. The reliability of peer assessments of quality of care. *Journal of American Medical Association,* 17**,** 3-21.

GOWENSMITH, W., MURRIE, D. & BOCCACCINI, M. 2012. Field Reliability of Competence to Stand Trial Opinions: How Often Do Evaluators Agree, and What Do Judges Decide When Evaluators Disagree? *Law and human behavior,* 36**,** 130-9.

GRAZ, B., WIETLISBACH, V., PORCHET, F. & VADER, J. P. 2005. Prognosis or "curabo effect?" Physician prediction and patient outcome of surgery for low back pain and sciatica. *Spine,* 30**,** 1448-1452.

GUPTA, M., SCHRIGER, D. L. & TABAS, J. A. 2011. The presence of outcome bias in emergency physician retrospective judgments of the quality of care. *Annals of Emergency Medicine,* 57**,** 323-328.e9.

GUTHRIE, C., RACHLINSKI, J. J. & WISTRICH, A. J. 2007. Blinking on the bench: How judges decide cases. *Cornell L. Rev.,* 93**,** 1.

HELLER, R. F. & SALTZSTEIN, H. D. 1992. Heuristics in Medical and Non-medical Decision-making. *The Quarterly Journal of Experimental Psychology Section A,* 44**,** 211-235.

HEMMERICH, J. A., ELSTEIN, A. S., SCHWARZE, M. L., MOLISKI, E. G. & DALE, W. 2012. Risk as feelings in the effect of patient outcomes on physicians' future treatment decisions: A randomized trial and manipulation validation. *Social Science and Medicine,* 75**,** 367-376.

HOFER, T. P., BERNSTEIN, S. J., DEMONNER, S. & HAYWARD, R. A. 2000. Discussion between reviewers does not improve reliability of peer review of hospital quality. *Medical Care,* 38**,** 152-161.

HOFMANN, W., FRIESE, M. & WIERS, R. W. 2008. Impulsive versus reflective influences on health behavior: A theoretical framework and empirical review. *Health Psychology Review,* 2**,** 111-137.

HUTCHINSON, A., COSTER J, E., COOPER K, L., MCINTOSH, A. & WALTERS S, J. 2010. Comparison of case note review methods for evaluating quality and safety in health care. *Health Technology Assessment,* 14**,** 170.

KAHNEMAN, D. 2011. *Thinking, fast and slow*, Macmillan.

KAHNEMAN, D., SIBONY, O. & SUNSTEIN, C. R. 2021. *Noise: a flaw in human judgment*, Little, Brown.

KAHNEMAN, D., SLOVIC, P. & TVERSKY, A. 1982. Judgement under uncertainty: Heuristics and biases. Cambridge University Press, Cambridge.

KEIJZERS, G., FATOVICH, D. M., EGERTON-WARBURTON, D., CULLEN, L., SCOTT, I. A., GLASZIOU, P. & CROSKERRY, P. 2018. Deliberate clinical inertia: using meta-cognition to improve decision-making. *Emergency Medicine Australasia,* 30**,** 585-590.

KELLER, N., COKELY, E. T., KATSIKOPOULOS, K. V. & WEGWARTH, O. 2010. Naturalistic Heuristics for Decision Making. *Journal of Cognitive Engineering and Decision Making,* 4**,** 256-274.

KLEIN, G. A. 1999. *Sources of power: How people make decisions*, MIT press.

KOO, T. K. & LI, M. Y. 2016. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *Journal of chiropractic medicine,* 15**,** 155-163.

KUHN, T. S. 2012. *The structure of scientific revolutions*, University of Chicago press.

LARRICK, R. P. 2004. Debiasing.

LILFORD, R., EDWARDS, A., GIRLING, A., HOFER, T., DI TANNA, G. L., PETTY, J. & NICHOLL, J. 2007. Inter-rater reliability of case-note audit: A systematic review. *Journal of Health Services Research and Policy,* 12**,** 173-180.

LOEWENSTEIN, G. F. & PRELEC, D. 2000. Preferences for sequences of outcomes. *Choices, Values, and Frames***,** 565-577.

LUDOLPH, R. & SCHULZ, P. J. 2018. Debiasing Health-Related Judgments and Decision Making: A Systematic Review. *Medical Decision Making,* 38**,** 3-13.

MACKENZIE, E. J., STEINWACHS, D. M., BONE, L. R., FLOCCARE, D. J. & RAMZY, A. I. 1992. Inter-rater reliability of preventable death judgments. The Preventable Death Study Group. *The Journal of Trauma,* 33**,** 292-302, 302-3.

MAKARY, M. A., WICK, E. & FREISCHLAG, J. A. 2011. PPE, OPPE, and FPPE: complying with the new alphabet soup of credentialing. *Archives of Surgery,* 146**,** 642-644.

MAMEDE, S., DE CARVALHO-FILHO, M. A., DE FARIA, R. M. D., FRANCI, D., NUNES, M. D. P. T., RIBEIRO, L. M. C., BIEGELMEYER, J., ZWAAN, L. & SCHMIDT, H. G. 2020. 'Immunising' physicians against availability bias in diagnostic reasoning: a randomised controlled experiment. *BMJ Quality & Safety*.

MANASEKI-HOLLAND, S., LILFORD, R. J., TE, A. P., CHEN, Y.-F., GUPTA, K. K., CHILTON, P. J. & HOFER, T. P. 2019. Ranking hospitals based on preventable hospital death rates: A systematic review with implications for both direct measurement and indirect measurement through standardized mortality rates. *The Milbank Quarterly,* 97**,** 228-284.

MANDOLESI, L., POLVERINO, A., MONTUORI, S., FOTI, F., FERRAIOLI, G., SORRENTINO, P. & SORRENTINO, G. 2018. Effects of physical exercise on

cognitive functioning and wellbeing: biological and psychological benefits. *Frontiers in psychology,* 9**,** 509.

MARCIN, J. P., ROMANO, P. S., DHARMAR, M., CHAMBERLAIN, J. M., DUDLEY, N., MACIAS, C. G., NIGROVIC, L. E., POWELL, E. C., ROGERS, A. J. & SONNETT, M. 2018. Implicit review instrument to evaluate quality of care delivered by physicians to children in emergency departments. *Health services research,* 53**,** 1316-1334.

MAREWSKI, J. N. & GIGERENZER, G. 2012. Heuristic decision making in medicine. *Dialogues in Clinical Neuroscience,* 14**,** 77-89.

MARTEAU, T. M., HOLLANDS, G. J. & FLETCHER, P. C. 2012. Changing human behavior to prevent disease: the importance of targeting automatic processes. *science,* 337**,** 1492-1495.

MAZZOCCO, K. & CHERUBINI, P. 2010. The effect of outcome information on health professionals' spontaneous learning. *Medical Education,* 44**,** 962-968.

MICHIE, S., VAN STRALEN, M. M. & WEST, R. 2011. The behaviour change wheel: a new method for characterising and designing behaviour change interventions. *Implementation Science,* 6**,** 42.

MIRON, A. M. & BREHM, J. W. 2006. Reactance theory-40 years later. *Zeitschrift für Sozialpsychologie,* 37**,** 9-18.

MOREWEDGE, C. K., YOON, H., SCOPELLITI, I., SYMBORSKI, C. W., KORRIS, J. H. & KASSAM, K. S. 2015. Debiasing decisions: Improved decision making with a single training intervention. *Policy Insights from the Behavioral and Brain Sciences,* 2**,** 129-140.

MUMMA, G. H. & WILSON, S. B. 1995. Procedural debiasing of primacy/anchoring effects in clinical-like judgments. *Journal of clinical psychology,* 51**,** 841-853.

NATIONAL QUALITY BOARD UK 2017. National Guidance on Learning from Deaths: A Framework for NHS Trusts and NHS Foundation Trusts on Identifying, Reporting, Investigating and Learning from Deaths in Care.

NORMAN, G. R., MONTEIRO, S. D., SHERBINO, J., ILGEN, J. S., SCHMIDT, H. G. & MAMEDE, S. 2017. The causes of errors in clinical reasoning: cognitive biases, knowledge deficits, and dual process thinking. *Academic Medicine,* 92**,** 23-30.

PALUCK, E. L. 2009. Reducing intergroup prejudice and conflict using the media: a field experiment in Rwanda. *Journal of personality and social psychology,* 96**,** 574.

PETTIGREW, L. M. & MAYS, N. 2021. What role is there for 'nudging' clinicians? *British Journal of General Practice,* 71**,** 82.

RITOV, I. & BARON, J. 1995. Outcome Knowledge, Regret, and Omission Bias. *Organizational Behavior and Human Decision Processes,* 64**,** 119-127.

ROBERT A. CAPLAN, M. K. L. P., PHD; FREDERICK W. CHENEY, MD 1991. Effects of Outcome on Physicians' Judgment of Appropriateness of Care. *JAMA - Journal of the American Medical Association,* 266**,** 794.

ROGERS, B. A. & PLEAT, J. 2010. Is there adequate information on operation notes? The application of the Royal College of Surgeons of England guidelines. *Journal of perioperative practice,* 20**,** 339-342.

SALAS, E. & KLEIN, G. A. 2001. *Linking expertise and naturalistic decision making*, Psychology Press.

SANDARS, J. 2009. The use of reflection in medical education: AMEE Guide No. 44. *Medical teacher,* 31**,** 685-695.

SAPOSNIK, G., REDELMEIER, D., RUFF, C. C. & TOBLER, P. N. 2016. Cognitive biases associated with medical decisions: a systematic review. *BMC Medical Informatics and Decision Making,* 16**,** 138.

SCHROEDER, S. A. & KABCENELL, A. I. 1991. Do bad outcomes mean substandard care? *Jama,* 265**,** 1995-1995.

SHEERAN, P., GOLLWITZER, P. M. & BARGH, J. A. 2013. Nonconscious processes and health. *Health Psychology,* 32**,** 460.

SHI, Q., MACDERMID, J. C., SANTAGUIDA, P. L. & KYU, H. H. 2011. Predictors of surgical outcomes following anterior transposition of ulnar nerve for cubital tunnel syndrome: A systematic review. *Journal of Hand Surgery,* 36**,** 1996-2001.e6.

SIMINOFF, L. A. & FETTING, J. H. 1989. Effects of outcome framing on treatment decisions in the real world: Impact of framing on adjuvant breast cancer decisions. *Medical Decision Making,* 9**,** 262-271.

SIMON, H. A. 1979. Rational decision making in business organizations. *The American economic review,* 69**,** 493-513.

SULLIVAN, P. A. & ANDERSON, J. M. 2010. The Health Care Debate: If Lack of Tort Reform is Part of the Problem, Federalized Protection for Peer Review Needs to be Part of the Solution. *Roger Williams UL Rev.,* 15**,** 41.

TANAKA, Y., FUJINO, J., IDENO, T., OKUBO, S., TAKEMURA, K., MIYATA, J., KAWADA, R., FUJIMOTO, S., KUBOTA, M., SASAMOTO, A., HIROSE, K., TAKEUCHI, H., FUKUYAMA, H., MURAI, T. & TAKAHASHI, H. 2015. Are ambiguity aversion and ambiguity intolerance identical? A neuroeconomics investigation. *Frontiers in Psychology,* 5.

TE, A., QUINN, L., MANASEKI-HOLLAND, S., MANNION, R., LILFORD, R. & VLAEV, I. 2017. Systematic review of the cognitive biases, heuristics and human factors influencing the review and assessment of hospital medical records: a narrative synthesis (PROSPERO 2017 CRD42017060142) [UNPUBLISHED].

THALER, R. H. & SUNSTEIN, C. R. 2008. *Nudge: Improving decisions about health, wealth, and happiness*, Yale University Press.

THE ROYAL COLLEGE OF PHYSICIANS 2016a. National Mortality Case Record Review Programme (NMCRR): Using the structured judgement review method - A clinical governance guide to mortality case record reviews.

THE ROYAL COLLEGE OF PHYSICIANS. 2016b. *RCP hosts pioneering national Mortality Case Record Review Programme* [Online]. Available: https://www.rcplondon.ac.uk/news/rcp-hosts-pioneering-national-mortality-case-record-review-programme [Accessed 29/04/2016 2016].

TVERSKY, A. & KAHNEMAN, D. 1974. Judgment under uncertainty: Heuristics and biases. *Science,* 185**,** 1124-1131.

VLAEV, I. & DOLAN, P. 2015. Action change theory: A reinforcement learning perspective on behavior change. *Review of General Psychology,* 19**,** 69-95.

VLAEV, I., KING, D., DOLAN, P. & DARZI, A. 2016. The theory and practice of "nudging": changing health behaviors. *Public Administration Review,* 76**,** 550-561.

WEBB, T. L. & SHEERAN, P. 2006. Does changing behavioral intentions engender behavior change? A meta-analysis of the experimental evidence. *Psychological bulletin,* 132**,** 249.

WEBER, E. U. 2010. Risk attitude and preference. *Wiley Interdisciplinary Reviews: Cognitive Science,* 1**,** 79-88.

WEBER, E. U. & STERN, P. C. 2011. Public Understanding of Climate Change in the United States. *American Psychologist,* 66**,** 315-328.

WELLINGTON, H. H. 1982. The nature of judicial review. *The Yale Law Journal,* 91**,** 486-520.

WIKIPEDIA. 2020. *List of Cognitive Biases* [Online]. Available: https://en.wikipedia.org/wiki/List_of_cognitive_biases [Accessed 26/08/2020].

WILSON, J. VII—Internal and external validity in thought experiments. Proceedings of the Aristotelian Society, 2016. Oxford University Press, 127-152.

WOLFSON, A. M., DOCTOR, J. N. & BURNS, S. P. 2000. Clinician judgments of functional outcomes: How bias and perceived accuracy affect rating. *Archives of Physical Medicine and Rehabilitation,* 81**,** 1567-1574.

ZIMMERMANN, C., BALDO, C. & MOLINO, A. 2000. Framing of outcome and probability of recurrence: Breast cancer patients' choice of adjuvant chemotherapy (ACT) in hypothetical patient scenarios. *Breast Cancer Research and Treatment,* 60**,** 9-14.

CHAPTER 3: INVESTIGATING CASE-NOTE CHARACTERISTICS AND REVIEWER DEMOGRAPHICS AND ATTITUDES FOR THEIR INFLUENCE ON CASE-NOTE REVIEWER GLOBAL CARE QUALITY JUDGEMENTS IN THE HISLAC STUDY

"…in order to have anything like a complete theory of human rationality, we have to understand what role emotion plays in it."

-- Herbert Simon, cognitive scientist, Nobel Laureate

3.1. Chapter Preface

In continuation of the Chapter 2's theme of investigating reviewer's psychology, Chapter 3 defends and applies the psychological attitudes(FitzGerald and Hurst, 2017), demographics and individual differences (ID) (Eisenberg, 1979) to case-note reviewer care quality judgements. These components have the potential to explain observed variation 'error' for global care quality judgements as indicated in the below equation:

Demographics are population-based factors such as age, race, and sex. Demographics and psychological attributes have been widely used to study behaviours and decisions in healthcare. For instance, psychological factors such as locus of control and anxiety appear to be important determinants of medication on-adherence.(Davis et al., 2012) Demographics were used to explain physician decisions.(Swets et al., 2000) For our purpose, the use of demographics is reasonable as we seek to find any role demographics have in the observed variability between case-note reviewer care quality judgements.

Individual differences (ID) are defined as enduring psychological characteristics.(Baumeister and Vohs, 2007) ID help to distinguish one person from another and help to define each person's individuality which include traits such as intelligence, personality traits, and values. ID are neither a fiction nor a nuisance as they are enduring psychological features that

contribute to the shaping of behaviour and to each person's sense of self. And attitudes are "things" or possessions individuals acquire in response to their life experiences, personal chosen values, and information from society.(Dasgupta, 2013) Attitudes are extremely positive to extremely negative attitudes and both positive and negative attitudes can be held towards the same object in a range of contexts. Attitudes themselves are the feelings and reactions a person has towards a certain object.(Wood, 2000) For this thesis, the attitudes concern the feelings and reactions towards the process of appraising care quality using case-note reviews. Notably, individual differences and attitudes can sometimes be conflated especially if an attitude endures.(Burrus and Carney, 2015)

In summary, Chapter 3 investigates the influence of physician reviewer characteristics, their attitudes and case-note reviewer characteristics upon case-note reviewer care quality judgements using a multi-level model analysis.

3.2. Abstract

Objective: To test whether patient case-note characteristics, reviewer characteristics and attitudes significantly influence case-note reviewer care quality and confidence in care quality judgements to guide feasible approaches to reduce the variability around reviewer care quality judgements.

Data sources: Survey of the case-note reviewers involved in the High-Impact Specialist Led Acute Care (HiSLAC) study Phase II data (2015-2018) merged with the HiSLAC case-note review dataset.

Study design: In HiSLAC, four thousand randomly selected case records were reviewed by 79 senior doctors with a range of specialities to score for quality of care and confidence in this score. Following HiSLAC study completion, attitudinal surveys (Need for Cognition, Personal Need for Structure, Anxiety due to Uncertainty) were completed on RedCap electronic data capture system, by these HiSLAC reviewers.

Principal Findings: Of the 79 reviewers, 72 completed the surveys and had their data analysed. The mean care quality score was 4.17 (s.d. 0.82), range 1-5. The mean confidence in quality score was 82.6 (s.d. 18.6), range 0-100. Case-note reviewers accounted for 20% of total variance of care quality judgement. There was no relationship between the measured reviewer attitudes and their care quality judgements. Higher scores on the Personal Need for Structure and Anxiety due to Uncertainty scales were associated with modest reductions in reviewer confidence in the quality score they gave a case (-3.4[s.d. 1.4] and -2.6[s.d.] per standard deviation in scale score). The internal consistency (Cronbach α) of each of the three scales were above 0.7, corresponding with the literature. Other reviewer and patient characteristics examined did not explain any of the reviewer variance.

Conclusions: Case-note reviewers contribute about 20% of the variance in the care quality judgement score. The anxiety due to uncertainty and personal need for structure scores influence reviewer confidence in care quality judgement scores. While these results provide little guidance for improving selection or training of reviewers, the reliability of expert case note review for quality of care could still be improved by relatively simple strategies to calibrate reviews for the systematic differences found in ratings of individual reviewers.
Keywords: judgment, quality improvement, individuality, medical records [E05], hospital records [E05]

What was already known

- Reviewer case-note quality judgements have low inter-rater reliability for single case-notes, in part contributed by reviewer-specific variance
- Physicians' medical decisions are influenced by personal characteristics and attitudes
- Overconfident physician ratings of their self-competence are often inaccurate
- Case-note reviewer judgements have rarely been studied

What this study adds

- Systematic differences in the scores by case-note reviewer contributes one fifth of the variability of care quality judgement, a significant proportion of which may be amenable to modification.
- Care quality judgements are not influenced by three attitudinal measures (Personal Need for Structure, Need for Cognition, and Anxiety Due to Uncertainty, nor to the personal characteristics of reviewers

3.3. Introduction

Retrospective case-note review is a common method for assessing care quality in hospitals.(Manaseki-Holland et al., 2019b) The case-note reviewer with expertise in the clinical area concerned will typically evaluate care quality using a 5 or 6–point Likert scale.(Manaseki-Holland et al., 2019b) While it is known that care quality ratings from physician review of sets of case-notes have low reliability,(Goldman, 1992, Lilford et al., 2007, MacKenzie et al., 1992, Goldman, 1994, Hutchinson et al., 2010c) it is important to point out that the variability of judgements that leads to low reliability includes inherent uncertainty about clinical practice itself, as physician practice patterns show substantial variation that is attributed by some to the "diversity of accepted opinion on the need and value of alternative treatments."(Wennberg, 1987) This variability can only be resolved by improvements in the scientific evidence base for clinical practice.

However, there are plausible explanations put forward for these variations in clinical practice. The physicians' medical decisions can be influenced by factors unrelated to a patient's clinical presentation, including their personal characteristics and attitudes(Smith et al., 2008, Sabin et al., 2008), which themselves are influenced by experience (Albarracín and Wyer, 2000) and cognitive biases(Wallsten, 1981, Blumenthal-Barby and Krieger, 2015a, Saposnik et al., 2016). It seems more likely than not that any characteristics that affect management decisions by physicians in their own clinical practice would also affect their judgements about medical management by other clinicians. If so, these clinician reviewer characteristics could contribute to the lack of consistency in judgements observed in case-note reviews (Armstrong et al., 2007, Hutchinson et al., 2010a, Goldman, 1992, Manaseki-Holland et al., 2016, Goldman, 1994). The reliability of case-note reviews might then be improved, through modifications to the measurement procedure and analysis or by training and selection of

clinician reviewers, assuming we can identify reviewer factors affecting care quality judgements.

A wide array of factors is known to influence decision-making generally and medical decision-making. These include factors like age (Finucane et al., 2005), past experience(Juliusson et al., 2005), cognitive biases(Stanovich and West, 2008), individual differences (Parker et al., 2007) and salience for the individual (Acevedo and Krueger, 2004).(McKinlay et al., 2002, Burns et al., 1997, Forrest et al., 2006, Hajjaj et al., 2010) In an effort to focus on those factors most salient to case-note review, we previously conducted an expert consensus exercise to identify the set of cognitive biases and heuristics (Saposnik et al., 2016, Blumenthal-Barby and Krieger, 2014, Gilovich et al., 2002, Tversky and Kahneman, 1975) that are most likely to influence quality judgements made as part of case-note review.(Te. A et al., 2021) To our knowledge there is no literature that describes the influence of such factors on care quality judgements from case-note reviews of hospital patients.

In this study we have used case record reviews of care quality from the High Intensity Speciality Led Acute Care (HiSLAC) project(Bion et al., 2020), one of the larger case-note review projects undertaken, to investigate the relationship between several personal and attitudinal characteristics of reviewers and their care quality judgements and confidence in those judgements.

3.3.1. The HiSLAC Project

The HiSLAC project was funded by the National Institute of Health Research (NIHR-HS&DR Programme Commissioned call 12/128: Organisation & delivery of 24/7 healthcare).

HiSLAC was an independent, professionally led study which evaluated a key component of NHS England's policy drive for 7-day services: the intensity of specialist-led care of emergency medical admissions, with a focus on weekend provision. This research was important for patients and for the NHS care delivery strategy because it offered a unique opportunity to evaluate the impact of the transition to seven-day working, and to understand factors likely to impede or enhance the effectiveness of care following this change in practice.

In addition to examining the impact on patient-centred outcomes, the project also undertook a health economics analysis of the impact of increasing specialist provision across the NHS. (Watson et al., 2018) HiSLAC provided useful information across the NHS about the cost-effectiveness of investing in consultant and other specialist staffing in implementing the drive to 7-day service provision.

The HiSLAC had pooled 80 case-note reviewers, one of the largest populations studying case-note reviewers. We endeavoured to explore the influence of three key reviewer attitudinal measures, their demographics and the case-note review patient characteristics.

3.3.2. HiSLAC study characteristics

Two different epochs (time periods from which the case-notes were created) of emergency patients admitted to hospitals at weekends and on weekdays have no consequences upon the analysis as the admissions have been randomly selected across each epoch from the Trust electronic patient records. The study employed a difference-in-difference within Trust

analysis which compared the average change over time in the ‘global care quality judgement’ and ‘confidence in the global care quality judgement’ from that same Trust. Epoch-specific effects are inconsequential here as all notes are allocated and reviewed within a short period (approx. 3 months). For the three attitudinal measure constructs used, they have each been demonstrated to persist for period time greater than 3 months.(Gerrity et al., 1995, Sadowski and Gulgoz, 1992, Neuberg and Newsom, 1993)

3.4. Methods

3.4.1. Design and setting

The High Intensity Specialist-led Acute Care (HiSLAC) project was a cross-sectional two epoch case record review of 20 acute hospital Trusts in England to evaluate England’s seven-day acute care services.(Bion et al., 2017, Chen et al., 2016) Four thousand randomly assigned non-operative emergency admissions were reviewed by 79 reviewers with 800 case-notes reviewed in duplicate. Each reviewer reviews approximately 60 case notes giving a global care quality rating on a 5-point Likert scale (1 being lowest quality care and 5 being the best care quality) and self-rated their level of confidence in the global rating on a 0-100 continuous scale.

We extended this analysis by conducting a survey of the HiSLAC reviewers as they finished their reviews for the main project, and we then examined the relationship between case-note care quality judgements and reviewer attitudes and demographic characteristics as well as patient demographic and case characteristics. The nested nature of the data required the use of multilevel modelling.

There are several reasons for choosing a multilevel model and not a traditional multiple linear regression technique. First, the reviewers are themselves a group and a level because reviewers will each review different case-notes. And thus, the reviewer must be considered as one source of variation for the dependent outcomes. The case-notes will differ, in their characteristics, and will be one source of variation for the dependent outcomes. So, they too must be considered a level of their own. The method allows us to piece apart the contribution of the reviewer and the contribution of the case-note into separate variances. This is a more realistic understanding of care quality judgements as case-note reviews characteristics and reviewer characteristics could influence care quality judgement variance in different ways. The method of multiple linear regression, without the multiple levels, would not erroneously assume that the variance is the same for the reviewer and the case-notes. The reason this is not the case is because one would expect effects from the reviewers to be different to those from the case-notes. A traditional multiple linear regression assumes that all our reviewers would have the same variance and gradient of change across a certain variable. But the situation is that the ordinary least squares (OLS) assumption would treat the reviewer and the case-note as having identical variances and slopes (gradients) in relation with any other variable. This is well illustrated in a paper which employs multilevel and OLS methods on the same dataset and seeks to find the differences between them.(Jones et al., 2012) One consequence of failing to recognise hierarchical structures is that standard errors of regression coefficients will be underestimated, with confidence intervals being too narrow. Thus, this leads to an overstatement of statistical significance, and the higher likelihood of finding a false positive association. Standard errors for the coefficients of higher-level predictor variables will be the most affected by ignoring grouping.(Van den Noortgate et al., 2005)

In not being able to prove that the reviewer and the case-notes have different variances, the variances of the reviewer and the gradients are erroneously calculated. In brief, the multi-level model was needed to be able to separate reviewer effects from the case-note effects. The hypotheses we derived for use in conjunction with a multi-level model had the foreknowledge that reviewer and case-note effects on care quality judgements were likely to be different. This approach to understanding variance required careful design to appreciate the different sources of variance to the care quality judgement.

The multilevel model helps to elaborate on the reality of the sources of variability (i.e., attitudes, individual differences, demographics) for care quality judgements. The design of the HiSLAC study allowed us access to a large cadre of case-note reviewers and case-notes. The study has helped contribute a more accurate understanding of the sources of variation that influence care quality judgements. The multi-level model also allows us to partition the variance contribution from reviewer and from the case-notes, which in effect has allowed us an insight into the noise contributed at these respective levels. This variability can be explored through further attitudinal measures and factors as evidenced from any extant literature. This study has started to explore this variability in the reviewer judgements by sourcing this upon attitudinal measures, individual differences, and demographics. It is important to continue to establish the sources of this variance to better employ a suitable policy to review hospital deaths and make better use of limited clinician time and healthcare resourcing.

3.4.2. Multilevel Modelling

In the single level model multiple regression model, it is assumed that the measured units (i.e., reviewer ratings of the case notes) are independent and that the unexplained or residual[28] variations (e_{ij}) are not correlated with each other. However, the reviewer and the case-notes are expected to both contribute to the variance observed in these residual variations. In a multi-level model, each rating will have a distinct case note (i) and reviewer (j) component as represented in Eq.1. Furthermore, the case notes are clustered within reviewers. Multi-level modelling, otherwise called hierarchical linear models, can account for this clustering and hierarchy by splitting up the case note and reviewer variance components. Specifically, a null (empty) random intercept model for care quality scores by case-note (i) and reviewer (j) is expressed as:

$$y_{ij} = ß_0 + \mu_j + e_{ij} \quad (Eq.1)$$

y_{ij} = global care quality rating on Likert scale on the i^{th} case-note by j^{th} reviewer, $ß_0$ = mean rating across all reviewers, μ_j = random intercept for reviewer care quality score, e_{ij} = difference between individual reviewer rating and $ß_0$.[29]

[28] The residual for each observation is the difference between the observed value of the care quality score and the value of care quality score predicted by the model. In other words, it is difference between what the model accounts in the real score and the real score itself.

[29] which have an i and j variance component

This null model would require a multi-level model if intra class correlation, ρ, is small (<0.1).[30] However, this criterion alone is insufficient as we also need to consider the contribution of data structure to any differences across groups.(Nezlek, 2008)

Multiple patients (or case-notes) were reviewed by one physician case-note reviewer. Our dataset has multiple case-notes nested within reviewers. In addition, there were eight hundred case-notes which were duplicate reviews by two independent physicians. A standard logistic regression model assumes that the data is not nested and is found on a single level. A single level model assumes that there is no statistical bias from clustering effects, which can have important effects on the statistical results.

3.4.3. Sample size considerations

For testing the effect of level-one variables, the level-one sample size (~4,800) is of main importance; for testing the effect of a level-two variable it is the level-two sample size (~80 reviewers in the example). The average cluster sizes (n=60 at the reviewer-level) are not very important for the statistical power of such tests. This implies that the sample size at the reviewer level (n=80) is the main limiting characteristic of the design. In our case, there is no hindrance to determining a sufficiently powered analysis. It is almost always more informative to have a sample of 80 case-note reviewers with 4,800 case-notes than one of 40 case-note reviewers and 4,800. For example, a sample of 600 schools with a total of 3,300

[30] Namely, less than 10% of variation in global care quality rating is attributable to reviewer differences; a low ρ suggests negligible clustering within groups and the extent to which the variance in y_{ij} is attributable to reviewer differences

pupils would even be much better with respect to power, despite the low average number of students (5.5) sampled per school.

Small cluster sizes would mean we have less power for testing random slope variances at the case-note reviewer level, i.e., between-reviewer variances of effects of case-note variables. With a cluster size of 60 case-note for each reviewer, this is not a small number and sufficient for our purposes. Furthermore, for the examination of any cross-classified structure, the duplicate review cluster size is recommended to be ~10. We have 800 total duplicate reviews/72 case-note reviewers which is greater than 10.(Snijders and Bosker, 1999)

3.4.4. Understanding the effects of clustering at the single level

Clustering can be accounted for by a *design effect*. The *design effect* of this single-level clustering is expressed in *Eq.2*:

$$\text{Design effect } = [1 + \rho(\text{nclus} - 1)] \quad (Eq. 2)$$

where ρ is the intra-class correlation, n_{clus} is the common cluster size. The intra-class correlation represents the degree of similarity between patients within the same case-note reviewer. When the design effect is 1, the cluster is 1 which indicates that each case-note is independent with no clustering effects within reviewers, not any other variable.(Kish, 1965) When the design effect is greater than 1, this indicates that there is clustering at a single level. This then leads to underestimates of the sampling variance. This is because a single-level model calculates standard errors on the assumption that the individuals in the sample provide, say, 2,000 independent pieces of information. When outcomes are clustered, there will be fewer than 2,000 independent observations. This subsequently leads to erroneous standard errors.

Standard errors (SE), commonly referred to as standard error of the mean, indicates the expected difference between the sample mean and the population mean. It can be calculated using *Eq.3*:

$$Standard\ error = \frac{\sigma}{\sqrt{n}} \quad (Eq. 3)$$

Where σ is the standard deviation and *n* are the number of (independent) observations or sample size. As you can see, as the number of independent observations fall with clustering effects, the denominator term decreases. Standard errors are underestimated because any clustering (dependence) effects will reduce the effective sample size (n_{eff}) to less than *n*. For instance, if the σ was 3 and n = 2,000, we would obtain a standard error of 0.0671. However, if we were to have any clustering effects with the same σ and a n_{eff} = 1500, we would have a standard error of 0.0775. When we do not account for clustering effects (and the effective sample size) the standard errors are underestimated (i.e., 0.0671 (n_{eff} = 2000) < 0.0775 (n_{eff} = 1500)

The underestimation of standard errors subsequently leads to underestimated confidence intervals. Standard errors are proportionally correlated with confidence intervals, as indicated in *Eq. 4*:

$$Confidence\ interval = mean\ \pm (zscore\ \times\ standard\ error) \quad Eq. 4$$

The mean is the arithmetic mean, the z-score represents the z-statistic for the chosen percentile point (1.96 for intervals between 2.5 (lower bound) and 97.5 (upper bound) and the standard error represents the dispersion between the sample mean and the expected population mean). Correspondingly, p-values are also narrower. This will entail that we are

more likely to incorrectly reject the null hypothesis. This is known as a type I error or "false positive."

3.4.5. Clustering at multiple levels

Clustering is often found at multiple levels across datasets (i.e. patients within physicians and children within schools).(Rasbash, 2008) A multi-level model can account for these effects in a statistical model which represent the dependencies and hierarchical structure in the data. The difference between a single level regression model and a multi-level model is that the latter separates the variance components at the individual and the group level whilst the former does not. For this dataset, Figure 4. represents some of the relationships between the individual units (case-notes/patients) and the group units (case-note reviewers). Case-notes ($C_{1,2,3\ldots n}$) are nested within distinct case-note reviewers ($R_{1,2,3\ldots n}$). For illustrative purposes, I use a lower C-to-R ratio to demonstrate the type of relations over the actual C-to-R relations that are prohibitive for any heuristic purpose.

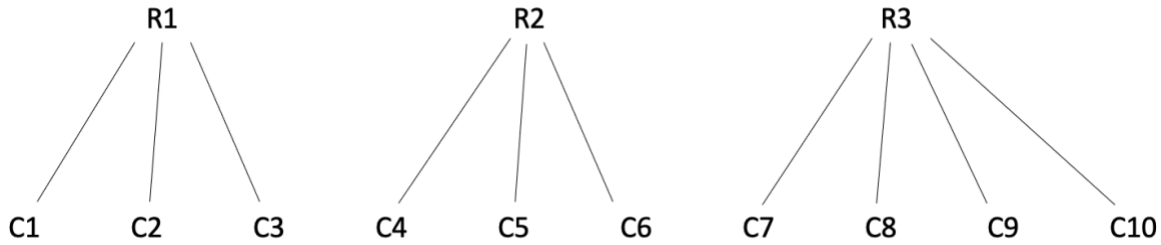


Figure 4. Simple Multilevel model relations

This is more simply represented in Figure 5, where the direct relationship between case-notes and reviewer are indicated in this simple multi-level model.

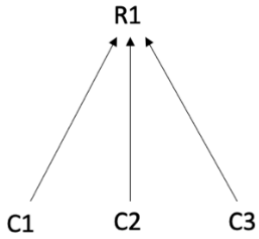


Figure 5. Our classification diagram for the simple multi-level model structure

However, this simple multi-level relation is not entirely representative of our dataset. to realistically represent the proportion of duplicate reviews by R-pairs, we need another model to account for this cross-classified structure.

3.4.6. Cross-classified models

In cross-classified data, lower-level units do not belong to one and only one higher level unit but belong to pairs or combinations of higher-level units formed by the crossing of two or more higher level classifications with each other. A common example in health services research arises in studies of hospital patient outcomes. Hospitals and general practitioners (GPs, i.e., family doctors) are cross classified as GPs tend to refer their patients to different hospitals depending on patient need while hospitals typically treat patients who have been referred by many different GPs.

In our dataset, we observe that multiple case-notes are nested within a reviewer, however, a select number of individual case-notes are each reviewed by two reviewers (i.e., in other words, a reviewer pair is nested within a case-note). This is what we observe when we have reviewer pairs (R-pairs) each reviewing the same case-note. This leads to a new non-hierarchical level, which can be considered in a cross-classified model. Figure 6. represents

the cross-classified relations of the 800 duplicate reviews ($D_{1,2,3....n}$) across R-pairs in addition to the simple multi-level model from Figure 5.

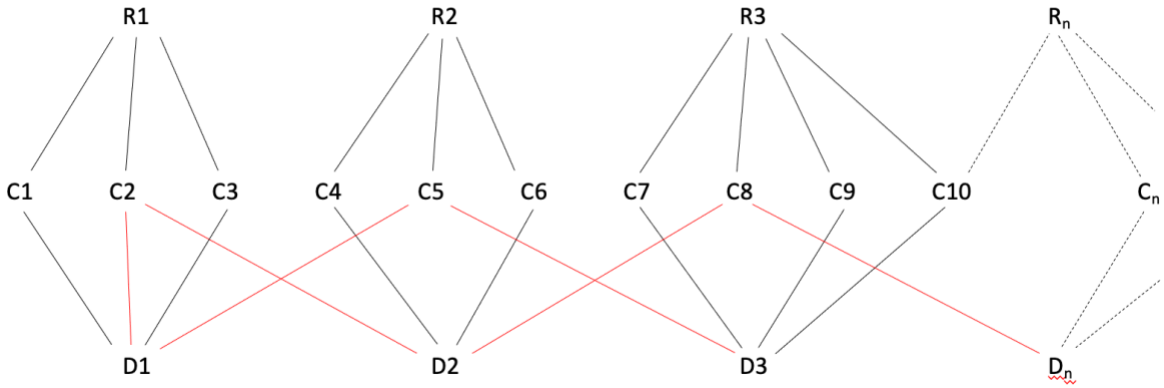


Figure 6. Cross-classified model relations

For instance, we could observe that the same case-note is nested within more than one reviewer. This is indeed what we observe with 800 case-notes (of the total 4,800) which have been duplicate reviewed by different pairs of reviewers. For simplicity, the classification diagram for the two-way (and higher) classified structure is found in Figure 7.

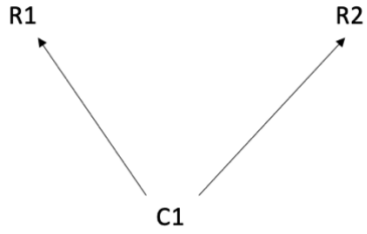


Figure 7. Our classification diagram for the two-way classified structure

Higher than two-way classifications are possible, but this is not represented in our dataset. With more highly cross-classified data structures, the cross-classified structure is more likely to undermine the assumptions of the simple multi-level model.(Rasbash, 2008) Some of the variation that we attribute to reviewers may be better characterised as review-occasion

differences. The review occasion, namely that occasion of when different reviewers review the same case-note at different places and points in time, may be a factor which leads to the imprecise attribution of variances at the case-note and the reviewer level. The naive fitting of the nearest equivalent hierarchical model to cross-classified data will lead to the misattribution of the care quality score variation to the case-note and reviewer levels.(Van den Noortgate et al., 2005) By incorrectly modelling the dependency in the data we will likely obtain biased standard errors for the predictor variables, especially those measured at the reviewer level (higher levels). A simple cross-classified model is represented in Eq.5:

$$y_{ijk} = ß_0 + \nu_k + \mu_j + e_{ijk}. \quad \text{(Eq.5)}$$

where $yijk$ - is the care quality score of the i^{th} case-note (i) nested within j^{th} reviewer (R1) to which the k^{th} reviewer-pair (R2) is nested within i^{th} case-note, $ß_0$ - is the mean care quality score across all reviewers, ν_k - is the effect of k^{th} reviewer-pair *k* (R2), μ_j - is the effect of the j^{th} reviewer *j* (R1), and e_{ijk} - is the case-note-level score residual error term with variance components segregated into *i, j* and *k*.[31]

If cross-classification structures are not considered, this could lead to the underestimation of standard errors. This will raise the likelihood of a type 1 error. The random allocation of these duplicate case-notes to reviewers ensure that the effect of cross-classification are accounted for.(Bion et al., 2017) To ensure this assumption has been met, a cross-classified

[31] The random effects and residual errors are assumed independent of one another and normally distributed with zero means and constant variances.

model is run to compare variance partition coefficients with the simple multi-level model. The variance partition coefficients measure the proportion of total variance that is due to differences between reviewers. And any remaining coefficients account for the differences within reviewers.

3.4.7. Data cleaning

First, the excel file containing the case note review data was loaded into STATA version 16. The codebook command was run to ascertain the variables and its properties. There were no missing values. Each of the scales has different sum totals and so raw scores would be incommensurable. And so standardized scales were created to allow for the comparison between different scales.

In a separate excel file, the reviewer data from the RedCAP survey was loaded into STATA. The data at the variable level was cleaned for duplicates and the variable structure was tidied removing redundant variables. The within-variable information was recoded for simplicity and ease of analysis and interpretability later.

The case-note review data was then loaded into STATA and inspected. The variables names were renamed for interpretability. The data was frequency tabulated to check completeness at the variable level. Variables which were not useful for the analysis were dropped. Rogue data which did not fit within their scales were re-scaled using *mydecode* such as with the length of stay (LOS) variable with magnitude '-1.' Lastly, the three datasets were then merged using "merge."

3.4.8. Model-building

These standardized scales were then had reliability correlations run across them to ensure that the scales were not measuring the same construct, which is otherwise known as co-variance. If two instruments measure the same underlying construct (or a very similar construct) and are both including in a linear regression model, then there is the high chance of observing co-variance, which is the dependency of one variable upon the other. This is known as multiple collinearity. This can have unseen consequences when investigators are trying to make causal inferences or discuss implications from their results. The graph matrix function was used to generate these correlation matrices. This generated scale correlations. The strongest correlation is between PNS and ADTU (0.34) but there are no high correlations between the scales.

An empty multi-level regression (i.e., without dependent variables) model was run to produce an estimate of the average case-note reviewer care quality judgement and the confidence intervals. An intra-class correlation was found which shows the percentage of variation in quality score which is attributable to the reviewer or the patient records within reviewer.

The variables retained within the final dataset were based on pre-developed hypotheses and the completeness of each of the variables. The patient level predictors were individually placed into the empty multi-level model to test whether there was a statistically significant contribution to the care quality score. These included patient gender, LOS, and patient condition pre-admission. It was felt that patient gender may have a differential impact on the types of presenting patients (diagnostic code) which could be reflected in the care quality score. I considered LOS as a possible proxy for the complexity of the patient, which could

differentially impact the care quality judgement; I was looking for how longer lengths of stay could lead to a biased decision which rated the care quality worse than one other would have done if the stay was shorter. And the patient pre-admission variable is a categorical description of the patient's health status prior to admission; it is plausible to indicate that the starting condition or point of a patient may heavily shape the perceptions of the reviewers. In this, I was looking for a bias of judgement like a desire to preserve the weak and infirm over those who were not so to begin with; it is more a type of emotional bias.

At the reviewer level variables, which is the second level of aggregation, the variables separately considered in this model were gender and reviewer years since graduation. I envisaged that gender may have an effect on care quality judgement for their broad gender effects on the manner and style of reviewing.(McKinlay et al., 1996) And the reviewer years since graduation is a proxy for experience (and for age) which has been shown previously to influence medical decision-making.

Scales scores were pictorially represented in graphs depicting the scale score against the care quality score; this would make it easier to visualize the effect of these constructs on the scores. Next, the multi-level models were then populated with these variables; as indicated the models were built according to the plausible influence hypothetically or as informed by the literature. Different models were run with different independent variables (using '*qui mixed*' (to run the model), '*est store*' (to store) and '*lrtest*' to evaluate the model; these individual models were stored in STATA's local memory for access later. Once all the models had been run and stored, these were then compared using the estout function to bring all the models and respective lrtest outputs for comparison in a single table. The lr_test output with a statistically significant lrtest_p-value indicates that the model is superior and

significantly so compared to the other models within the test parameters typically denoted with a θ.(Koch, 2013) This was undertaken for both the quality score and the confidence in the quality score as dependent variables. When testing for the main effect of reviewer-specific variables such as the 3 attitudinal measure scores, it is only possible to use variables which are at the same level (i.e., at the reviewer level). Thus, there were five models which were available for each attitudinal measure (i.e., the empty model, the fixed effects models and the three other random effects (multi-level) models). It must be noted that these tests do not determine whether the coefficient for attitudinal measure changes between models but only whether one model is superior to the other. Finally, a *marginsplot* (using the margins function) is generated which creates a plot of set attitudinal measure cut points with confidence intervals against the dependent variables (i.e., care quality score, confidence in care quality score).

3.4.9. Attitudinal measures

The two behavioural psychologists (Ivo Vlaev and Kelly Ann Schmidtke) provided guidance on the academic literature and potential instruments to guide the selection of the attitudinal measures. These sources were primarily from supervisory input and Measurement Instrument Database for the Social Sciences (MIDSS). These instruments were longlisted, and criteria were mutually agreed to whittle these down into a shortlist. These criteria were:

- previous use and evidence of its use in medical decision-making,
- the total time taken to complete the instrument,
- acceptability to survey audience
- potential contribution to a new, burgeoning field of inquiry

Based on these criteria the following instruments were first longlisted:

- MSTAT I(McLain, 1993), The Physicians' Reaction to Uncertainty Scale(Gerrity et al., 1995)
- Risk Attitudes Scales(Weber et al., 2002)
- Munich Chronotype Questionnaire(Juda et al., 2013)
- Morningness-Eveningness Questionnaire(Horne and Östberg, 1976)
- Need for Cognition scale(Lins de Holanda Coelho et al., 2018)
- Personal Need For Structure(Neuberg and Newsom, 1993)
- Proactive Coping Inventory(Greenglass et al., 2014)
- Physician Anxiety due to Ambiguity(Gerrity et al., 1995)
- Overconfidence vignettes(Meyer et al., 2013)
- Outcome Bias Instrument(Gupta et al., 2011)
- Framing effect survey(Bui et al., 2015)

Following further review with the behavioural psychologists and clinical professionals, with no involvement with HiSLAC, we shortlisted the following instruments:

- Need for Cognition
- Personal Need for Structure
- Physician Anxiety due to Ambiguity

These tools were chosen specifically for their ease for reviewers to complete, the minimal time taken to complete and their well-validated status.

The overall step-wise process is shown in List 1 below:

List 1: Psychometric tool selection process

1. Individual differences (ID) and psychological traits (PT) are proposed to influence the reviewer global care quality judgements.
2. I consulted the Society of Judgement and decision-making for their inventory of over 200 psychological measures, which is a comprehensive and authoritative inventory of psychological measures.
3. I then consulted two behavioural scientist/psychologist (Ivo Vlaev and Kelly Ann Schmidtke) on the tools that, by their expert judgement, would be involved. Recommendations were to focus on measuring risk constructs, as this is considered essential in the evaluation of patient care.
4. AT constructed a psychometric database profiling feature for each test. These were discussed with the supervisory team.
5. The shortlisted subscales were then included; these included 5 MSTAT-I, Physician Reactions to Uncertainty, Overconfidence, Risk attitude scales, Outcome bias, DOSPERT
6. This was presented for use in the HISLAC study as an amendment. The HiSLAC Principal Investigator were critical of these instruments and their applicability to the medical context.
7. Following this feedback, these tools were then selected following input from my research team for relevance to the time-constrained medical reviewers. The constructs of these tools were considered pertinent for reviewing and them being possible influencers of care quality judgements.

8. The following three tools were selected, physician anxiety due to ambiguity, need for cognition, personal need for structure.

In table 4. there are several features that describe the various features of each psychological characteristic in the survey, including psychometric measure, the number of instrument items, the response options, Cronbach alpha, its validation, and the approximate time to complete the scale. Table 9. outlines the key for the properties of these various measures whilst Table 10. presents the psychometric properties of the attitudinal measures. Table 11. presents the attitudinal measure items.

Table 9. Key of Psychometric properties

Feature	**Definition**
Psychometric measure	The purported construct being measured
Instrument items	The number of individual items that constitutes the whole instrument
Response options	The type of response option(s) (i.e., Likert, numerical, categorical)
Cronbach alpha	A measure of internal consistency and reliability of the instrument.(Tavakol and Dennick, 2011, Bland and Altman, 1997)
Validation	Psychometric tools are considered more robust once validated across different contexts.

Time to complete	The time to complete the feature is an important consideration given the little disposable time clinicians typically have.

Table 10 describes the psychometric properties of these three attitudinal measures (Need for cognition, Personal Need for Structure and Physician Anxiety due to ambiguity).

Table 10. Psychometric properties of attitudinal measures						
Name	Psychometric Measure	Number of Items/ Point Ratings	Response options	Cronbach α	Validity	Completion Time (mins)
Need for Cognition subscale (derived from Rational-Experiential Inventory (REI))	preferences for information processing	5	5 point Likert Scale, (1 = completely false, 5 = completely true)	0.75-0.86(Reyna and Ortiz, 2016)	Yes(Björklund and Bäckström, 2008, Marks et al., 2008) Robust **convergent** and **discriminant** validity (i.e. Scores are said to 'converge' with similar measures and 'diverge' from dissimilar ones.)(Björklund and Bäckström, 2008)	3

					Concurrent validity (i.e., whether a measure correlates significantly with relevant criteria) was negatively correlated with risky choice framing effects in Kahneman and Tversky's Asian disease task.(Marks et al., 2008) Yes(Monacis et al., 2016, Reyna and Ortiz, 2016, Calder et al., 2012) Weak **predictive validity** (i.e., aimed at predicting behaviour) for recreational activities.(Reyna and Ortiz, 2016)	

					Strong **construct validity** as factors do correlate with suspect construct. Moderate **convergent validity** with similar and related measures, and the **discriminant validity** associating with two thinking styles in Italian students.(Monacis et al., 2016) **Face validity** with emergency physicians.(Calder et al., 2012) Not yet validated in health care professionals.	

Personal Need for Structure (PNS)	preferences for "structure and clarity in most situations, with ambiguity and grey areas proving troublesome and annoying."	12	6 point Likert scale, where 1 = strongly disagree and 6 = strongly agree	0.77(Neuberg and Newsom, 1993)	Yes(Thompson et al., 2001, Selltiz et al., 1976, Neuberg and Newsom, 1993) Reasonably good convergent and discriminant validity.(Neuberg and Newsom, 1993) Good **convergent validity** for psychology students.(Selltiz et al., 1976, Thompson et al., 2001) Not yet validated in health care professionals.	6

Anxiety due to uncertainty subscale (Revised Physicians' Reactions to Uncertainty scales)	developed to measure physician's affective reactions to uncertainty.	5	6 point Likert scale, where 1 = strongly disagree and 6 = strongly agree	0.79(Gerrity et al., 1995)	Yes(Gerrity et al., 1995, Schneider et al., 2007) Moderate to strong **structural validity** (correlations between the items were as expected) of the individual **constructs** Anxiety due to Uncertainty, Concern About Bad Outcomes, Disclosing Uncertainty to Patients, and Disclosing Mistakes to Physicians.(Gerrity et al., 1995) Strong **item-discriminant validity** (i.e., extent to which an item measures what it is not supposed to measure)	3

					between the scales 'anxiety due to uncertainty' and 'concern about bad outcomes' with overlap between both constructs.(Schneider et al., 2007) Validated in health care professionals.	

Table 11. The attitudinal measure items

	Need for Cognition subscale (from Rational-Experiential Inventory (REI) - (1 = completely false to 5 = completely true)
1.	I would prefer complex to simple problems.
2.	I like to have the responsibility of handling a situation that requires a lot of thinking.
3.	Thinking is not my idea of fun.
4.	I would rather do something that requires little thought than something that is sure to challenge my thinking abilities.
5.	I really enjoy a task that involves coming up with new solutions to problems.
6.	I would prefer a task that is intellectual, difficult, and important to one that is somewhat important but does not require much thought.
	Personal Need for Structure (PNS) (1 = strongly disagree to 6 = strongly agree)
7.	It upsets me to go into a situation without knowing what I can expect from it.
8.	I'm not bothered by things that upset my daily routine.
9.	I enjoy having a clear and structured mode of life.

10.	I like a place for everything and everything in its place.
11.	I like being spontaneous.
12.	I find that a well-ordered life with regular hours makes my life tedious.
13.	I don't like situations that are uncertain.
14.	I hate to change my plans at the last minute.
15.	I hate to be with people that are unpredictable.
16.	I find that a consistent routine enables me to enjoy life more.
17.	I enjoy the exhilaration of being put in unpredictable situations.

18.	I become uncomfortable when the rules in a situation are not clear.
	Anxiety due to uncertainty subscale (From Revised Physicians' Reactions to Uncertainty scales) (1 = strongly disagree to 6 = strongly agree)
19.	I usually feel anxious when I am not sure of a diagnosis.
20.	I find the uncertainty involved in patient care disconcerting.
21.	Uncertainty in patient care makes me uneasy.
22.	I am quite comfortable with the uncertainty in patient care.
23.	The uncertainty of patient care often troubles me.

Having introduced the HiSLAC study, the rationale for demographics and attitudinal measures and the selection process of key demographics and the attitudinal measures, I turn

to its corresponding publication manuscript, which is pending submission to Health Services Research.

3.4.10. Survey of HISLAC reviewers

The survey invitation was sent out in June 2019 timed for one month after the individual reviewer had completed their HiSLAC case-note reviews. The month gap provided time to allow HiSLAC reviewers to answer data queries from the research team on their case-note reviews. As with HiSLAC's original case-note review data, the survey was completed on RedCap(Harris et al., 2009), an electronic data capture (EDC) system. There was no incentive offered for completing the survey. Non-completers were followed up by email (Appendix 10).

Involving the study team which included practicing physicians, clinicians, health system experts and two psychologists, we posited that attitudinal scales measuring "*risk*" and "*orderliness*" would be the most pertinent to case record reviews. I decided to include a risk measure because medical treatments are widely accepted as having some element of risk, as indicated by professional medical bodies. (The Royal College of Anaesthetists, 2020, The Royal College of Obstetricians and Gynaecologists, 2015) and medical researchers.(BMJ Best Practise, 2002)

On orderliness, given the various presentation styles of case-notes, the clinician's personal response to the "orderliness" of the case-note could influence their final care quality judgement. The UK Royal College of Physicians has recommended a set of guidelines for the clinical structure and content of clinical patient records which are essential for safe, high

quality care and have wider implications for its research value and quality improvement potential.(The Health and Social Care Information Centre, 2013) For example, physicians with a lower desire for orderliness could give higher care quality judgements than those with higher desire for orderliness. We can measure any influence a physician's personal desire for "orderliness", which we used the "personal need for structure" to measure because their definitions overlap, could have upon their care quality judgements. We can do this because of the random allocation of case-notes which, via *ceteris paribus*, equalises the presentational forms of case-notes.

To source the instruments, the Decision-Making Individual Differences Inventory (DMIDI) was recommended by two behavioural science doctoral supervisors as a comprehensive repository of general psychological instruments. Other population-specific instruments were sourced from supervisors and obtained from academic articles. From the DMIDI and we then selected a short-list of 8 attitudinal scales and cognitive biases related to these domains.(Society of Judgement and Decision-making, 2020) which satisfied our two attitudinal dimensions, "risk" and "orderliness." These five scales were selected for on two dimensions: the time (minutes) taken to complete the scale and its pertinence to the research question.

The rationale for selecting the "risk" and "orderliness" dimensions was because risk was a pervasive component of medical decision-making and the perception of orderliness was a

critical component of medical decision-making.[32] The reason why we chose the DMIDI as our source of attitudinal measure was because it offered a very large, comprehensive, and easy-to-access resource of psychological instruments, above and beyond the capacity of other resource compendia. It was firmly recommended by my two supervisors, both cognitive psychologists. In consultation about the salience of the measures to case-note review with the HiSLAC study leadership and given their desire to minimise the burden of the questionnaire on busy physicians who had already spent a considerable amount of time on the reviews for the parent study, we chose the following three measures.

Need for Cognition (NFC) short-scale was selected to measure the reviewer's tendency towards cognitive work given that case-note reviewing is a cognitively demanding task.(Petty and Cacioppo, 2012) NFC is a sub-scale of a larger scale from the rational-experiential framework which seeks to capture the individual spectrum of rational to experiential influences. The sub-scale has 6 items with a scale score range of 6 to 30. A high score indicates higher need for cognition. Prior work by the developers showed that across two groups of online participants from the general population, one in the United States (n=521) the mean of this scale was 22.11 (s.d. 0.21) and second in the United Kingdom (n=476) the mean of this scale was 21.44 (s.d. 0.24). Given that the evaluation of case-note reviews is cognitively demanding, those with higher NFC scores are expected to deliberate with greater effort over the case-notes. Consequently, those with lower NFCs are expected to give less variable care quality judgements than those with higher NFC because they will be less

[32] See immediate section above "Attitudinal measures" for more details on the initial longlist and shortlist of tools.

assiduous to the case-note details contributing to overall care quality judgement. Thus, those with lower NFCs will give a lower range of care quality scores, with less variability, than those with high NFCs. Also, we predict that lower NFC are associated with lower confidence in care quality scores because studies have shown that lower NFC results, representative of lax cognitive processing, are correlated with higher confidence in their own judgements.(Harvey, 1994, Peterson and Pitz, 1986)

Personal Need for Structure (PNS) measures the desire someone has for correct structure and order; this is relevant as medical case-notes are frequently very variable in the degree to which they are ordered and legible.(Neuberg and Newsom, 1993) Work by the University of Groningen showed that across all participants faculties (n=53, response rate 85.3%) the mean scale score was 3.18 (mean s.d. 0.58 across individual items) in investigating how PNS affect the perceived usefulness of structured and unstructured evaluations.(Slijkhuis et al., 2013) This scale has 12 items with score range (12-72). A high scale score indicates higher desire for structure. We consider Personal Need for Structure (PNS), our proxy for measuring the reviewer's desire for orderliness, given the similarity between desire for order and desire for structure. We predict that physicians with higher PNS will give lower care quality scores and be less confident of their scores for poorly documented records. As case-notes with poor documentation would be evenly distributed across reviewers by randomization, we would then expect systematically lower scores for reviewers with high PNS.

Anxiety Due to Uncertainty (ADTU) captures a person's anxiety when under situations of imperfect or unknown information, a common feature of case-notes. It is one of four sub-scales which comprise the Revised Physicians' Reactions to Uncertainty scales and is composed of five items (Gerrity et al., 1995) with a scale score range of 5 to 30). Prior work

by the developers showed that across all residents, fellows and faculty at the University of Indiana (n=269, response rate 79%) the mean total of this scale was 18.8 (s.d. 4.7).(Gerrity et al., 1995) Reviewers who scored higher on this scale could be expected to have lower quality ratings, potentially due to their ambiguity intolerance when required to make judgements based on incomplete or illegible records. The psychometric properties of the instruments are listed in Table 12.

Table 12. Characteristics of the Psychometric Tools

Name	Psychometric Measure	Number of Items/ Point Judgements	Response options	Cronbach α	Validated	Estimated Completion Time (mins)
NFC (from Rational-Experiential Inventory (REI))	Preferences for information processing(29)	5	5 point Likert Scale, (1 = completely false, 5 = completely true)	0.75-0.86(25)	Yes(24, 30) (Internal) Yes(25, 31) (External)	3
PNS	Preferences for "structure and clarity in most situations, with ambiguity and grey areas proving troublesome and annoying."(27)	12	6 point Likert scale, (1 = strongly disagree, 6 = strongly agree)	0.77(32)	Yes(27) (Internal)	6

ADTU (from Revised Physicians' Reactions to Uncertainty scales)	Physician's affective reactions to uncertainty.(28)	5	6 point Likert scale (1 = strongly disagree, 6 = strongly agree)	0.79(28)	Yes(28) (Internal) Yes (External validity with primary care physicians)(33)	3

The attitudinal instrument items are listed in Table 13. Additional questions in the survey measured reviewer demographics including 'gender' as well as level of training by 'time since the reviewer graduated', factors that have previously been shown to influence clinical decision-making and thus might influence quality judgements about clinical decisions. (Juliusson et al., 2005, Acevedo and Krueger, 2004, Hayward et al., 1997, Eisenberg et al., 1983, Cher et al., 1997, Hayes et al., 1999)

Table 13. Attitudinal measures with their items (22 items)

	NFC subscale (from Rational-Experiential Inventory[a] (REI)
1.	I would prefer complex to simple problems.
2.	I like to have the responsibility of handling a situation that requires a lot of thinking.
3.	Thinking was not my idea of fun.
4.	I would rather do something that requires little thought than something that was sure to challenge my thinking abilities.
5.	I really enjoy a task that involves coming up with new solutions to problems.
6.	I would prefer a task that was intellectual, difficult, and important to one that was somewhat important but does not require much thought.
	Personal Need for Structure (PNS)[b]
7.	It upsets me to go into a situation without knowing what I can expect from it.
8.	I'm not bothered by things that upset my daily routine.
9.	I enjoy having a clear and structured mode of life.
10.	I like a place for everything and everything in its place.
11.	I like being spontaneous.

12.	I find that a well-ordered life with regular hours makes my life tedious.
13.	I don't like situations that are uncertain.
14.	I hate to change my plans at the last minute.
15.	I hate to be with people that are unpredictable.
16.	I find that a consistent routine enables me to enjoy life more.
17.	I enjoy the exhilaration of being put in unpredictable situations.
18.	I become uncomfortable when the rules in a situation are not clear.
	ADTU subscale (From Revised Physicians' Reactions to Uncertainty scales)[b]
19.	I usually feel anxious when I am not sure of a diagnosis.
20.	I find the uncertainty involved in patient care disconcerting.
21.	Uncertainty in patient care makes me uneasy.
22.	I am quite comfortable with the uncertainty in patient care.
23.	The uncertainty of patient care often troubles me.

[a] = 5-point Likert scale, [b] = 6-point Likert scale

Patient level variables obtained from administrative data or case note reviews.

Patients were randomly assigned to reviewers so differences in patient characteristics would not contribute to differences in average reviewer ratings in our design, but we were interested in describing how case-note reviewer care quality judgements and their confidence in those judgements would vary within reviewer across patients (case-notes).Thus, we included patient gender, length of stay and patient level of functioning pre-admission.

3.5. Analytic approach

3.5.1. Key HiSLAC features relevant to the analysis

As mentioned, the case-notes were assigned randomly across reviewers and period (epoch) which was the essential feature of the original study that balanced patient attributes across reviewers allowing us to compare ratings between reviewers. In the original study there were 3967 case-notes that had analysable quality ratings with 793 reviewed twice for a total of 4760 reviews by 79 reviewers. For the 72 reviewers who answered our survey we had 4372 reviews of 3705 patient case-notes with quality ratings.

For the purposes of our analysis each case-note review served as a measurement of the reviewer's severity, or tendency to rate quality higher or lower than other reviewers. As we had multiple measurements for each reviewer, we used a multilevel analysis. This allowed us to estimate the variance attributable to differences in reviewer average quality rating and a residual variance due to both patient differences in quality and occasion-to-occasion variability of reviews of the same patient. There were 800 duplicate reviews of patient case-notes by design of which 667 remained when restricted to our responding reviewers. Randomly assigned duplicate reviews offer the possibility of doing an analysis that accommodates the cross-classified structure of review occasion within reviewer crossed with patient that allows further partitioning of the variance components into reviewer, patient, and review occasion within patient. But as this further partition is not essential to our analysis and the low number and proportion of duplicate reviews were small (800 of 4,800), cross-classification is ignored for the primary analysis. A secondary cross-classified analysis was undertaken for one of the primary models although the sparsity of cross-classified data and

the extent of the cross-classification (i.e. n>2 replicate reviews for a single case-note) makes these models difficult to estimate.(Bion et al., 2017)

The data was analysed in STATA 16.0. First, as the scales were originally designed, we created the scale scores as sums of the items and present means and standard deviations. We then standardized the scores and assessed their correlations. Second, we estimated the intra-class correlation (ICC) of reviews within reviewer, a measure of the reviewer contribution to the outcome variable (care quality judgement). Third, we estimated the relationship of the primary variables of interest with the outcome variables. Finally, we examined the demographic and patient variables to see if they predicted either of the scores. We used maximum likelihood tests to compare models. As a sensitivity analysis, we ran a cross-classified model with an empty model to see if the total variance attributable to the reviewer differed from that estimated ignoring the cross-classified structure (i.e., simple multi-level model).

3.6. Results

3.6.1. Survey response and reviewer characteristics

Of 79 HiSLAC reviewers, 72 responded to all items of the three attitudinal scales for a response rate of 91%. The patient and reviewer characteristics are tabulated in Appendix 11 along with the mean quality ratings and the reviewer's confidence in those judgments. The baseline characteristics of case-notes and reviewer are provided in Appendix 12. The description of the level-1 and level-2 predictors and outcome variables are given in Appendix 13.

3.6.2. Scales measuring domains relevant to case note quality scores

The average scale scores and standard deviations for the three scales are shown in Table 14.

Table 14. Scale scores (unstandardized)

	Average reviewer scale score (range)	Standard deviation
NFC	21.7 (12-28)	3.2
PNS	40.0 (20-69)	10.1
ADTU	19.7 (7-30)	5.5

We examined correlations of the standardized scales to look for evidence of any strongly overlapping constructs. We found relatively low correlations between most scale items with the highest correlation estimated of 0.34 between PNS and ADTU (Figure 8). The coefficients changes controlling for each attitudinal measure for both care quality judgements and confidence in care quality judgement are presented in Appendix 14.

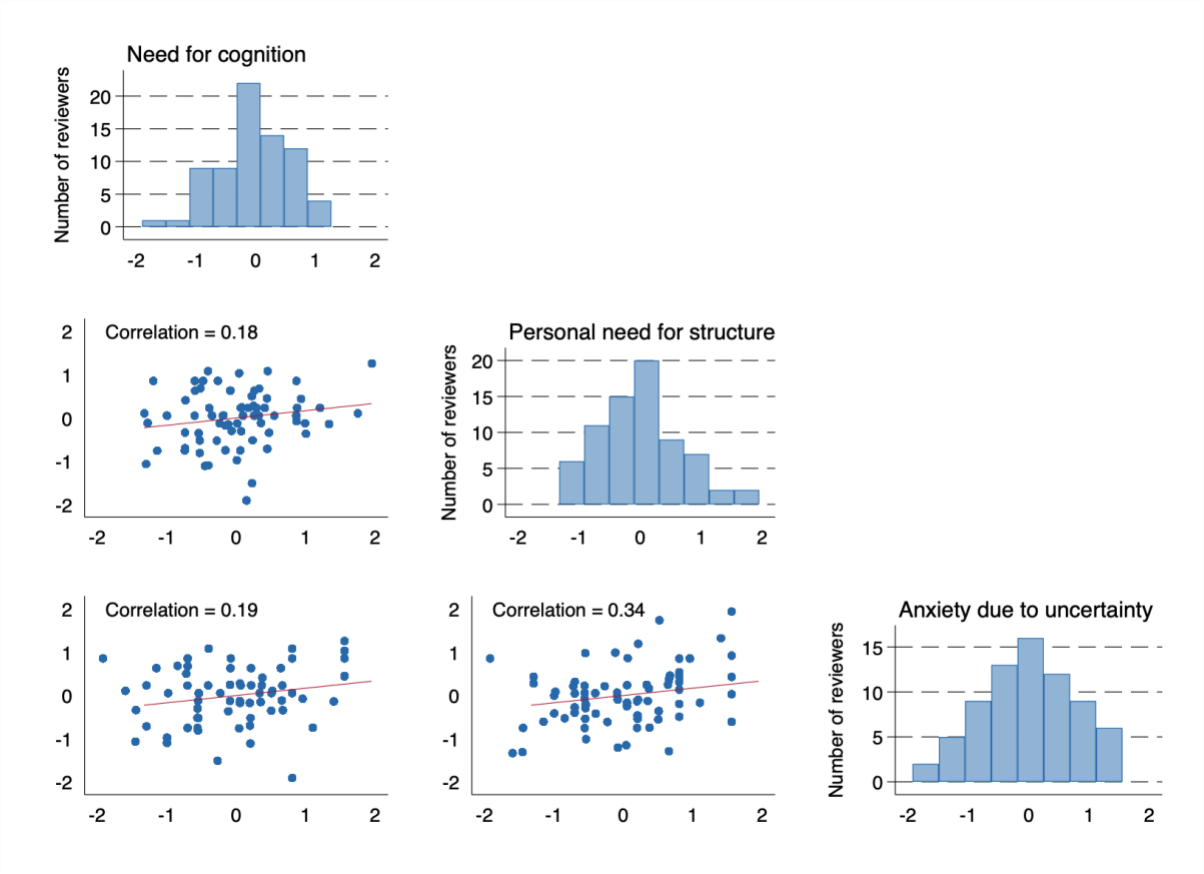


Figure 8. Scale score distribution

We found low correlations (r) between the measures (r=0.18-0.34 range) with the highest between PNS and ADTU (r=0.34).

The scale internal consistencies were obtained from the survey data with coefficient alpha's (α) and were comparable to the published estimates. We obtained an α=0.71 for the NFC scale, comparing favourably to the published α for the full scale (α range=0.75-0.86); we used a sub-scale of the published scale which likely explains the lower consistency.(Reyna and Ortiz, 2016) Alpha was 0.89 for PNS scale, which is higher than the published α=0.77.(Neuberg and Newsom, 1993, Thompson et al., 2001, Selltiz et al., 1976) Finally α=0.89 for the ADTU scale, which is higher than the published α=0.79.(Gerrity et al., 1995)

3.6.3. Models predicting care quality score

The mean care quality judgement was 4.17 (s.d. 0.82) (Table 15) indicating a relatively high average quality score. Of the total variance in the care quality score, 20% (CI 15-27%) was attributable to the reviewer, also representing the ICC for case-note reviews within reviewer. The remaining variance is due to differences in the true quality of care received by the patient and case-note review occasion-to-occasion variation.

Table 15. Reviewer level fixed effects in null model

Reviewer mean care quality judgement (i)			Reviewer mean confidence in care quality judgement (ii)		
CQJ	SD	Intra class correlation	CCQJ	SD	Intra class correlation
4.17	0.82	0.205 [15.3-26.6]	82.6	18.6	0.19[0.14-0.26]

Figure 9 (left column) shows the relationship of the quality score with each of the three scales, averaging over all reviewer effects other than the scale score, with a quadratic term to allow some flexibility in the modelled relationship. All three appear to have a weak relationship with the quality-of-care score with little change in quality rating across the scale. Tested with a linear predictor, all three have a non-significant negative slope of -0.10 (on the 1-5 scale of quality) or less per standard deviation in the attitude score.

Figure 9. Estimations of the relationship of attitudinal scores to care quality judgement and confidence in care quality judgement

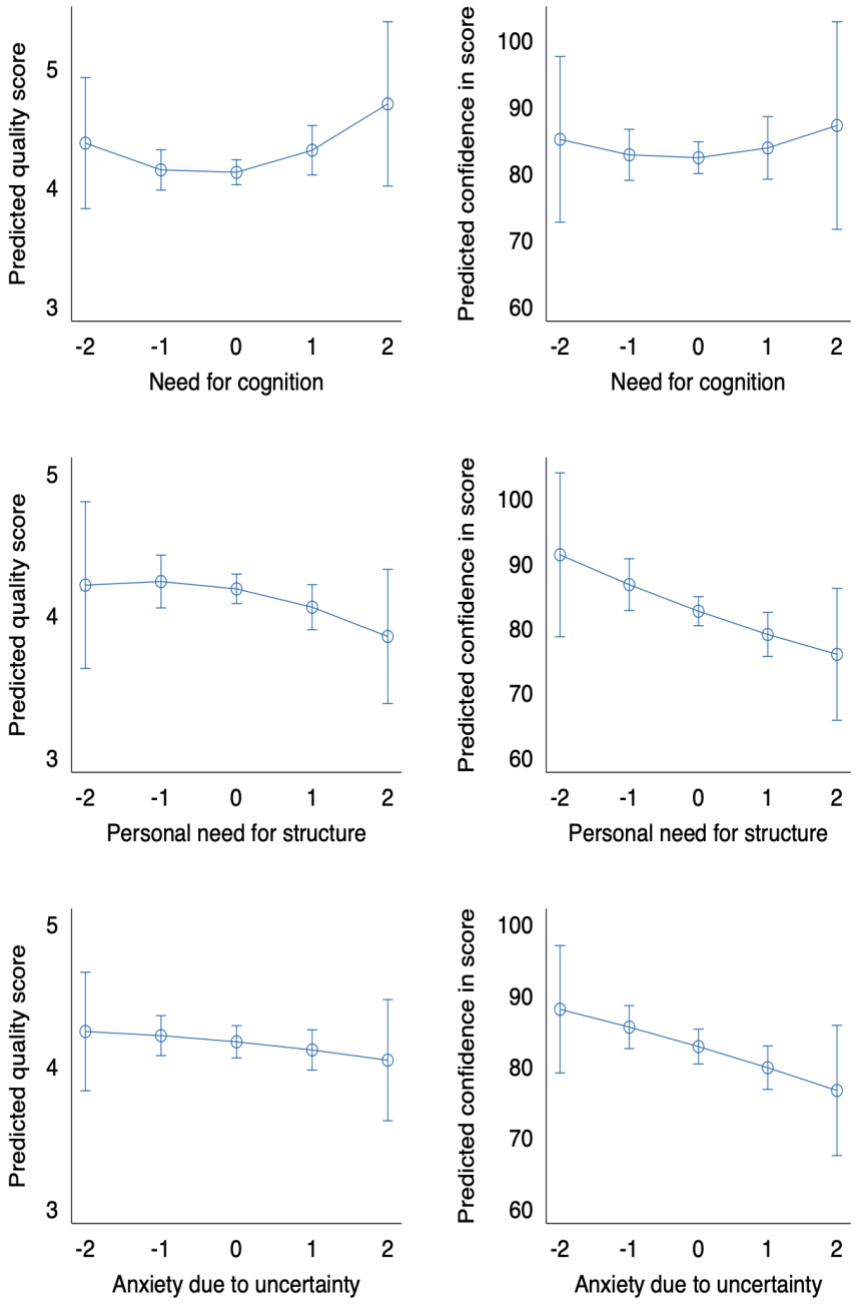


Figure 9 (left column) shows the relationship of the quality score with each of the three scales with a quadratic term to allow some flexibility in the modelled relationship. Tested with a

linear predictor, all three have a non-significant slope of 0.10 per standard deviation in the attitude score or less. The right columns show the relationships of the three attitudes to the confidence in quality judgment with a significant downward trend for both PNS and ADTU.

3.6.4. Models predicting confidence in the care quality score

The mean confidence in care quality judgement was 82.6 (s.e. 18.6) on a 0-100 scale indicating a moderately high mean confidence in care quality judgement. Like the quality score, 19% [14-26%] of the variance in the confidence score was attributable the reviewer. The relationships of the three attitudes to the confidence in quality judgment are shown in the right column of Figure 9. While there is no clear relationship of the confidence score with the NFC scale the PNS and ADTU scales were inversely related to the quality score with the linear predictor having slopes of -3.8 (s.e. 1.4) and -2.8 (s.e. 1.2) points per standard deviation of the scale score.

3.6.5. Other reviewer and patient predictors tested

Quality scores did not differ by patient gender but did decrease a small amount for patients who had a length of stay over seven days (-0.09[-0.15, -0.04]). There was lower quality reported for patients requiring more assistance prior to admission (-0.10[-0.16, -0.04] and -0.12[-0.20,-0.05]) for requiring help with some or most activities relative to patients who were identified as independent. The quality score was substantially lower (-0.52[-0.60,-0.44]) for those patients for whom it was not possible to determine their need for help prior to admission (see Appendix 14).

Reviewer gender did not predict quality scores, nor was their evidence that the amount of variation attributable to the reviewer differed by gender. Similarly, the number of years since graduation had little relationship to the care quality score. The reviewer variance component was the same magnitude (20% of the variance) in a model that accounted for the cross-classification of reviewers and patients as in our simpler model used in the above results (see Appendix 15).

A cross-classified model was run to test for the presence of a cross-classified effects of when duplicate reviews are shared between reviewers, rather than there being an exclusive nesting of case-notes within reviewers (Appendix 15).

3.7. Discussion

Overall, we found that in the HISLAC study about 20% of the variance in quality scores was attributable to the reviewer. This figure is in agreement with the 15-20% of variance reported in the few studies that have measured reviewer variation contributions.(Manaseki-Holland et al., 2016, Manaseki-Holland et al., 2019b) The general implication of this is clear; if a care quality judgement of 3 is given on a 6 points Likert scale, then it is quite possible that the rating of 3 could in fact be either 2 or 4. This is because a fifth of the variability in the score is attributable to the reviewer and make contribute in a one-sided manner to raise the score or lower score. This has implications on hospitals; firstly, reviewing case-notes are mandated in the UK and other high-income countries but not all case-notes are reviewed, but almost all are screened before reviewing. This screening process is legitimately influenced by this 20% variance from reviewers and could be the difference between getting included for case-note review or not at all. In effect, it applies to all levels of

any clinician-led judgement process. Until it is known exactly which way (increase or decrease the scale score), then it is likely to have a significant role for clinicians, patients, and healthcare policymakers. Secondly, there is the implication of possibly misattributing, on a 5+ scale, by a single point. Practically, this could be the difference between being escalated and handled more seriously or put up for discussion to take notes on for the future if it can swing between a 2-point score or 4-point score. These are simple illustrations to express the possible significance of this reviewer variance. The precise implications of this 20% variance contribution from reviewers to the care quality score is shrouded in a little mystery as we cannot pinpoint the exact source, that is the single culprit variable or factor, responsible. Our method does not allow it because the contributions from all the independent variables were, overall, slight, and non-significant so the source cannot be known. However, the contribution is not negligible and warrants further empirical study with the reviewer-specific variables employed in our models.

This 20% figure does not compare well to other hospital care quality measures favourably because it is the first time it has been applied to case-note reviewer care quality judgements. This is because most other hospital measures are not derived directly from an individual, but from an algorithm or method for collecting or processing data. However, it may be worth considering that, there have been instances where the hospital metrics such as hospital mortality statistics and other various metrics of care quality (i.e., complication measures, mortality measures, readmission measures and local patient safety indicators (PSI)). In a study bringing these four metrics together, the investigators conducted within-group correlations and Cronbach α found only strong correlations of a common construct for mortality and readmission, and not between any of the others. This indicates that these four metrics may well be measuring different constructs to each other. Such a study has not been done to compare the correlational relationship between these metrics and care quality

judgements; the reason being that care quality judgements cannot be broken down in component parts but stands alone in its assessment.(Hu et al., 2017) There have been no other studies which have investigated the within-reviewer variance of case-note reviewers for care quality judgements (nor their confidence in their care quality judgements).

While the level of reviewer variance may seem modest it is important to note that it is of a similar magnitude to the amount of signal that we are looking for with case-note review, that is the variance in true patient differences in quality of care received. The proportion of variation due to patient differences in quality found in prior studies ranges most commonly from 0.2 to 0.4.(Hofer et al., 2000) The remaining variance (about 80%) is challenging to try to reduce being composed of idiosyncratic interactions between a "specific reviewer and case-note" as well as "occasion-to-occasion" variation within patient and reviewer. Thus, removing systematic reviewer variation may be one of the easiest ways to increase the signal-to-noise ratio for detecting the true quality of care received by a patient.

However, contrary to our expectations, the care quality judgements did not differ by the level of our three attitudinal measures. The trends were in the anticipated directions for the most part, with higher scale scores for PNS and ADTU being associated with lower quality ratings, but the magnitude of the effects would be considered small even at the upper limits of the confidence intervals for the estimated slopes. These slopes were unchanged after adjusting for other patient or reviewer characteristics. The scores measuring the reviewers' confidence in their ratings vary more substantively, with higher levels on the PNS and ADTU scales each significantly associated with lower confidence score. However, these effects still represent less than 0.25 standard deviations of the outcome variable usually considered as a relatively small effect size.(Cohen, 2013) The relationship of the attitudes with the confidence in

quality score was also unchanged by adjusting for our other patient and reviewer characteristics.

3.7.1. The possibility of reducing care quality judgement disagreement and the application of these findings

Reducing the reviewer variance could improve reliability of case-note reviews by about 20% by removing one of the most obvious components of extraneous variance in detecting differences in patient quality of care.[33] The two possible ways of reducing reviewer variance are first through selection or training and second through calibration. The former requires understanding the reviewer and environmental factors producing reviewer judgement variation and based on our results, this seems less promising as still we have no clear idea of the sources of reviewer variation.

A cadre of reviewer scores can be calibrated to remove the reviewer variance if we establish an ongoing process to monitor the average care quality judgement score of each reviewer as compared to the average for the population of reviewers and subtract the population mean deviation from all their scores. This has previously been done in the analysis of only a few research studies that use case-note review(Smith et al., 1997a, Hofer et al., 2004, Manaseki-

[33] Assuming patient variance of 0.2-0.4, reviewer variance of 0.2 and a residual variation, then the reliability =0.3/(0.3+0.2+0.5) = 0.3, which represents the variance components of reviewer, patient and general noise using the equation: $\sigma^2_{patient}/(\sigma^2_{patient}+\sigma^2_{reviewer}+\sigma^2_{review\text{-}occasion})$. If you remove the reviewer component (0.2) then the reliablity will increase by 25%, as the reviewer-excluded reliability = 0.3/(0.3+0.5) = 0.375] using this equation: $\sigma^2_{patient}/(\sigma^2_{patient}+\sigma^2_{review\text{-}occasion})$. Subsequently, 0.375/0.3 = 1.25, which equates to 25% higher reliability with the reviewer component accounted for.

Holland et al., 2016). Operationally this can be done if an adequately sized subset of case-notes is assigned to multiple reviewers in a random fashion and the analysis is designed to account for the replicate reviews. Alternatively, systematic differences across case-note reviewers can be periodically estimated using a fixed set of 20-35 case-note "calibration" reviews interspersed in the reviewer's normal caseload.

3.7.1.1 Identifying further attitudinal measures

It is unclear how we are going to proceed selecting future attitudinal measures to determine their capacity to explain some of the reviewer contributions to the total care quality judgement variance given that none of our variables explained much of this variance. The first option would be to continue through the list of measures shortlisted, and test these when such a study with a large cadre of case-note reviewers becomes available in the future. There must be factors which explained this unexplained variance, and it most likely resides within the psychology of individuals, and thus attitudinal measures and psychological measures would be one of factors speculated as one its sources.

3.8. Limitations and Methodological Reflections

3.8.1. Limitations

We must assume that the predictor variable measures remain unchanged in the intervening period between the conclusion of the reviewing process and the completion of the survey. The time elapsed from survey invite to last review was five calendar months, however, some reviewers did complete all their case-notes before some had even commenced their reviews. Due to the small intervening period and because attitudinal measures and demographic measures are persistent, these constructs were not expected to vary over this period.

Furthermore, the PNS (Neuberg and Newsom, 1993, Hess, 2001), NFC (Cacioppo et al., 1996) and ADTU (Gerrity et al., 1995, Carney et al., 2004) are widely considered as stable personality traits.

Other than ADTU, the attitude scales were not originally designed to measure factors that affect expert assessments of clinical care. However, as described above, they were selected for their relevance to two dimensions, risk, and orderliness, thought to be major potential factors in reviewing case-notes for quality of care and to build on existing theory that applies to decision-making more broadly. Cognitive biases, which are systematic deviations from rational norms, and heuristics, mental shortcuts aiding cognition, have been known to influence medical decision-making.(Blumenthal-Barby and Krieger, 2015a, Saposnik et al., 2016) Thus, we had suspected that they might explain some of our unexplained reviewer variance (see Chapter 2). However, the selection of a biases or heuristic is unclear as there are no studies or systematic discussions available concerning their plausible influence upon clinician peer review of patient records. Further research is needed to highlight the key sources contributing to any reviewer variance in care quality score. This will also help develop robust study designs suited for these investigations.

As the findings from the study do not shed light on the factors driving this 20% variance within reviewers, it would be worth exploring alternative methods to explore this hidden variance. One option is to employ semi-structured qualitative interviews from a wide-audience including case-note reviewers, non-reviewers, board members and former patients to establish where their sentiments are on case-note reviews and what factors they feel affect both its fidelity and reliability as a care quality instrument. Through this, we find other reviewer variables which have not been considered in this study. Thus, it could prove

insightful to run the qualitative interviews and have them inform any future variable selection process. These selected features can then inform further multi-level models which have care quality judgements as the dependent variable.(Britten, 1995)

An alternative approach for characterising and measuring psychological constructs called the theory of conjoint measurement is concerned with the individual characterisation of psychological phenomena captured by individual psychological instruments.(Krantz and Tversky, 1971) It is found to be superior to the current standards of psychological measurement but has the weakness of being mathematical derived and heavy for the those unaccustomed to its use of symbolic representation. The theory of conjoint measurement is much less susceptible to random noise because of the highly specified nature of the hypotheses and the precise characterisation of the factors (or constructs) demand that the methodological rigour is precise and exacting to reveal any quantitative relationship with care quality judgements.(Michell, 1999) This method was found to be viable after the study has been completed, nevertheless, there are opportunities to develop programmes under this approach.

Our approach to selecting variables that might explain reviewer variation may have overestimated the degree to which reviewer judgements are due to internal, dispositional factors rather than situational factors (Jones and Harris, 1967). Situational factors have been demonstrated to influence decisions (Ross and Nisbett, 2011) as well as human factors.(Croskerry and Chisholm, 2005, Smith et al., 2008, Leonard et al., 2004) Thus further studies might identify contextual factors (i.e. hospital culture, incentivisation, mandatory time allocations), physical reviewing conditions (i.e. lighting, noise levels), or number of cases

reviewed per session that could explain some of the reviewer-specific variation contribution to the total care quality judgement variation.

One type of variation not investigated is the actual content of the case-note itself. The effect of the linguistic and textual (and handwritten) format of the case-notes on reviewer judgements was not studied due to the lack of this type of digital infrastructure and knowledge around this format. It is perfectly plausibly for this to be a source of variability in this study.

3.8.2. Methodological Reflections: Part I – The strengths of the study methodology

First, the internal validity was preserved by having the HiSLAC case-note reviewers complete more or less immediately after the completion of the HiSLAC study to ensure fidelity free from any temporal element. Collecting the attitudinal measure data occurred some 6 months after the case-note reviewers had completed their study. One limitation from this could have been that their attitudinal measures could have changed over this elapsed period, either through the study itself, or some non-HiSLAC related stimulus, exposure, or belief. However, evidence suggests that attitudinal measures are persistent over time, especially for those with firmly established databases (i.e., their beliefs and feelings regarding their experiences).(Erber et al., 1995) Unfortunately, the demographic data collected in this study were partially complete which precluded the testing of certain hypotheses. Losing 7 of a total 79 possible reviewers as non-responders in our attitudinal survey posed no significant problem for the generalisability of our findings. In sum, there are unlikely to be issues with internal validity.

Second, the external validity of the study applies to all such case-note reviewers as the HiSLAC reviewing cohort did not markedly differ in their experience, specialty from any other case-note reviewers.[34]

Third, the reliability of the study was directly attributable to the co-ordination and support of the HiSLAC investigators that ensured the case-notes were redacted of any identifiable information, cogent, legible, of were recruited and trained to undertake case-note review during the study conditions. Thus, the reviews can be assessed to have been rigorously and reliably conducted. The case-note reviews were completed by reviewers who understood the significance of the attitudinal measures; to ensure this, the attitudinal instruments were piloted with a cadre of case-note reviewers from the University Hospitals Birmingham to sense-check the measures.

Fourth, the objectivity of this study was maintained as the attitudinal surveys and hypotheses were pre-specified before receiving any of the HiSLAC data; objectivity was also preserved using validated psychological instruments scales and guidelines for its recommended use to ensure the correct identification of psychological constructs. Also, the difference-by-

[34] We do note that their familiarity with hospital systems, cultures and clinical styles may differ in other settings. However, I feel that the case-note reviewing process is near immune to any influence from such factors given the condensed, focused, desk-bound nature of the activity.

differences[35], and stratification variables ensured that temporal changes had no significant bearing upon the study(Bion et al., 2017); the robustness of which was tested and confirmed using the cross-classified multi-level, that corroborated with the simple multi-level model results. In summary, the HiSLAC study design was robust.

3.8.3. Methodological Reflections: Part II - the Quantification of Psychology Phenomena

The theory of conjoint measurement is a methodology reflecting on the quantitative measurement of psychological phenomena. It is a plausible model with which we can use to appreciate the deeper assumptions of psychology and those who practise it. This theory is considered because cognitive psychologists should be aware of the core assumptions made in their research.

The dual process theory assumes cognitive mechanisms occur by one process, then these mechanisms in reverse or specific mediation of these active mechanisms can improve the quality of cognitive processes. In other words, the dual process theory assumes the possibility of the mitigation of these biases and heuristics is possible. Thus, the systematic review assumed that the mitigation of these cognitive biases and heuristics was possible. But I offer a critique of this approach. However, differential psychology studies the psychological processes behind behaviour and the nature of psychological attributes. This will shape how cognitive biases and heuristics can be mitigated. Mainstream psychological schools tend to

[35] Which compares the average change over time in the outcome variable (i.e., specialist staffing levels on weekends) for before the HiSLAC intervention compared to the average change over time for after-HiSLAC intervention group.

commit to the assumptions that their phenomena of study are quantitative in nature (i.e. differential psychology and experimental psychology). I elaborate further on this by discussing the implications of "*the theory of conjoint measurement.*"

Quantitative psychology attempts to measure distinct psychological quantities. However, how do we know these are quantitative entities? When is it justified to treat a psychological construct as a quantity? I use the theory of conjoint measurement to reveal the assumptions of the measurement of attributes in psychology. This section may appear technical, but it conveys an important assumption within psychology, which they themselves are not aware of.

3.8.4. The assumption of quantification

A survey of psychological publications reveals that psychologists believe they can measure psychological attributes, such as cognitive abilities, personality traits, social attitudes, and sensory intensities. These attributes are psychological as they form part of psychology's subject matter and are not normally measurable using the methods of the physical sciences (see, for example, Jerrard & McNeill, 1992; Sena, 1972).(Jerrard and McNeill, 1992, Sena, 1972) While these psychological attributes do not form part of the physical sciences, it is evident that quantitative psychology was initially modelled on quantitative physics (Fechner, 1860). In both disciplines, attributes are assumed to have a quantitative structure. But what then is required for an entity to be a quantity (or have a quantitative component)? The following section will discuss just this. I outline what is required for a continuous quantity.

3.8.5. How do psychologists define measurement?

Even though quantitative psychologists hypothesize that the attributes they use are quantitative and, so, commit themselves to the concept of scientific measurement, the definition of measurement endorsed by most of them is typically different. This definition is the one formulated by Stanley Stevens (1946)(Stevens, 1946):

measurement is the assignment of numerals to objects or events according to rule.(Stevens, 1946)

3.8.6. Should we be concerned with this definition of measurement in psychology?

Stevens' definition of measurement entered quantitative psychology at a particularly crucial stage of its history. The definition was accepted by psychologists out of ignorance of the truth about quantity and measurement, as outlined in 6.2.2.2: it was accepted and became entrenched because it appeared to deal with a conceptual problem which had existed since Fechner's time and which in the 1940s was particularly pressing. The acceptance of this definition involved a quite deliberate turning away from traditional concepts and it resulted in a sustained blind spot, one which has persisted to the present day. These are claims that I will support shortly by considering how psychological history has developed in time.

If a quantitative scientist (1) believes that measurement consists entirely in making numerical assignments to things according to some rule and (2) ignores the fact that the measurability of an attribute presumes the contingent hypothesis that the relevant attribute possesses an additive structure, then that scientist would be predisposed to believe that the invention of appropriate numerical assignment procedures alone produces scientific measurement. This is exactly the situation that exists in quantitative psychology, a situation that Stevens' definition serves to justify. These two facts, the widespread acceptance of Stevens' definition of

measurement amongst psychologists and the failure of books on psychological measurement to discuss the character of quantitative attributes have meant that the true nature of scientific measurement and the empirical content of the hypothesis that an attribute is quantitative are almost universally overlooked within psychology. For instance, a psychologist could measure three constructs, however, but not supply how precisely these three relate to one another and the quantity (Q) of interest. Positing that one construct influences Q one way and another construct in another way is not enough. If the constructs measure the same Q, then they must have an additive relationship; this is neglected in most psychological research. Next, I survey the some of the historical developments in psychology which has led to this neglect of additivity for quantitative measurement.

3.8.7. The theory of conjoint measurement

Prior to the discovery of conjoint measurement theory, it was not clear how derived measurement worked. To illustrate this point, let us consider the case of the property, density. Physicists knew that the density of something was the ratio of its mass to its volume, but what was not clear was the kinds of observations sustaining such a relationship. Campbell (1920) took it to be so that because the ratio of mass to volume is a constant for each different kind of stuff, the constant identifies a quantitative attribute. The theory of conjoint measurement explains why this is so. It is because density and volume trade off against one another relative to mass. For example, if a brick of pure gold weighs the same as a block of pure aluminium, then relative to effects upon mass, the increase in volume between the aluminium block (VA) and the gold brick (VG) equals the increase in density between the gold brick (DG) and the aluminium block (DA). Identifying equal ratios directly via such trade-offs means that VA/VG = DG/DA. If density is a quantitative attribute, then double cancellation must obtain

for sets of such volumes of densities. This means that the known relationship between density, mass and volume is not an arbitrary stipulation, but is a testable, scientific hypothesis. This same logic applies to all instances of derived measurement in physics or any other strictly quantitative physical science.

Extensive measurement relies upon locating a concatenation operation upon the relevant effects of which depend almost entirely upon a single attribute. For example, in the length measurement of rigid straight rods, the operation of joining two rods end-to-end, linearly, is one where an outcome depends pretty much only upon the lengths of the rods used. Our capacity to locate concatenation operations suitable for extensive measurement of the attributes that interest us depends upon the existence of a special class of causal relations, as well as upon our sensory-motor capacities and how we, as observers, relate to these attributes. Thus, this concatenation operation must extend to psychological constructs if we desire to treat them as quantitative variables.

Let me demonstrate what I mean. Suppose persons Paul (P) and Valerie (V) perform at exactly the same level on the test even though they differ from one another in motivation and ability. P does as well as V because of a higher level of motivation and L compensates for P's higher level of motivation by possessing a higher level of ability. That is, P's level of motivation (MP) minus L's level of motivation (MV) equals L's level of ability (AV) minus P's level of ability (AP) in terms of effects upon performance. Putting it in quantitative terms, relative to performance, a difference in motivation equals a difference in ability, i.e., MP − MV = AV − AP. The basic idea is that levels within either of the two attributes (motivation and ability) can be 'traded off,' as it were, against one another, relative to effects upon a third attribute (performance). If this is possible for one pair of differences, then it is possible for

adjoining differences. Additivity between differences can be indirectly identified via adjoining component differences. This prediction provides a specific test of the hypothesis that the attributes are quantitative: if they are so, then this prediction follows; if not so, then it does not, and the attributes are not physically relatable.[36]

Another example can be demonstrated with verbal ability, which is the ability to do well in verbal exercises. Invariably, the best we can do in science is to identify something via its effects, but this never justifies defining it as one disposition to produce those effects, for some unjustified reason assuming it has no intrinsic character, and only effects.(O'Neil, 1944) A necessary step in applying conjoint measurement theory to Spearman's theory on verbal ability is that of hypothesising more about abilities than just their likely effects upon test performance. Only when theorists in this area are prepared to hypothesise about the intrinsic character of abilities and the nature of the psychological construct can the issue of whether abilities (or any other psychological variable) can be quantitatively investigated.

The theory of conjoint measurement is only concerned with the individual characterisation of psychological phenomena captured by individual psychological instruments. It can be argued that it is not necessary to have a quantitative measure of psychological phenomena to understand and satisfactorily reduce the unexplained variability between case-note reviewer care quality judgements. To the contrary, I argue that to make proper use of the psychological instruments attempting to explain this variability, we must understand the true nature and

[36] Realistically, one would reflect on the model, the rigor and the assumptions made if there is no clear trade-off between the attributes.

relationship these psychological attributes have with care quality judgements. For instance, a quantitative paradigm will be able to determine an accurate contribution of a range of true psychological instruments. By understanding their specific contributions to the care quality scores, these true quantitative psychological instruments could predict the influence of psychological constructs upon case-note reviewer care quality judgement. Of course, this proposal presumes that the psychological constructs have been quantitively derived according to the above definition of a continuous quantity. It is of little use to spuriously indicate the effects of a construct and not the process; without the causal process and knowledge of the additive relationship, very little can be practically done. The unfolding of the process gives us understanding about the process, otherwise, we would not know how to effectively implement changes for the better.

Given the novelty and scant research exploring the variability sources for the care quality judgement, psychologists and healthcare researchers and clinicians would do well to heed this discussion and continue developing an appropriate methodological framework to quantitively measure, where appropriate, the contributions of psychological attributes to this care quality score.

3.8.8. What is the theory's relevance to this thesis?

If there are any quantitative influencers upon the care quality judgement, it is important to discover the components which comprise this care quality judgement. Once this is done, it would be helpful to characterise whether there is a quantitative character to the factors studied (i.e., attitudinal measures, cognitive bias, or heuristic) with the care quality judgement score of case-note reviewers. Then, and only then, will there be a veritable demonstration of

the influence of these factors upon the care quality judgements. The theory of conjoint measurement is much less susceptible to random noise because of the highly specified nature of the hypotheses and the precise characterisation of the factors (or constructs) demand that the methodological rigour is precise and exacting to reveal any quantitative relationship with care quality judgements. It would be compelling to consider the viability of this approach given that medical decision-making and care quality judgements could be classified as naturalistic, as it often incorporates complex, contextual factors and assumptions in any evaluation.

In the instance when a psychological approach by individual differences is not conducive to explaining the reviewer-sourced variation to quality scores, a different approach is required. A psychological approach may not faithfully report on the cause of this inter-reviewer variation because psychological constructs are often vaguely-defined and are themselves a fluctuating mass of attributes.(Michell, 2019) This is because psychological measures are "*premised upon the proposition that psychological attributes are quantitative but are devoid of serious attempts to consider relevant evidence for that premise*." (Michell, 2008) The psychological constructs are validated by their own self-fulfilling standards. This limitation needs to be borne in mind when interpreting the implications for the significance of psychological testing and their results. However, in practise, this has involved giving the reviewers the same test items which has failed to consider whether the care quality standard the reviewers have employed in their minds was indeed different or not. Standardised instructions can be used to constrain divergences, but the effects remain unpredictable. Any precision, thus, is called into question. To explore the significance of the reviewer's concept of "care quality", we envisage that '*properly basic beliefs'* has the potential to elicit these

further.[37] Furthermore, individual difference factors may have an influence upon the reviewer-sourced variation upon their care quality judgements (Kelly, 2005, Plantinga, 1993) but this has yet to be formally elucidated experimentally.

3.8.9. The contributions of the psychological approach

There are a few advantages of using a psychological approach to understand case-note reviewer care quality judgements. A list of the benefits with its corresponding result are given below (Table 16):

Table 16. Advantages of attitudinal research and examples of this from the literature

Advantages of psychological research	**A concrete example sourced from the literature**
Can help find correlations between attitudes and helpful or harmful behaviours across special settings. This helps to give potential predictive power for the effects of an attitude on defined behaviours.	Through understanding the tendency for certain individuals and the behaviour it is correlated with. For instance, it was shown that attitudinal factors, such as institutional trust and its usage, determine the use of electronic health care records systems.(Ortega Egea and Román González, 2011)
The parties influenced by attitudes may be unaware of the prevalent effect on their behaviours. Thus, this	In understanding the influence of certain attitudes on different types of people, the greater awareness of a group proclivity such as when attitudinal

[37] A properly basic belief if a belief which can be justifiably held without the justification of other beliefs.

research has the potential to shed more light on their issue and bring more self-awareness to the problem in its context.	factors help explain the reticence (and willingness) to report workplace internet violations. This may appear unimportant, but if certain attitudes shape whether people report violations, then there is great benefit in understanding this further.(Campbell et al., 2016)

3.9. The utility of data science for studying variability in judgements

There is a case for employing data science to explore the variability around these care quality judgements. Variability has two components. Firstly, there is the need to make sense of this variability, and secondly, the other component that is simply unexplained by any applicable method. Given today's glut of information in the form of case-note data, there need to be modern and robust methods for understanding this variability. And data science may fit the need as an approach to understand just that.

Data science is an approach to extract actionable insights from a large and ever-increasing volume of data collected and created in today's modern information age. When we consider the case-note as a data aggregated into a physical form, then this data can be extracted to identify important variables (typically called feature engineering) that can be studied systematically using the methods of data science. Methods that interrogate the text of the case-notes, from the data that is available from it, may reveal subtle and hidden relationships of the case-note themselves and how the case-note reviewer interacts with the case-note also.

The assumption made in this chapter (and chapter 2) assumed that the variability was sourced within people alone and that it was a characteristic of either the reviewer or the patient (case-note) that was responsible for the variability observed in the case-notes. Thus, it became the primary perspective of how I viewed the source of the reviewer care quality judgements disagreements and variability. However, the case-note itself must be thought of for a time. Assuming a reviewer or patient source neglects the fact that the case-note is intrinsically a text document and, which is predominantly in the form of text, may itself be a source of this variability. Are there un-identified textual features that can explain this variance? And to help healthcare organisations better apply learning from case-note reviews, are there features (in the case-note) that may be predictive of certain healthcare outcomes? Data science through narrative mining, which is the extraction of insights from textual data, may hold Features of these case-note may also influence the care quality judgements, but there has been no study into the stable features of a case-note that would influence a clinical judgement. Applying data science to case-notes is a virgin field but shows some promise that it is a matter that is being explored for predictive purposes. For instance, the study that mine clinical phrases from clinical notes to discover the risk factors for patient deterioration.(Korach et al., 2020) There is the untapped potential for the utilisation of data science methods and narrative mining with respect to identifying salient features in the case-notes.

3.10. Conclusions

Our multilevel model indicates that reviewers contribute about 20% to the total variation observed around the care quality judgements which is comparable to amount of patient variation in quality (estimated at 20-40% of the variation across studies). For health care quality measurement, the reviewer variance is ‘noise’, that is unexplained by factors included

in our model, but we did not find that reviewer attitudinal, demographic, or patient characteristics explained much of this variance. While future research may identify other attitudinal measures or cognitive biases, in the interim, reviewer calibration seems to be the most promising way to improve the reliability of expert clinical reviews of the quality of care.

3.10. Conflict of Interest Disclosures

No conflicts were reported.

3.11. Acknowledgements

We would like to thank Gavin Rudge, Jianxia Sun, Cassie Aldridge, Emma Lancashire, Chris Beet, Olivia Brookes and the HiSLAC (High-intensity Specialist-Led Acute Care) reviewers for their support and consent to be involved in using the study's data.
Ethical approval was granted through a HiSLAC amendment on 15 March 2019. The HiSLAC ethics amendment to the NHS HRA [Integrated Research Application System (IRAS) Project ID 139089 was sent on 11 February 2019]) The ethics concerning storage are subsumed under the existing HiSLAC study.

3.12. References

ACEVEDO, M. & KRUEGER, J. I. 2004. Two egocentric sources of the decision to vote: The voter's illusion and the belief in personal relevance. *Political Psychology,* 25**,** 115-134.

ALBARRACÍN, D. & WYER, R. S., JR. 2000. The cognitive impact of past behavior: influences on beliefs, attitudes, and future behavioral decisions. *Journal of Personality and Social Psychology,* 79**,** 5-22.

ARMSTRONG, D., LILFORD, R., OGDEN, J. & WESSELY, S. 2007. Health-related quality of life and the transformation of symptoms. *Sociology of Health and Illness,* 29**,** 570-583.

BAUMEISTER, R. F. & VOHS, K. D. 2007. *Encyclopedia of social psychology,* Sage.

BION, J., ALDRIDGE, C., GIRLING, A. J., RUDGE, G., SUN, J., TARRANT, C., SUTTON, E., WILLARS, J., BEET, C. & BOYAL, A. 2020. Changes in weekend and weekday care quality of emergency medical admissions to 20 hospitals in England during implementation of the 7-day services national health policy. *BMJ Quality & Safety*.

BION, J., ALDRIDGE, C. P., GIRLING, A., RUDGE, G., BEET, C., EVANS, T., TEMPLE, R. M., ROSEVEARE, C., CLANCY, M. & BOYAL, A. 2017. Two-epoch cross-sectional case record review protocol comparing quality of care of hospital emergency admissions at weekends versus weekdays. *Bmj open,* 7**,** e018747.

BJÖRKLUND, F. & BÄCKSTRÖM, M. 2008. Individual differences in processing styles: validity of the Rational–Experiential Inventory. *Scandinavian journal of psychology,* 49**,** 439-446.

BLAND, J. M. & ALTMAN, D. G. 1997. Statistics notes: Cronbach's alpha. *BMJ,* 314**,** 572.

BLUMENTHAL-BARBY, J. & KRIEGER, H. 2014. Cognitive biases and heuristics in medical decision making a critical review using a systematic search strategy. *Medical Decision Making***,** 0272989X14547740.

BLUMENTHAL-BARBY, J. S. & KRIEGER, H. 2015. Cognitive biases and heuristics in medical decision making: A critical review using a systematic search strategy. *Medical Decision Making,* 35.

BMJ BEST PRACTISE. 2002. *Understanding risk* [Online]. Available: https://bestpractice.bmj.com/info/toolkit/practise-ebm/understanding-risk/ [Accessed 19/02/2021].

BRITTEN, N. 1995. Qualitative research: qualitative interviews in medical research. *Bmj,* 311**,** 251-253.

BUI, T. C., KRIEGER, H. A. & BLUMENTHAL-BARBY, J. S. 2015. Framing effects on physicians' judgment and decision making. *Psychological Reports,* 117**,** 508-522.

BURNS, R. B., FREUND, K. M., MOSKOWITZ, M. A., KASTEN, L., FELDMAN, H. & MCKINLAY, J. B. 1997. Physician characteristics: Do they influence the evaluation and treatment of breast cancer in older women? *The American Journal of Medicine,* 103**,** 263-269.

BURRUS, J. & CARNEY, L. 2015. Social Beliefs and Attitudes: Personality and the Isms. *In:* WRIGHT, J. D. (ed.) *International Encyclopedia of the Social & Behavioral Sciences (Second Edition).* Oxford: Elsevier.

CACIOPPO, J. T., PETTY, R. E., FEINSTEIN, J. A. & JARVIS, W. B. G. 1996. Dispositional differences in cognitive motivation: The life and times of individuals varying in need for cognition. *Psychological Bulletin,* 119**,** 197.

CALDER, L. A., FORSTER, A. J., STIELL, I. G., CARR, L. K., BREHAUT, J. C., PERRY, J. J., VAILLANCOURT, C. & CROSKERRY, P. 2012. Experiential and rational decision making: a survey to determine how emergency physicians make clinical decisions. *Emergency Medicine Journal,* 29**,** 811-816.

CAMPBELL, M., STYLIANOU, A. C. & SHROPSHIRE, J. 2016. The impact of attitudinal factors on intention to report workplace Internet abuse. *Journal of Information Privacy and Security,* 12**,** 68-83.

CARNEY, P. A., ELMORE, J. G., ABRAHAM, L. A., GERRITY, M. S., HENDRICK, R. E., TAPLIN, S. H., BARLOW, W. E., CUTTER, G. R., POPLACK, S. P. & D'ORSI, C. J. 2004. Radiologist uncertainty and the interpretation of screening. *Medical Decision Making,* 24**,** 255-264.

CHEN, Y. F., BOYAL, A., SUTTON, E., ARMOIRY, X., WATSON, S., BION, J. & TARRANT, C. 2016. The magnitude and mechanisms of the weekend effect in hospital admissions: A protocol for a mixed methods review incorporating a systematic review and framework synthesis. *Systematic Reviews,* 5**,** 84.

CHER, D. J., MIYAMOTO, J. & LENERT, L. A. 1997. Incorporating risk attitude into markov-process decision models:: importance for individual decision making. *Medical Decision Making,* 17**,** 340-350.

COHEN, J. 2013. *Statistical power analysis for the behavioral sciences,* Academic press.

CROSKERRY, P. & CHISHOLM, C. What does human factors ergonomics need to know about front-line medicine? International Conference on Healthcare Systems Ergonomics and Patient Safety: Human Factor, a Bridge Between Care and Cure, HEPS 2005, 2005 Florence. 77-80.

DASGUPTA, N. 2013. Chapter Five - Implicit Attitudes and Beliefs Adapt to Situations: A Decade of Research on the Malleability of Implicit Prejudice, Stereotypes, and the Self-Concept. *In:* DEVINE, P. & PLANT, A. (eds.) *Advances in Experimental Social Psychology.* Academic Press.

DAVIS, D. P., JANDRISEVITS, M. D., ILES, S., WEBER, T. R. & GALLO, L. C. 2012. Demographic, Socioeconomic, and Psychological Factors Related to Medication Non-adherence Among Emergency Department Patients. *The Journal of Emergency Medicine,* 43**,** 773-785.

EISENBERG, J. M. 1979. Sociologic influences on decision-making by clinicians. *Annals of internal medicine,* 90**,** 957-964.

EISENBERG, J. M., KITZ, D. S. & WEBBER, R. A. 1983. Development of attitudes about sharing decision-making: a comparison of medical and surgical residents. *Journal of Health and Social Behavior,* 24**,** 85-90.

ERBER, M. W., HODGES, S. D. & WILSON, T. D. 1995. Attitude strength, attitude stability, and the effects of analyzing reasons. *Attitude strength: Antecedents and consequences***,** 433-454.

FINUCANE, M. L., MERTZ, C., SLOVIC, P. & SCHMIDT, E. S. 2005. Task complexity and older adults' decision-making competence. *Psychology and aging,* 20**,** 71.

FITZGERALD, C. & HURST, S. 2017. Implicit bias in healthcare professionals: a systematic review. *BMC medical ethics,* 18**,** 19.

FORREST, C. B., NUTTING, P. A., VON SCHRADER, S., ROHDE, C. & STARFIELD, B. 2006. Primary care physician specialty referral decision making: patient, physician, and health care system determinants. *Medical Decision Making,* 26**,** 76-85.

GERRITY, M. S., WHITE, K. P., DEVELLIS, R. F. & DITTUS, R. S. 1995. Physicians' reactions to uncertainty: refining the constructs and scales. *Motivation and Emotion,* 19**,** 175-191.

GILOVICH, T., GRIFFIN, D. & KAHNEMAN, D. 2002. *Heuristics and biases: The psychology of intuitive judgment,* Cambridge university press.

GOLDMAN, R. L. 1992. The reliability of peer assessments of quality of care. *Journal of American Medical Association,* 267**,** 958-960.

GOLDMAN, R. L. 1994. The reliability of peer assessments of quality of care. *Journal of American Medical Association,* 17**,** 3-21.

GREENGLASS, E., SCHWARZER, R., JAKUBIEC, S. D., FIKSENBAUM, L. & TAUBERT, S. 2014. The Proactive Coping Inventory (PCI): A Multidimensional Research Instrument. Paper Presented at the 20th International Cracow, Poland, 1999. *Conference of the STAR (Stress and Anxiety Research Society).*

GUPTA, M., SCHRIGER, D. L. & TABAS, J. A. 2011. The presence of outcome bias in emergency physician retrospective judgments of the quality of care. *Annals of Emergency Medicine,* 57**,** 323-328.e9.

HAJJAJ, F. M., SALEK, M. S., BASRA, M. K. A. & FINLAY, A. Y. 2010. Non-clinical influences on clinical decision-making: a major challenge to evidence-based practice. *Journal of the Royal Society of Medicine,* 103**,** 178-187.

HARRIS, P. A., TAYLOR, R., THIELKE, R., PAYNE, J., GONZALEZ, N. & CONDE, J. G. J. J. O. B. I. 2009. Research electronic data capture (REDCap)—a metadata-driven methodology and workflow process for providing translational research informatics support. 42**,** 377-381.

HARVEY, N. 1994. Relations between confidence and skilled performance.

HAYES, R. P., STOUDEMIRE, A. S., KINLAW, K., LYNN DELL, M. & LOOMIS, A. 1999. Changing attitudes about end-of-life decision making of medical students during third-year clinical clerkships. *Psychosomatics,* 40**,** 205-211.

HAYWARD, R. S. A., GUYATT, G. H., MOORE, K. A., MCKIBBON, A. & CARTER, A. O. 1997. Canadian physicians attitudes about and preferences regarding clinical practice guidelines. *Canadian Medical Association Journal,* 156**,** 1715.

HESS, T. M. 2001. Ageing-related influences on personal need for structure. *International Journal of Behavioral Development,* 25**,** 482-490.

HOFER, T. P., ASCH, S. M., HAYWARD, R. A., RUBENSTEIN, L. V., HOGAN, M. M., ADAMS, J. & KERR, E. A. 2004. Profiling quality of care: Is there a role for peer review? *BMC Health Services Research,* 4**,** 9.

HOFER, T. P., BERNSTEIN, S. J., DEMONNER, S. & HAYWARD, R. A. 2000. Discussion between reviewers does not improve reliability of peer review of hospital quality. *Medical Care,* 38**,** 152-161.

HORNE, J. A. & ÖSTBERG, O. 1976. A self-assessment questionnaire to determine morningness-eveningness in human circadian rhythms. *International journal of chronobiology*.

HU, J., JORDAN, J., RUBINFELD, I., SCHREIBER, M., WATERMAN, B. & NERENZ, D. 2017. Correlations among hospital quality measures: what "Hospital Compare" data tell us. *American Journal of Medical Quality,* 32**,** 605-610.

HUTCHINSON, A., COSTER, J., COOPER, K., MCINTOSH, A., WALTERS, S., BATH, P., PEARSON, M., RANTELL, K., CAMPBELL, M. & NICHOLL, J. 2010a. Assessing quality of care from hospital case notes: comparison of reliability of two methods. *Quality and Safety in Health Care***,** qshc. 2007.023911.

HUTCHINSON, A., COSTER J, E., COOPER K, L., MCINTOSH, A. & WALTERS S, J. 2010b. Comparison of case note review methods for evaluating quality and safety in health care. *Health Technology Assessment,* 14**,** 170.

JERRARD, H. & MCNEILL, D. 1992. Systems of units. *Dictionary of Scientific Units.* Springer.

JONES, E. E. & HARRIS, V. A. 1967. The attribution of attitudes. *Journal of Experimental Social Psychology,* 3**,** 1-24.

JONES, K., WRIGHT, C. & BELL, A. 2012. Do multilevel models ever give different results. *Or why bother with multilevel models*.

JUDA, M., VETTER, C. & ROENNEBERG, T. 2013. The Munich chronotype questionnaire for shift-workers (MCTQShift). *Journal of biological rhythms,* 28**,** 130-140.

JULIUSSON, E. Á., KARLSSON, N. & GÄRLING, T. 2005. Weighing the past and the future in decision making. *European Journal of Cognitive Psychology,* 17**,** 561-575.

KELLY, T. 2005. The epistemic significance of disagreement. *Oxford studies in epistemology,* 1.

KISH, L. 1965. Survey sampling.

KOCH, K.-R. 2013. *Parameter estimation and hypothesis testing in linear models*, Springer Science & Business Media.

KORACH, Z. T., YANG, J., ROSSETTI, S. C., CATO, K. D., KANG, M.-J., KNAPLUND, C., SCHNOCK, K. O., GARCIA, J. P., JIA, H. & SCHWARTZ, J. M. 2020. Mining clinical phrases from nursing notes to discover risk factors of patient deterioration. *International journal of medical informatics,* 135**,** 104053.

KRANTZ, D. H. & TVERSKY, A. 1971. Conjoint-measurement analysis of composition rules in psychology. *Psychological review,* 78**,** 151.

LEONARD, M., GRAHAM, S. & BONACUM, D. 2004. The human factor: the critical importance of effective teamwork and communication in providing safe care. *BMJ Quality & Safety,* 13**,** i85-i90.

LILFORD, R., EDWARDS, A., GIRLING, A., HOFER, T., DI TANNA, G. L., PETTY, J. & NICHOLL, J. 2007. Inter-rater reliability of case-note audit: A systematic review. *Journal of Health Services Research and Policy,* 12**,** 173-180.

LINS DE HOLANDA COELHO, G., H. P. HANEL, P. & J. WOLF, L. 2018. The Very Efficient Assessment of Need for Cognition: Developing a Six-Item Version. *Assessment***,** 1073191118793208.

MACKENZIE, E. J., STEINWACHS, D. M., BONE, L. R., FLOCCARE, D. J. & RAMZY, A. I. 1992. Inter-rater reliability of preventable death judgments. The Preventable Death Study Group. *The Journal of Trauma,* 33**,** 292-302, 302-3.

MANASEKI-HOLLAND, S., LILFORD, R. J., BISHOP, J. R., GIRLING, A. J., CHEN, Y.-F., CHILTON, P. J., HOFER, T. P., ALNER, V., BLEASDALE, J. & BOSANKO, N. 2016. Reviewing deaths in British and US hospitals: a study of two scales for assessing preventability. *BMJ Quality & Safety,* 26**,** 408-416.

MANASEKI-HOLLAND, S., LILFORD, R. J., TE, A. P., CHEN, Y.-F., GUPTA, K. K., CHILTON, P. J. & HOFER, T. P. 2019. Ranking hospitals based on preventable hospital death rates: A systematic review with implications for both direct measurement and indirect measurement through standardized mortality rates. *The Milbank Quarterly,* 97**,** 228-284.

MARKS, A. D., HINE, D. W., BLORE, R. L. & PHILLIPS, W. J. 2008. Assessing individual differences in adolescents' preference for rational and experiential cognition. *Personality and Individual Differences,* 44**,** 42-52.

MCKINLAY, J. B., LIN, T., FREUND, K. & MOSKOWITZ, M. 2002. The unexpected influence of physician attributes on clinical decisions: Results of an experiment. *Journal of Health and Social Behavior,* 43**,** 92-106.

MCKINLAY, J. B., POTTER, D. A. & FELDMAN, H. A. 1996. Non-medical influences on medical decision-making. *Social Science & Medicine,* 42**,** 769-776.

MCLAIN, D. L. 1993. The MSTAT-I: A new measure of an individual's tolerance for ambiguity. *Educational and Psychological Measurement,* 53**,** 183-189.

MEYER, A. N. D., PAYNE, V. L., MEEKS, D. W., RAO, R. & SINGH, H. 2013. Physicians' diagnostic accuracy, confidence, and resource requests: A vignette study. *JAMA Internal Medicine,* 173**,** 1952-1961.

MICHELL, J. 1999. *Measurement in psychology: A critical history of a methodological concept,* Cambridge University Press.

MICHELL, J. 2008. Is psychometrics pathological science? *Measurement,* 6**,** 7-24.

MICHELL, J. 2019. The fashionable scientific fraud: Collingwood's critique of psychometrics. *History of the Human Sciences***,** 0952695119872638.

MONACIS, L., DE PALO, V., DI NUOVO, S. & SINATRA, M. 2016. Validation of the rational and experiential multimodal inventory in the Italian context. *Psychological reports,* 119**,** 242-262.

NEUBERG, S. L. & NEWSOM, J. T. 1993. Personal need for structure: Individual differences in the desire for simpler structure. *Journal of Personality and Social Psychology,* 65**,** 113.

NEZLEK, J. B. 2008. An Introduction to Multilevel Modeling for Social and Personality Psychology. *Social and Personality Psychology Compass,* 2**,** 842-860.

O'NEIL, W. 1944. Factors and faculties. *Australasian Journal of Psychology and Philosophy,* 22**,** 55-69.

ORTEGA EGEA, J. M. & ROMÁN GONZÁLEZ, M. V. 2011. Explaining physicians' acceptance of EHCR systems: An extension of TAM with trust and risk factors. *Computers in Human Behavior,* 27**,** 319-332.

PARKER, A. M., DE BRUIN, W. B. & FISCHHOFF, B. 2007. Maximizers versus satisficers: Decision-making styles, competence, and outcomes. *Judgment and Decision making,* 2**,** 342.

PETERSON, D. K. & PITZ, G. F. 1986. Effects of amount of information on predictions of uncertain quantities. *Acta Psychologica,* 61**,** 229-241.

PETTY, R. E. & CACIOPPO, J. T. 2012. *Communication and persuasion: Central and peripheral routes to attitude change,* Springer Science & Business Media.

PLANTINGA, A. 1993. *Warrant and proper function*, Oxford University Press.

RASBASH, J. 2008. Module 4: Multilevel structures and classifications.

REYNA, C. & ORTIZ, M. V. 2016. Psychometric study of the Rational Experiential Inventory among undergraduate Argentinean students. *Revista de Psicología*, 34.

ROSS, L. & NISBETT, R. E. 2011. *The person and the situation: Perspectives of social psychology*, Pinter & Martin Publishers, London.

SABIN, J. A., RIVARA, F. P. & GREENWALD, A. G. 2008. Physician implicit attitudes and stereotypes about race and quality of medical care. *Medical Care***,** 678-685.

SADOWSKI, C. J. & GULGOZ, S. 1992. Internal consistency and test-retest reliability of the Need for Cognition Scale. *Perceptual and Motor Skills,* 74**,** 610-610.

SAPOSNIK, G., REDELMEIER, D., RUFF, C. C. & TOBLER, P. N. 2016. Cognitive biases associated with medical decisions: a systematic review. *BMC Medical Informatics and Decision Making,* 16**,** 138.

SCHNEIDER, A., SZECSENYI, J., BARIE, S., JOEST, K. & ROSEMANN, T. 2007. Validation and cultural adaptation of a German version of the Physicians' Reactions to Uncertainty scales. *BMC health services research,* 7**,** 81.

SELLTIZ, C., WRIGHTSMAN, L. S., COOK, S. W., BALCH, G. I., HOFSTETTER, R. & BICKMAN, L. 1976. Research methods in social relations.

SENA, L. A. 1972. *Units of physical quantities and their dimensions*, Mir Pub.

SLIJKHUIS, J., RIETZSCHEL, E. F. & VAN YPEREN, N. W. 2013. How evaluation and need for structure affect motivation and creativity. *European Journal of Work and Organizational Psychology,* 22**,** 15-25.

SMITH, M., HIGGS, J. & ELLIS, E. 2008. Factors influencing clinical decision making. *Clinical reasoning in the health professions,* 3**,** 89-100.

SMITH, M. A., ATHERLY, A. J., KANE, R. L. & PACALA, J. T. 1997. Peer review of the quality of care: reliability and sources of variability for outcome and process assessments. *Journal of American Medical Association,* 278**,** 1573-1578.

SNIJDERS, T. A. Power and sample size in multilevel modeling.

SOCIETY OF JUDGEMENT AND DECISION-MAKING. 2020. *Decision Making Individual Differences Inventory (DMIDI)* [Online]. The Society for Judgment and Decision Making. Available: http://www.sjdm.org/dmidi/ [Accessed 2018].

STANOVICH, K. E. & WEST, R. F. 2008. On the relative independence of thinking biases and cognitive ability. *Journal of Personality and Social Psychology,* 94**,** 672-695.

STEVENS, S. S. 1946. On the theory of scales of measurement.

SWETS, J. A., DAWES, R. M. & MONAHAN, J. 2000. Psychological Science Can Improve Diagnostic Decisions. *Psychological Science in the Public Interest,* 1**,** 1-26.

TAVAKOL, M. & DENNICK, R. 2011. Making sense of Cronbach's alpha. *International journal of medical education,* 2**,** 53-55.

TE. A, SCHMIDTKHE, K., MANASEKI-HOLLAND, S., LILFORD, R., BION, J., VLAEV, I. & HOFER, T. 2021. The influence of physician reviewer characteristics and attitudes on care quality judgements during case note reviews: a multilevel analysis. *Health services research*.

THE HEALTH AND SOCIAL CARE INFORMATION CENTRE 2013. Standards for the clinical structure and content of patient records.

THE ROYAL COLLEGE OF ANAESTHETISTS. 2020. *Risk explained* [Online]. [Accessed 19/02/2021].

THE ROYAL COLLEGE OF OBSTETRICIANS AND GYNAECOLOGISTS. 2015. *Understanding how risk is discussed in healthcare* [Online]. Available:

https://www.rcog.org.uk/en/patients/patient-leaflets/understanding-how-risk-is-discussed-in-healthcare/ [Accessed].

THOMPSON, M. M., NACCARATO, M. E., PARKER, K. C. & MOSKOWITZ, G. B. The personal need for structure and personal fear of invalidity measures: Historical perspectives, current applications, and future directions. Cognitive social psychology: The Princeton symposium on the legacy and future of social cognition, 2001. 19-39.

TVERSKY, A. & KAHNEMAN, D. 1975. Judgment under uncertainty: Heuristics and biases. *Utility, Probability, and Human Decision Making.* Springer.

VAN DEN NOORTGATE, W., OPDENAKKER, M.-C. & ONGHENA, P. 2005. The effects of ignoring a level in multilevel analysis. *School Effectiveness and School Improvement,* 16**,** 281-303.

WALLSTEN, T. S. 1981. Physician and medical student bias in evaluating diagnostic information. *Medical decision making,* 1**,** 145-164.

WATSON, S. I., CHEN, Y. F., BION, J. F., ALDRIDGE, C. P., GIRLING, A., LILFORD, R. J. & THE HI, S. C. 2018. Protocol for the health economic evaluation of increasing the weekend specialist to patient ratio in hospitals in England. *BMJ Open,* 8.

WEBER, E. U., BLAIS, A. R. & BETZ, N. E. 2002. A Domain-specific Risk-attitude Scale: Measuring Risk Perceptions and Risk Behaviors. *Journal of Behavioral Decision Making,* 15**,** 263-290.

WENNBERG, J. E. 1987. The paradox of appropriate care. *Journal of American Medical Association,* 258**,** 2568-2569.

WOOD, W. 2000. Attitude change: Persuasion and social influence. *Annual review of psychology,* 51.

CHAPTER 4: THE BARRIERS AND FACILITATORS OF RETROSPECTIVE MORTALITY CASE-NOTE REVIEW FOR QUALITY IMPROVEMENT (QI): EVIDENCE FROM A UK HOSPITAL CASE STUDY

Chapter Preface

Having addressed the question of whether reviewer attitudes, their demographics and patient characteristics influence global care quality judgements and confidence in such care quality judgements in the previous chapter, Chapter 4 is concerned with organisational culture, embedding, information flow and use of case-note review for quality improvement (QI). Social science methods, such as qualitative case studies, can illuminate case-note review's information flow, use, embedding and how it informs quality improvement processes. Case-note review is one of many quality improvement interventions and is a qualitative methodology well-suited to investigating case-note review application and use.(Stake, 2005) It is to this subject matter that we now turn to in this chapter. Chapter 4 will consider the embedding of the case-note review intervention in the Trust (against the backdrop of existing interventions and systems), barriers and facilitators to its uptake and use, the perceived contributions toward hospital care quality improvement and the implications this could have for other NHS Trusts and healthcare institutions.

This chapter is organised as follows: first, it provides information concerning the method that was used in undertaking this research as well as a justification for the use of this method. Second, it describes the various stages of the research, which includes the selection of participants, the data collection process, and the process of data analysis. Third, it discusses the role of the researcher in qualitative research in relation to the concept of reflexivity. Finally, there is a discussion of validity and reliability in qualitative research and of how these two requirements were met in the current study. The Chapter 4 content has been prepared for journal submission.

4.1. Abstract

Objectives: To determine the barriers and facilitators to QI from case-note reviews. Our subsidiary aim was to identify barriers and facilitators for case-note review's embedding and information flow, within quality improvement activities.

Methods: Using a single case-study method, we conducted semi-structured interviews with case-note reviewers, managers, board members and non-executive members and documentary analysis in an English NHS hospital setting. Patients and the public were not involved in this research. We used the COREQ checklist. Normalization Process Theory provided a conceptual framework for data collection, analysis, and interpretation. An inductive framework method was used to analyze factors related to information flow and quality improvement. Results were triangulated from interviews and documents.

Results: Case-note review was well-embedded in hospital processes and information flowed well from board-to-ward, but less so from ward-to-board. Information flow was facilitated by tools, policy guidelines, information technology, administrative support, and formal committee/officer roles. Barriers to information flow included poor information technology, overlapping purpose of patient safety interventions, limited learning from case-note reviews, poor administration of case-note review information, specialty-silos, and the lack of reviewing time. Perceived QI facilitators from MCNRs, included adherence to mandates, support from patient advisory liaison services, continued professional development requirement, a supportive organizational culture and case-note reviewer investment into their patients. Barriers included being unaware of training, scepticism about the benefit of training, pressures of work and negative attitudes to reviews.

Conclusions: We identified a range of facilitators and barriers to mortality case-note review information flow, embedding and perceived facilitators and barriers to quality improvement. We found some evidence, suggesting that mortality case-note reviews led to QI around complex, deteriorating patients.

4.2. Introduction

4.2.1. What is Quality?

Quality is a concept describing features of a product or service to which value is ascribed. Consequently, the nature of quality can vary between products and services, individuals, and organisations. In this thesis, it is discussed in relation to its significance within the hospital healthcare setting. Given the subjective nature of quality, defining clearly 'high quality' healthcare provision is challenging. It is unsurprising that various healthcare organisations differ in their ways of defining quality.(Raleigh and Foot, 2010) The Institute of Medicine (IoM) views quality as 'the degree to which health services for individuals and populations increase the likelihood of desired health outcomes and are consistent with current body of professional knowledge.'(Lohr, 1991) However, in order to more clearly define quality by the development of expectations or standards (for benchmarking purposes), it is necessary to ascribe dimensions of quality, also known as domains.(Healthcare and Workforce Improvement, 2017) Different organisations opt for various numbers and combinations of these domains. In the report *Crossing the quality chasm* by the Institute of Medicine expanded on its previous statement in listing six fundamental domains of quality: safety, patient experience, effectiveness, efficiency, equity and timeliness.(Wolfe, 2001) In Lord Darzi's Next Stage Review, there was a call for high quality care for all that is 'personal, effective and safe.'(Health, 2008) The US Quality Assurance Project went further, defining

nine domains of quality: access, technical performance, effectiveness, efficiency, interpersonal relationships, continuity, safety, choice, and physical infrastructure and comfort.(Healthcare and Workforce Improvement, 2017) And most recently for the UK, The Royal College of Physicians (RCP) adopted a definition of quality which comprises patient experience, safety, effectiveness, efficiency, equity, timeliness and sustainability.[38] Despite the apparent variability in the number and nature of the domains ascribed, certain areas seem consistently important. A document released in March 2006 by the Organisation for Economic Co-operation and Development drew together domains of quality from the frameworks of six member countries and three international health organisations. The report found that "Effectiveness, safety, patient experience, efficiency, equity and accessibility" were the most frequently incorporated domains.(Kelley and Hurst, 2006)

In any dynamic social, economic, and political contexts of healthcare delivery, certain domains may be considered more pertinent at certain times. Interactions between different domains are inescapable and, while some of these will be synergistic, attempts to fulfil all of them may lead to tensions developing. It will be necessary to make decisions regarding the relative importance of certain domains. Oftentimes, however, the organisation is not at all aware of these tensions within their own subculture or the organisation. Next, we turn to some conceptions and models of quality improvement which have been used to make sense of QI work and improve its work. Given that quality is a complex concept, we would benefit from discussing the concept and some of its key domains.

[38] There is more detailed discussion of this already provided in the Introduction to this thesis.

4.2.2. Quality Improvement: the concept of quality improvement

Quality is not a static concept because there is a need for continued improvement across time. The key elements of QI are a change (e.g., an improvement) and a method, specifically techniques and tools). Improvement is about change and action based on experience. And this experience is a product of the culture; this culture is laden with assumptions and must be unpicked to reveal what is seen as acceptable and what is not. For instance, a safety-I culture will seek to reduce errors in care and practise defensive medicine, whilst an improvement culture informed by safety-II will look to understand why and how the good care has arisen in the first place.(Hollnagel et al.)

Within this culture, individuals and teams combine a threshold for action with a readiness to develop and test ideas, and to make changes to protocol. Improvement is aspirational, future-focused and applied in complex clinical systems, unlike pure research, where variables are controlled. There are several methods, all of which are founded on a science concerned with understanding variation and the application of statistical methods and behavioural principles.(Marshall, 2009) Given the range of operational cultures and environments in healthcare, successful and effective changes must be context-specific. Each team works differently, and implementation requires a strategy specific to the setting. And the implementation strategy must consistently align with the culture of safety and work in that specific context for it to be effective. Having outlined some of the dimensions of quality concept, we are now well placed to discuss the components and approaches on measuring and improving this quality.

4.2.3. Quality Improvement: Components and Approaches

Quality improvement describes a systematic process to improve quality with areas identified for improvement, the problem understood with solutions tested and the impact of any change evaluated and measured. One example could be the improvement of waiting times to access a mental health service or reducing the length of stay and bed occupancy or reducing medication error in the emergency department. There is no single definition of quality improvement per se, and no single approach can be said to be more successful than any other. However, there are some which have been employed more widely than others and they will be discussed in this section.

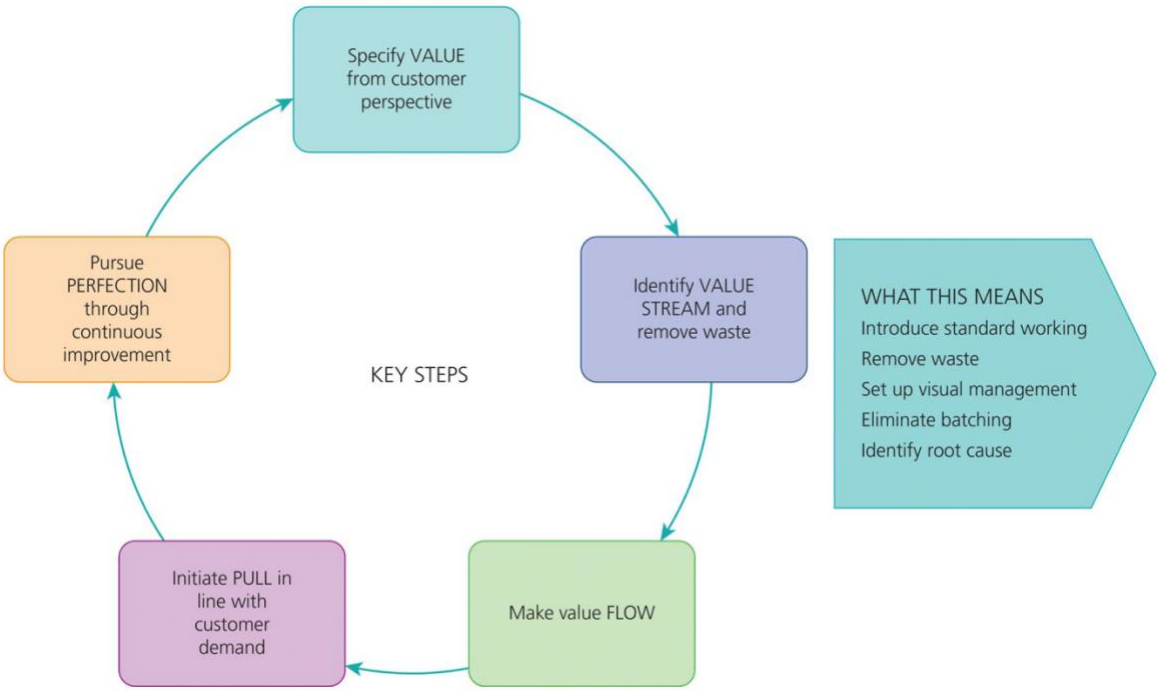
4.2.3.1. Lean

The most common model adopted as an organisation-wide approach in healthcare is 'Lean', based on the Toyota Production System. The appeal of this methodology to the healthcare context was the prevailing philosophy within Toyota about empowering frontline workers to improve processes to develop an enhanced and more effective product, more efficiently. In essence, Lean is the continuous and systematic elimination of waste, with waste being defined as anything that does not add value to the patient or process. Lean methodology describes five key tasks that, if addressed, deliver what the patient wants, and needs, at the highest quality and safety level possible with the lowest associated costs, provided by a highly motivated workforce (see. Figure 10).

Healthcare organisations in the UK are learning from overseas organisations and are increasingly adopting an organisational approach to continuous quality improvement. An organisation-wide deployment of continuous improvement principles engages frontline staff

and embeds a scalable methodology for the growing and the coordinating of improvement activities. This approach is fast becoming the normative approach of commissioners and regulators of healthcare and, as a result, many organisations have built improvement and transformation teams with specific expertise in their chosen methodology. Organisations include the Virginia Mason Institute (USA), Jonkoping County Council (Sweden), Southcentral Foundation (Alaska) and the UK's Salford Royal NHS Foundation Trust have demonstrated that an organisation-wide approach to quality improvement, adopting a single model approach, creates a sustainable culture of improvement which enhances patient experience and safe care delivery.

Figure 10. Key iterative steps in the LEAN methodology



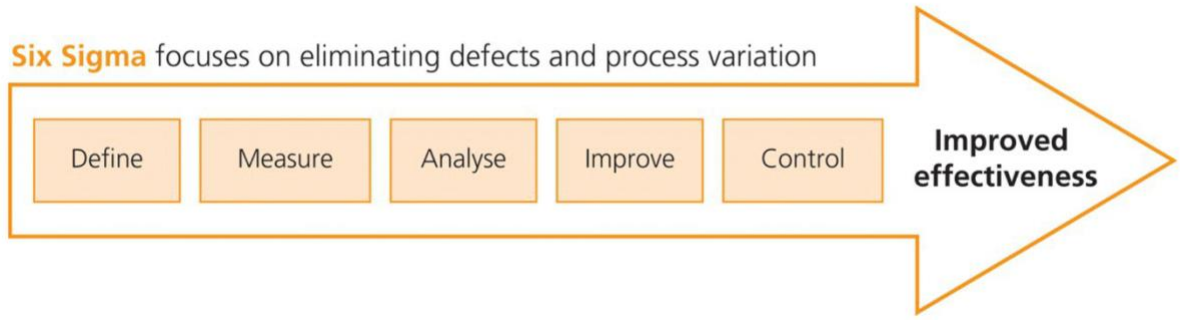
There are several documented examples that demonstrate where this Lean approach has improved patient satisfaction, reduced waiting times and improved productivity. Improvements have been recorded in the processing of paperwork and the scheduling of

appointments. Health services have successfully deployed Lean in various areas of operational processes but increasingly the approach is being adopted across entire organisations. One notable example where Lean has been adopted wholesale is by the US hospital and integrated care system, Virginia Mason. This Institute is now supporting several Trusts in the UK to adopt a Lean approach. There are several instances from this systematic review which illuminate how and why the lean method is less successful than expected.(Moraros et al., 2016) A realist review is more favourable on the effectiveness of lean interventions, there, nevertheless, was a need to more deeply involve senior management, pursue value with patients and nurture a long-term journey of continual improvement.(Mazzocato et al., 2010)

4.2.3.2. Six Sigma

Six Sigma is an improvement methodology focused on understanding variation and then reducing in-process variation to improve results. It is used to complement Lean methodology or can be used independently. Six Sigma has not been widely adopted in healthcare settings independently – perhaps because of cost, time constraints and contributions by a handful of experts rather than mass participation by staff and/or patients – but it is beneficial due to its focus on measurement and understanding variation (Figure 11).

Figure 11. Key steps of the Six Sigma philosophy for improved effectiveness of healthcare interventions



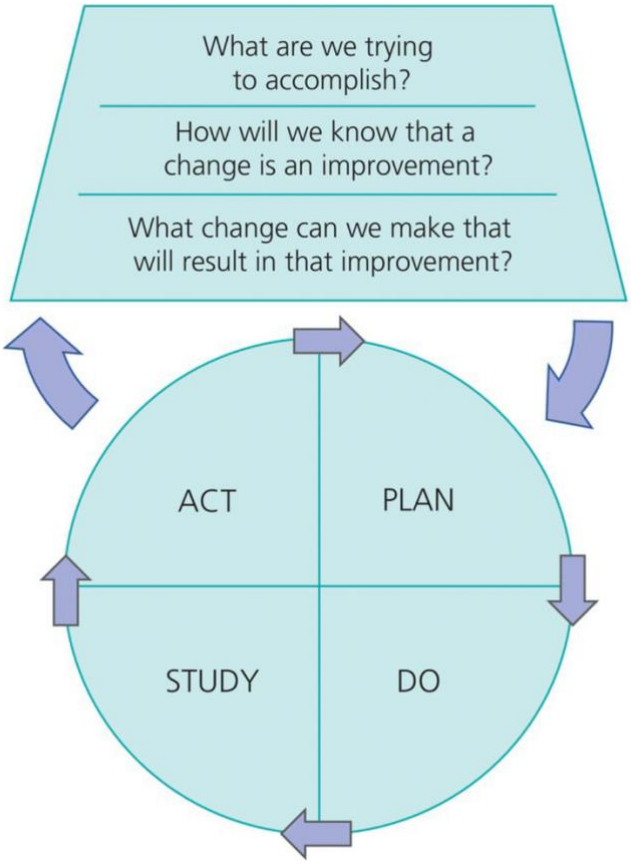
4.2.3.3. Experienced-based Co-design

Experience-based Co-design (EBCD) is a tool that was designed to develop simple solutions to improve patient experience. It uses a qualitative, story-telling approach that enable staff and patients to co-design services and care pathways in partnership. It focuses on developing a deeper understanding of how patients and staff experience a service. It does this by gathering the perspectives of patients and staff through in-depth interviewing, observations, and group discussions. There are emotionally significant points of the experience, known as key touch points, which are then identified, and the group assigns positive or negative feelings to each of these touch points.(Swanwick and Vaux, 2020)

4.2.3.4. Model for Improvement

The Model for Improvement(Langley et al., 1996) was originally developed by the Institute of Healthcare Improvement (IHI) in the USA and provides a relatively simple framework for delivering change at pace. The model is based on asking three fundamental questions to clarify the problem, understand what will be different if the change has been a success and to clarify the actual change required. It then goes on to provide a framework, known as a Plan Do Study Act (PDSA) cycle, for testing change(s) and reviewing impact (Figure 12).

Figure 12. Key questions filtering into the PDSA cycle



4.2.3.5. Safety-I and Safety-II

Culture comes before the concept of improvement. And concept comes before any co-ordinated actions. Thus, it is important to find common ground in a culture. Culture matters to quality improvement because the way individuals and groups perceive their reality will shape their concepts and their actions. For instance, a culture intent on minimising errors will focus its efforts to understand how the delivery of care has gone poorly to improve the outcome. But it will not develop a complete picture as it neglects how and why the outcome from "good care" has gone well. A possible consequence of an error-minimising culture could develop into a context where individuals do not desire to be investigated for their mistakes, intentional or not, which forms a barrier to true and open reflection on care. This is

likely to have unintended consequences as it unintentionally degrades the resources and procedures needed to make things go right. However, this neglects why and how care has gone well.(Hollnagel et al.)

With Safety-II, the emphasis on investigations is to understand how things usually go right, as this forms the basis of the explanation for why things do not go right. This approach normalises the study of "everyday" work and seeks to understand and better respond to inevitable surprises. This approach is adaptive and seeks to build upon existing good work without the consequences of having to overhaul the very foundations of care, which are on the most part, exemplary.(Finkel, 2020)

4.2.3.6. Single method vs 'pick and mix'

There is growing evidence that organisations that have committed to one improvement approach achieve better outcomes and a more coherent culture of work. These organisations have developed training programmes that staff are strongly encouraged to attend as an improvement "language" becomes familiar across the organisation. In short, these organisations are developing a culture of improvement – '*the way we do things round here*' – the culture – are being more actively of their culture and their assumptions which automatically generates further interest and engagement in the improvement methodology. (Swanwick and Vaux, 2020) Having discussed concepts and approaches to quality improvement imported into healthcare from non-healthcare settings, we will look at a model of some factors which are healthy indicators of success for quality improvement in a healthcare setting.

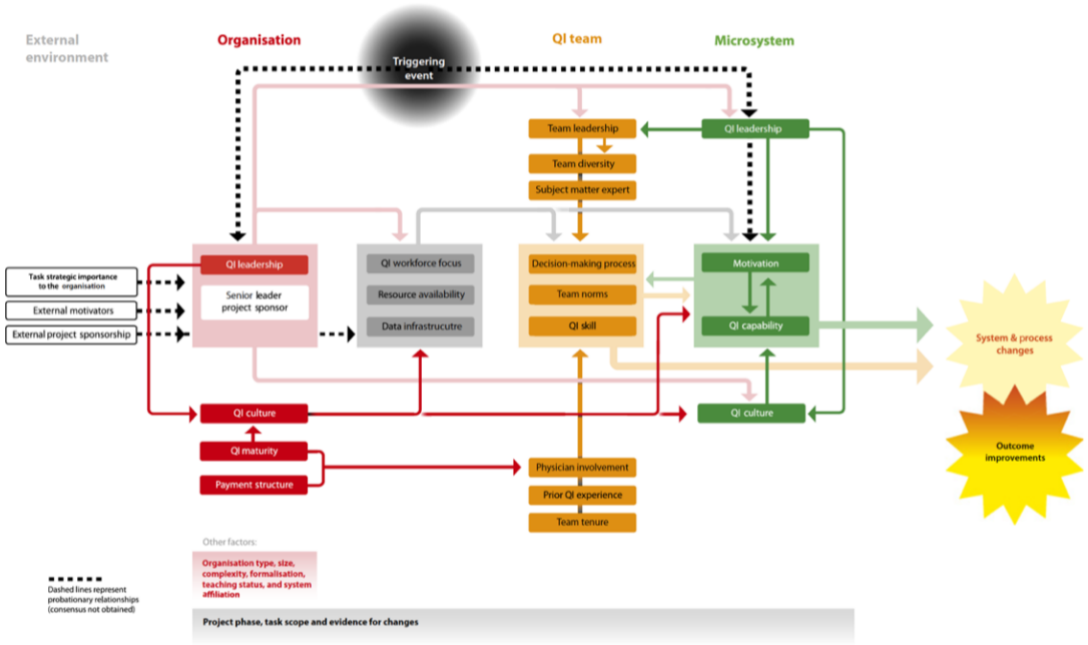
4.2.4. Conceptualization of Quality Improvement

4.2.4.1. The Model for Understanding Success in Quality (MUSIQ)

There have been various conceptualisations of what leads to enhanced quality improvement. However, most of them are insensitive to contextual factors. The Model for Understanding Success in Quality (MUSIQ) is different as it is a model which is designed to help organisations and QI researchers to understand and optimize contextual factors affecting the success of a QI project.(Kaplan et al., 2012) This current model includes broad-brushed categories revolving around the hospital's external environment, factors within the organisation, QI support and capacity, the microsystem (within-hospital services) and the QI team itself. Quality improvement is a complex social intervention for which there are high levels of variance in the context, content, and application. Figure 13 highlights the interactions and pathways of MUSIQ. This model is chosen because it reflects the complex interactions of quality improvement better than other less granular theories of quality improvement. This will help us to appraise in the discussion to what extent these factors are satisfied, and the reasons why, in the discussion.

This case study could complement and/or inform MUSIQ. For instance, if there is independent support from a non-MUSIQ framework, then this provides additional support for MUSIQ's relevance to case-note reviews. If this is so, then it would be beneficial for hospitals to employ MUSIQ categories when implementing and operationalising case-note reviews. This is a way to triangulate salient components for case-note reviews. In the next section, I detail the theoretical approach I will take.

Figure 13. Model for Understanding Success in Quality (MUSIQ).



MUSIQ shows how the organisational context influences the success of individual quality improvement (QI) projects. Contextual factors are organised based on the level of the healthcare system in which they are believed to operate, including the microsystem (green), organisational or macrosystem (red), and environmental levels (white). Factors relating to aspects of QI support and capacity (grey), or characteristics of the QI team (orange) are identified as existing across system levels. Factors not assigned to a specific system level are also in white.

Bearing MUSIQ in mind, I move on to now consider a theory to help us to understand the embedding of this intervention in this Trust. There is a need to first consider whether the intervention is embedded within organisation before we can examine the success of achieving an improved outcome with this intervention.

4.2.4.2. The study: the context

The case study examined the embedding, information flow and the quality improvement contributions from case-note reviews. Previous research has demonstrated that case-note reviews have the potential to support quality improvement (Kobewka et al., 2017) but the underlying organisational processes behind this are unknown. The main objective of this study was to assess whether mortality case-note reviews lead to hospital care quality improvement.

The retrospective review of hospital patient mortality case-notes is an important reflective process that has been used to inform the improvement of care quality over several years (Buck et al., 1987, NCEPOD, 2018, Raja and Thomas, 2019) in the National Health Service (NHS). Case-note reviews are independent evaluations of patient notes with no conferring allowed between medical peer reviewers. Such reviews are also commonly used in the United States for medico-legal processes(Sullivan and Anderson, 2010), hospital credentialing (Makary et al., 2011) and adverse event reviews. (Edwards, 2013)

This case study aimed to investigate whether case-note review is an inefficient use of clinician and hospital staff time and resources. Specifically, the study sought to explore whether case-note reviews lead to discernible differences in hospital quality improvement

and the role case-note review embedding and information flow serves in supporting this contribution to quality improvement. The next sections outline the choice of research design, selection of informants, data collection procedures, pilot-testing, research instruments employed, and the data analysis.

We explored whether MCNR identify care quality issues which lead to hospital quality improvement. MCNRs evaluate the overall quality of care of deceased patients through the use of a Likert scale (Manaseki-Holland et al., 2019b), see (Manaseki-Holland et al., 2016, Lilford et al., 2007, Manaseki-Holland et al., 2019b, Goldman, 1992, Goldman, 1994). Case-note reviews have been shown to contribute to hospital quality improvement(Sari et al., 2007, Hogan et al., 2014, Sujan et al., 2017, Zwaan et al., 2010, Girling et al., 2012b) but this finding has yet to be demonstrated in national programmes such as the MCNR.

To address questions around MCNR QI, we need to know first whether MCNR is embedded in hospital activity and routinely used by staff. To explore this, we used Normalisation Process Theory (NPT), which is a theoretical framework that can be used to explore the extent of the *embedding* of a technological or organisational intervention (Murray et al., 2010, May et al., 2009). And second whether the information derived from MCNRs is used by staff. MCNR embedding and information are pre-requisites for any MCNR QI. Our aim is to identify the barriers and facilitators to quality improvement, embedding, information flow from case-note review, which involve the retrospective review of hospital inpatient notes (Manaseki-Holland et al., 2019b) in compliance with England's learning from deaths national policy initiative. (National Quality Board UK, 2017, Raja and Thomas, 2019, McGrath et al., 2015, O'Reilly et al., 2017)

4.3. Methods

4.3.1. Research Philosophy

I describe and justify the use of different research philosophies. This includes a presentation of the various theoretical paradigms used in the undertaking of this case study. Axiology (*the values assumed in our work*), ontology (*which entities are involved and their nature*) and epistemology (*what and how we can know*) to this research practice. Paradigms (of research) are defined as "*basic belief systems based on ontological, epistemological and methodological assumptions*."(Guba and Lincoln, 1994) Each paradigm can subsequently be broken down into the following dimensions: axiology, ontology and epistemology. These dimensions will justify the use of these research philosophies.

4.3.2. Axiology

Derived from the Greek word *axios*, meaning value, axiology studies the nature, types, and criteria of value judgements. For my purposes, my axiological orientation is established on Avedis Donabedian's quality of care components.(Donabedian, 2002) These include efficacy, effectiveness, efficiency, optimality, acceptability, legitimacy and equity. These are each desirable features that need to be maximised for optimal clinical care.

Several values are helpful to consider in the context of healthcare quality improvement. *Efficacy is* the ability of the science and technology of health care to bring about improvements in health when used under the most favourable circumstances. "*Effectiveness"* is the degree to which the greatest possible improvements in healthcare at that moment in time, are in fact attained. "*Efficiency*" is the ability to lower the cost of care without diminishing improvements in health, *ceteris paribus*. *Optimality* is concerned with the

balancing of improvements in health against the cost of such improvements. *Acceptability* is defined as conformity to the wishes, desires, and expectations of patients, and to some extent their kith and friends. By *accessibility* I mean the ease with which persons can obtain care. The patient-practitioner relationship must be considered within this rubric; however, the case-notes will not be totally transparent on these clinician-patient relational dimensions (i.e., congruence, adaptation and flexibility, mutuality, stability, maximum autonomy). An idealised case-note would be able to capture some if not all these dimensions. Case-notes are plagued by poor information, inaccuracies, and the non-chronological ordering of care events. Next, we discuss an important aspect of any framework, the substance and nature of it, through its ontologies.

4.3.3. Ontology

Ontology is a branch of philosophy which explores the existence and essence (*essentia*) of objects.(Bricker, 2014) Aristotle called ontology the first philosophy.(Cohen, 2000) Ontology is focussed on the nature of being or the kinds of things that have existence with questions including '*what exists?*' '*what is true?*' and '*how can we sort existing things?*' Qualitative research is more interested in what is called 'social ontology', that is '*the analysis of various entities in the world that arise from social interaction*.' We intend to study this social ontology to describe and understand the organisational dynamics around hospital mortality case-note reviews.(Epstein, 2018)

The ontological position in this research is described as 'realist' in that it was assumed that there is the existence of an independent reality that is entirely separate from any individual's cognitive model (i.e., perceptions and beliefs) about the world. In essence, truth exists, and is

independent of our human understanding of it.(Crossan, 2003) This is contrary to the idealist (and relativistic) stance which posits that an external objective account of reality is contested. Reality is understood as including both physical and social phenomena (whilst the materialist accepts the former, the idealist accepts the latter and the relativist accepts any mixture of the two). However, allegiance to a purely realist stance is naïve as our participants will inevitably have different axiological positions regarding case-note reviews. This is not an inviolable problem as this variant of realism allows for some philosophical latitude. My ontological position allows multiple subjective accounts using material sourced on an objective external reality. To demonstrate, mortality case-note review is either effective or not effective at producing quality improvement assuming, but the questions remains whether there is sufficient evidence to test this hypothesis. Any mechanisms behind it being effective (or not) can be subject to a critical realist lens, namely through the empirical, actual and real domains.(Fletcher, 2017)

Qua critical realism, the researcher focusses on the manipulation of physical and social phenomena, from any participants perspective which is ultimately grounded in a single external reality. Hereon in, the critical realist stance upholds reality as accessible and knowable through the individual's socially constructed world and that the knowledge of this reality, though incomplete, approximates to probabilities of high confidence if the methods employed are adequate for eliciting these dimensions.(Huberman and Miles, 2002) Critical realists strive for objectivity; yet perfect objectivity from any researcher/participant perspective is not obtainable. A consequence of the realist approach is that there is a correct approach to any mortality case-note review, in any time or place, where quality improvement lessons can be obtained and that, given the interaction of the features at each of the three

levels (see Appendix 16 for a discussion of critical realism's assumptions and capacity to frame the case-note review process).

4.3.4. Epistemology

Epistemology is the study of the nature of knowledge and the ways it can be acquired.(Bricker, 2014) In terms of my own role in gathering knowledge for this case-study, I reject the positivist position because that would entail that the researcher has direct, incorrigible access to the uncontested 'facts' of an independent, external reality. This is evidently not possible, as I can only be given access to a limited set of these facts given the complexity of the hospital setting. I assume that the external reality is the same for all participants. But individuals can come to construct their own reality within this objective, external reality. It is the subjects of the external reality which matters most, and these can vary considerable from person-to-person. This position accounts for individual choice and representation of constructed reality which is anchored on an independent external reality. In other words, this external reality is knowable, however, we do not have full access to the sum of data equivalent to direct knowledge of external reality. Thus, the agent has the freedom to (and must) socially construct the datum presented to them.

This classically leads to the 'under-determination' of evidence(Stanford, 2009) which is the situation where the evidence supports the vindication of more than one theory explaining the phenomenon of interest. In underscoring this position, the agents may have access to the same or different datum (*evidence*). Furthermore, this data is then coupled with their own cherished values (*axios*) which can lead to multiple instances of socially constructed positions despite the employment of the same data. In other words, the evidence is the raw

material used by subjects to be interpreted according to their own preferences. This may appear to undermine the objectivity of qualitative findings. However, this is not the case. Objective observer-independent reality is as real as the social reality of subjects interpreting the objective reality. Just as the objective reality is governed by physical laws, the social reality is governed by their social laws and mechanisms. The aim of this research is to describe and understand the mechanisms behind the different presentations of social reality in this hospital Trust.

To characterise a social epistemology, we recourse to sentences and propositions as real-world truth makers to establish a relation to external reality. This is developed from the correspondence tradition of truth which states that the truth or falsity of a statement is determined only by how it relates to the world and whether it actually describes that particular world.(Hanna and Harrison, 2004) This is construed and defended in Alvin Goldman's '*Knowledge in a Social World*' which establishes an objective philosophical grounding for institutional structures relevant to healthcare organisations.(Goldman, 1999)

4.3.5. Research Design

4.3.5.1. Case-study method and justification

The case study method was used because it is well-suited to answering the elicited research questions. The strength of the case study method is that it allows for the in-depth examination of the phenomenon using various kinds of evidence obtained from participant interviews, direct observation of events and the analysis of documents and artifacts (Yin, 2003). Also, the case study was used because the study aim was to describe rather than predict quality improvement from case-note reviews, and case-note review is not easily isolated away from

its study context (Merriam, 1988). In addition, the case study allows for empirical inquiry around case-note reviews within its real-life context, especially when the boundaries between phenomenon and context are not clear and cannot be easily pieced apart from each other (Yin, 2003). An objection to the use of single-case study research design is its lack of external validity or applicability to other cases. However, a single-case study's ability to apply or generalize the results of this study are dependent on several factors, with sample size being one of these. To preserve replicability, I adhere to this primary mechanism for establishing the generality of an empirical finding, which is a common convention of the natural sciences. The basis of its generality is founded on its applicability in theory, which is supported by prior observation or the evidence from the literature. In this way, the single case-study design is analytically generalisable to other NHS Acute care Trusts in the UK for following reasons: firstly, the United Kingdom's National Health Service Acute Care Trusts offer a common set of core clinical services which are generally supported by similar central business and administrative systems and so the context from this case-study is extensible to other NHS Healthcare Trusts.[39] Secondly, the mandate for reviewing has been cascaded to all NHS Trusts(National Quality Board UK, 2017) and the requisite training provided.(The Royal College of Pathologists, 2016) Thirdly, this NHS Trust has seen successive improvements in its overall Care Quality Commission rating, which signifies that this Trust's context may offer more lessons for other NHS Trusts and contexts to learn from than those

[39] There may be a case for its applicability beyond the NHS realm, however, this needs to explore further in relation to other healthcare organisations and their systems, and their own contexts

with other CQC quality trends.[40] However, this is assuming that our theoretical approach to understanding quality improvement in healthcare has verisimilitude with the intervention context, content, application and outcomes.(Walshe, 2007)

This approach enables us to study the complexity of a complex social intervention such as a retrospective mortality case-note review, one of a constellation of quality improvement interventions. In first approaching the hospital, we were able to describe to participants the case study aims, the process and the likely implications of this case study. I discuss in section

4.3.8. Analytic generalizability

Conceptually, case-note reviews are a retrospective tool used by clinicians to identify care quality issues and/or lapses in care. The overall purpose of case-note reviews is unanimously lauded as an important contributor to care quality improvement. However, the efficiency and effectiveness of the discrete steps part of case-note reviews has not been studied in any detail. For instance, the effects of how case-notes are processed, are distributed to reviewers, and collected again for collation as case-notes and their re-distribution are largely unknown. There have been no previous studies of this for case-note reviews. There are several unknowns about how case-note reviews are operationalised in the hospital setting despite case-note reviews becoming an established method, certainly in the United Kingdom, for

[40] Specifically, at the time when the case-study was being conducted, the CQC report indicated that was 'very good' in 2017-2018 which has improved upon its previous CQC overall rating. As of December 2019, the Trust now has a CQC overall rating of 'outstanding.' The Trust has gone on to improve with each successive CQC inspection.

retrospective learning from deaths for care quality improvement. For instance, is there an optimal guidance for how case-note reviews are effectively used in the hospital, and specifically within its various specialties? And which elements of the operationalisation of case-note reviews are most critical for its success within hospitals as a function of its specialties? And importantly, why does this result even obtain? Thus far, no literature exists on the operationalisation of case-note reviews. The question must be asked; why are case-note reviews used the way they are? And is this indeed the best way to use them? And if not, what could be changed to improve the quality improvement from case-note reviews. This case-study is intended to begin the conversation around these questions.

4.3.6. Normalisation Process Theory

The SSI questions (and themes) were derived using Normalisation Process Theory (NPT) and its associated constructs. NPT is a template commonly used to identify factors promoting or inhibiting the routine incorporation of a complex intervention.(May et al., 2009, Finch et al., 2012, Murray et al., 2010) The data-gathering and the interpretation steps were informed by NPT. This theory is widely used in health service research to evaluate the embedding of complex interventions.(O'Reilly et al., 2017, de Brún et al., 2016b, Chambers et al., 2019) These four NPT constructs (i.e. coherence, cognitive participation, collective action and reflexive monitoring) were applied to explore the embedding of these MCNRs. As case-note review was neither a simple nor uniform hospital intervention, we consider NPT a useful template for such exploration. The questions were adapted from Murray's table 1. "*NPT in developing a complex intervention*." (Murray et al., 2010) The NPT constructs satisfy the multi-dimensionality of case-note reviews which was used across multiple specialties and both clinical and non-clinical staff and thus can be captured by NPT's coherence, cognitive

participation, collective action, reflexive monitoring constructs (see Appendix 17). A further fifth construct for organisation culture was added to further reflect the significance of organisational context above and beyond what NPT theory fails to scope over; it is acknowledged by the key developer of NPT, Carl May, that it does not provide a theory of organisational culture with its concomitant structure and behaviour. This is because NPT is an implementation theory and does not assume any causal roles from organisational entities.(May et al., 2016)

NPT was selected as the framework to guide the development of the semi-structured interview questions. This was done because the embedding of an intervention is the primary goal of any complex intervention; if the intervention is not embedded and widely adopted, then it is meaningless to say has contributed to QI. By adopting this NPT approach in this single case study, there are likely to be generalisable lessons for understanding the embedding of mortality case-note reviews from this single-case study. The single case study was chosen. How extant theories from the literature conceive CNR and how it is conceived by those who use it may differ, and so any discrepancies between these must both be reconciled and understood in the rapprochement process. It would contribute to a more realistic understanding of the barriers and facilitators for the key components (embeddings, information flow and contributions to QI) towards QI from case-note reviews. There has been no literature that has expanded on this aspect of mortality case-note reviews, and it would be auspicious for researchers and clinicians to confer further on this topic given the highly topical nature of mortality case-note reviews and the recent national mandate to learn from these types of reviews.

This case study is likely to realistically represent the intentions of users of MCNR and help researchers to understand how MCNR works (or falters) in the naturalistic hospital setting. Quality improvement concepts could benefit from a deeper understanding of how case-note review is operationalised and used; there stand to be benefits for understanding similar interventions as NPT is a generic intervention embedding theory that scopes over many quality improvement type interventions such as incident reporting and chart review-type interventions which serve as potential smoke-detectors alerting hospital staff of any untoward care. Thus, it is generalisable to other quality improvement concepts as it assumes similar MUSIQ models which comprehensively considers elements involved in quality improvement.

4.3.7. Framework Method

Using an analytical framework called the "framework method", a set of codes were organised into categories that had been jointly developed by researchers (AT and EF) involved in analysis that were used to manage and organise the data.(Spencer et al., 2004) The framework created a new structure for the data (rather than a complete account given by participants) that is helpful to reduce the dimensionality of the data in a way that can begin to answer the research questions. The process of deriving themes from the framework method is given in "Coding and analysis" of the Methods section (see Appendix 18 for more information on the Framework method theory and process).

4.3.8. Analytic generalizability

An English hospital, from an acute National Health Service Trust, was selected based on the size (bed capacity of the main acute hospital), hospital service, geographical characteristics,

and receptivity to our approach.(Appendix 19) Within this, the source of data collection was solely from the central hospital site, where interviews were also predominantly conducted. This moderately-sized hospital was situated in a rural-urban setting, typical of most non-teaching hospitals. Our findings are analytically generalizable, in other words this single case-study setting provides rich, contextual data from in-depth semi-structured interviews which identify factors otherwise unobtainable from quantitative methods e.g., questionnaires. Purposive sampling was adopted which is a method of sampling where the researcher deliberately chooses who to include in the study based on their ability to provide necessary data.(Etikan et al., 2016) The rationale for choosing this approach was that the researcher was seeking knowledge about staff knowledge and perceptions of case-note review process which the participants would provide by virtue of their direct experience. The researcher worked in conjunction with two senior clinical leaders, in choosing participants, based on the proximity of their involvement in case-note reviewing and the processing of any associated information. There were 25 potential participants, of whom twenty-four participated in the study. No participants were unavailable for interview. In-person interviews were generally conducted and recorded in quiet, neutral locations with participants not subject to any danger, intimidation, or coercion.

4.4. Setting and recruitment

Hospital ethical clearance was granted for this service evaluation which is found in Appendix 20. This NHS acute hospital provided a range of hospital and community health services to 220,000 urban and rural populations.(Yin, 2017) This research/service evaluation was granted ethical approval by the participating hospital. Patients and the public were not involved in this research. To produce reliable and meaningful results, the researcher-

interviewer met two key informants, who were senior clinicians and managers, to carefully plan and select the appropriate research topics to then help develop the semi-structured interview questions. This helped to pre-identify set themes to explore which could form the basis for the corroboration via triangulation of codes and themes from interviewee transcripts. For instance, the informants noted the importance of case-note review for learning from deaths but that this might compete with their frontline clinical work with patients. And so, the questions were suitably developed to identify competing concerns reviewers may have around their limited time to review. The informants were provided with more details in-person and how and what case-note review related information circulated within the hospital; this helped to focus the interview questions on the content of these documents with case-note review information. All this was done to help constrain the scope of the qualitative interview data and so help to raise the likelihood of successful triangulation of case-note reviews; this information guided the development of the semi-structured interview questions and to help achieve some commensurability between interview data and document data.

At the interviews, each interviewee received a consent form, an information sheet, and a copy of the list of semi-structured interview questions for their future reference; they were told at the interview to read through all the information given to them before commencing with the interview. The information sheet contained important information on the study context and the main research questions of the researcher's PhD thesis, which was intended to allow the interviewee to have a better understanding of the researcher's focus on case-note review as a source for quality improvement.

Twenty-four participants were interviewed with seven having repeat interviews to clarify details or semantics from the collected data. Repeat interviews, phone calls and email

correspondence were used to clarify any further points discussed. One non-participant was approached but not interviewed because of review non-involvement. All interviews were conducted with participants on the hospital site. Interview length ranged between 28 and 70 minutes. Upon interviewee request, six transcripts were returned to the researcher following interviewer consent. Due to the semi-structured, focussed nature of these questions, this has helped to constrain the topic area and promote our ability to triangulate across interviews and documents. Where required, the interviewer developed lines of inquiry pertinent to their overall research questions.

In high-income countries (HIC), the MCNR process typically involves nurses and doctors as case-note reviewers with additional support from quality officers, managers, and Trust board members. The selection of the NHS Trust was considered by the following criteria of likely learning opportunities from the collaboration, the Trust size, the Trust's receptivity to our engagement and the turnaround time for ethical procedures. We interviewed two key senior clinical managers as informants, both doctors, to initially identify participants involved in the MCNR process.(Yin, 1981)

4.4.1. Pilot Test

A pilot-study was conducted with 2 senior clinical managers interviewed at their workplace. The interview was audio-recorded to ensure reliable data collection. The five basic components of NPT were used to inform these pilot test interview questions. During the exercise, attention was given to body language and non-verbal responses and the manner of asking questions. As the researcher was the main data collection instrument, the pilot-study provided an insight into the phenomenon studied and increased the researcher's exposure to

the research setting and whilst also raising the experience in interviewing whilst enhancing interpersonal skills. Also, errors in interviewing skills were rectified and not repeated in the main study. The pilot interview questions were changed to help answer the research questions (see Appendix 21 for pilot interview question template).

4.4.2. Formal Case Study Recruitment

At the outset, the informants were engaged directly with senior clinical leaders in the hospital, who helped to initiate the case study. These participants were contacted by email and recruited into our case-study using a snowball participant selection process.(Coleman, 1958) (Appendix 22: consent and participant information forms) Participants were selected for their ability to provide information to answer the research questions. Participants were interviewed and documents collected in the hospital (see Table 17).

After obtaining written consent, the semi-structured interviews (SSI) were securely recorded on voice recorder. Interview questions were developed from NPT. The lead investigator (AT) conducted all interviews between June 2018 and April 2019 with the consistent reflexive monitoring and discussions with senior team members. Each interview lasted between 30-70 minutes and were transcribed verbatim by AT; when required, participants were repeat interviewed to obtain more information. Post-interview, participants were invited to receive a typed transcript of the interview and feedback to researchers. Participant feedback was received and shared with the research team to determine any further required actions. AT kept a log of handwritten “field” notes during and after interviews.

The details of the collected documents are provided in Appendix 23. Prior to interview commencement, each received AT's academic credentials with physical handouts of the research aims and his contact details. Only AT conducted interviews with participants using the Interview Questions from Appendix 21. However, participants were allowed the freedom to comfortably talk about their experiences beyond the discrete interview topics. All doctors had reviewed case-notes. Standard procedures for semi-structured interviews and their analysis were used with field notes made and transcripts returned, if willing (Appendix 24: COREQ checklist).

Table 17. Participants and Document sources

PARTICIPANTS		
Hospital group	Number of participants	Participant details
Board Member	4	Three board members were clinicians. Two boards members were senior doctors.
Support Officer	4	Three safety officers were clinicians. Two were case-note reviewers. Two safety officers were nurses.
Doctor	9	All doctors were case-note reviewers. Nine different specialties were represented across these nine doctors from emergency medicine to anaesthetics.
Nurse	7	One nurse was a board member. Two nurses were dedicated case-note reviewers with non-clinical duties.
DOCUMENTS		
	Number of Documents	Time periods
Annual General Meeting	2	July 2018, March 2019
Board of Directors Meeting Agenda	10	Jan - Oct 2018

Board Meetings	9	Jan - Oct 2018
Performance Dashboard	10	Jan - Oct 2018
Board Conflict of Interests	2	April - May 2018
Mortality Surveillance Committee	3	January 2018 – June 2019
MCNR Policy	1	Aug 2017- Aug 2022
Patient Safety Newsletter	4	February 2018 – February 2019
Annual Quality Accounts	3	2017-2019 inclusive
Specialty-specific documents (i.e., gastroenterology, general medicine)	7	

KEY: B=Board member, D=Doctor, N=Nurse, SO=Safety Officer, *telephone-interview (all others face-to-face)

4.5. Coding and analysis

Coding analysis was undertaken alongside conducting the semi-structured interviews and the collection of document data. Data were managed using NVivo 11 qualitative analysis software. A '*framework method*' guided thematic analysis identified codes and key themes from data.(Fereday and Muir-Cochrane, 2006) This framework method employed several steps.

4.5.1. Implementation of the framework method

Step One: The interviews were all transcribed by AT with the first four, randomly selected, which were then vetted by a senior researcher and EF in the research team for fidelity to the original voice-recording; the structure of participant responses (i.e., length of pause, laughing, nervousness etc.) were included in these first three transcriptions, but were abandoned as they did not significantly contribute to the research questions.

Step Two: AT became familiar with these interview transcripts having transcribed them, and EF with all transcripts through having heard voice-recordings and seen them. It was from this that AT and EF independently coded these transcripts. This allowed AT and EF to familiarise with the interview transcripts.

Step Three: The transcripts were independently coded by AT and EF in groups with initial impressions placed into the margins for further references to help later transcript reconciliation. Four transcripts were randomly chosen. Coding was given in the left-hand margin of the word document with the transcript positioned-central and notes and ideas positioned on the right-hand side.

Step Four: After the four transcripts has been coded, EF and AT convened to discuss the labels assigned to each passage; the same passages were mostly highlighted, however, there could be different interpretations of the content and so discussion is needed. After discussion, there is an agreed set of codes, and each with its definition, which would go on to form the analytical framework. AT and EF would continue to code three additional transcripts using this initial framework and carefully note the emergence of any new codes. The process of refining and applying the analytical framework was repeated until no new codes were generated.

Step five: The final analytical framework was then applied to all the interview transcripts (including those which has already been coded beforehand but not with the final analytical framework matrix) in the QSR package NVivo 11. These indexed scripts were shared with EF and the wider research team.

Step six: The data is charted into the final analytical framework and is then summarized into a matrix form. For convenience, the first-level arrangement is by NPT construct to which individuals with their quotes can be my multiply instantiated. This is done to preserve the original NPT constructs to have its order preserved rather than have a large and sparsely populated grid with more than 20 rows (interview participants) and more than a dozen columns (themes).

Step seven: The data was interpreted with the themes defined by the research team with codes relating to it, with a summary of the raw data (interview and document with representative instances), a discussion of any deviant cases, and further points for consideration and comparison. To help guide data interpretation, these were the sub-headings for each of the themes (to demonstrate an example theme of "duty" is presented in Appendix 25: The "Duty" theme as expressed under the Framework Theory. Overall the qualitative researchers (An Te and Erica Ferris) and the research team discussed the themes using an inductive framework

method, a common method to help structure the identification of codes from the data(Ritchie et al., 2013), to then agree a final framework of themes from these codes (Appendix 26: Inductive labels), which were interpreted by each theme.

A triangulation protocol (Farmer et al., 2006) (Denzin, 1978) was used to identify convergence and divergences between key codes, which is found in Appendix 27. These research activities align well with critical realism as we both assume that the personal knowledge of "phenomena" can be multiply instantiated through interviews and documents and so can come to confirm or disconfirm hypotheses. The overlap of instances, or "codes and themes" as we refer to them in this section, raises the likelihood that a social reality exists apart from our own awareness. Codes were regularly brought to the research team and regularly reviewed to discuss when data saturation had been achieved(Faulkner and Trotter, 2017) it was decided that there would be no further need for data collection. This framework was re-applied to interpret all data 'information flow' and 'contribution to QI' objectives. We endeavoured to explore case-note review's embedding in the hospital. During repeat interviews, preliminary results were shared with participants for their reflection. Full qualitative coding details of NPT, information flow and quality improvement domains are recorded in Appendix 28, Appendix 29 and Appendix 30, respectively. NPT constructs are intended to guide the investigators to explore the embedding of mortality case-note reviews in this Trust using interviews, documents and investigator notes and annotations. Information flow refers to the extent to which information contained and/or derived from case-note reviews are found in other places within the organisation; our approach to measure this is to use the interview content, collect documents from mortality and patient safety meetings most pertinent to the mortality case-note review discussion and to then bring all this together to

broadly determine the overlap and disconnections between these two sources.[41] It is from this that an information flow diagram is derived. The contribution of quality improvement from case-note review is broadly focussed on the instances and the types of clinical episodes, which can be discrete and singular or grouped at a higher level, which indicate that learning has been found from these case-notes sourced as evidenced across the interviews and documents. For this, there is no formal diagram for information flow. However, the findings will be discussed and explored further as they arise from the data.

4.5.1. Ethical considerations

This study posed zero-to-minimal risk to participants. The probability and magnitude of harm or discomfort anticipated in the research will not be greater than any experience not ordinarily encountered in daily life, or during the performance of routine physical or psychological examinations or tests. Participants fully understood the nature of the study and the fact that participation is voluntary. For example, all data collected was anonymised by replacing the participants' names with alpha-numeric identifiers (D1, N1 etc). Participants

[41] It must be noted that information may flow up-down-up and down-up-down the organisation. Such things are observable if trackers were placed on discrete episodes in real-time. As investigators of quality improvement, we sought to broadly characterise the information flow from mortality case-note reviews from ward-to-board and separately, the board-to-ward, and its disconnections. As this is a complex social context, a higher-level explanation is both helpful and convenient for our purpose of elaborating clear description and theory about the flow of case-note review information in this Trust.

were informed that they could withdraw from the study at any point during or after the interview.

4.5.2. Researcher Reflexivity

Throughout the research process, I have interacted with informants or participants or documents and there is the possibility that this could have influenced the methodology or findings of the study. It is impossible to completely distance oneself across all extraneous interactions. Hence, I need to state my 'position' in relation to the research topic which is described as 'reflexivity'. Reflexivity involves a self-scrutiny on the part of the researcher; a self-conscious awareness of the relationship between the researcher and an 'other'.(Bourke, 2014) In short, it is a process of continuous self-analysis in which I reflect more deeply on the experiences I encountered when undertaking the research.

The researcher is a young and single male domiciled with parents; he worked in the public heath profession. At the time, he is a co-contributor to the running of the household and assists in maintaining the life at home. He is religious and holds strongly to his cultural roots but is sympathetic to a broad-spectrum of views. The researcher acknowledged that he could identify with the participants as he made sure that he did not impose his values or opinions on the participants during the interviews. The researcher was able to put aside his own understanding of the subject of investigation and open his mind to attentively perceive and document what was communicated to him by the participants.

4.6. Results

In 2017-18 and 2018-19, according to the Trust's annual report and accounts from a total of 749 case-note reviews (including MCNR), across all specialities four deaths were formally attributed to sub-standard care through case-note reviews (see Table 18), which were collected through an on-line mortality review e-proforma (Appendix 31). In 2019-20 from 244 case-note reviews (including MCNR), one death across all specialities was attributable from case-note reviews. Reviews were expected by a nurse board member to yield QI from "*ten-to-five percent*" of all MCNRs. There has been some QI benefit from case-note reviews in terms of capacity building infrastructure and the improvement of attitudes and culture towards QI, with all investigative approaches applied and information sources collected. However, there is some confirming evidence that MCNRs directly contribute to QI. In terms of quality improvement techniques and methods, there were no mentions in the data of any improvement methodologies in relation to MCNR (e.g., lean, Six Sigma, PDSA). Importantly, despite the distinction between CNR and MCNR in this thesis, all interviewees confirmed (and re-confirmed through emails where necessary) that they were alluding to MCNR in their interviews. This does not complicate matters but simply provided more data to allow the investigators to understand the QI potential of MCNR, a national policy-driven initiative, in context.

Table 18. Trust case-note reviews by annum

YEAR	QUARTER 1	QUARTER 2	QUARTER 3	QUARTER 4	CASE-NOTE REVIEWS TOTAL	TOTAL IN-HOSPITAL DEATHS	CASE-NOTE REVIEW ATTRIBUTED TO SUB-STANDARD CARE
2017-18	148	106	105	59	391	842	2
2018-19	118	79	119	42 (excl. Mar)	358	745	2
2019-20	92	90	49	13	244	784	1

4.6.1. Embedding of mortality case-note review

To achieve hospital QI from an intervention, that intervention must be first embedded in the hospital. For instance, a well-embedded intervention is when an intervention has become part of the daily running of the hospital and does not continue to disrupt the running of the hospital. The NPT constructs of coherence, reflexive monitoring, collective action, cognitive participation with an extra organisational construct (to consider organisational features) were consistently represented across participants. Coherence is concerned with how the work is conceptualised by those using the intervention. Reflexive monitoring is concerned with new ways of working – or attempts to introduce them – and how these are interpreted by those involved. Furthermore, evaluation of these new ways of working is important to consider. Collective action is described as '*how the work is defined and organised in enacting a practise.*' And cognitive participation concern those who do the work and how they decide whether to engage and hope to achieve.

Coherence was lacking around the purpose of case-note reviews which was reflected in their diverse use, however, there was a consistent emphasis on these reviews not being of significant use for quality improvement. Reflexive monitoring was inconsistently represented with MCNR consistently well-received, but case-note reviewer workload hindered its undertaking. For collective action, reviews were being undertaken, however, CNRs were applied using tools such as the "*global trigger tool*", as one doctor indicated, which supporting specialty documents corroborated. The Global trigger tool is a methodology for retrospective review of patient records which uses triggers or "clues" to identify possible AEs. The tool includes a list of known adverse event triggers as well as instructions for selecting records, training information, and appendices with references and common

questions.(Institute for Health Improvement (IHI), 2009) Collectively, there were issues around the timeliness of MCNR because of its lengthy processing time (~8 weeks). Organisationally, case-note reviews were favoured by some specialties, but not others. We reflect on these results.

We found that MCNRs were well-embedded and used alongside CNRs. Appendix 28 summarises quotes for the NPT constructs. Undertaking case-note reviews was personally expressed as:

> *"It's an extra duty that they have to do. I think it's done to fit in around some of their other priorities. So, if they've got a bit of time at lunchtime or a bit of free time at the end of the day, they might do some" senior consultant orthopaedic surgeon (D4).*

For this same surgeon, it was simultaneously a "*professional duty*" with the tension evident between their preference not to review and their professional duty towards the patient. Reviews were done to "*learn from mistakes*" but this orthopaedic surgeon indicated that much of the time "*there is nothing*" to learn about. A senior quality manager indicated that their purpose was to identify "*deteriorating patients*" amongst other types of patients (SO2). And this was noted by a doctor, practising in general medicine, with a typical delay of "*around eight weeks*" to fully prepare the case-notes for the reviewer. (D5, Consultant Acute/General Medicine Doctor) A nurse case-note reviewer stated that "*something else*" usually gained priority over these reviews (N1). Case-note reviewers had different reasons for their involvement with longstanding non-executive director (and vice-chairman) suggesting that MCNRs were a good activity because it encouraged more "*multi-disciplinary*" work. A senior surgeon positively affirmed reviews as being able to clearly "*show (to clinicians)*...

what good quality is" (D12). Furthermore, a senior safety officer that clinicians were "*keen to review*" especially if a death was unexpected (SO1). A senior nurse, who also was a long-term board member, observed a change in the MCNR purpose which ranged from initially finding quality of care "*failings*" with it then later becoming a tool for understanding the "*qualitative*" aspects of a good death (N4). Organisationally, the senior quality manager indicated that there were diverse "*insights.*" But, a senior clinician, who is also a board member, expressed that "*new learning from [mortality-review] cases are rare.*" And as an organisation, being a senior board member, who was an ex-nurse, expressed that being busy was a difficult situation to be in as:

> *"It is a competing demand on people who are very busy, and I think that goes throughout the organisation if you're talking about consultants or senior nurses or patient safety team, members of the board, a lot of pressure on their time, it does impact" (SO2, patient safety manager)*

which was the main reason for not completing case-note reviews in the first place.

4.6.2. Information flow from case-note review

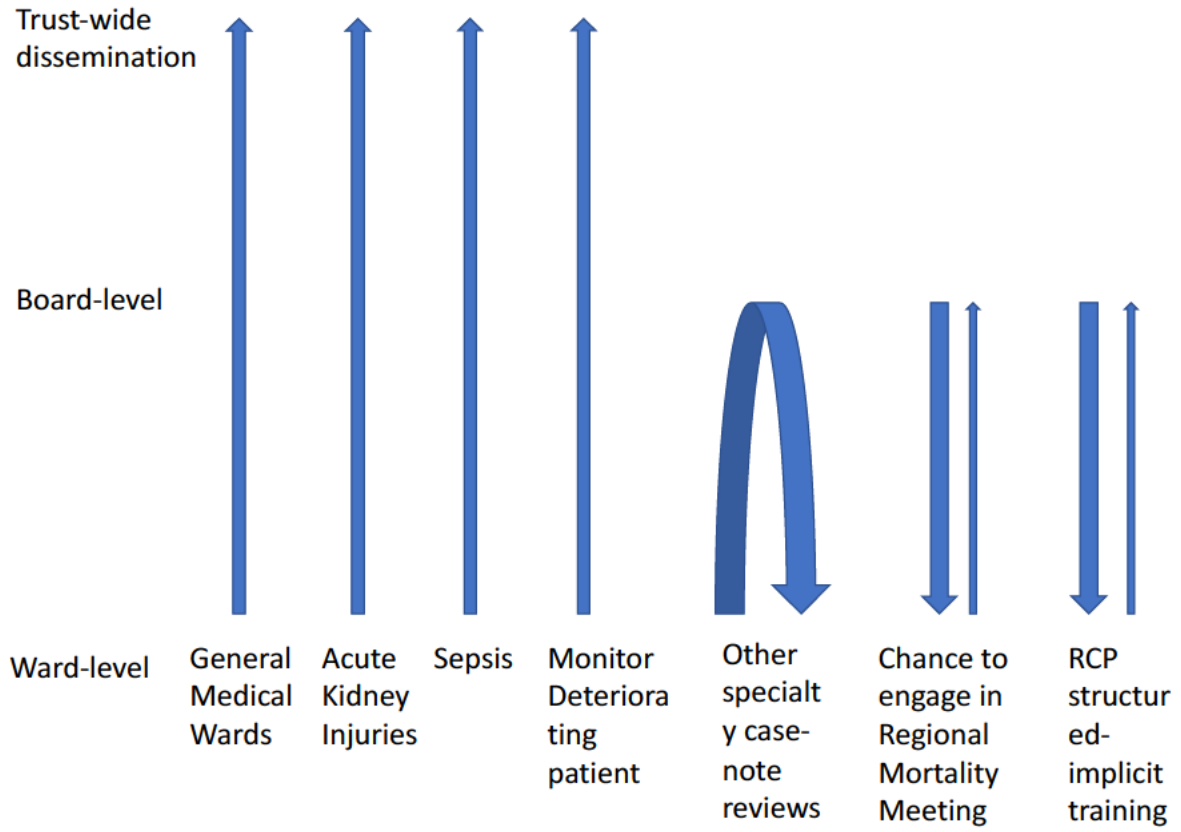
There can be no quality improvement from an intervention if no information is flowing out of this intervention. In terms of items flowing from the board-to-ward, these include the Board of Directors meeting agendas, their minutes, board conflict of interests, the annual quality accounts, and performance dashboards. In terms of the items flowing from the ward-to-board, these include completed MCNR review proformas, quota of MCNRs undertaken per fiscal quarter, electronic Datix forms, patient complaints and specialty-specific documents. If information flow is impeded, it can negatively influence the quality of care by inhibiting QI through the organisation. The inductive themes are presented without their association to the

NPT constructs, from which the questions were originally derived. NPT was essential for us to establish the intervention's embedding. But once this is known then further themes on dependent sections (i.e., information flow and contributions to quality improvement) supervene on the intervention being embedded. In other words, one must have case-note review embedding before one can have any information flow or quality improvement; the latter two are dependent on the former. This dependency is assumed throughout Chapter 4 .

Information flow was bi-directional (Figure 14), in other words, the information derived from case-note reviews flows from either the board-to-ward or the ward-to-board. Examples of information flow from the ward-to-board include the case-note review information from mortality case-note reviews in general medicine were more widely presented and discussed at the Mortality Surveillance Committee (MSC), a board-level platform for discussing mortality, and across the Trust as a whole, with much of the Trust being recipients of this information. Another instance was when acute kidney injuries (AKIs) were detected through a series of mortality case-note reviews five years ago, which has since led to a keen sense of purpose, ownership, and openness on issues especially amongst the anaesthetists in the intensive care unit (ITU). These anaesthetists have openly shared their findings to the MSC and to the patient safety team which has been aired through the patient safety letters. But overall, very little of the total case-note review information makes its way to the board. For information flow from the board-to-ward that was informed by the ward initially, there was a concerted encouragement from the board to promote sepsis awareness, AKI workups and packages to help support preventative programs for this condition and the need to better monitor patients who quickly deteriorated in their first 24 hours of admission to hospital. There were independent board decisions to share the mandate to review deaths with all their specialties which released by the National Quality Board in 2017. There was also a call in

2016 to all practising case-note reviewers, for at least one reviewer from each specialty, to attend structured implicit review training provided by the Royal College of Physicians. And there has also been information provided through the board to the ward about the work occurring at the Regional Mortality Meetings, which in brief allowed those who were innovators and champions of patient safety to present any compelling patient safety findings from their own daily clinical practise; the response from the ward level has been lukewarm with not all specialties having had those trained in the structured-implicit case-note review methodology. It is worth adding that the lack of presence for some members (i.e., the lack of quorum to satisfy the gathering of members and invited guests to a monthly MSC) not a serious enough issue to be flagged as a cause for concern for a handful of the monthly MSC meetings (in the year 2018) was. In brief, the lack of concern over absence of key individuals in the MSC likely suggested a lukewarm attitude to the NQB MCNR policy mandate.

Figure 14. Dynamics of information flow and its sources and who ultimately receives this information.



It evident from Figure 14 that the information flow within certain specialties of their specific-types of medical issues is better throughout the organisation than compared to the flow of information from case-note reviews. This is the first time these reviews have been studied in this way and warrants further exploration through both qualitative and quantitative means.

Figure 14. outlines the flow of case-note information from the source, which lies at the ward-level, and travels upwards towards the board and beyond to the wider Trust. There were four key streams identified where there was uninterrupted flow of the case-note from the ward level via the board to the wider Trust, with these being the General Medical Wards, cases pertaining to Acute Kidney injuries (AKI), sepsis patients and the monitoring of deteriorating patients. In contradistinction, the information flow of other specialties and topics did go to the board level but then largely remained at the ward-level. Regarding the representation of

case-notes at the regional mortality meetings, the wards had few opportunities to attend these mortality meetings hosted by Health Education England to spread their intramural learning obtained from their case-note reviews. It was evident that these regional meetings had more learning passed down to this Trust (and other regional Trusts). The question remains as to how the wards can be given the voice to share their learning at these regional meetings. And the topic of RCP structured-implicit training, furthest to the right on the figure, represents the strong emphasis placed on this training from the board-level for the training to be completed and implemented. However, the information flow at the ward-level of the RCP review form is poor which represents the poor uptake and implementation at this level.

Our findings reveal that there are four main factors which facilitate information flow from case-note reviews. First, a senior quality manager shared that:

> *"There is a Trust policy which is available to the public and everybody else on the public facing page of the Trust, internet page. That sets how we will address and look at mortality across the organisation." (SO2, Quality Safety Manager)*

Second, a senior quality manager expressed that external mortality policy mandates (Trust, 2017, National Quality Board UK, 2017) shaped hospital policy and practise with concomitant effects on information flow of case-note reviews and is now "*available to the public and everybody else on the public-facing page of the Trust*" (SO2). Third, a senior quality manager indicated that existing QI tools such as the global trigger tool and case-note reviews were online or "*paper*" for all reviewers (Institute of Healthcare Improvement (IHI), 2008, Hogan et al., 2012b, Hogan et al., 2014) that:

"*There are a number of streams of work which are happening in this place at the moment. If we start with mortality reviews, per se, for the adult in-patient group and the child in-patient group. The deaths are recorded within the bereavement office and the information team, and the clinical coding team review those notes in terms of coding. Then the quality team then get a list of those deaths from the information team and the associate medical director (AMD) for governance distributes those deaths out across the organisation for peer review*" (SO2, Senior Quality Manager).

Third, a Patient Safety Officer indicated that administrative staff would both "*physically get the notes to the consultants*" and also "*chase these notes.*" The administrators would also help to circulate patient safety "*newsletters*" and take minutes in meetings. An information flow barrier was noted by a Consultant in Acute Medicine noted that "*I would say the only information you get would the one that went downstream from mortality surveillance committee; the information goes from the department to the MSC and then to the mortality newsletter. It is not something that goes across from department to department. I cannot access [as a consultant] the learning points from other departments but as governance lead I can. One of my peers will not be able to access the mortality reviews, of say, cardiology*." (D5, Consultant Acute Medicine Doctor)

Barrier factors include poor administration, with a nurse reviewer stating that case-notes were not "*being scanned in any sensible order at all*" by administrators. The poor administration of case-notes misplaced proved "*quite time-consuming and frustrating*" to locate for a patient safety officer working closely with clinicians. Significantly, the culture seemed to affect the information flow. A blame culture of "*finger-pointing*" was expressed as a potential influencer on information flow where people became more reticent to share their learning

within the organisation, but there was no tangible evidence of this culture in this case-study. And the low number of completed reviews due to MCNR's…

> *"time-consuming and cognitively demanding… compete with existing... I think it is important, but I think when you have a certain amount of time to deliver, you tend to deliver to the living rather than the dead." (D11, Consultant Cardiologist)*

This lowers the information available with the concomitant effects for information flow. It must be stated that the adherence to the critical realist approach helped us to further explore the possible reasons for our findings; for instance, indicators of information flow were obtained but we went further and tried to identify this through multiple methods and means (i.e., semi-structured interviews, documents and emails) to establish where exactly information flow channels were open, and where they were less open, and importantly, why this was occurring at all (Appendix 32). This critical realist stance gave us the framework and help poise us to consider the influence of other underlying factors and other reasons for the results we had obtained. For instance, the specialty itself was a unit of case-note review information as often information would not be shared with other specialties or even the mortality surveillance committee. A Patient Safety Officer, who had close involvement with all case-note reviews, shared a likely reason that it was because specialties reviewed "*slightly differently*" to each other and occasionally had incommensurable purposes which were bespoke to their own specialty "*because they do what fits for them [their own specialty]*" (SO1, Patient Safety Officer). When administration was poor, information flow was less good as indicated by the number of documents and mentions of a key event, for instance learning from a particular subject area was apparent from emergency medicine than it was for general medical wards. But it was not known precisely whether this was because of the lack of

motivation or lack of resourcing or whether it was something peculiar to the specialty itself. The lack of reviewer time and their emphasis on conducting reviews meant that less information flowed from case-note reviews; each specialty appreciated the importance of reflecting on care using case-note reviews, but it was more common than not that those other competing clinical reasons were cited as a reason for not undertaking reviews.

4.6.3. Facilitating and inhibitive factors perceived to influence QI arising from mortality case-note reviews

We present the perceived facilitative and inhibitive factors influencing QI from these reviews. We did not set out to provide direct evidence based on a positivistic understanding of cause and effect, but we discussed complex social interactions from interview and document data's which were conducive (and were not) for QI from MCNRs. For clarity, we dichotomise these factors into facilitators and inhibitors. Within these, we applied three discrete domains: systems, organisational culture, and individual HCP/staff factors. These domains are sourced from clinical handover research. Clinical handover shares many task-specific features with case-note reviews[42] and so was considered an appropriate filter through which MCNR can be meaningfully discussed. MNCR, like clinical handover, have information flow, culture and staff factors as essential factors.(Australian Council for Safety Quality in Health Care, 2005, Frankel et al., 2012, Turner et al., 2006) Systems factors were concerned with hospital level processes(Benson and Harp, 1994, Australian Council for Safety Quality in Health Care, 2005), organisational culture with Trust specialty and organisational cultures.(Ibrahim et al., 2019, Suserud and Bruce, 2003) Individual HCP/staff factors were concerned with specific individual-level phenomena modifying case-note review derived QI. (Appendix 31)

[42] This includes information management, systems to store the medical record/case-notes, staff to convey the information contained in the notes, the safe and effective storage of medical records, the need for clinical judgement and decision-making.

4.6.3.1. System factors

System factors are comprised of sub-systems operating at different levels of the health system (e.g., outside of the hospital, hospital and/or program level, point of care) each with specific goals, resources (human, financial, equipment) and processes (formal and informal.) It is important to identify how they influence each other to strengthen patient safety and incident management in health systems.(Gluyas, 2018)

4.6.3.2. National Care Quality Board "learning from deaths" Mandate

Several participants alluded to England's National Care Quality Board's policy with a nurse seconded to "*writ[e] the mortality case-note review policy.*" A doctor and a board member acknowledged the hospital did "*follow*" this "*board*" policy (see Appendix 31).

4.6.3.3. Case-note reviews facilitated through Medical Examiners model

The Medical Examiner system, as of 2019, is a live initiative in England and Wales which assigns training to physicians who then independently scrutinise each hospital death by recording their own view of the cause of death through a rapid case-note review (distinct from MCNR), escalating concerns with patterns and trends with Medical Examiner office collaboration, supporting the bereaved and thereby being an arbitrator who escalates accordingly with special cases referred for case-note review.(The Royal College of Pathologists, 2016) A doctor alluded to the "*possibility"* that *"including a Medical Examiner model*" had helped to embed case-note reviews and establish a more co-ordinated case-note review programme. One safety officer opined that "*with the Medical Examiner [system] coming up… that's the thing driving the [need to reform the case-note review] process at the moment*", as one of the 'institutional' forces driving the case-note review adoption. Supported

by mortality surveillance documents, the Medical Examiner system was perceived by a safety officer as a "*national driver*" which should "*re-assure*" policymakers of its complementarity with case-note review programmes.

4.6.3.4. Patient Advice Liaison Service (PALS)

In 2008, PALS was shown to enhance culture change, service improvement, patient experience and help hospitals deliver QI.(Evans et al., 2008) For instance, a safety officer recalled when PALS had co-ordinated mortality case-note review information "*so that was picked up after PALS/bereavement,… And we pick up things from PALS, bereavement team as well.*" A doctor noted that PALS often helped "*feedback [case-note review information] to (those) other specialties.*"

4.6.3.5. Continued professional development (CPD)/validation requiring MCNR

Participants considered case-note reviews a normal part of their daily work. A doctor and board member noted that all forms of case-note reviews were "*surely part of*" continued professional development (CPD) with a safety officer indicating that part of "*appraisals and re-validations...*" for "*appropriate specialties*" (SO3) normalised case-note reviews.

4.6.3.6. National Programmes using Case-note Review

There were various national programmes facilitating specialty specific case-note reviews. For example, the international Surviving Sepsis Campaign (SSC) has sought to raise awareness of the burgeoning sepsis problem in hospitals since 2016.(Rhodes et al., 2017, Slade et al., 2003b, Slade et al., 2003a, Willson et al., 2008, Heyworth et al., 2009, Levy et al., 2010) This awareness appears to have percolated to this Trust as sepsis was a focal area during case-note

reviews. Thus, case-note reviews and their use has been partly shaped by the legacy of sepsis with one safety officer indicating that "*the case-note reviews have helped support … a particular drop in (annual) sepsis (mortality).*" A doctor and board member noted that "*other (non-sepsis) reviews*" yielded less learning than from "*deteriorating patient groups around sepsis*" of which a nurse indicated there had been "*quite a lot of learning*" from. Bespoke "*sepsis care bundles*" have been used during case-note reviews to identify lapses in care.(Vorwerk et al., 2009, Cronshaw et al., 2011) Sepsis was mentioned by >10 participants to be informed by case-note reviews. The greatest change in their approach has been the greater vigilance of the initial health status of the patient at admission and, with it, the monitoring of more susceptible and fragile patients which are likely to then deteriorate in their health condition rapidly. Unlike before when they were not cogniscient of a patient deteriorating in real-time, the clinicians who had reviewed the case-notes of rapidly deteriorating patients came to understand the typical hallmarks of a patient and thus come to also understand why they came to deteriorate so rapidly. For instance, the Trust did not previously have care bundles to spotlight these patients, but case-note reviews have specifically paved the way for the development of acute care bundles, of which there is one for sepsis, that has been implemented consistently and has overtly reduced the patient harm risk and their mortality from sepsis over the last 3 years. In short, the identification of deteriorating patients has improved significantly due to the insights obtained from a focussed mortality case-note reviews of such patients.

The most likely reason why case-note reviews were undertaken was because of the higher than usual rates of deteriorating patients and it was from this that sepsis patients, who should have had sepsis, then provided learning and insights into the review process. Though clinicians may have been exposed to sepsis as an issue from the SSC, the focus on sepsis and

the drop in mortality from sepsis was attributable a greater vigilance from the Trust's clinicians on rapidly deteriorating patients. It is from this change in outcome, here mortality rates from sepsis, that this precedent of focussing on sepsis as a programme that is closely informed by case-note reviews. There are checklists employed to prevent unnecessary patient harm from sepsis, however, most harm is retroactive and cannot be captured well in a metric but needs to be subject to a retrospective reviewed. Given the complex nature of care and the unreliability of clinician recall from memory, the case-notes are the best available information source to determine whether a sepsis diagnosis was preventable or not.

In addition, case-note reviews specifically fed into the Acute Kidney Injury (AKI) Programme (UK Renal Registry (UKRR), 2014) which was informed by UK guidance.(National Institute for Health and Care Excellence (NICE), 2007) Though there was a national programme for identifying and mitigating AKIs, it was case-note review activity which actually led to a diminution in mortality from AKIs. The lessons were themselves sourced from Trust investigations around the localised rise in mortality from rapidly deteriorating patients, much like with the sepsis learning from case-note reviews.

Case-note reviews complemented the aims of certain national audit and surveillance drives in certain specialties. For instance, the National Emergency Laparotomy Audit (NELA) was stated by a doctor and board member to be founded on "*collective responsibility*" to extract information from case-note reviews and other information sources to improve the quality of emergency laparotomy surgery.(Poulton et al., 2019) A cardiologist case-note reviewer stated hospital QI was enhanced by the "*national audit*" infrastructure, informed by case-note reviews, which improved mortality rates in cardiology "*compared to the mortality of patients*

elsewhere (from other specialties)." The documents did not mention the national programmes alluded to above.

4.6.3.7. Organizational Culture Factors

Organisational culture is a shared "*pattern of basic assumptions*" which group members have acquired over time as they continually learn to cope with internal and external organizationally relevant problems.(Schein, 1985) Given the complexity of the healthcare context, hospitals are best viewed as having multiple subcultures, which may be driving forces for change or may also, curiously, undermine quality improvement initiatives.(Mannion and Davies, 2018b) Organisational culture and outcomes have been shown to affect healthcare outcomes as shown by multiple studies done across many settings and countries.(Braithwaite et al., 2017) Thus, organisational culture in hospitals may have a bearing on healthcare quality of care.

4.6.3.8. Specialty learning culture from case-note reviews

Case-note reviews were perceived by participants as facilitating specialty learning culture. A board member and doctor noted certain cultures were supported by the "*mortality case-note review policy.*" A doctor and board member noted that an open *specialty culture* on QI was "*important*" as "*a culture where people (can) report incidents*" and where "*they (the specialties) can learn from them [case-note review].*" The medical and surgical specialties were "*in-grained*" to derive QI from any form of case-note review. While a safety officer indicated "*everyone [all specialties*]" used reviews (SO3). Contrarily, an emergency department doctor indicated these case-note reviews were "*not particularly useful*" in their specialty.

4.6.3.8. Specialty culture of case-note reviewer investment into their patients

The case-note reviewer culture related to the extent to which case-note reviewers felt involved and invested in their patient and their quality of care was noted as a driver for case-note review derived QI. A board member noted the case-note reviewer ownership of patients enabled QI from case-note reviews which "*make[s] it [,case-note review for QI,] work.*" A doctor indicated "*nephrologists really own their patients [and their case-notes]*" and so are "*really keen to learn from [their patient deaths] through case-note review.*"

The lack of ownership of case-note reviews was a barrier to review-led QI. A doctor noted that clinicians with poor "*ownership*" of their patients often suffered from poor communication with inevitable QI implications. A safety officer highlighted that the monthly MSC failed to meet for "*non-quoracy*" reasons for several months because, "*it [the Mortality Surveillance Committee][43] was not important.*"

4.6.3.9. Patient expectations driving quality improvement culture

On changing culture of patient expectations and litigation behaviours, a board member indicated the "*last 10-15 years of expectations*" have resulted in "*current*" unrealistic care quality standards. Patient expectations exceed the current healthcare performance, and which, in some way, modify clinician attitudes to case-note reviews. There is a sense in which they must communicate clearly to the patient's relatives the facts of the matter according to the

[43] The organisational platform where clinicians and board members and non-executive directors discuss matters relating to trust mortality and patient safety.

notes, but it must be noted that the high care quality expectations of patients can place a barrier for clinicians to be completely forthright about the care quality given.

4.6.3.10. Policy awareness

Some were aware of the hospital mortality policy. Two doctors, one also a board member, indicated "*there is mortality case-note review "policy" which the [hospital] does follow.*" A case-note ex-reviewer, now a board member, stressed case-note reviews helped drive "*policy*" led to QI "*work streams.*"

4.6.3.11. Limited knowledge and availability of training

There was mixed consensus around case-note reviews training opportunities. A nurse noted that case-note reviewers "*haven't received training*" because they "*didn't even know training was available.*" A safety officer and doctor stressed that the RCP training is "*limited*" to a number of people, mostly those who have had "*root cause training*", go on to become case-note reviewer trainers.

4.6.3.12. Competing processes

There was variation around the contribution of hospital interventions to national and local quality improvement. A board member indicated that the interventions were "*connected*" whilst, a doctor stated that, "*quality improvement outcomes were the same*" as other interventions. They perceived other QI interventions interfered with QI outputs from case-note reviews.

4.6.3.13. Specialty QI culture

The case-note specialty modifies any QI outputs. For instance, the specialty culture historically influences outcomes. (Karassavidou et al., 2011) This was also stressed by a doctor who expressed that:

"*I think some clinical groups by the nature of their work have had more mortality reviews and have had patient safety ingrained in their specialty... I'm thinking of anaesthetics perhaps. General surgery because of the nature of that specialty…*."

This indicates that not all specialties stand to benefit from MCNR in the same way, with general surgery and anaesthetics likely to benefit from lessons learnt than other specialties. This person also conveyed the importance of reviewing by expressing that:

"*I think some specialties have come a bit or has come a bit later (to heavily use case-note reviews) to, but I think everyone, I would hope, understand the process and the reasons behind it*."

Another safety officer noted that there was "*no consistency*" in case-note review method. Some interviewees specialties perceived more QI from case-note review for their own specialty than others. But it is difficult for individuals to make comparative judgements when they have not experienced all the specialties. But each specialty does differ in its QI outputs from case-note reviews and a comparison cannot be make because they are not like-for-like but incommensurable. A consultant emergency medicine doctor did reflect on the varying utility of reviews for specialties by saying that:

> *"The juniors I think do a very efficient thorough job of investigating a lot of cases that we learn nothing from in terms of the amount of depth that they go in for out of hospitals or cardiac arrests. I would say it's not*

particularly useful but actually they do a very good job, and they write lots of stuff about all of these patients who came in and probably had a cardiac arrest. I don't think the group of patients we are looking at is useful for (our department). I'm not saying the whole thing isn't effective. The way it's done here and the group of patients we focus on in (our department) is not useful."

An acute medicine nurse considered:

"Whether [case-note reviews] could be done quicker or differently" who also "noticed quite a lot that other specialties won't bring up the subject of what their plans were… what their plans were if their heart were to stop beating. There's not a lot of input from other specialties. Quite often end up being an acute medical physician that brings up the subject of respect form and plans for the death. That's what we've identified mostly."

This clearly shows furtiveness on some medical issues such as end-of-life care. This could be having repercussions for the under-representation of the potential end-of-life lessons learnt from MCNR for the Trust.

4.6.3.14. Busyness

Being busy was *the* main reason for not completing reviews. A doctor and board member indicated it was that reviewing is not appealing but that:

"*… it's not so much about incentive but that they have the time to do it. Yeah, we don't have the time within our busy schedule. We need some time to do it, that's all*."

A nurse case-note reviewer acknowledged the busyness but that it would be in the Trust and patient's interest to undertake these reviews:

"*I know everybody's really busy, and everyone's got targets and the rest of it, but we need to learn from what we do. It might make it easier to do what we do instead of saying 'I haven't got time for that.' In fact, we have to make time and learn and make it better through that process*."

4.6.3.15. Attitudes toward reviewing

A senior doctor and board member noted case-note review was "*onerous*" and did "*not find anything particularly new* with "*[any] new learning is relatively small*" implying reviews did not lead to substantial QI. An experienced doctor case-note reviewer commented, "*there is nothing [no learning]*" from the sickest patients. A young doctor remarked geriatric medicine "*doesn't seem to [have] that many lessons*" from "*the [relatively large number of] case-note reviews [compared to other specialties]*."

4.6.3.16. Attitudes toward training

A senior doctor and board member commented on the government "*minister's*" desire for case-note reviews as a benchmarking tool rather than its preferred use as "*smoke detector*" and so "*didn't attend the training.*" A doctor indicated that case-note review training was "*a lot of…common sense"* not compelling enough to attend. Conversely, a doctor indicated "*structured training*" was limited whilst others stressed the selectivity of training. This attitude reduces the coherence of reviewing as a QI tool.

4.6.3.17. Challenges toward dissemination of learning from case-note reviews

There is cultural inertia from specialties for information to "*stay within their division*" as discussed by a Patient Safety Officer who works closely with all specialties and their case-note reviews (S01). This is attributable to the fact that there are no clear systems in place which was indicated by an ex-nurse board member that:

> *"a couple of the acute medical wards where we were seeing a delay in escalation of patients who were having high NEWS scores and led us to looking at electronic observation systems and we didn't have one until that point. And that absolutely drove our commitment to drive the escalation of patient's observations." which led to a tangible "delay in the escalation of patients" (B3, Ex-nurse and board member)*

And so, any information from case-note reviews must have a corresponding system to share this information through. Another major problem plaguing case-note reviews was that much of the time, there was no tangible learning from case-note reviews; the signal is much too small and the system also insensitive enough to discern between relevant quality improvement information and non-relevant information. Another issue to overcome was about having a Trust-wide infrastructure for learning, which was important given that the mandate to review came from the board-level, which was able to receive the case-note review information from specialties which used these reviews to varying degrees; a patient safety officer indicated that:

> *"There's no consistency in how they do their reviews which in a way is good because they do what fits for them, for trauma and orthopaedics and endoscopy or A&E. A&E, when they do their deaths, they go through all*

their deaths and the questions they ask are slightly different to the other specialties. So that's kind of… I look at the proforma side of it whereas the specialties, they do an in-depth review with more of a discussion with their colleagues and see what their colleagues think and kind of get that peer review… which I think stays within the division" (SO1, Patient Safety Officer)

This lack of structure and purpose for reviewing from the top-level led to a variety of practical consequences for the reviewers. The most important consequence was perhaps the patchy exposure to training provided to reviewers. One consultant gastroenterologist indicated that he "*didn't even know training was available on it [case-note reviewing]*" (D7, Consultant Gastroenterologist and reviewer) and even scepticism about the real use of case-note reviews as a "*benchmarking tool and its publication as data*" (D1B2, Senior clinician and board member).

4.7. Discussion

MCNRs are well-embedded within the Trust with information flow most occurring downwards from board-to-ward rather than upwards, from ward-to-board. Though there were limited instances of learning from MCNR or case-note reviews, there was substantial learning for deteriorating patients, especially in the sepsis and acute kidney injury (AKI) areas. Some specialties obtained less benefit from the reviews than others and indicated their reasons for this. Compellingly, there were no mentions of other quality improvement techniques (i.e., lean, Six Sigma, PDSA) used in conjunction with mortality case-note reviews. This omission need not imply an absence of any QI techniques; however, it signifies that MCNR may not require a quality improvement technique or is not important for MCNR for QI purposes; it thus appears to standalone. Thus, the mandate to undertake MCNR is not universally beneficially for all hospital specialties, and thus needs to respect the specific healthcare conditions it is used in. We suggest that hospitals, within their specialties and existing clinical work programmes, develop systems to better record and collect process and outcome measures around MCNRs. This would help enable investigators to assess whether these reviews are a good source of QI within various clinical areas. The facilitators and barriers to QI are now explored.

4.7.1. Perceived facilitators and barriers to QI from mortality case-note reviews

We identified several factors facilitating or hindering MCNRs. We discuss the major factors leaving out the minor themes due to their lesser relevance.

In terms of systems, we found national mandates do support QI from case-note reviews such as in MCNR sepsis mandates.(Cooke and Iwashyna, 2014) The RCP national training programme have formally reported QI from case-note reviews.(The Royal College of

Physicians, 2018) However, this is likely influenced to some extent by cherry-picking and the omission of non-QI instances. Mandatory MCNR may yield specialty-level learning as this has been demonstrated already for CNR before the UK mandate to review became established and widely employed in hospitals.(Kobewka et al., 2017, Tiwari et al., 2020) The opportunities for MCNR learning are most apparent in specialties with complex patient presentations who are susceptible to rapid deterioration. Behind the scenes and in conjunct with MCNR, the Medical Examiner system is likely to have helped co-ordinate a system for case-note reviews such as international sepsis(Cooke and Iwashyna, 2014) and national new early warning systems.(Smith et al., 2013) Additionally, certain patient services considered patient voice with the Patient Advice and Liaison Service (PALS), which raised the profile of patient voice in conversations with relatives and patients around care quality.(Evans et al., 2008) Better engagement with relatives are another element which the Medical Examiner programme is expected to help with.

In terms of organizational culture, the hospital-wide and specialty culture do appear to influence QI.(Walshe and Offen, 2001) As indicated in the medical sociology literature, the quality of leadership, work culture mix and the wider organizational culture influence hospital care quality.(Mannion and Davies, 2018b) Our case-study found that the greater use of MCNRs were correlated with more instances of reported QI. Previous evidence supports this which there then being 347 new acute non-invasive ventilation recommendations from 37 NCEPOD case-note reviews.(Outcome and Death, 2017) Clinician expectations influence culture.(Mountford and Shojania, 2012)

On the individual level, the clinician participants who exhibited greater ownership of patients tended to give more QI examples. Also, the best performing healthcare institutions exhibited local QI ownership and pro-active responses to quality issues.(James and Savitz, 2011,

Bohmer, 2011) Thus, greater ownership likely raised the clinician's attentiveness to QI opportunities from all forms of case-note review, but this attribute is hard to quantify well. More concern for their patients may entail the more assiduous evaluation of case-notes, but this not necessarily entail that latent learning is found in these case-notes. If learning is present, then a more assiduous reviewer would be more likely to find any lessons from the case-notes. Conversely, negative attitudes toward case-note reviews could lower the QI yielded from case-note reviews. The speciality itself influenced the identification of care quality issues from case-note reviews i.e., the emergency department cited little learning whilst acute medicine or cardiology found comparatively more learning for QI. Thus, clinicians should bear in mind specialty-specific dimensions when evaluating in their minds the QI cost-benefit from MCNRs.

Over three years, patient safety training in an English hospital was found to reduce hospital mortality which involved clinical observation, medication safety, infection control.(Wright et al., 2006) Case-note review training could help clinicians reinforce clinical knowledge and improve their care provision (Cabitza et al., 2019, Boonyasai et al., 2007) with clinicians more likely to attend if non-monetary rewards and CPD are associated with MCNR. Competing processes hindered the possible QI contribution from MCNR which can be given protected time if there is clearer QI evidence to support its case. To obtain QI from case-note reviews, the whole process of case-note reviews needs to first be embedded in the healthcare organization. Then, the information from reviews must flow through the organization for any QI to occur.

4.7.2. Embedding of mortality case-note reviews and case-note reviews

NPT was appropriate to explore mortality case-note review embedding given the recent policy mandate to review in-hospital deaths. Given its novelty and potential open-endedness, it was appropriate to use NPT to study the influence of this MCNR policy for any QI contributions. The embedding was strong demonstrating that a resolute emphasis was placed on the duty to undertake MCNR across interviews and documents. Case-note reviews were used before the 2017 NQB mandate which specifically involved the use of the global trigger tool(Institute of Healthcare Improvement (IHI), 2008) as confirmed by one case-note reviewer (D7). In terms of safety culture, a safety-I culture dominated which is a patient safety culture which emphasizes that things must go wrong with all error and harm to be minimized. The narrative that learning takes place only after an incident is reported, when something goes wrong is testament to this Safety-I culture. In the Trust, incident reporting was a common conduit to case-note reviews, rather than case-note reviews being used as intrinsic tools for learning from case-note reviews. As a concept for further exploration, a safety-II culture is encouraged to understand not just what goes wrong but why and how care goes well. This transition from Safety-I culture to one that of a Safety-II culture will change the purpose of case-note reviewing (and any other quality improvement intervention) to reflect on all deaths – if this is not feasible then a prioritized list of reviews could be studied – as a window to understand why good care is considered "good" in this Trust. And so, the extent of embedding of case-note reviews was understood not as safety-II culture activity but a safety-I activity; this was duly reflected across most of the specialties (i.e., MCNRs were undertaken to highlight mistakes and errors rather than understand – and praising where apropos – good care.

4.7.3. Information flow of case-note reviews

Our critical realism approach is essential for discovering implicit influencers of information flow; for instance, many reviewers indicated that reviewing was a professional duty, but the practice of case-note reviews differed somewhat from their own descriptions of this activity. The critical realist would challenge this first assumption that "*reviewing was a duty*" by asking "*how if case-note reviewing was a clinical and mandated duty, how and why did so many reviewers say they had little time for it*?" In asking these questions, I was able to explore some of the deeper relations between clinician duty, their time they had available for reviewing, and the cultural dynamics within their specialty and the organization to understand more precisely the factors shaping how the duty of case-note reviewing received its expression. The case-study identified more information flow downwards from the board-to-ward than vice-versa. It was clear that board members, managers and administrators helped to facilitate case-note review-specific information flow and that it would help motivate case-note reviewers if a record of what had led to QI was formally reported on to inform clinicians and document QI of past lessons and trends. Hospital protocols help lay down pathways for this information to flow. External policy guidelines and mandates reinforced these information pathways. However, the lack of local empowerment and the absence of a clear overarching structure for information flow fully owned by the hospital specialties very likely led to the underwhelming general use of MCNR across the organization. This does not imply its non-use, but its use could have been more widely advocated than it was. But there may be several reasons for this in this case and these include the types of patients (and/or specialties) which could benefit had benefited as best as they could. Clinicians had inside knowledge of which patients were susceptible and would be able to determine, briefly and without any full implementation, that MCNR would not have generated much QI for them. It does remain to be seen whether those who chose not to employ MCNR widely that their judgement was the

optimal decision – optimal assuming maximal QI-cost benefit - as we have no way of confirming their hunch from this single case study. What we can do is triangulate on the perceptions and data of all participants and documents to arrive at why MCNR was used the way it was.

This emphasizes the importance of the complex interactions of socio-cultural dynamics of an intervention like case-note reviews; in terms of the MUSIQ alluded to earlier, the QI leadership was poorly matched with the various specialties' QI culture. It was a top-down, performative operation where there was a lack of true engagement with the needs of different specialties. It appears to be attributable to two factors. Firstly, the mandate to review did not fit some specialty aims well and thus the culture of reviewing did not pervade the entire Trust. And secondly, there were no concerted efforts from senior clinical and board members to support and help individual specialties to adapt the top-down initiated mandate to review case-note reviews for their own purposes; it was very much left for their own adoption and implementation – as the RCP guidance did encourage.(The Royal College of Physicians, 2016a) There were also other QI tools which laid down a precedent for case-note review information flow. And the local information flow was helped by efficient and rapid turnover of case-notes by administrative staff, which requires top-down led adoption. There is a need to support legislation for QI that was indicated in a literature study search on protocols and policy guidelines supporting data collection from medical records for better research data. (Jansen et al., 2005) Thus, protocols are a *sine qua non* for robust data collection. Similarly to our own results, a qualitative case study of the UK Medical Examiner system, which involves reviewing cases at the request of patient relatives, also found a lack of feedback and shared learning from their qualitative interviews with those involved in implementing the programme and acting as medical examiners.(O'Hara et al., 2021) In sum, there needs to be

more coordinated, tailored support for a complex intervention like MCNR at the policy, Trust-board, managerial and clinical levels for there to be a greater likelihood of success for the intervention. Its full potential has not been met, but these need to respect the specific factors of each context. This is the reason why the case-study method was a suitable method to understand this phenomenon.

4.7.4. Reflecting on the case-study findings using the MUSIQ lens

This section will look more closely at the barriers and facilitators to QI from MCNR using MUSIQ, which is a model for understanding success in quality (MUSIQ). The three MUSIQ meso-level categories of organization, QI team and microsystem are invoked from the model to help shape the discussion. This will represent the role that MCNR has in this case study's context.

4.7.4.1. Organisation

At the organizational level, the QI culture was strong across the organization, however, the QI leadership from the top (board level) was fragmented with key clinical leaders expressing both skepticism and cynicism about the motivations behind MCNR reviews and thus its use. There was also no incentivized payment structure for these MCNR as is the norm in the UK NHS for a significant health indicator or target. Tt was a mandated process with no remuneration. This could explain its lack of enthusiastic uptake for those who did not have it strongly reinforced in their specialty QI culture.

4.7.4.2. QI team

At the QI team level, MCNR was variably used across specialties due mostly to the perceived utility of MCNR toward their own specialty-specific objectives. The team leadership was justifiably strong for QI leaders for those in specialties with a precedent for gaining comparatively more learning than specialties with less learning from MCNR. It must be noted that the teams which had reviewers not just at the physician-level, but also the nurse-level, stood to use MCNR more and found greater benefit from completing these reviews. This was certainly the case for general medicine. However, it is unclear whether the MCNR potential for learning was always apparent with general medicine - as it was also the case that a more diverse set of clinicians were involved - or that there was a different stimulus for why it was a diverse team that employed MCNR more effectively to obtain more learning from this process. In the teams (i.e., specialties) more ardent in their use of MCNR, there were more subject matter experts and clinician specialists involved in the process of MCNR. Those specialties (i.e., general medicine, gastroenterology) exhibited more transparency and exhibited a more involved form of decision-making with more optimism for MCNR; this could be reinforced by the fact that a more open, deliberative communication style lent itself to making the most from MCNR for QI. But with such things, *ceteris paribus* (CP) conditions cannot be elicited and neither is it worth testing these in these complex social conditions, which is the reason for the selection of a case study method that considers the ineluctable contextual conditions surrounding MCNR.[44]

---

[44] CP conditions are befitted to highly controlled acultural conditions such as the physical sciences. Team norms were important for relating behavior around MCNR to their specialty-specific goals. In other words, team norms were important to help establish MCNR

4.7.4.3. Microsystem

Microsystems concern the intersection between the motivational and capability aspects modifying QI processes.(Kaplan et al., 2012) The QI leadership has been poor especially at the highest level of management with internal conflicts concerning the central purpose of MCNR. This conflict had consequences for the motivation of mid-level managers and QI practitioners who sought to develop specific QI programmes with MCNR (i.e., pilot of MCNR in acute general medicine). A senior clinical leader was skeptical about the prospect of significant quality improvement coming from MCNR; it is unclear how the rest of the leadership and management are to manage the situation when a senior clinical leader is skeptical about the soundness of the methodology of MCNR as a genuine QI activity. These power dynamics are clearly critical for how this MCNR becomes embedded and could be explored further in-depth through future qualitative research.

---

as a working programme within their specialty. Linked to team diversity and the presence of a subject matter expert, QI skill was better if the former two were good also; the greatest potential for QI from MCNR lies in diverse, topic-knowledgeable and those with QI skill in making changes. Teams that have worked longer together in unchanging teams were strongly expressed in gastroenterology where a core cadre of physicians has undertaken QI together for over a decade; this long team tenure naturally had a positive impact upon MCNR use in this Trust.

4.7.4.4. QI support and capacity

Though the review mandate message was the same from the board-to-ward, the QI workforce focus on review training was underwhelming; for instance, there was little concerted agreement from the board for all reviewers to be trained in the RCP structured-implicit format for reviewing and the uptake by reviewers was sporadic save for those already passionate about the retrospective learning from case-note reviews. In terms of resource availability, there was some financial support to pilot the new process for MCNR, however, this foundered due to a lack of senior clinical backing at key junctures of its evaluation. It has been shown that the clash in "subcultures" between mid-level to junior members of staff and senior clinical members of staff do undermine the adoption of MCNR, for better or ill and has been showcased convincingly over the many decades of medical socio-cultural studies.(Mannion and Davies, 2018b) Data collection was conducted through a hospital-wide mortality proforma, with the information managed at both the specialty and the hospital-level[45]; the key platform for case-note review information facilitation was the mortality surveillance committee, which filtered pertinent case-note review information from specialties for discussion and distributed their lessons more widely – its primary function. However, there were information flow gaps to the MSC, with limited relevance of most learning from MCNR at the specialty-level. This does show that there is little tangible learning from MCNR, especially concerning its relevance for those of other specialties.

[45] However, some information did not flow from the specialty to the hospital data repositories due to its lack of perceived relevance or for other unmentioned reasons.

4.7.4.5. Environmental levels

Other than the mandate to complete MCNRs, there are no external motivators to undertake MCNR. The sponsorship of MCNR is represented by the RCP and its provision of training for reviewers, however, this was only voluntarily taken up, and inconsistently so, across the hospital. In terms of the strategic importance of reviews to the hospital and its four values (safe, effective, compassionate, and trusted), there is more direct alignment of MCNRs for the three "Trust" values of "safe" and "effective." MCNRs were also adapted for a third value "trusted" because the board perceived MCNR as a re-assurance activity for the public. For the board – on behalf of the Trust - the MCNR was more of a quality assurance activity rather than a quality improvement or learning activity.

4.7.4.6. Process improvements

Some system and process changes include a newly adapted mortality review proforma, the need to collect and report the number of MCNR reviews online to the Department of Health, and the closer engagement with the Patient Liaisons Service (PALS) to detect MCNRs earlier. These were all implemented during the cascade of MCNR and its training to the Trust.

4.7.4.7. Outcome improvements

The outcome improvements include the reduction in mortality of fragile and complex co-morbid patients due to the greater sensitivity to the sources of harm for deteriorating patients sourced predominantly from sepsis and acute kidney injury case-note reviews. This was

attested to by multiple interviewees and documents as evidenced in the results section of Chapter 4 .

4.7.5. How can MCNR contribute to QI?

Given all that has been discussed using MUSIQ, there are clear areas where MCNR support quality improvement in this case-study. Firstly, there is evidently learning obtained from MCNR which do not just have immediate repercussions for the area from which the learning originated (i.e., sepsis and AKI patients). This benefit was had for the identification and broader management of rapidly deteriorating patients. There is a likely positive contribution to sepsis and AKI patients from MCNR because of the complexity and non-trivial nature of the presenting patients. There is typically not one parameter which indicates susceptibility to sepsis or AKI, but it is likely that careful retrospection over the case-note by a clinician can help unpick some of the clinical features involved in the rapid deterioration of that patient's health. (Godin et al., 2015) It is with hindsight that more complex patterns become clear. One reason for the MCNR learning obtained from the AKI topic is due to the topic's amenability for contextualized evaluation. This is so on two fronts; the first is to limit the severity and duration of AKI (e.g. nephrotoxin stewardship) and secondly prevent avoidable AKI complications by enhancing the monitoring of complications and their medication regimens (e.g. BMP, bicarbonate/phosphorus measurement).(Kashani et al., 2019) The second is to do with the variety of presenting complications known to be associated with AKI which are identifiable as risk factors or concerns after the fact. On a separate point, the efforts by NHS England and Improvement, with guidance from the National Institute of Health and Care Excellence (NICE), have sought to better identify early signs of AKI and prevent undue harm. There was a contemporaneous national patient safety

alert for acute kidney injury across the NHS(Selby et al., 2015) which showed that AKI patients have an association with longer hospital lengths of stay[46] and 30-day readmission rates. Thirdly, the recent national standardization of the AKI definition may be helping clinicians better manage AKIs by enabling them to have clearer alerts around AKI (Thomas et al., 2015) of which there were previous calls from clinical leaders in renal medicine to raise awareness of.(Mehta et al., 2007)

Now, we turn to the topic of sepsis. A reason that sepsis has received much attention from policymakers over recent years is due to the elevated public profile of sepsis in acute care settings through the Surviving Sepsis Campaign with the national mandates for sepsis care and its public reporting.(Dellinger et al., 2004, Dellinger et al., 2008, Dellinger et al., 2013) This spotlighting of sepsis may have helped raise the profile of sepsis during MCNR, however, there was no mention in the data of any adherence to this international guidance.

The benefit of more time and hindsight is apparent for AKI and sepsis when reviewing their case-notes. Secondly, the MCNRs overlap well functionally with existing quality improvement processes, which include incident reporting systems (i.e., Datix), clinical audit and specialty-specific retrospective case-note reviews. Whilst the intervention, the medical examiner's (ME) programme, was not fully implemented nationally - nor in this Trust - at the time when the data for this case-study was collected, the ME programme is expected to make further use of MCNR to deliver a better service to the bereaved and

[46] The longer lengths of stay lead to larger case-note reviews, which signifies the importance of case-note reviews being used to distil the critical AKI risk signals from the more voluminous case-notes of these patients. MCNR is a possible method to identify the signal from the noise of more information.

improve the quality of mortality data.(NHS Improvement, 2020a) There is every reason to expect that MCNR will continue to complement this ME programme as the they will use case-note reviews to examine unexpected deaths and answer the safety queries of bereaved relatives.(NHS Improvement, 2020a)

MCNRs were shown to conflict with other quality improvement interventions, especially at the specialty-level. This might appear to conflict with the previous statement about its sound integration, but this point seeks to express whether there is a clear purpose and use for MCNR in each specialty. For instance, the emergency medicine (ED) department found little utility in terms of learning obtained from MCNR, and so it was adapted as a training tool for more junior doctors and repurposed as an additional audit tool. They thought it would most likely interfere with the subculture and its assumption, but nevertheless it must be undertaken because of the review mandate. Thus, the ED specialty devised a workaround where junior doctors were tasked to undertake MCNR as part of their training, despite it yielding little learning for overall QI. It was employed as an educational activity for junior doctors. ED incidents do not often present much MCNR learning because of them being accidental and one-off presentations. The MCNR is not problematic for ED consultants to undertake but it yields a lower perceived QI cost-benefit compared to other specialties. To help accommodate for a hospital specialty's subculture, it may help hospital administrators and senior medical staff to gather perspectives and discuss the use and desired contribution from MCNR at the outset. If a plan of action for MCNRs can be rationalized, have buy in and then be implemented, this would help integrate the MCNR process into daily work that is both effective and minimally disruptive to the daily clinical work. Contrarily, other specialties like acute medicine and general surgery have robust procedures in place which co-align with the process of MCNR; case-note reviews and the audit of medical records were

commonly discussed in their mortality and morbidity (M&M) meetings and so it is not surprising to see the use of MCNR in both their specialty practice.

In sum, for any case-note review intervention to take root in a specialty both its subculture and a certain type of complex-presenting patient can help yield more learning from MCNR. Specialties ought to agree what the likely contribution of MCNR before using it. This would both help MCNR gain traction where otherwise it would not have otherwise been adopted, and for those specialties which stand to gain little QI dividends for their efforts. The success of an intervention is closely tied to the individual's assiduity to review, a subculture of learning from reviews – especially with more complex patients - and the prevailing healthcare policy context.

4.7.6. Challenges to MCNR contribution to QI

Despite the tangible quality improvement for certain domains of care (i.e., sepsis, AKI, and the deteriorating patient), MCNR have several challenges regarding its contribution to QI across the healthcare organization. First, there is a large resource requirement of time and effort to review the case-notes for busy clinicians. Despite their own words indicating that MCNRs are an intrinsic duty as a clinician, the time-sensitive environment of their work is, in the most part, not conducive to the establishment of special protected time intervals exclusively for MCNR duties. One way to overcome this is to have a dedicated, protected timeslot for MCNR to happen. Another is to offer an incentive for reviewers to undertake MCNR as part of their professional re-validation. Second, the culture of MCNR is that of a safety-I culture, which emphasizes not just how and why care went well but how and why it went badly. Overall, the culture of MCNR sought to elicit negative instances of care and rectify this issue by change to the system and the behavior of individuals (safety-I culture),

rather than the positive, praiseworthy work of staff commonly found on a week-to-week basis (safety-II culture). The sustained undertaking of MCNR under a safety-I culture will highlight inadequacies in care which neglect the total care that was overall "exemplary". It is safety-I's myopic vision of care quality – looking at the error and the mistake in isolation – that will limit the openness of specialties and the sharing of MCNR findings. It is in fact what is likely to have been observed here in our case study. To help with this, a transition to a safety-II culture that esteems the good and the bad aspects of care would help to develop a more open culture for MCNR information to circulate between wards, boards, and the organization as a whole; it would take the sting out of sharing mistakes, when the culture of sharing is normalized by the greater discussion of care going well. Encouragingly, the heart of the MCNR method is designed to emphasize good aspects of care as indicated by a senior clinician and board member:

"I was just going to mention the RCP tool that it gives you the opportunity to talk about what has gone well in the patient. Care as well which is again a much more positive way of looking at the review. So, I think clinicians will value that aspect to it as well." (SO3, Senior Clinician and Board Member)

Furthermore, there is a need to explore the integration of learning approaches within the organization(Sujan et al., 2017) and its respective subcultures.(Mannion and Davies, 2018a) And thirdly, there is a poor feedback loop for information which flows from the specialty to the board and then back to the specialty. There is poor receptivity from the board to receiving information from the specialty. There is a lack of a formal process on how to process the information for wider dissemination for the entire organization. This poor infrastructure manifests as a poor feedback loop to reviewers. Did the care raise pertinent questions? If so, how? And importantly, what should we, as a specialty, and a healthcare organization take

away from this information? This form of feedback was not available in any standardized form which could help reviewers reflect meaningfully on their practice and, in a wider sense, the work with other specialties. There are number of things to consider for its improvement as identified across US hospitals through a process of open-ended qualitative interviews like our own. They found that effective data feedback efforts are more often than not established when clinicians are motivated by the data's salience, time has to be given for the feedback data to gain credibility, the data source is reliable and timely, the benchmarks are meaningful, clinical leaders reinforce the feedback loops, prohibit the individual profiling of a clinician's practice as it can be seen as punitive and lastly, that the data feedback loop must persist so sustain improved performance over time.(Bradley et al., 2004)

4.7.7. Strengths & Limitations

This case-study was important and unique because we examined the extent of embedding and information flow which likely influenced QI contributions from MCNR. We considered perceived facilitators and barriers for QI from MCNR and confirmed our findings through independent judgement via documentary source triangulation. An ethnographic method was not chosen because this would have been prohibitively resource-intensive and costly compared to the semi-structured interview questions which were tailored for case-note reviews through the pilot interviews with the key informants. QI facilitators could well be inhibitors in different contexts; we caution any interpolation of these findings without consideration of its context. The results apply and so thus does any theory which has been developed out of this case study; this is what is intended by analytic generalizability. We employed the professional distancing of peer debriefing the supervisory panel members to resolve the final framework matrix.(Spillett, 2003) Member checks were not strictly required

as the lead investigator conducted, transcribed all interviews and communicated with participant post-interview to clarify any troubling meanings. Single case studies may appear to present a methodological limitation; however, we have defended our selection of this approach in the methods section with its capability of generating analytic theories and rival hypotheses. Single case studies provide more detail and offer "better stories" which are helpful in describing rich, complex phenomena and because the employment of multiple case studies was not required given the development of deviant cases (contrast and compare between specialty cultures) within this single case study. There are also journal reporting space constraints, which would encourage the use of extensive tables which would risk mimicking quantitative data – which lacks the contextualised richness required to understand the setting of this Trust. It is not the aim of this case-study to generate quantitative data. This would strip the rich, contextual data from case presentations, which is a major advantage of the case study.(Baker, 2011b) The case study showed that there is QI from MCNRs and future research platforms to explore this important issue. It could be further explored why certain specialties have more potential for identifying lessons from case-note reviews than others.

The qualitative methods employed in this case study were able to convey more detail and insight on the processes contributing to care quality improvement compared to pure quantitative research methods. Quantitative research methods often employ strategies which greatly reduce the contextual richness of information. In other words, quantitative measurement must be strictly defined which means the contextual richness surrounding this data can be lost. For instance, if quantitative research methods were employed, questions through a survey form would likely have been posed for participants to complete in the process – and most likely those same people who were recruited into the case study – to

answer questions pertaining to the case-note review process and its outcomes. Some example quantitative questions would have been included "*how long does it take you to complete a MCNR?*", "*Please circle from the following the factors which compete with the completion of MCNRs: workload, poor case-note formatting, delays in receiving the notes etc*" and "*Please select from the below which factors influence MCNR information flow?"* It might be apparent that the quantitative questions are ill-equipped to demonstrate the wider factors which influence reviewing. Therefore, a framework like NPT is helpful to help us to look further into the embedding's influence upon the case-note reviews. Also, quantitative research implies that the investigators have identified the factors at play but want to elicit their extent of influence; our research questions assumed no factors, as we were open to what the data presented to us through the two frameworks employed; NPT as the investigative and orienting tool, and MUSIQ as the lens with which to understand the findings. The main reason for this is because quantitative questions can elaborate only on discrete information not relating to why and how something has occurred. To receive an explanation or elaboration of the MCNR process, a qualitative approach for this case study was apropos to reveal more of the contextual and process-specific information around embedding, information flow and contributions to MCNR QI. In other words, the richness of the data required to answer the research questions is more adequately satisfied by a qualitative research approach. This is the reason for the adoption of semi-structured interviews and the amassing of collected documents. I go on to elaborate on their contribution to the initial qualitative research questions.

4.7.7.1. Contribution of the case-study method

The advantages of using a case study to explore the embedding, information flow and quality improvement derived from MCNR are manifold. A list of the benefits with its corresponding result are given below (Table 19):

Table 19. Advantages of case-studies and their examples

Advantages of case-study	**A concrete example sourced from the case study**
identification of new or omitted variables and hypotheses	Through deduction and induction, it became apparent that there were a variety of reasons for specialties benefitting from mortality case-note reviews and others not. It became evident complex presenting patients were a higher source of learning than those from sheer accidents alone. Through deduction, from the lack of quorum for MSC meetings, there was an evident lack in the concerns raised at monthly meetings despite agendas being raised each month to be discussed. The lower priority placed upon MCNR, and mortality issues is evident from their actions.

Examining intervening variables in individual cases to make inferences on which causal mechanisms may have been at work	That more information remained within the specialty indicated either that the review information was perceived as less important or was not likely to be prioritised at the higher hospital levels despite its importance.
Attaining high levels of construct validity	Through the triangulation process, the data became saturated between interviews, between documents, and between interviews and documents.
Using contingent generalizations to model complex relationships such as path dependency and multiple interaction effects	The board set the agenda that MCNR is very much a “reassurance” activity and not per se about the detection of care quality issues for improvement. This senior hospital level position was reflected in the decisions of senior level clinicians and their non-openness to other narratives from less senior managers.

I selected a case study design because of the nature of the research problem and the questions being asked. The case study approach is the best plan for answering the research questions. The case study offers a means of investigating complex social units consisting of multiple variables of potential importance in understanding the embedding, information flow and QI from MCNR. Anchored in real-life situations, the case study results in a rich and holistic account of the MCNR process. It offers insights and illuminates meanings that expand its readers' experiences which could not be provided by quantitative research methods. These

insights can be considered as tentative hypotheses that help structure future research and so therefore, case studies play an important role in advancing our understanding about MCNR's contribution to quality improvement. Because of its strengths, this case study is an appealing design for applied fields of study such as education, social work, administration, health and more. An applied field's processes, problems, and programs can be examined, in their own context, to bring about a truer understanding which can affect and improve practice. The case study has historically proven useful for studying educational innovations, evaluating programs, and informing policy. Thus, its extension to healthcare quality improvement is a sensible use of the case study.

As this case study focuses on a single unit, the issue of generalizability appears to loom large. However, much information and theory can be gleaned from a particular case. Readers can learn vicariously from an encounter with the case study through the researcher's thick narrative description. (Stake, 2005), Gertz). The rich description in a case study can create an image: "*a vivid portrait of excellent teaching, for example--can become a prototype that can be used in the education of teachers or for the appraisal of teaching*" (Eisner, 2017)p. 199) Further, Erickson (1986) argues that since the general lies in the particular, the single, what we learn in a particular case can be transferred to similar situations. It is the reader, not the researcher, who determines what can apply to his or her context. And Stake (2005, p. 455) explains how this knowledge transfer works specifically that case researchers "*will, like others, pass along to readers some of their personal meanings of events and relationships--and fail to pass along others. They know that the reader, too, will add and subtract, invent and shape--reconstructing the knowledge in ways that leave it...more likely to be personally useful.*" (Stake, 2005) It is in this sense that this single case study has led us to insights which

are more directly practicable for those reading about these results, where the personal significance of relationships is preserved between events.

In a helpful discussion on the value of case study research, Flyvbjerg (2006) sets up five "misunderstandings" about case study research, which are rebutted, substituting a more true and constructive re-statement about the erroneous assumptions posed by the misunderstanding.(Flyvbjerg, 2006) These misunderstandings and their rebuttals are displayed in Table 20. The second misunderstanding, for instance, "*that one cannot generalize on the basis of a single case is usually considered to be devastating to the case study as a scientific method*" (p.224). However, citing single cases, experiments, and experiences of Galileo, Newton, Einstein, Bohr, Darwin, Marx, Freud, and other prominent scientists, Flyvbjerg makes the point that both human and physical sciences can be advanced by a single case (study).

Table 20. Five misunderstandings about case study research

Misunderstanding	Restatement
1. General knowledge is more valuable than context-specific knowledge.	Universals can't be found in the study of human affairs. Context-dependent knowledge is more valuable.
2. One can't generalize from a single case, so a single case doesn't add to scientific development.	Formal generalization is overvalued as a source of scientific development; the force of a single example is underestimated

3. The case study is most useful in the first phase of a research process; used for generating hypotheses.	The case study is useful for both generating and testing of hypotheses but is not limited to these activities.
4. The case study confirms the researcher's preconceived notions.	There is no greater bias in case study toward confirming preconceived notions than in other forms of research.
5. It is difficult to summarize case studies into general propositions and theories.	Difficulty in summarizing case studies is due to properties of the reality studied, not the research method.

With misunderstanding 1, general knowledge about case-note review is not very helpful given the variety of factors which influence case-note review embedding, information flow and its subsequent QI contribution. This prioritisation of universals cannot be found in the study of case-note review because it must be studied within, not without, its context; this contextualised knowledge is far more useful as knowledge because case-note review must be applied and must differ in different contexts. And so, by observing its nature in this hospital case-study setting, we can generate theory preserved in its context that then can be translated to other similar contexts. Its extension to dissimilar contexts is possible given the development of analytic theory, which can explain similarities and dissimilarities. This is what is meant by analytic generalisability or theory that is applicable to other contexts.

With misunderstanding 2, the assumption that generalising from one case to another is unscientific is unwarranted. There are paradigmatic cases where new phenomena have established entirely new ways of perceiving and understanding physical and social reality. For instance, Galileo's rejection of Aristotle's law of gravity was not based on observations

"across a consummate range," and observations were not "carried out in some numbers." The rejection consisted primarily of a conceptual experiment and later of a practical one. These experiments, with the benefit of hindsight, are self-evident. Nevertheless, Aristotle's view of gravity dominated scientific inquiry for close to 2,000 years before it was falsified. In his experimental thinking, Galileo reasoned as follows: if two objects with the same weight are released from the same height at the same time, they will hit the ground simultaneously, having fallen at the same speed. If the two objects are then stuck together into one, this object will have double the weight and will, according to the Aristotelian view, therefore fall faster than the two individual objects. Galileo's view continued to be subjected to doubt, however, and the Aristotelian view was not finally rejected until half a century later, with the invention of the air pump. The air pump made it possible to conduct the ultimate experiment, known by every pupil, whereby a coin or a piece of lead inside a vacuum tube falls with the same speed as a feather. After this experiment, Aristotle's view could be maintained no longer. What is especially worth noting, however, is that the matter was settled by an individual case because of the clever choice of the extremes of metal and feather. Naturally, it has since been replicated many times since. One might call it a critical case; for if Galileo's thesis held for these materials, it could be expected to be valid for all or a large range of materials. This single case study helped to redefine our understanding of the physical world. This shows the explanatory power of a single case study.(Flyvbjerg, 2006)

In misunderstanding 3, the case study is considered only for hypothesis-generating purposes. However, this is strictly not true because hypotheses can be generated and proven (or disproven) by the astute employment of rival hypotheses. The data lines may not be as co-ordinated in case studies given their contextual richness, however, the multiple lines of information (e.g., media, documents, interviews, physical objects, photographs) can all be

collated to see if hypotheses are confirmed, disconfirmed or are insufficient for either of the former two.

In misunderstanding 4, the case study is accused of being a product of the investigator's preconceived notions. A form of confirmation bias if you will. However, there is no difference in confirmatory bias than in any other field of enquiry. A notable 16th century philosopher of science noted that confirmation bias is not exclusive to just the social sciences, but pervades the natural sciences in equal degree in this quote:

> *"The human understanding from its peculiar nature, easily supposes a greater degree of order and equality in things than it really finds. When any proposition has been laid down, the human understanding forces everything else to add fresh support and confirmation. It is the peculiar and perpetual error of the human understanding to be more moved and excited by affirmatives than negatives."(Bacon, 1858)*

It is a figment of our humanness that we are drawn to certain ideas and perspectives. However, this does not entail that this innate "bias" is any way inferior for scientific use. The scientific method must first employ ideas already available to them to test the veracity of their hypotheses; it is a common corollary of the hypothetico-deductive method; however, it is often unrecognised as operating in the physical sciences and is swiftly seen as a flaw for social science and the humanities.

And finally, in misunderstanding 5, there is a general understanding that case studies cannot be reduced into theories, like they are two incompatible entities. In other words, the one-off event cannot have anything general said about it. The rule formulation that takes place when researchers summarize their work into theories is characteristic of the culture of

research, of researchers, and of theoretical activity, but such rules are not necessarily part of the studied reality. Thus, case researchers, thus, tend to be sceptical about erasing phenomenological detail in favour of conceptual closure. And the following quote embodies the irreducible richness of case studies in conveying direct answers to real world questions:

> *"It is correct that summarizing case studies is often difficult, especially with matters concerning case process. It is less correct as regards case outcomes. The problems in summarizing case studies, however, are due more often to the properties of the reality studied than to the case study as a research method. Often it is not desirable to summarize and generalize case studies. Good studies should be read as narratives in their entirety."(Flyvbjerg, 2006)*

In retrospect, the case study is an important research tool for investigating context-rich phenomena to explore the dynamics of case-note review process in its real-world hospital setting. It has gone beyond what would have been possible with quantitative research methods, which cannot elicit the *how's* and *why's* in this complex hospital setting.

Organisational research, in particular case studies, offer methods to improve understanding of organisational and microsystem contexts for improving care and the development of theories to possibly guide improvement strategies. The organisational arrangement within this case study, and the variability around the perceived contribution of MCNR at the specialty level have shaped the adoption and use of MCNR for QI purposes. The high levels of medical professionalisation around MCNR, through independent bodies and government mandates, was not sufficient to ensure that MCNR would be successfully implemented across the entire Trust. The introduction of MCNR was slowed at the first

instance at the pilot stage when there was disagreement concerning the exact role of MCNR between senior clinical managers and middle managers implementing the pilot. This finding has found some overlap with existing case studies which has expected fast adoption across settings but slowed significantly due to disagreements across different professional group and various factions.(Ferlie et al., 2005) The adoption of MCNR depended on the specialty's perceptions of the likelihood for MCNR to provide care quality insights, and the existing patient safety culture and processes rather than any resourcing or training. This finding is like a case study exploring the first adoption of minimally invasive cardiac surgery for coronary bypass graft or valve replacement in US hospitals. This study found that team learning processes, with its organisational micro-culture, were most important for the adoption of the surgical innovation.(Edmondson et al., 2001)

In sum, the case study has contributed where there was no known contextual study of case-note review in the hospital setting. This case study has contributed to a deeper understanding of the local and Trust-wide dynamics which influence embedding, information flow and QI. This phenomenon of case-note review has not been studied in this manner before. The case study has provided additional context to the factors affecting the use of case-note review type methods and interventions for QI purposes, which include but are not restricted to "micro-culture, microsystems, competing work, clinical duty, reviewer-perceived lack of utility, unresolved disagreements between clinicians and its overlap with extant QI procedures in both purpose and function."

4.7.8. Reflecting on qualitative research methods

The case study in my thesis used qualitative research methods. During the case study and its data collection phase, I used a physical notebook to capture the gestures, the possible

meanings, the clarifications needed and the ideas relating to their influence on case-note reviews or quality improvement in general. It was essential, I felt, to reflect the significance of the interviewee's concepts of case-note reviews and quality improvement to describe and explain case-note review's contribution to quality improvement. These were best captured by the consideration of rival hypotheses, which are the possible hypotheses which explain an outcome of interest.(Yin, 2017) The concept was not new, however, the extent to which one needed discipline to prune less plausible hypotheses in favour of others where there was some evidence of support or to one's where support was plausible given a different method or perspective on data collection. For instance, could incident reporting and specialty-specific case-note reviews be responsible for most of the quality improvement given that it itself directs mortality case-note reviews? Or were case-note reviews themselves the source of quality improvement? It is uncertain the direction of the causal arrow; only a multiple time-series or a difference-in-difference study design can derive whether one caused the other. We found that case-note reviews do identify trends, otherwise imperceptible to clinicians, when undertaken in small batches e.g. 10-25; incident reporting is not able to demonstrate how care is deficient, but it acts much like a smokescreen detector for further investigation.[47] It cannot be known, in our Trust, whether quality improvement is sourced from case-note reviews which is an instrument capable of describing and explaining poor hospital care and/or health outcomes. These things were considered and a pen, notebook and the returning train helped

[47] Mortality indicators are justifiably useful as smokescreens given their inability to reliably distinguish between the features of deteriorating patients from patients receiving poor care. Case-note reviews are better suited for (accurately and reliably) identifying the nature and causes of care quality concerns, rather than detecting the first signs of poor care quality.

me to reflect and sift the data signal from the noise. From study conception to the final report, I have come to appreciate more the contributions of the qualitative element and cognitive processing elements over quantitative results.

Qualitative research quality can be assessed using credibility, transferability, dependability, and confirmability as criteria. First, credibility is a key quality of qualitative research that refers to the research faithfully representing reality. Lincoln and Guba indicated that credibility is preserved by prolonged engagement, triangulation, persistent observation, peer debriefing, negative case analysis, member checking and referential adequacy.(Lincoln and Guba, 1985) I spent a substantial period at the case study site which allowed me to adjust to the surroundings, the culture, the systems and the daily running of the hospital. I spent some lunches and informal meetings on the hospital site at the behest of certain interviewees and other non-interviewing members of staff.

The use of triangulation has been shown to increase credibility by drawing on more than one data source to develop understanding of the phenomena, and there are four types according to researchers.(Patton, 1999, Denzin, 2017) There are four types of triangulation. Methodological triangulation was not used given the limited capacity and in-depth work required for the researcher to confirm or refute hypotheses, however, implications are provided for use in future case studies. Further extensive cross-methodological checks were not feasible given the time and resource limitation. This method would not have been beneficial to undertake compared to other triangulation methods.

Another method of triangulation is method triangulation. This concerns the involvement of more than one researcher who conducts the research process or the review of it. Resource and

ethical constraints meant that one researcher collected, coded and analysed the data, which is a common occurrence in qualitative research.(Barry et al., 1999) However, the supervisory team and second reviewer were closely involved in reviewing select manuscripts for transcription and thematic coding fidelity.

Theoretical triangulation concerned the use of multiple theories to interpret data. This was achieved in our case-study as certain methods were better suited in their theory than other theories. For instance, the NPT was suited to answer the question of the embedding of the case-note reviews, whilst the more open framework matrix was suited to the information flow and QI contribution.

Source triangulation, another of the triangulation methods, was undertaken. The work explored the case study context on an individual basis using semi-structured interviews, board room and group meetings, internally sourced documents obtained by request via the interview, external documents obtained through publicly available means (websites, online databases, HSCIC) and the notes of an immersed researcher. All these activities have increased the credibility of these findings.

Another method for improving creditability is peer debriefing. This involves researcher discussing interpretations with other individuals to air and challenge assumptions, explore plausible hypotheses, and encourage researchers to reflect on the reflexive nature of the research. The supervisors were the sources of peer debriefing, as were some of the interviewees, throughout the case selection, data collection, analysis and interpretation stages given the strict service evaluation and disclosure on involving non-Trust and researcher personnel. However, Lincoln and Guba would consider them '*interested*' with the

supervisor's credentials wide-ranging (non-qualitative) and sufficiently far-removed to facilitate this process.[48]

When not interviewing or collecting source materials, I would continue to observe the daily operations of the hospitals, specifically the interviewees. Though this case-study is not strictly an ethnographic study, I did note people's body language, gestures, and the general goings-on before and during the interviews. This provided rich data some of which could not be explored due to the nature of the research design. There was need to brief peers as the sole researcher on site, however, I would brief my qualitative supervisor and lead supervisor on developments in the field and draw on their research intuitions and develop the research accordingly. Negative case analysis in the guise of 'rival hypotheses' were applied to data to which multiple hypotheses plausibly obtained with the most plausible one selected given the totality of data.

Negative case analysis[49] that identify disconfirming cases is one more strategy for enhancing credibility. Information that seems to refute mainstream emerging theories were explored in-depth until both theories could be reconciled into an interpretation. A specific instance was when the general levels of attendance at Mortality Surveillance Committee (MSC) meetings were alluded to by an interviewee outside the meeting. There had been poor attendance levels

---

[48] In not being involved at all stages and having never visited the site nor met the interviewees, the peers in the debriefing could be considered as impartial and interested members.

[49] Otherwise known as "disconfirming cases."

at MSC meetings. There are several explanations for this. Lack of enthusiasm, insufficient data due to good care quality, clinicians being especially busy during these times and poor communication could each have been reasons for this low attendance. With further interview information, it was identified as a combination of lack of enthusiasm and absence of poor care quality cases to discuss at the meeting.

Member checks, or respondent validation, are used to establish the credibility of the transcription process. This involves asking interviewees to review the data and its interpretation. Upon meeting and introducing the service evaluation, member-checking was indicated verbally and in writing on the consent form. Member checks were used and taken up by a handful of interviewees. It was considered worthwhile to engage them more deeply in the research process and obtain their trust concerning this study as I had the opportunity to interview participants more than once to pursue rival hypotheses/negative case analysis. Member checks were not essential as the investigator AT had personally interviewed all participants and transcribed the audio-interviews verbatim; any further clarifications were conducted over the phone with relevant participants to determine the exact meaning of phrases and terms.

The last approach enhancing credibility is referential adequacy(Lincoln and Guba, 1985) where a portion of unanalysed data is set to one side and is then later used as a reference to check analysed or interpreted data. This approach was not used given the valuable and important contribution of each interview, adoption of the snowballing approach to recruitment of the data and the number of rival hypotheses generated.

Further approaches to deliver credibility involved the encouragement of the honest sharing of comments from interviewees. Some interviewees established good researcher rapport to give their personal contact and phone number when further conversations were required. All this was within the Trust policy and terms of the contract of the service evaluation. At the outset of each interview, I would outline my background stressing my non-medical affiliation and in so doing lower the likelihood of any Hawthorne effect from external study and scrutiny from an external authority.(Sedgwick and Greenwood, 2015)

Given there were no qualitative service evaluations nor case studies into hospital-wide case-note review as a quality improvement intervention (some qualitative literature exists for QI study from specialty-specific or harm analysis from case-note review(Lipitz-Snyderman et al., 2011, Lau and Litman, 2011, Mitchell et al., 2014) but these are not hospital-wide studies).

There are three further components complementing credibility. Transferability is akin to the quantitative researcher's external validity. It expresses the degree to which the research findings can be generalized to another context. 'Thick' descriptions(Geertz, 1973) are full and contextually rich narrative accounts which can be used to determine their transferability to different contexts. One of the key findings was that QI-receptive specialty micro-culture with a presence (or absence) of national QI/audit framework, reviewer awareness of both policy and training and patient ownership is sufficiently rich to apply to other healthcare organisation settings. Yet, it must be indicated the research methodology was not able to

deduce a hierarchy of the most important factor[50], from the above, nor the elucidation of this relationship.[51] The minority, especially the dissenting groups, were not completely forthcoming about their views considering their lack of voice and power in influencing the status quo.(Mannion and Davies, 2015) More consideration from the researcher and the team are needed to broach issues of voice and power for its realistic representation and its transferability.

The second is dependability. This is akin to the quantitative researcher's reliability. Given the contextually rich and unique nature of the research, the co-production of findings with and between interviewees, documents, and researcher, it is most unlikely to be reproduced elsewhere in a similar manner. Qualitative researchers emphasise a different standard of "*accurate and adequate documentation of changes, surprise occurrences*" to determine the dependability of the results.(Lincoln and Guba, 1985) Three criteria have been provided of which this research seeks to match.(Shenton, 2004) These are that reporting should be faithful to the research design and implementation, transparently noting deviations. Furthermore, the details concerning data gathering are critical.

Thirdly, reflective appraisal and reflexivity are important qualitative aspects to consider. According to Alvesson and Skoldberg, any reflexive exercise involves interpretation and reflection.(Alvesson) Interpretation asserts that any research finding is based on an

---

[50] If this is indeed the right approach to take in a naturalistic context.

[51] It is in this instance where a realist review approach could have yielded more value in identifying the likely success factors and leverage points in this research case study.

interpretation of data by the researcher and, being assimilated and re-created by the researcher, cannot be taken as a perfect representation of the phenomenon. The researcher's findings are constructed from facts. The reflective element is an introspective exercise requiring the researcher to examiner one's own psychological, intellectual, professional, sociocultural, and spiritual influences to bring about an interpretation of these "interpretative findings." A critical and systematic approach to this reflexive exercise is essential for improving the research validity and rigour. My reflexive reflections are provided further on in this section.

Lastly, confirmability is like objectivity in positivistic research. It is self-evident objectivity is not possible for qualitative research, nor is it desirable, as the skill of the researcher is to negotiate human viewpoints in leading readers to a convincing narrative of the phenomena.(Patton, 1999) Bias must be limited as best as possible by the researchers, however, it is inevitable the researcher's attitudes and those of the participants will influence the data. It is here that researchers must explicitly acknowledge and report their self-interest and influence upon the research.(Huberman and Miles, 2002) To this end, I have maintained a reflexive stance throughout the research, noting down ideas and my rationale in a notebook before coming to analyse the data. I have consulted my supervisory team (a process otherwise known as 'peer debriefing') to identify any self-interested motivations and to challenge any of my own entrenched views.(Spillett, 2003) It has become apparent that there are wider contextual factors that influence case-note review use. This next, I discuss a theory which can help us to better understand how wider contextual factors.

4.7.8. Receptive contexts: how interventions gain traction in real-world situations

I wanted to use conceptual framework on how change comes about called “receptive contexts” which can be an area of future exploration. Receptive contexts are defined as situations where there are features of context, and of management action, that ‘*seem to be favourable, associated with forward movement*’. On the other hand, non-receptive contexts are those situations where a combination of conditions effectively creates blockages or resistances to change. Pettigrew et al. suggested eight key factors which created a receptive context for the changes at the heart of their study. This concept of culture and receptivity are closely intertwined. A sense of openness to innovation has previously been studied (Pettigrew et al., 1992, Bate et al., 2002) indicating that components of a receptive context include the following:

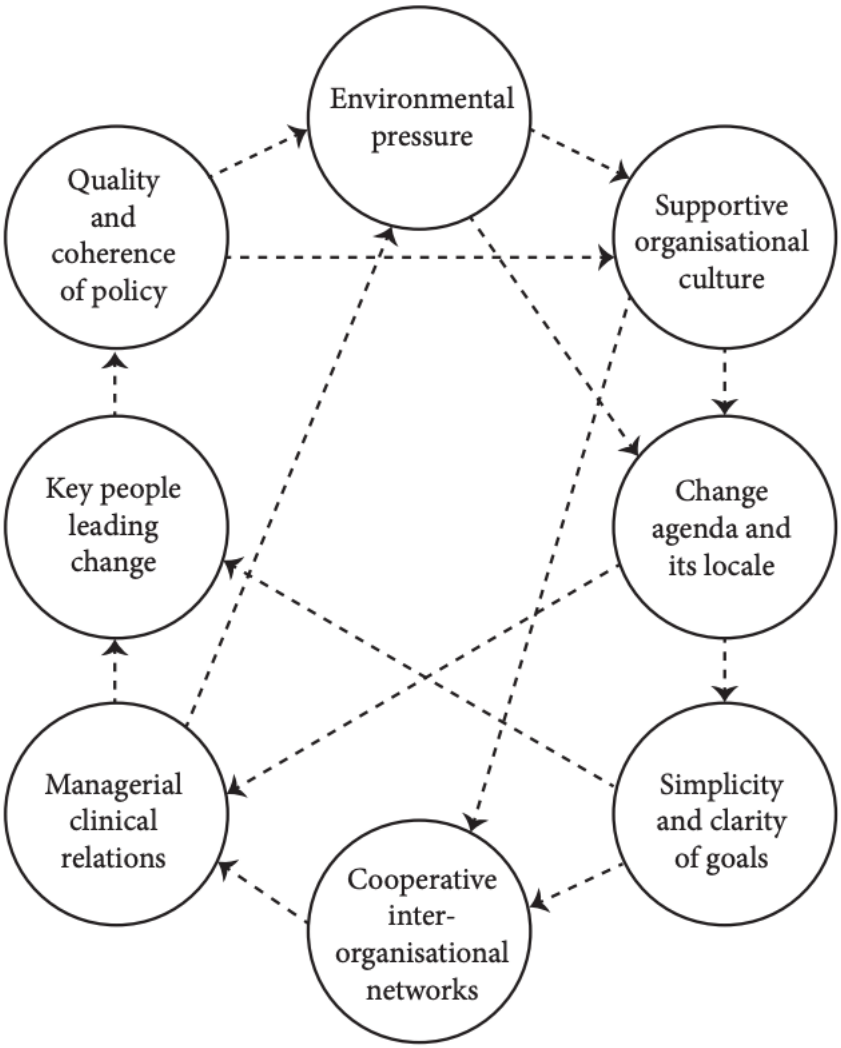
1) the role of human factors and environmental pressure to drive changes
2) the presence and leadership of key visionary individuals
3) good managerial and clinical relations
4) a supportive organisational culture
5) the quality and coherence of ‘policy’ generated at a local level
6) the development and management of co-operative inter-organisational network
7) simple and clear goals
8) the change agenda and its locale

Taking this concept of receptive context and seeing it extend to this case study, key factors such as local specialty micro-culture, board and managerial support and the specialty-specific types of patients are the most important contextual factors affecting mortality case-note review’s (MCNR) potential contribution to care quality improvement.

The eight factors and associated model (Figure 15) have subsequently been tested in empirical studies. In one case study of change within the UK primary health care sector, the most significant pattern of association was between the quality and the coherence of policy, key people leading the change, supportive organisational culture and effective managerial clinical relations.(Newton et al., 2003)

Figure 15 conveys the key factors and their relationships involved in a receptive context for change. For instance, the environmental pressures from without (outside of the NHS Trust) can influence the extent to which the organisational culture is supportive or not. This organisational culture is central to establishing long term change in the Trust indicated. It is this determination to change that necessitates the need for simple and clear goals. And undoubtedly both organisational culture and the simplicity of these goals is essential for cooperative inter-organisational networks within the Trust. The managerial and clinical relations are an important subset of the network and will need to co-operate effectively for effective clinical work and the management of human resourcing and digital documents. There is the need for key motivated people to drive this change with clear and simple goals. But it is not sufficient for leaders to do this alone but to enshrine this in policy which then drives the environmental pressure to continue to support organisational change and help co-operation towards the goal of quality improvement. It is a figure which illustrates the multi-faceted nature of practical change and the importance of the discrete factors that may lend itself to a deeper understanding of the mechanisms that drive quality improvement in one setting. This diagram is not representative of all healthcare organisations, but this of course requires re-adaptation to the local setting. Overall, it is a helpful starting point for any research-motivated individual to understand the underlying dynamics that drive change, whether it be for better or worse, in complex organisations such as acute NHS Trusts or any other healthcare entity.

Figure 15. Factors of receptive context for change



The role of the receptive context is required for any effective use of an intervention. The receptive context is important, and it is not true that an effective organisational healthcare intervention is always effective across other settings. There is a tangle of receptive context factors to consider in understanding the functional status of any organisational intervention. It is clear from our case study, and this precedent exhibited by previous case studies, in eliciting the need to consider wider contextual factors such as board and managerial support, specialty micro-culture and local specialty-specific features (i.e., types of commonly presenting patients). The value of this method is to help practitioners and

clinicians understand the key factors in MCNR adoption, and how it can be better implemented in other contexts given the barriers and facilitators to MCNR uptake and its QI use in this Trust.

4.7.9. Information infrastructure

With the NHS case study, I could have considered the information flow and contribution to QI from case-note reviews from a more technological and information perspective. In other words, I could have considered the use of review e-copies and their dissemination through internal networks. However, given the interchangeability of case-note review format, it would not have been useful to distinguish between electronic and paper case-notes as each served the same function. This would have required a level of acquaintance with the hospital information technology infrastructure not available to me. Thus, the desire to embark on organisational culture and information flow rather than a precise information flow analysis on case-note review use for quality improvement.

It is more than interesting to note that despite the digitisation of a significant proportion of case-notes in this Trust, there is no processing of this information that itself drives learning from case-note reviews. Though the digitisation of case-notes does not necessarily entail that data science methods – to obtain computer-aided insights from data - must be employed out of obligation, it does pose the question why this has not occurred yet. The central aim of case-note reviews is to learn and deliver insights from the case-notes. The question is posed. Might not the application of data science methods to case-notes not yield some benefit for reviewers and the patients that they serve? In 2022, Ben Goldacre and his team undertook a comprehensive review of the use of NHS data on the challenge of curating, managing,

cleaning, and preparing healthcare data so it is useable in well-designed, efficient, and secure platforms. The report found there was a distinct lack of organisation and structure for the effective and efficient use of analytical methods for use in data science.(Goldacre, 2022) It is a little surprising that the methods of applying text mining procedures have not received more concerted work from NHS leaders and innovators given the capacity for data science to derive insights from unstructured data such as case-notes and the inordinate hours required for the completion of these reviews by highly trained clinical staff. These are suggestions outside of the current known use of case-note reviews, but these techniques do exist but have yet to be applied. In sum, the aim of the case-note review is to obtain learning from the case-note reviews and this better understanding of the case-note could help identify this learning – and expedite this process - for quality improvement purposes. The information infrastructure of this Trust (and the NHS) could better organised and curated to not only track this information flow but distil hidden patterns and insights from it soon. All this requires the correct information infrastructure and data pipelines.

For the information infrastructure based on external hospital representation, I considered but did not employ patient and public involvement (PPI) with local patient and stakeholder groups. This included end-of-life care specialty-specific groups who were regionally sourced or nationally represented. This would have yielded a narrative perspective that would have informed a large swathe of patients' needs and wishes regarding the hospital care which case-note reviews could look to scrutinise. However, several studies and groups have already identified hospital care issues such as the Patient Liaison Service mentioned in Chapter 4 . Looking back and considering the extra work required for this, I believe that investing in a PPI group discussion would re-invigorate my own perspective. It could have informed the national mandate to learn from death reviews had i) traction with the public, ii) public policy,

iii) best way to review, not singly from an internal reviewer/medical standpoint but also from the public eye. In hindsight, these benefits were evident. However, the one challenge to all this is that patients and relatives knew little about how case-note reviews were used.

4.8. Conclusion

We found some evidence that mandatory mortality case-note reviews lead to hospital quality improvement. This learning was restricted mostly to specialties with a relative preponderance of acute and complex patients, who were particularly susceptible to rapid deterioration. Mortality case-note review was well-embedded with information flow weighted towards downward flows, board-to-ward rather than upwards from ward-to-board and side-to-side flow between colleagues especially those within the same specialty. Senior-level board support can help enhance the embedding, information flow and QI from MCNR by empowering mid-level staff and create localized specialty plans for its use. The specialty, clinician culture and patient complexity influenced quality improvement across all forms of case-note reviews. Researchers and policymakers should aim to better distinguish between the implementation of mortality case-note review to separately evaluate any quality improvement contributions pre- and post-programme. The purpose of MNCR should be better characterised and described in specialties and healthcare organisation to maximise the utility of MCNR.

4.9. Acknowledgements

We would like to thank the hospital that gave us privileged access to undertake this study. The qualitative researchers, AT (male) and EF (female), were both PhD students and had formal master's level qualitative research training and were assigned four randomly selected

interviews and six hospital-produced documents to identify codes and themes. I would like to thank the hospital board members and the participants for their time and support on this project.

4.10. References

ALVESSON, M. Skoldberg,(2000) Reflexive methodology: New vistas for qualitative research. London: Sage.

AUSTRALIAN COUNCIL FOR SAFETY QUALITY IN HEALTH CARE 2005. Clinical handover and patient safety: Literature review report.

BACON, F. 1858. *The physical and metaphysical works of Lord Bacon*, HG Bohn.

BAKER, G. R. 2011. The contribution of case study research to knowledge of how to improve quality of care. *BMJ quality & safety,* 20**,** i30-i35.

BARRY, C. A., BRITTEN, N., BARBER, N., BRADLEY, C. & STEVENSON, F. 1999. Using reflexivity to optimize teamwork in qualitative research. *Qualitative health research,* 9**,** 26-44.

BATE, S., ROBERT, G. & MCLEOD, H. 2002. Report on the'Breakthrough'Collaborative approach to quality and service improvement within four regions of the NHS. A research based investigation of the Orthopaedic Services Collaborative within the Eastern, South & West, South East and Trent regions.

BENSON, R. & HARP, N. 1994. Using systems thinking to extend continuous quality improvement. *The Quality letter for healthcare leaders,* 6**,** 17-24.

BOHMER, R. M. 2011. The four habits of high-value health care organizations. *N Engl J Med,* 365**,** 2045-2047.

BOONYASAI, R. T., WINDISH, D. M., CHAKRABORTI, C., FELDMAN, L. S., RUBIN, H. R. & BASS, E. B. 2007. Effectiveness of teaching quality improvement to clinicians: a systematic review. *Jama,* 298**,** 1023-1037.

BOURKE, B. 2014. Positionality: Reflecting on the research process. *Qualitative Report,* 19.

BRADLEY, E., HOLMBOE, E., MATTERA, J., ROUMANIS, S., RADFORD, M. & KRUMHOLZ, H. 2004. Data feedback efforts in quality improvement: lessons learned from US hospitals. *BMJ Quality & Safety,* 13**,** 26-31.

BRAITHWAITE, J., HERKES, J., LUDLOW, K., TESTA, L. & LAMPRELL, G. 2017. Association between organisational and workplace cultures, and patient outcomes: systematic review. *BMJ open,* 7.

BRICKER, P. 2014. Ontological commitment.

BUCK, N., DEVLIN, H. B., LUNN, J. N. & VICKERS, M. D. 1987. *The report of a confidential enquiry into perioperative deaths*, King's Fund Publishing Office.

CABITZA, F., ELLINGSEN, G., LOCORO, A. & SIMONE, C. 2019. Repetita Iuvant: Exploring and Supporting Redundancy in Hospital Practices. *Computer Supported Cooperative Work (CSCW),* 28**,** 61-94.

CHAMBERS, D., CANTRELL, A. & BOOTH, A. 2019. Factors that facilitate the implementation of interventions to reduce preventable hospital admissions with a focus on cardiovascular or respiratory conditions: an evidence map and realist synthesis. *Health Services and Delivery Research*.

COHEN, S. M. 2000. Aristotle's metaphysics.

COLEMAN, J. S. 1958. Relational analysis: the study of social organizations with survey methods. *Human organization,* 17**,** 28.

COOKE, C. R. & IWASHYNA, T. J. 2014. Sepsis mandates: improving inpatient care while advancing quality improvement. *Jama,* 312**,** 1397-1398.

CRONSHAW, H. L., DANIELS, R., BLEETMAN, A., JOYNES, E. & SHEILS, M. 2011. Impact of the Surviving Sepsis Campaign on the recognition and management of severe sepsis in the emergency department: are we failing? *Emergency Medicine Journal,* 28**,** 670-675.

CROSSAN, F. 2003. Research philosophy: towards an understanding. *Nurse researcher,* 11**,** 46-56.

DE BRÚN, T., O'REILLY-DE BRÚN, M., O'DONNELL, C. A. & MACFARLANE, A. 2016. Learning from doing: the case for combining normalisation process theory and participatory learning and action research methodology for primary healthcare implementation research. *BMC health services research,* 16**,** 346.

DELLINGER, R. P., CARLET, J. M., MASUR, H., GERLACH, H., CALANDRA, T., COHEN, J., GEA-BANACLOCHE, J., KEH, D., MARSHALL, J. C. & PARKER, M. M. 2004. Surviving Sepsis Campaign guidelines for management of severe sepsis and septic shock. *Intensive care medicine,* 30**,** 536-555.

DELLINGER, R. P., LEVY, M. M., CARLET, J. M., BION, J., PARKER, M. M., JAESCHKE, R., REINHART, K., ANGUS, D. C., BRUN-BUISSON, C. & BEALE, R. 2008. Surviving Sepsis Campaign: international guidelines for management of severe sepsis and septic shock: 2008. *Intensive care medicine,* 34**,** 17-60.

DELLINGER, R. P., LEVY, M. M., RHODES, A., ANNANE, D., GERLACH, H., OPAL, S. M., SEVRANSKY, J. E., SPRUNG, C. L., DOUGLAS, I. S. & JAESCHKE, R. 2013. Surviving Sepsis Campaign: international guidelines for management of severe sepsis and septic shock, 2012. *Intensive care medicine,* 39**,** 165-228.

DENZIN, N. K. 1978. The research act: An introduction to triangulation–Sociological methods. New York: McGraw-Hill.

DENZIN, N. K. 2017. *Sociological methods: A sourcebook,* Routledge.

DONABEDIAN, A. 2002. *An introduction to quality assurance in health care*, Oxford University Press.

EDMONDSON, A. C., BOHMER, R. M. & PISANO, G. P. 2001. Disrupted routines: Team learning and new technology implementation in hospitals. *Administrative science quarterly,* 46**,** 685-716.

EDWARDS, M. T. 2013. A longitudinal study of clinical peer review's impact on quality and safety in US hospitals. *Journal of Healthcare Management,* 58**,** 369-384.

EISNER, E. W. 2017. *The enlightened eye: Qualitative inquiry and the enhancement of educational practice*, Teachers College Press.

EPSTEIN, B. 2018. Social Ontology.

ETIKAN, I., MUSA, S. A. & ALKASSIM, R. S. 2016. Comparison of convenience sampling and purposive sampling. *American journal of theoretical and applied statistics,* 5**,** 1-4.

EVANS, D., BOOKER, S., DAYKIN, N., POWELL, J., TAYLOR, P. & WEIL, S. 2008. National Evaluation of Patient Advice and Liaison Services (PALS): Final Report. *University of the West of england, Bristol*.

FARMER, T., ROBINSON, K., ELLIOTT, S. J. & EYLES, J. J. Q. H. R. 2006. Developing and implementing a triangulation protocol for qualitative health research. 16**,** 377-394.

FAULKNER, S. L. & TROTTER, S. P. 2017. Data saturation. *The international encyclopedia of communication research methods***,** 1-2.

FEREDAY, J. & MUIR-COCHRANE, E. 2006. Demonstrating rigor using thematic analysis: A hybrid approach of inductive and deductive coding and theme development. *International journal of qualitative methods,* 5**,** 80-92.

FERLIE, E., FITZGERALD, L., WOOD, M. & HAWKINS, C. 2005. The nonspread of innovations: the mediating role of professionals. *Academy of management journal,* 48**,** 117-134.

FINCH, T. L., MAIR, F. S., O'DONNELL, C., MURRAY, E. & MAY, C. R. J. B. M. R. M. 2012. From theory to'measurement'in complex interventions: methodological lessons from the development of an e-health normalisation instrument. 12**,** 69.

FINKEL, M. 2020. *On flexibility*, Stanford University Press.

FLETCHER, A. 2017. Applying critical realism in qualitative research: methodology meets method. *International Journal of Social Research Methodology,* 20**,** 181-194.

FLYVBJERG, B. 2006. Five misunderstandings about case-study research. *Qualitative inquiry,* 12**,** 219-245.

FRANKEL, R. M., FLANAGAN, M., EBRIGHT, P., BERGMAN, A., O'BRIEN, C. M., FRANKS, Z., ALLEN, A., HARRIS, A. & SALEEM, J. J. 2012. Context, culture and (non-verbal) communication affect handover quality. *BMJ quality & safety,* 21**,** i121-i128.

GEERTZ, C. 1973. Thick description: Toward an interpretive theory of culture. *Turning points in qualitative research: Tying knots in a handkerchief,* 3**,** 143-168.

GIRLING, A. J., HOFER, T. P., WU, J., CHILTON, P. J., NICHOLL, J. P., MOHAMMED, M. A. & LILFORD, R. J. 2012. Case-mix adjusted hospital mortality is a poor proxy for preventable mortality: A modelling study. *BMJ Quality and Safety,* 21**,** 1052-1056.

GLUYAS, H. 2018. Understanding the human and system factors involved in medication errors. *Nurs Stand*.

GODIN, M., MURRAY, P. & MEHTA, R. L. Clinical approach to the patient with AKI and sepsis. Seminars in nephrology, 2015. Elsevier, 12-22.

GOLDACRE, B. 2022. Better, broader, safer: using health data for research and analysis. Department of Health and Social Care, UK.

GOLDMAN, A. I. 1999. *Knowledge in a social world*, Clarendon Press Oxford.

GOLDMAN, R. L. 1992. The reliability of peer assessments of quality of care. *Journal of American Medical Association,* 267**,** 958-960.

GOLDMAN, R. L. 1994. The reliability of peer assessments of quality of care. *Journal of American Medical Association,* 17**,** 3-21.

GUBA, E. G. & LINCOLN, Y. S. 1994. Competing paradigms in qualitative research. *Handbook of qualitative research,* 2**,** 105.

HANNA, P. & HARRISON, B. 2004. *Word and world: Practice and the foundations of language*, Cambridge University Press.

HEALTH, S. O. S. F. 2008. *High quality care for all: NHS next stage review final report*, The Stationery Office.

HEALTHCARE AND WORKFORCE IMPROVEMENT. 2017. *Dimensions of Quality* [Online]. Available: http://www.qaproject.org/methods/resdimension.html [Accessed].

HEYWORTH, J., KEEP, J. & DANIELS, R. 2009. On behalf of the Clinical Effectiveness Committee of the College of Emergency Medicine. Sepsis pack 2009.

HOGAN, H., HEALEY, F., NEALE, G., THOMSON, R., BLACK, N. & VINCENT, C. 2014. Learning from preventable deaths: Exploring case record reviewers' narratives using change analysis. *Journal of the Royal Society of Medicine,* 107**,** 365-375.

HOGAN, H., HEALEY, F., NEALE, G., THOMSON, R., VINCENT, C. & BLACK, N. 2012. Preventable deaths due to problems in care in English acute hospitals: A retrospective case record review study. *BMJ Quality and Safety,* 21**,** 737-745.

HOLLNAGEL, E., WEARS, R. L. & BRAITHWAITE, J. From Safety-I to Safety-II: a white paper.

HUBERMAN, M. & MILES, M. B. 2002. *The qualitative researcher's companion*, Sage.

IBRAHIM, H., MUNKHBAYAR, U., TOIVGOO, A., HUMPHRIES, C., OCHIR, C., NARULA, I. S., LILFORD, R. & MANASEKI-HOLLAND, S. 2019. Can universal patient-held health booklets promote continuity of care and patient-centred care in low-resource countries? The case of Mongolia. *BMJ quality & safety,* 28**,** 729-740.

INSTITUTE FOR HEALTH IMPROVEMENT (IHI) 2009. IHI Global Trigger Tool for Measuring Adverse Events: Second Edition. http://www.ihi.org/_layouts/15/ihi/login/login.aspx?ReturnURL=%2fresources%2fPages%2fIHIWhitePapers%2fIHIGlobalTriggerToolWhitePaper.aspx: Institute for Health Improvement (IHI).

INSTITUTE OF HEALTHCARE IMPROVEMENT (IHI) 2008. IHI Global Trigger Tool for Measuring Adverse Events (UK version). http://www.ihi.org/resources/_layouts/download.aspx?SourceURL=%2fresources%2fKnowledge+Center+Assets%2fTools+-+IHIGlobalTriggerToolforMeasuringAdverseEvents_df8a18b6-52cc-4674-8258-030941832115%2fIHIGlobalTriggerTool_UKVersion2008.pdf: Institute of Healtcare Improvement (IHI).

JAMES, B. C. & SAVITZ, L. A. 2011. How Intermountain trimmed health care costs through robust quality improvement efforts. *Health Affairs,* 30**,** 1185-1191.

JANSEN, A. C. M., VAN AALST-COHEN, E. S., HUTTEN, B. A., BÜLLER, H. R., KASTELEIN, J. J. P. & PRINS, M. H. 2005. Guidelines were developed for data collection from medical records for use in retrospective analyses. *Journal of Clinical Epidemiology,* 58**,** 269-274.

KAPLAN, H. C., PROVOST, L. P., FROEHLE, C. M. & MARGOLIS, P. A. 2012. The Model for Understanding Success in Quality (MUSIQ): building a theory of context in healthcare quality improvement. *BMJ quality & safety,* 21**,** 13-20.

KARASSAVIDOU, E., GLAVELI, N. & ZAFIROPOULOS, K. 2011. Assessing hospitals' readiness for clinical governance quality initiatives through organisational climate. *Journal of health Organization and Management*.

KASHANI, K., ROSNER, M. H., HAASE, M., LEWINGTON, A. J., O'DONOGHUE, D. J., WILSON, F. P., NADIM, M. K., SILVER, S. A., ZARBOCK, A. & OSTERMANN, M. 2019. Quality improvement goals for acute kidney injury. *Clinical Journal of the American Society of Nephrology,* 14**,** 941-953.

KELLEY, E. & HURST, J. 2006. Health care quality indicators project: conceptual framework paper.

KOBEWKA, D. M., VAN WALRAVEN, C., TURNBULL, J., WORTHINGTON, J., CALDER, L. & FORSTER, A. 2017. Quality gaps identified through mortality review. *BMJ quality & safety,* 26**,** 141-149.

LANGLEY, G. J., NOLAN, K., NORMAN, C., PROVOST, L. & NOLAN, T. 1996. A practical approach to enhancing organizational performance. New York: Jossey-Bass.

LAU, H. & LITMAN, K. C. 2011. Saving lives by studying deaths: using standardized mortality reviews to improve inpatient safety. *The Joint Commission Journal on Quality and Patient Safety,* 37**,** 400-AP2.

LEVY, M. M., DELLINGER, R. P., TOWNSEND, S. R., LINDE-ZWIRBLE, W. T., MARSHALL, J. C., BION, J., SCHORR, C., ARTIGAS, A., RAMSAY, G. & BEALE, R. 2010. The Surviving Sepsis Campaign: results of an international guideline-based performance improvement program targeting severe sepsis. *Intensive care medicine,* 36**,** 222-231.

LILFORD, R., EDWARDS, A., GIRLING, A., HOFER, T., DI TANNA, G. L., PETTY, J. & NICHOLL, J. 2007. Inter-rater reliability of case-note audit: A systematic review. *Journal of Health Services Research and Policy,* 12**,** 173-180.

LINCOLN, Y. S. & GUBA, E. G. 1985. Naturalistic inquiry. CA: Sage.

LIPITZ-SNYDERMAN, A., STEINWACHS, D., NEEDHAM, D. M., COLANTUONI, E., MORLOCK, L. L. & PRONOVOST, P. J. 2011. Impact of a statewide intensive care unit quality improvement initiative on hospital mortality and length of stay: retrospective comparative analysis. *Bmj,* 342.

LOHR, K. N. 1991. Medicare: a strategy for quality assurance. *Journal of quality assurance: a publication of the National Association of Quality Assurance Professionals,* 13**,** 10-13.

MAKARY, M. A., WICK, E. & FREISCHLAG, J. A. 2011. PPE, OPPE, and FPPE: complying with the new alphabet soup of credentialing. *Archives of Surgery,* 146**,** 642-644.

MANASEKI-HOLLAND, S., LILFORD, R. J., BISHOP, J. R., GIRLING, A. J., CHEN, Y.-F., CHILTON, P. J., HOFER, T. P., ALNER, V., BLEASDALE, J. & BOSANKO, N. 2016. Reviewing deaths in British and US hospitals: a study of two scales for assessing preventability. *BMJ Quality & Safety,* 26**,** 408-416.

MANASEKI-HOLLAND, S., LILFORD, R. J., TE, A. P., CHEN, Y.-F., GUPTA, K. K., CHILTON, P. J. & HOFER, T. P. 2019. Ranking hospitals based on preventable hospital death rates: A systematic review with implications for both direct measurement and indirect measurement through standardized mortality rates. *The Milbank Quarterly,* 97**,** 228-284.

MANNION, R. & DAVIES, H. 2018a. Understanding organisational culture for healthcare quality improvement. *BMJ (Online),* 363.

MANNION, R. & DAVIES, H. 2018b. Understanding organisational culture for healthcare quality improvement. *Bmj*, 363.

MANNION, R. & DAVIES, H. T. 2015. Cultures of silence and cultures of voice: the role of whistleblowing in healthcare organisations. *International journal of health policy and management*, 4**,** 503.

MARSHALL, M. 2009. Applying quality improvement approaches to health care. British Medical Journal Publishing Group.

MAY, C. R., JOHNSON, M. & FINCH, T. 2016. Implementation, context and complexity. *Implementation Science*, 11**,** 1-12.

MAY, C. R., MAIR, F., FINCH, T., MACFARLANE, A., DOWRICK, C., TREWEEK, S., RAPLEY, T., BALLINI, L., ONG, B. N. & ROGERS, A. 2009. Development of a theory of implementation and integration: Normalization Process Theory. *Implementation Science*, 4**,** 29.

MAZZOCATO, P., SAVAGE, C., BROMMELS, M., ARONSSON, H. & THOR, J. 2010. Lean thinking in healthcare: a realist review of the literature. *Quality and Safety in Health Care*, 19**,** 376.

MCGRATH, B., WILKINSON, K. & SHAH, R. K. 2015. Notes from a small island: lessons from the UK NCEPOD Tracheotomy Report. *Otolaryngology–Head and Neck Surgery*, 153**,** 167-169.

MEHTA, R. L., KELLUM, J. A., SHAH, S. V., MOLITORIS, B. A., RONCO, C., WARNOCK, D. G. & LEVIN, A. 2007. Acute Kidney Injury Network: report of an initiative to improve outcomes in acute kidney injury. *Critical care*, 11**,** 1-8.

MITCHELL, C., LAWTON, E., MORTON, C., MCCAIN, C., HOLTBY, S. & MAIN, E. 2014. California pregnancy-associated mortality review: mixed methods approach for

improved case identification, cause of death analyses and translation of findings. *Maternal and child health journal,* 18**,** 518-526.

MORAROS, J., LEMSTRA, M. & NWANKWO, C. 2016. Lean interventions in healthcare: do they actually work? A systematic literature review. *International Journal for Quality in Health Care,* 28**,** 150-165.

MOUNTFORD, J. & SHOJANIA, K. G. 2012. Refocusing quality measurement to best support quality improvement: local ownership of quality measurement by clinicians. *BMJ quality & safety,* 21**,** 519-523.

MURRAY, E., TREWEEK, S., POPE, C., MACFARLANE, A., BALLINI, L., DOWRICK, C., FINCH, T., KENNEDY, A., MAIR, F. & O'DONNELL, C. 2010. Normalisation process theory: a framework for developing, evaluating and implementing complex interventions. *BMC medicine,* 8**,** 63.

NATIONAL INSTITUTE FOR HEALTH AND CARE EXCELLENCE (NICE). 2007. *Acutely ill adults in hospital: recognising and responding to deterioration* [Online]. Available: https://www.nice.org.uk/guidance/cg50 [Accessed 12/08/2020].

NATIONAL QUALITY BOARD UK 2017. National Guidance on Learning from Deaths: A Framework for NHS Trusts and NHS Foundation Trusts on Identifying, Reporting, Investigating and Learning from Deaths in Care.

NCEPOD 2018. Themes and Recommendations to all hospital specialties.

NEWTON, J., GRAHAM, J., MCLOUGHLIN, K. & MOORE, A. 2003. Receptivity to change in a general medical practice. *British journal of management,* 14**,** 143-153.

NHS IMPROVEMENT. 2020. *The Natiional Medical Examiner System* [Online]. National Health Service, UK. Available: https://improvement.nhs.uk/resources/establishing-medical-examiner-system-nhs/ [Accessed 03/07/2020].

O'REILLY, D., MCPHERSON, S., SINCLAIR, M. & SMITH, N. 2017. Lessons from a national audit of acute pancreatitis: A summary of the NCEPOD report 'Treat the Cause'. *Pancreatology,* 17**,** 329-333.

O'HARA, R., COSTER, J. & GOODACRE, S. 2021. Qualitative exploration of the Medical Examiner role in identifying problems with the quality of patient care. *BMJ open,* 11**,** e048007.

O'REILLY, P., LEE, S. H., O'SULLIVAN, M., CULLEN, W., KENNEDY, C. & MACFARLANE, A. 2017. Assessing the facilitators and barriers of interdisciplinary team working in primary care using normalisation process theory: an integrative review. *PloS one,* 12**,** e0177026.

OUTCOME, N. C. E. I. P. & DEATH 2017. Inspiring Change. NCEPOD London.

PATTON, M. Q. 1999. Enhancing the quality and credibility of qualitative analysis. *Health services research,* 34**,** 1189.

PETTIGREW, A., FERLIE, E. & MCKEE, L. 1992. *Shaping Strategic Change: Making change in large organisations, the case of the NHS,* Sage Publications.

POULTON, T., MURRAY, D. & TEAM, N. E. L. A. P. 2019. Pre-optimisation of patients undergoing emergency laparotomy: a review of best practice. *Anaesthesia,* 74**,** 100-107.

RAJA, S. & THOMAS, C. 2019. Standardising care of patients requiring non-invasive ventilation in response to NCEPOD–a chief registrar quality improvement project. *Clinical Medicine,* 19**,** 79.

RALEIGH, V. S. & FOOT, C. 2010. Getting the measure of quality. *Opportunities and challenges*.

RHODES, A., EVANS, L. E., ALHAZZANI, W., LEVY, M. M., ANTONELLI, M., FERRER, R., KUMAR, A., SEVRANSKY, J. E., SPRUNG, C. L. & NUNNALLY,

M. E. 2017. Surviving sepsis campaign: international guidelines for management of sepsis and septic shock: 2016. *Intensive care medicine,* 43**,** 304-377.

RITCHIE, J., LEWIS, J., NICHOLLS, C. M. & ORMSTON, R. 2013. *Qualitative research practice: A guide for social science students and researchers,* Sage.

SARI, A. B.-A., SHELDON, T. A., CRACKNELL, A. & TURNBULL, A. 2007. Sensitivity of routine system for reporting patient safety incidents in an NHS hospital: retrospective patient case note review. *Bmj,* 334**,** 79.

SCHEIN, E. H. 1985. Organisational culture and leadership: A dynamic view. *San Francisco*.

SEDGWICK, P. & GREENWOOD, N. 2015. Understanding the Hawthorne effect. *Bmj,* 351.

SELBY, N. M., HILL, R. & FLUCK, R. J. 2015. Standardizing the early identification of acute kidney injury: the NHS England national patient safety alert. *Nephron,* 131**,** 113-117.

SHENTON, A. K. 2004. Strategies for ensuring trustworthiness in qualitative research projects. *Education for information,* 22**,** 63-75.

SLADE, E., TAMBER, P. S. & VINCENT, J.-L. 2003a. The Surviving Sepsis Campaign: raising awareness to reduce mortality. *Critical Care,* 7**,** 1.

SLADE, E., TAMBER, P. S. & VINCENT, J.-L. 2003b. The Surviving Sepsis Campaign: raising awareness to reduce mortality. BioMed Central.

SMITH, G. B., PRYTHERCH, D. R., MEREDITH, P., SCHMIDT, P. E. & FEATHERSTONE, P. I. 2013. The ability of the National Early Warning Score (NEWS) to discriminate patients at risk of early cardiac arrest, unanticipated intensive care unit admission, and death. *Resuscitation,* 84**,** 465-470.

SPENCER, L., RITCHIE, J., LEWIS, J. & DILLON, L. 2004. Quality in qualitative evaluation: a framework for assessing research evidence.

SPILLETT, M. A. 2003. Peer debriefing: Who, what, when, why, how. *Academic Exchange Quarterly,* 7**,** 36-41.

STAKE, R. E. 2005. Qualitative case studies.

STANFORD, K. 2009. Underdetermination of scientific theory.

SUJAN, M. A., HUANG, H. & BRAITHWAITE, J. 2017. Learning from incidents in health care: Critique from a Safety-II perspective. *Safety Science,* 99**,** 115-121.

SULLIVAN, P. A. & ANDERSON, J. M. 2010. The Health Care Debate: If Lack of Tort Reform is Part of the Problem, Federalized Protection for Peer Review Needs to be Part of the Solution. *Roger Williams UL Rev.,* 15**,** 41.

SUSERUD, B.-O. & BRUCE, K. 2003. Ambulance nursing: Part three: Bjorn-Ove Suserud and Karin Bruce examine the hand over of patients to emergency wards.(feature). *emergency nurse,* 11**,** 16-22.

SWANWICK, T. & VAUX, E. 2020. *ABC of quality improvement in healthcare,* John Wiley & Sons.

THE ROYAL COLLEGE OF PATHOLOGISTS 2016. Medical examiners. The Royal College of Pathologists.

THE ROYAL COLLEGE OF PHYSICIANS. 2016. *National Mortality Case Record Review (NMCRR) programme resources* [Online]. Available: https://www.rcplondon.ac.uk/projects/outputs/national-mortality-case-record-review-nmcrr-programme-resources [Accessed 17/11/2016 2016].

THE ROYAL COLLEGE OF PHYSICIANS 2018. A review of the learning from the national mortality review programmes.

THOMAS, M. E., BLAINE, C., DAWNAY, A., DEVONALD, M. A. J., FTOUH, S., LAING, C., LATCHEM, S., LEWINGTON, A., MILFORD, D. V. & OSTERMANN,

M. 2015. The definition of acute kidney injury and its use in practice. *Kidney International,* 87**,** 62-73.

TIWARI, D., DONNELL, A., RENAUT, R., RICHARDSON, T. & ALLEN, S. 2020. Reducing hospital mortality: Incremental change informed by structured mortality review is effective. *Future Healthcare Journal,* 7**,** 143.

TRUST, S. W. N. A. C. 2017. Mortality Review Policy.

TURNER, P., WONG, M. C. & YEE, K. C. 2006. Understanding interactions of factors influencing clinical handover: insights for information technology. *HIC 2006 and HINZ 2006: Proceedings***,** 328.

UK RENAL REGISTRY (UKRR). 2014. *The Acute Kidney Injury (AKI) Programme* [Online]. Available: https://www.renalreg.org/projects/the-acute-kidney-injury-aki-programme/ [Accessed].

VORWERK, C., LORYMAN, B., COATS, T., STEPHENSON, J., GRAY, L., REDDY, G., FLORENCE, L. & BUTLER, N. 2009. Prediction of mortality in adult emergency department patients with sepsis. *Emergency Medicine Journal,* 26**,** 254-258.

WALSHE, K. 2007. Understanding what works—and why—in quality improvement: the need for theory-driven evaluation. *International Journal for Quality in Health Care,* 19**,** 57-59.

WALSHE, K. & OFFEN, N. 2001. A very public failure: lessons for quality improvement in healthcare organisations from the Bristol Royal Infirmary. *Quality in Health Care,* 10**,** 250.

WILLSON, A., HOPE, D. & SMITHIES, M. 2008. Welsh critical care improvement programme: the 'how to guide'for improving critical care rapid response to acute illness.

WOLFE, A. 2001. Institute of Medicine report: crossing the quality chasm: a new health care system for the 21st century. *Policy, Politics, & Nursing Practice,* 2**,** 233-235.

WRIGHT, J., DUGDALE, B., HAMMOND, I., JARMAN, B., NEARY, M., NEWTON, D., PATTERSON, C., RUSSON, L., STANLEY, P. & STEPHENS, R. 2006. Learning from death: a hospital mortality reduction programme. *Journal of the Royal Society of Medicine,* 99**,** 303-308.

YIN, R. K. 1981. The case study as a serious research strategy. *Knowledge,* 3**,** 97-114.

YIN, R. K. 2017. *Case study research and applications: Design and methods,* Sage publications.

ZWAAN, L., DE BRUIJNE, M., WAGNER, C., THIJS, A., SMITS, M., VAN DER WAL, G. & TIMMERMANS, D. R. 2010. Patient record review of the incidence, consequences, and causes of diagnostic adverse events. *Arch Intern Med,* 170**,** 1015-21.

CHAPTER 5: STUDY REFLECTIONS

5.1. Chapter Preface

In Chapter 5, I offer my critical reflections on the studies in Chapter 2, Chapter 3 and Chapter 4 . These reflections are on the PhD process. I reflect also on the possible research opportunities from these three studies. Finally, I then reflect on the overall integrity of these studies.

5.2. Reflections on the research process

Reflexive research practise involves two elements: interpretation and reflection.(Alvesson and Sköldberg, 2008, Finlay, 1998) The interpretation element acknowledges that any interpretation is produced by the researcher and indeterminately derived from the original data. Facts are not 'found' but engaged, understood, and interpreted by subjects. The reflective element requires the researcher to engage with his psychological, cultural, intellectual, and societal mores to conduct an "*interpretation of the interpretation.*" Before I start reflecting on my research process, I will begin with one reservation. The bias of post-hoc reflections are likely to plague any retrospective evaluation of one's methods and even intentions (otherwise termed hindsight bias). However, every researcher must begin outlining some of the processes that have come to influence their thinking.

5.2.1. Reflections on research conception

When I first heard of a possible research project around case-note review whilst still an administrative assistant, the topic did not immediately pique my interest as I had no previous exposure to reviewing case-notes nor its methodology. I had yet to consider the type of approach to make the project unique and my own. Yet, I became more interested when I did

consult the literature, especially around the error-susceptible processes of human decision-making. For the conceptual review (Chapter 2), there was a sampling phase which involved identifying cognitive biases and heuristics from the medical decision-making literature. These identified systematic reviews, which further led us to identify further supporting literature and their literature search terms. Once the search term saturation was reached (namely that no new search terms nor research terms were identifiable from new searches), the review team concluded that the search was sufficiently comprehensive for our purposes. In 2016, it became evident to me that case-note reviews were complex documents which varied in length and their quality of presentation. It was considered a field in which cognitive biases and heuristics had possible application. There was some plausibility in applying cognitive biases and heuristics, or a general psychological appraisal of the judgement of case-note reviews. Since then, I have come to better understand how to apply these cognitive biases to case-note review and healthcare quality improvement. However, I discovered during the consensus panel that the cognitive biases and heuristics were not as stable and distinct as I had first hoped they would be. There was considerable overlap between certain biases. For instance, the availability bias and the representativeness heuristic clearly inform one another as the latter is subject to the former in becoming available in the first instance.[52] The literature

[52] Given this overlap, an implicit assumption of the inquiry is that only one cognitive biases or heuristic would be operative at any one time for us to identify its activity. I do feel that this is a strong assumption and one far too strong to be pertinent for the daily complex decision-making of physicians. In fact, the judgements of clinicians mostly likely resemble a Bayesian approach, where new experiences (data) are used to consciously inform the cognitive schemas developed from prior experience.

represented biases of similar constructs but under different names, which causes can cause confusion for any developments of the concept in case-note review care quality judgements and medical decisions. (Shah and Oppenheimer, 2008, Blawatt Ken, 2016) There are currently no consistent and settled systematic taxonomy of classifying these biases and heuristics. There were such attempts, but no taxonomy has been ratified as authoritative and comprehensive. Furthermore, due to the novelty of the '*cognitive biases and heuristics*' paradigm, the panel of clinicians and health experts did not all, at first, fully comprehend the original psychological constructs of these biases and heuristics with their significance. However, this was not an insurmountable barrier as careful explanation of the definitions and the author's development of clinical scenarios helped to communicate any plausible effects of these biases.

The original definitions of these cognitive biases and heuristics were transformed into the applied definition. These applied definitions were then transformed into clinical scenario; one applied definition was tied to one clinical scenario. In panel terms, the members were diverse with reviewer methodologists, clinicians, inter-disciplinary behavioural psychologists, health system experts identifying and ranking the cognitive biases and heuristics and forming clinical scenarios. Their broad interests and specific expertise ensure the work did not systematically favour one professional opinion over another. The exercise involved inter-disciplinary engagement from behavioural scientists, clinicians and case-note trainers was auspicious for this task, as individuals could help others understand their different disciplines to broaden their own intellectual horizons to enable us to better understand some factors which likely contribute to the variability around care quality judgement of case-note reviews.

This leads us to Chapter 3 and the exploration of attitudinal and demographics factors influencing HiSLAC reviewers. My supervisors (RL, SMH) informed me about a large pool of reviewers for case-note review research purposes. Next, I considered the influence of reviewer attitudes on their care quality judgements (i.e., attitudinal measures).

5.2.2. Reflections on sampling and recruitment

In regards the quantitative study, the sampling had already been running and was conducted on subjects after the main study was complete. The sampling of validated instruments for measuring psychological risk constructs for reviewer completion involved me selecting, at first, plausible instruments for both their ease of completion and their previous use with medical professionals. The ease of completion became the preferred method of selection as very few instruments had been used on medical professionals, let alone validated with them in mind. With support from psychological experts, I whittled down a longlist of tools, which considered validated tools, two of which had not been validated in clinicians. These were shortlisted having consulted the HISLAC principal investigator and his team who developed with us clear reasons for their inclusion. One instrument had been validated with physicians (ambiguity due to uncertainty) and thus was automatically included whilst the others I had to develop hypotheses for given their previous use and validation details (need for cognition, personal need for structure). The analysis indicated that several of the hypotheses were sufficiently powered to determine a significant result. Recruiting the subjects for the quantitative study was done through the study investigators. The existing communication channels with reviewers were re-opened for their completion of the attitudinal measure survey.

In regards the qualitative case study, sampling involved interviewing the two key informants from the hospital trust to identify possible interviewees. These interviewees were asked to recommend possible interviewees. This was the optimal strategy given that not all staff were involved in the case-note reviewing process. This process continued until I perceived data saturation had been reached. Data saturation is when no significant insights are discovered from new data.(Fusch and Ness, 2015)

I recruited most participants through Trust managers who were board members which could possibly be coercive.(Qu and Dumay, 2011) A grassroots bottom-up recruitment procedure would have been interesting to undertake, however, it was not possible under these circumstances given the sensitive nature of the topic and importance of first engaging with healthcare leaders in the organisation for consent. The consideration of power differences that gatekeepers hold over other staff members may be present as previously shown in large organisations.(Khatri, 2009) I was forthcoming about my non-clinical background to minimise any power differences.(Karnieli-Miller et al., 2009) The qualitative study recruitment proceeded smoothly with data saturation self-evident during the 20th interview; I continued to follow through with the interviews arranged up until that point. I was supported by the supervisory team in my decision to stop interviews. I took care to apply active listening skills to clarify any hidden meaning of my supervisors words to understand the reasons for these implications.(Louw et al., 2011)

5.2.3. Reflections on data collection

The quantitative data collection required that several reviewers were sent several follow-up emails to encourage them to complete the survey, which added a few more weeks to the data

collection stage. These follow-ups were facilitated by the HiSLAC investigators. Special thanks must go to Gavin Rudge and Jianxia Sun for being available to answer my queries about the original HISLAC dataset. With the HiSLAC attitudinal survey, we were overwhelmed and delighted that 72 of the 79 reviewers (>90%) had completed their surveys. I had no face-to-face contact with reviewer participants and was only contactable if there were any issues or specific queries with the survey.

The qualitative data collection proceeded smoothly. During the interviews, I was forthcoming to participants about my own relationship with this study and the objectives of this study in relation to the PhD itself. I assumed I had clearly communicated the research rationale to participants when they no longer had questions for me. An important feature of conducting qualitative interviews is building rapport. During and after the interviews, I would establish trust by being respectful, empathic, light-hearted and humble in their opinions and thoughts.(Ritchie et al., 2013) I found I was able to offer much time to this process and it is something, which the data may not explicitly report, but has influenced the quality and credibility of the qualitative data. To my surprise, there were interviewees who were very keen to express their views. Reflecting on this, I sense that this may have had a disproportionate impact on my analysis (coding and thematic analysis stages). However, qualitative analysis does not emphasise the quantity of expression for the quality found in the content. The interview conditions were mostly optimal for clear and uninterrupted conversation; there was one instance where a small, busy café was the place for an interview. This participant selected this location and so it was decidedly considered fit-for-purpose. There was much background noise, and I would often ask for clarification from the interviewee. I used a secondary voice recording device to ensure fidelity to the interviewee's words. I did think twice about committing to interview in public locations with relatives,

carers, and other healthcare staff in the audible vicinity. However, I considered it unimportant given the very few people who would be concerned about this niche topic and that if the interviewee had selected the location that it was very likely to not be an issue. In the end, I adopted a spirit of transparency and rapport-building with the interviewee by trusting in their own good judgement on the matter. Despite the interruptions and noisy backdrop, the interviewee was relaxed and spoke at length about case-note reviews. These interruptions, fortunately, did not in the end affect the data quality. However, I have come to terms with qualitative research as a journey rather than a rigid process.(Thomas and Magilvy, 2011)

Given that more information is conveyed through one's body language than one's words, I now reflect on my communication style.(Mehrabian, 1972) As a novice qualitative researcher, I found following all the guidance difficult at first instance. During pilot interviews for instance, I found it instinctive to convey my assent (or dissent) to their ideas and thoughts, by uttering words such as "Ok", "I see" and "Good" as a form of encouragement to allow the participant to speak more freely. However, I realised at the 5th interview that this could be encouraging the participant's dialogue along certain lines, rather than allow them to convey their personal views. This concern was identified by my supervisors and through written transcripts of the recordings which showed that often the discussion was closed too quickly to the detriment of the expression of the participant's actual views. I also had the minor regret of not pressing my participant's further to disclose more information behind their words. I did not find that this significantly impacted the results when I did decide to press further for more information. I feel employing a consistently greater deep of questioning could have yielded further relevant data towards triangulating the findings especially concerning unanswered questions over "poor mortality surveillance committee meeting attendance." After receiving their feedback, I started to use more open

questions and fewer closed questions to promote the discussion and clarification of their ideas. Closing remarks and questions were used only when time was short or sufficient data had been collected to divert to another topic. My researcher judgement has improved since the first interviews but, of course, it could be improved further with more practise and exposure.

5.3. Reflections on data analysis and theory building

5.3.1. Case-note reviews as a "historical" document

In hindsight, I reflect on the nature of case-note reviews. During the research process, I did not have a theory that stated the nature of case-note reviews. The general definition is that it is a compilation of information retrieved concerning the patient's health condition and care given during time in a health institution. However, this definition does not give us any insight into the nature of who creates the notes, who can access the notes, how a judgement is made from it about the patient, what would make its way into a case-note and what would simply be left out? Seeing as the psychology of the case-note reviewer and the reviewing and relatable to the thesis itself process is the focus of Chapter 2 and Chapter 3 , it would have been helpful to characterise the interaction reviewers had with the case-notes as a type of historical document. It seems quite natural to consider case-notes as a type of historical text. In brief, the case-note is a document containing past information. And so, this makes it an historical artifact.(Collingwood and Collingwood, 1994) It has been created by health professionals and administrative staff. How case-note reviewers come to their care quality judgement or decision is based on how they have reviewed the information contained in this case-note.

Historical documents have special properties.[53] Delving a little more deeply into the historical nature of case-note reviews could help us understand case-note reviews. First, I adopt the characterisation of history according to R.G. Collingwood's conception. I assume the case-note reviews *can* convey to case-note reviewers the actual care given to the patient. I call this '*conveyance*.'[54] Otherwise, the case-note reviewers nor anyone else could take the case-note seriously if it were a purely fabricated document. So, it is assumed to communicate some truth about the care quality received by the patient. The question remains for researchers to explore the factors that limit this communication of the true care quality.

Second, as the case-notes are the thoughts of people placed into a document, I assume the reviewer *can* obtain information of these very thoughts and ideas to *re-enact* in their mind the care quality given to the patient as captured in the case-note.[55] It is the process of re-thinking what others first thought. I call this '*retrievability*,' as you will need to be able to faithfully represent the thoughts of predecessors to derive an accurate care quality judgement.[56] If you

---

[53] This is done so now as it was perceived after the research studies had commenced.

[54] If this were not the case, any interpretation concerning the case-note would be acceptable. This is simply not accessible as a definite care standard was enacted upon a patient, and not many variations of care. It is the aim of the case-note reviewer, as historian, to unearth the *true* care quality given.

[55] Or else the act of reviewing incomplete case-notes and searching for further information is futile.

[56] The correct care quality judgement remains elusive given the innumerable factors which likely influence a case-note reviewer's care quality judgement from internal psychology to

could not obtain the information required to re-enact, then most likely, your interpretation concerning the information in the case-notes would be, at best, misleading and erroneous, and at worst, conducive to harm. And if you could not "re-enact" well in your mind, due to a physiological impairment or bias,[57] then there is research needed to understand the nature of these factors which impair this ability to "re-enact" the actual care quality.

Third, I assume it is possible for case-note reviewers to suspend, to some extent, their motivational and epistemic attitudes when evaluating the care quality of the case-notes.[58] I call this '*composure.'* Without composure, this would entail that the historical process is construed solely upon the reviewer's motivations, epistemic and cultural concerns and not the features intrinsic to the case-notes.[59] In sum, the historian [case-note reviewer] misreads their own motivations in place of the historical subject [case-note]. Simply put, it would be a fabrication. *Composure* essentially relates to the integrity of the reviewer's cognitive

---

external environmental cues. The list is endless, but the quest must go on, for the patient could have been poorly dealt with. The desire for truth and justice for the patient ought to be the prime motivator for case-note reviewers.

[57] This also includes cognitive biases and heuristics.

[58] This has implications for availability of mitigation strategies to combat the influence of certain undesirable influencers which the reviewer most likely brings. I refer to these influencers loosely, for they are legion. Specifically, as a care quality judgement is a cognitive activity, later in the thesis, I explore the plausible influence of cognitive biases and heuristics on reviewer care quality judgements.

[59] Environmental pressures and nudges could impose themselves upon reviewer judgements.

processes for sound reasoning and judgement. All these elements bear further investigation. Research is needed to understand the significance of motivations and their effects upon case-note reviewer care quality judgements.

This would have revealed possible distinctions concerning the nature of case-note review and how the reviewer's mind could interact with case-notes during their evaluations of care. At first glance, conveyance, retrievability and composure are not practicable concepts. However, they are crucial for understanding case-note reviews as a document which is perceived and judged by a reviewer's mind. This area remains open for further practical developments.

5.3.2. Quantitative

With respect to quantitative elements, I would like to reflect on the process of data analysis and theory building. The use of multi-level modelling was appropriate given the nested nature of the reviews; the case-notes were nested within reviewers.(Van den Noortgate et al., 2005) We considered that the reviewers provided their own 'variance' to the global care quality judgement (and confidence in global care quality judgement) and that multi-level modelling was appropriate for separating the sources of this variance and by a host of independent variables (age, gender, length of stay etc.). The analysis was first undertaken by Professor Hofer with my contributions to the entire write-up. I have made every effort, challenging as it has been, to understand each line of code and the multi-level design of the research. Professor Hofer has helped to answer my queries and guide me through the process. The analysis was robust as the hypotheses were pre-specified by me and supervisors before testing. I scoured the literature to find plausible hypotheses worth testing given the variables

list we had. In the end, the completeness of the HISLAC dataset limited which hypotheses we could use.

5.3.3. Qualitative

I faced a challenge in co-ordinating the large amount of data produced from the qualitative data. I had amassed more than five hundred pages of transcripts, documents, and fields notes. With support from lectures, books, field guides, supervisors and most of all, time spent with the data, I could negotiate difficulties qualitative researchers often face in organising the data.(Flinders, 1997) I have since come to appreciate and relish the special perspective qualitative research offers.

As I have an analytic thinking style, I enjoyed using a computer-assisted qualitative data analysis software (CAQDAS) software package to manage the different file types and keep the data organised to enable me to approach the problem in an organised manner. I would identify a code or theme and seek to find others of its kind to justify the creation of a new category. This reduces personal anxiety as I tend to feel facing a wealth of unanalysed data. These categories are strengthened or weakened as the data is processed. Ultimately, this allows me to compare the codes and themes, even across methodologies (NPT vs. framework), for emergent features of the data I may have missed from the first approach. I was mindful that these categories may be far removed from the reality and thus employed, at all times when handling or considering the data, a self-critical reflexive stance and engaged with my supervisors for their corrective input.(Freeman, 2016) The NVivo function of key word searching and use of multiple windows allowed me to explore a range of possible and

plausible connections across the data. I was supported in this endeavour by NVivo, a computer-assisted qualitative data analysis software (CAQDAS) package.

Whilst transcribing the data, I adhered to an *ad verbatim* approach to remain as faithful to the actual words of the interviewees and minimise any influence from the researcher. Frequently, I found new ideas spring to mind which I noted down in my 'reflections' journal. I found this did not disturb my transcription flow but allowed a sense of perspective through regular engagement and disengagement from the data. In like vein, Richardson indicates that qualitative analysis concerns not only the discovery (and categorisation) of new phenomena but also the discovery of the researcher's relationship with this phenomena.(Richardson, 2003)

In developing these categories and continuing to collect data, I found that I became more boxed in by my own pre-conceptions of the data. These preconceptions were set which allowed me less latitude in terms of considering alternative concepts relating to the qualitative data. I feel, at least, that the conversations went down well-trodden paths. I had little to no space to manoeuvre another interrogative strategy to thoroughly test some of the rival hypotheses. Nevertheless, I reflected on the data and considered the bespoke and formalised nature of case-note reviews and concluded that data saturation was the most likely reason for the rapid exhaustion of new concepts. It was helpful to engage with the wider qualitative research team to test and develop ideas.

The final stage of the analysis aimed to provide an explanatory account of the data. This proved the most challenging aspect as I had become familiar with the data and the categories I had used. This stage demanded reflexivity to distinguish between the interviewee and my

contribution to data interpretation. It was critical at this stage to have involvement from seasoned qualitative researchers (RM, EF, and Antje Lindenmeyer) as guides and critical examiners of my work, to date. It has been noted that the oversight of qualitative research requires the direction of experienced researchers, to whom I am very much indebted. (Gale et al., 2013, Pope and Mays, 2006)

There were two methodological alternatives which could have helped in this case-study. First, from the documents, articles, and reports, this could have been developed out of a diverse set of resources. Upon reflection, if space, time and resource allowed, first I would have implemented a realist evaluation approach in determining "*what works, for whom, under what circumstances, and why?."*(Pawson, 2006, Greenhalgh et al., 2008) It is a method where the case study research (qualitative) and the research of individual reviewers (quantitative) can be merged into an account detailing the likely processes involved around QI from case-note reviews.

Second, if ethics and the hospital staff consented to it, I would have undertaken an ethnographic study which would have examined the inner workings, the culture, and the deeper structures of the organisation. This would have provided plenty of informative, *thick* descriptions to better understand how health interventions work within organisations.(Morgan-Trimmer and Wood, 2016) This would have allowed me to pursue some existing questions such as "*why was the mortality surveillance committee meetings cancelled month due to poor attendance levels?*" The use of participant observation would have breached the service evaluation ethics agreement with the hospital Trust because participant observation was far too invasive. Ethically, observing these clinical reviewers required access to clinical spaces which are privy only to Trust clinicians. And there was

sufficient focus and depth from the tailored semi-structured interview questions derived from the informant interviews and from the supporting field notes made.

Thirdly, I would have taken a more whole systems approach. Linear systems[60] tend not to consider the complexities of their interventions, in the very least, not map out the interactions that lead to particular outcomes, at small or large scales.(Peters, 2014) An approach called a '*systems approach*' could help to illuminate some of the factors at multiple levels. Single level analysis (i.e. at one level of a system such as reviewer psychology or decision-making or organisational) which is a holistic approach that focuses on the way a system's constituents interrelate and evolve over time(Erber et al., 1995), often in non-linear ways, which is addressed in the forthcoming sub-section.(Meadows, 2008) This is important for appraising well-defined linear research areas to closely derive the nature of the phenomena that is subservient to the larger whole. But one must be aware of the limitations of linear, mono-disciplinary perspectives that fail to consider the whole, as systems approaches do.(De Savigny and Adam, 2009)

Reflecting on the case-study, I would have liked to compare some of these different methods of interviewing and explored the less data-saturated qualitative themes. However, I realise the case-study methodology is adaptive to each new context and thus my expectation that method A is used in case A (i.e., our Trust), does not transpire that method A ought to be used with

[60] I have mentioned 'systems' as a category of analysis for the research case study, however, this has not been developed in the 'systems thinking' tradition. This is due to the researcher's lack of foresight, time, and resource with this method.

another case-study B. With the case-study undertaken and accruing more information, interviewing key informants, and obtaining consent and other materials, I envisage this could have been less similar in another case study than in my own. In short, I sense that the order in which I engage with one case may be different in method as I respond to the actual daily operating and structures of the organisation itself. Furthermore, the predominant data source are interviews and thus the methods will evolve in response to the available data. Case-study designs are undertaken not for their generalisability, which is a quantitative feature, but for the uniqueness and context-bound nature of that single case. Undertaking multiple parallel case-study designs may appear to invoke the comparability of the case-studies, to an extent, however, each case-study will have its own culture, beliefs, operations, functions which are only explainable on its own terms. Having not undertaken qualitative research prior to this research degree, I have reflected on the strength of its method to explain the single object or case of interest far more deeply and precisely than I had anticipated given my scientific background. I have learnt that the methodological rigour of qualitative research is sufficient to explain social phenomena.

CHAPTER 6: DISCUSSION

if the work of inquiry is to be carried on, it must be at once scientific and philosophic, that if, in particular, the scientist is not philosophic, he will fall into confusions, he will rebuff philosophic criticism-he will lack a theory of categories, of sorts of problem, of' method '-especially he will be carried away by practical interests, by interest in producing something or implementing a programme instead of in finding something out.

John Anderson, 1962, p. 183).(Anderson, 1962)

6.1. Introduction

Hospital care quality needs to be measured accurately and reliably for two reasons. First, to ensure that good quality care is indeed good care quality. Second, to ensure that changes to care can be well tracked to ensure there is "true improvement" and not the mirage of improvement. The measurement of hospital care quality is not a simple task as there needs to be good fidelity of any measuring instrument and the real construct of care quality. Hospital mortality statistics are one method to capture this quantity, however, they are seriously flawed by the constant risk fallacy and its questionable case-mix adjustment models.(Girling et al., 2012a, Nicholl, 2007) Case-note reviews are a common method to evaluate the care patients have received in hospitals. However, there is a substantial amount of disagreement between case-note reviewers evaluating the same set of case-notes.(Lilford et al., 2007, Goldman, 1994, Hutchinson et al., 2010c, Benning et al., 2011, Goldman, 1992, MacKenzie et al., 1992) The main source(s) of the disagreement are unknown. Chapter 2 and Chapter 3 had the aim of identifying the sources of disagreement at the reviewer level. At the organisational level, the exact uses and processes surrounding case-note review in hospitals has not been widely explored. Despite there being some promise of quality improvement

from case-note reviews in one study(Kobewka et al., 2017), there was the internal processing and perception of the purpose of case-note reviews within the hospital setting. Thus, I undertook the hospital case-study to explore the barriers and facilitators to case-note review embedding, its information flow and its perceived quality improvement through semi-structured interviews and documentary evidence. This thesis includes three studies which were designed to explore separate research objectives pertaining to the overall thesis.

A strength of this work is that the validity of the case-note reviewer care quality judgements and their organisational dynamics are studied closely. The fact that I have done mixed methods by exploring both the validity and the practise of case-note review QI enables me to reflect on the future trajectory of case-note reviewing for quality improvement processes. This helps preserve the story and ensure that key insights have not been omitted from my thesis. A detailed historiography of this research area can help to piece together a succession of event and research-specific discoveries, not necessarily on a causal pathway, but which map the past conversations which have shaped these theories. This is important to establish as ideas in their right and proper context could offer insights into future individual differences as influencers of case-note reviewer global care quality improvement.

This Chapter 6 is organised as follows (Table 21). First, there is a summary of the objectives and how they have been met. Second, there is a detailed summary of each chapter and its findings. Third, policy implications of this thesis are presented. Fourth, any opportunities for future research, which are linked to these implications, are presented. This information included in this chapter is summarised in the following research objective's table. This table forms the blueprint for the chapter's final discussion of case-note review validity and use

Table 21. Research objectives with their findings and implications

Research Objectives	Key findings and their implications
• Explore the measurement characteristics of preventable hospital deaths in high-income countries *(*Chapter 1*)*	• The literature provides limited information about the measurement characteristics of preventable deaths, suggesting that substantial numbers of case-note reviews may be needed to create reliable estimates of preventable deaths at the individual or hospital level. (Chapter 1 Results and Discussion) • Any program to detect care quality concerns within the mortality case-note reviews would require population-specific estimates of reliability to account for local variability. (Chapter 2 Discussion) • Preventable death rates are low, which is likely to make it difficult to use SMRs based on all deaths to validly profile hospitals. (Introduction, Chapter 1 Results and Discussion) • There is little information to guide improvements in the measurement procedures, and so more developments in this area are needed from both

	practitioners and methodologists to inform the best use for mortality case-note reviews. (Chapter 2 Discussion)
• Examine the plausible influence of cognitive biases and heuristics on case-note reviewer care quality judgements. *(*Chapter 2*)*	• Several biases, *ambiguity intolerance, outcome bias, availability bias, confirmation bias* and *omission bias* are likely the most important. (Chapter 2 Results) • The impact of these biases likely extends to patient care, clinician professional development. (Chapter 2 Discussion) • There are possible implications for medical regulation. The role these cognitive biases and heuristics may play across clinicians in general and in their different specialties is currently unquantified. In future, medical regulators could consider the activity of these biases and heuristics in their assessments and (re-)validation examinations, especially if the effects were both significant and malign. (Chapter 2 Conclusion)
• Identify the influence of case-note reviewer demographics and select attitudinal measures and	• Reviewers contribute one fifth (21%) of the total variation to the variability observed around care quality judgements. (Chapter 3 Results)

case-note features upon case-note reviewer care quality judgements and their care quality confidence judgements *(Chapter 3)*	• This reviewer variability is not explained by variables included in our models. Further research is needed to identify non-reviewer and reviewer sources of this variability to improve reviewer care quality judgement reliability. This implies that there are other sources for this variability which need to be explored further. (Chapter 3 Results and Discussion)
• Examine the potential quality improvement potential and any mechanisms from the use of case-note reviews in hospitals *(Chapter 4)*	• There is some clear evidence that mandatory mortality case-note reviews lead to hospital quality improvement, but case-note reviews in general have found few contributions to learning for QI. Thus, more research is needed to understand how and why these instances of learning appear where they do, and how case-note review methodologies can be tailored in future to better detect these "*rare*" learning events. (Chapter 4 Results) • Information flow from board-to-ward is better than the flow from ward-to-board. (Chapter 4 Results)

	• There are both conflicting and complementary purposes for case-note reviews perceived by hospital staff and board members. (Chapter 4 Results)
• Critically discuss study findings, limitations, and the research process *(Chapter 5 and 6)*	• More research with clinicians, psychologists and policymakers are needed to empirically confirm the influence of cognitive biases and heuristics in case-note reviewer's care quality judgements (Chapter 2 Discussion) • ~20% of the total variability around the case-note reviewer care quality judgement are not attributable to the study variables. More research is needed to identify these variables which factors are at play (Chapter 3 Results) • Mortality case-note reviews lead to care quality improvement but there was little tangible evidence given the small number of identified issues; historically, case-note reviews contribute to learning when there is a pro-

	active safety specialty culture, board support and the presence of a complex-presenting group of patients. (Chapter 4 Results) • Different research methodologies and replicate case-study designs are needed to confirm care quality improvement within individual specialties. (Chapter 4 Discussion)
• Implications for stakeholder groups *(Chapter 6)*	• Case-note reviewers could engage with our listed demographics and attitudinal factors and consider further what other factors could influence mortality case-note review care quality judgements. More specifically, case-note reviewer should bear in mind other likely factors that contribute to the reviewer-specific contribution to the variability in the global care quality judgement (see Chapter 3 Results). • A healthcare leader or manager would benefit from considering the many intra-organizational factors which affect the potential for quality improvement from mortality case-note reviews (see Chapter 4 Results)

	• Given the need to review mortality case-notes is mandated, healthcare leaders and healthcare policymakers could consider how, in future, to respond in implementing case-note review in every Trust successfully. It is in the best interest of every Trust to implement case-note reviews well. (Chapter 4 Results and Discussion) • Given the inordinate resources and staff time required for highly reliable replicate case-note reviews, behavioral scientists, psychologists, case-note reviewers, reviewer methodologists, hospital staff supporting in QI and QI policymakers could undertake more research together. This research could help us all to understand more sources of this reviewer-specific contribution to the variance around care quality judgements. This research could help identify the real-world contribution of our identified cognitive biases and heuristics in both this reviewer-specific contribution to the variance around care quality judgements and the

	inter-reviewer reliability (kappas) from the extant literature (see Chapter 1, Chapter 2, Chapter 3) • The behavioral scientists, psychologists, case-note reviewers, reviewer methodologists, hospital staff could work together to identify practical approaches to mitigate cognitive biases and heuristics in case-note reviewers in their specific contexts (see Chapter 2 Discussion, Chapter 4 on “receptive contexts”)

The main aim of this work has been to identify some causes of inter-reviewer care quality judgement variability and the barriers and facilitators for the use of case-note reviews in a hospital Trust. In short, the validity of case-note reviews and the use of case-note reviews are the main topics of discussion in this chapter.

In summary and on the validity of case-note reviews, I hypothesised that there would be two likely sources for the moderate-to-high inter-rater reliability across case-note reviewers for the same batch of case-notes. First, I appealed to cognitive biases and heuristics in a conceptual review and panel consensus of biases and heuristics likely to influence their care quality judgements. Second, I appealed to the influenced of attitudinal constructs on reviewer care quality judgements and their confidence in this judgement. I found that no attitudinal measures significantly influenced care quality judgements and that approximately 20% of the variation in care quality judgement scores were unexplained. And third, on the use of case-note reviews, the service evaluation was undertaken in a case-study to identify the barriers and facilitators to mortality case-note review embedding, information flow and any quality improvement, to which there was some clear evidence demonstrating quality improvement from mortality case-note reviews.

In this chapter, I discuss how successful this PhD has been in addressing the Research Objectives

As there has been a sustained focus upon the psychology of the reviewer in Chapter 2 and Chapter 3 , I will seek to discuss the implication this PhD has for the validity of case-note

reviews. Thereafter, I discuss the implications this PhD has for the use of case-note reviews. And last of all, I conclude by offering some areas of future research.

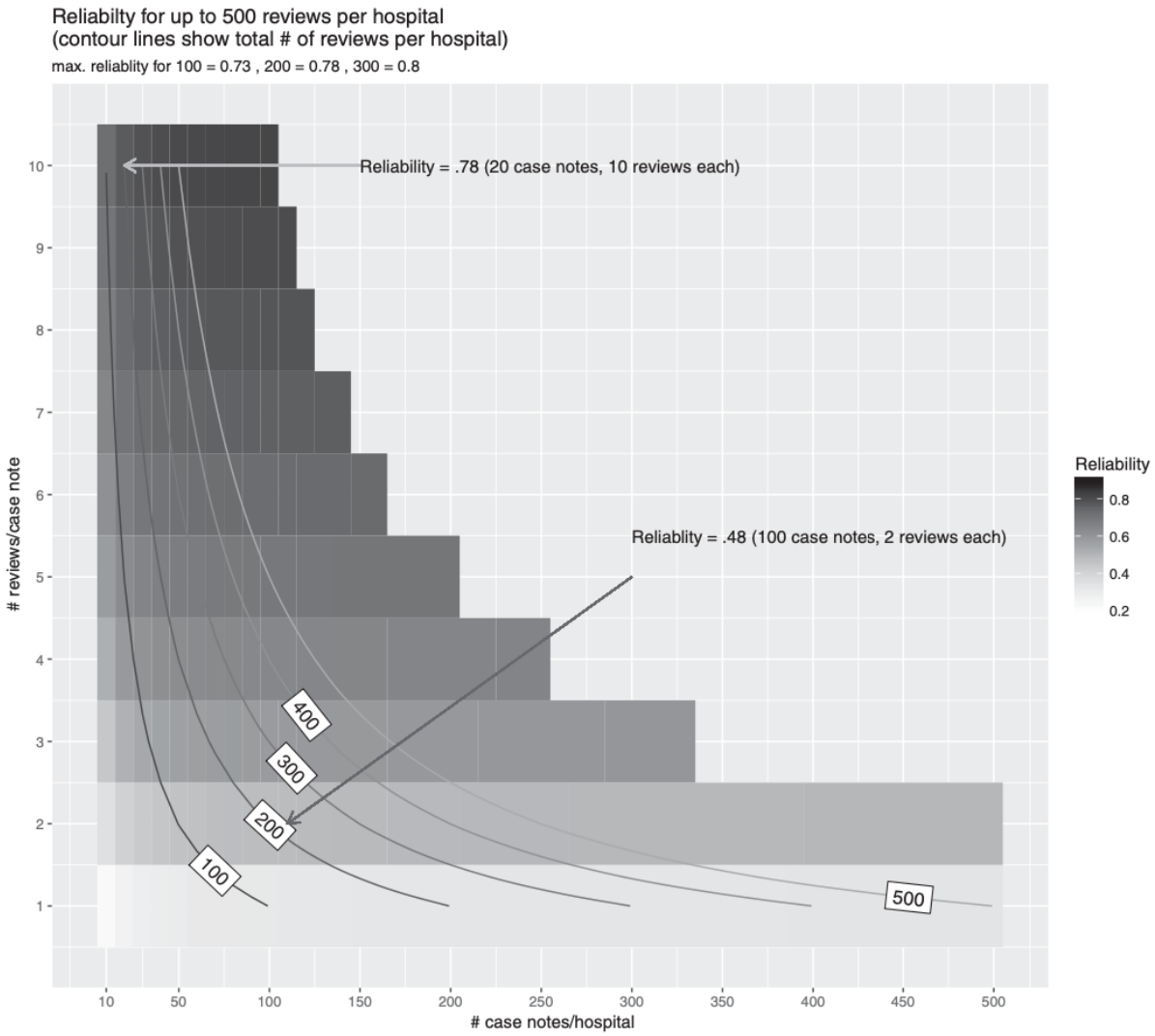
6.1.1. Objectives and their implications

6.1.1.1. Objective 1: Explore the measurement characteristics of preventable hospital deaths in high-income countries

This objective directly addresses the validity of case-note reviews. Chapter 1 found methodological weaknesses with hospital mortality statistics as a care quality measurement instrument. The case-note reviews were considered as a proxy measure for hospital care quality (Chapter 1).(Manaseki-Holland et al., 2019b) We conducted a systematic review of the literature supplemented by a re-analysis of authors' previously published and unpublished data and measurement design calculations. We conducted initial searches in PubMed, MEDLINE (OvidSP), and ISI Web of Knowledge in June 2010 and updated them in June 2012 and December 2017. Eligibility criteria included studies of hospital-wide admissions from general and acute medical wards where preventable death rates are provided or can be estimated and that can provide interobserver variations. The use of standardized mortality rates (SMRs) to profile hospitals presumes differences in preventable deaths, and at least one health system has suggested measuring preventable death rates of hospitals for comparison across time or in league tables. The influence of reliability on the optimal review number per case-note or hospital for such a program has not been previously explored (Figure 17). Figure 15 illustrates how the reliability changes as the numbers of reviews and reviewers per hospital vary in distinguishing one hospital from another. Holding the total number of

reviews constant, increasing the number of reviews per case increases reliability (e.g., 10 reviews per case for 30 cases) more than selecting more cases per hospital (e.g., 150 cases per hospital with two reviews per case). This implies that of a batch of 100 case-notes, ten independent case-note reviews are needed for each of the 20 distinct case-notes to yield a moderate-to-high reliability of kappa of 0.78 in determining the true care quality; in practical terms, better reliability is achieved with more reviews per patient and fewer patients (case-notes) overall. In practice, the hospital needs to balance its resources against number of replicate reviews that it can afford to undertake to reliably differentiate between hospitals. This is, of course, assuming the implementation of a summative system for the comparison of hospitals is in place, and that it is desired by all stakeholders from clinicians to healthcare policymakers and lobbyists. In the instance it is not desirable and used for formative purposes, these reliability estimates can continue to inform the reliability of single replicate reviews.(Manaseki-Holland et al., 2019a)

Figure 16. Reliability for up to 500 reviewer per hospital



Estimates for preventable death rates using implicit case note reviews by clinicians are quite low, suggesting that SMRs will not work well to rank hospitals, and any misspecification of the risk-adjustment models will produce a high risk of mislabelling outliers. Hospital level considerations have been considered; however, we expect there are reviewer-specific sources of variability in any global care quality judgement. Chapter 2 and Chapter 3 have explored the reviewer variability to explain this observation by identifying the potential sources of this reviewer disagreement.

The findings of this first objective have several implications. In practice, if a summative scenario were adopted, then the hospital clinicians would need to agree the

proportion of their time devoted to case-note reviews and decide how to prioritise their most important clinical work considering the significant time outlay needed for these case-note reviews. If a formative approach for reviews were adopted, the reliability of a single review will need to be calculated through a calibrated set of case-notes; thenceforth, the reliability level needs to be set (e.g., between 0.8-0.9) and then the number of replicate case-note reviews per case-note can be undertaken within that healthcare organisation. The more interesting question from all this is intra-organisational decision-making for their use of case-note reviews. This is typically explored only through qualitative research methods. This can include case-study and ethnographic work. The hospital with its healthcare policy context will have a role to play in whether this protocol is implemented, and if it is implemented, whether it is a summative or formative approach.

6.1.1.2. Objective 2: Examine the plausible influence of cognitive biases and heuristics on case-note reviewer care quality judgements (on the validity of case-note reviews)

This objective directly addresses the validity of case-note reviews. The first objective concerned the identification of cognitive and psychological mechanisms that plausibly contribute to the inter and intra-rater disagreement(Lilford et al., 2007, Goldman, 1992, MacKenzie et al., 1992, Goldman, 1994, Hutchinson et al., 2010c). Cognitive biases and heuristics were explored for this purpose of shedding some light on the sources of this inter-rater disagreement. This was encapsulated by "heuristic/biases" which commenced with a systematic search strategy to identify cognitive biases and heuristics that operate during medical decision making. This review yielded one study. Due to this paucity in the literature on this topic, I took the initiative to conduct a conceptual review with a diverse panel of

clinicians, review methodologists and psychologists to identify and rank cognitive biases and heuristics plausibly influencing case-note reviewer global care quality judgements. In descending rank order, we found that '*outcome bias'*, '*ambiguity aversion'*(Fox and Tversky, 1995), '*availability bias's*(Tversky and Kahneman, 1974), '*omission bias's*(Spranca et al., 1991) and *confirmation bias's*(Ross and Anderson, 1982) were considered by the consensus panel to be the most likely to influence reviewer global care quality judgements, in the opinion of clinicians and behavioural scientists. These plausible biases need to be tested empirically (Table 22.)

Table 22. Bias plausibly influencing care quality judgements judged by consensus panel

Biases	Percentage of Raters Reporting "likely" or "very likely" (%)
Ambiguity Intolerance	89
Outcome Bias	89
Availability Bias	78
Confirmation Bias	78
Omission Bias	78
Affect Heuristic	56
Bandwagon Effect	56
Commission Bias	56
Optimism Bias	44
Representativeness heuristic	44
Ambiguity Aversion	33

Anchoring Bias	22
Loss/gain framing bias	22
Order Effects: Primacy/recency bias	22
Sunk-cost effect	11

There are several implications for this process. As cognitive and biases can influence the care quality judgements, there needs to be empirical evidence of this during the case-note review care quality judgement procedure. Given the known inter-rater disagreements, there would need to be psychological, cognitive, or behavioural experiments which can elicit the action of these cognitive biases and heuristics. Assuming that some of our identified biases/heuristics are operative during case-note review care quality judgements, they can affect each reviewer in their specific context differently. This is not to say that it is entirely unpredictable, but that the influence of the bias/heuristic is likely a function of both the reviewer-intrinsic and external-environmental factors. If clinicians think these biases and heuristics are most likely operating in their environs and are subsequently compelled to learn more, the next step would be to have greater awareness of these heuristic/biases. However, it is simply not enough to know about them, but to develop strategies to identify and mitigate their action. Unlike addressing diagnostic decision-making, like in this one study which employed clinical vignettes and a knowledge intervention decreased the susceptibility of physicians to the availability bias(Mamede et al., 2020), the task of debiasing case-note review decisions is broader still than just diagnostic decision-making. There is this but also the case-note design and the case-note proforma design to consider as alluded to throughout Chapter 2. The implications for care quality judgements are currently an undefined quantity, but there is action which can be taken to raise the awareness and allow clinicians to develop

their own strategies to combat the more conspicuous biases we have shortlisted. Though the exact influence of specific heuristic/biases is unquantified, medical decision-making has been exhibited in medicine. Such debiasing strategies include cognitive/behavioural interventions involving education and training. Motivational training could help clinicians and staff remain more accountable to their profession and their patients. There are technological tools which could be employed within the clinician workflow, such as memory checklists and decision algorithms.(Hammond et al., 2021) Memory checklists can help remind reviewers of the key steps during the review process. Decision algorithms can help reviewers follow through routine-checking over the notes to ensure care has been provided for different patient types and presentations. These kinds of technologies could expedite effective decision-making for patients. This area can be explored further within their own individual settings. Artificial intelligence is a possible avenue to modify build on the decision algorithm work; this is discussed in a later section.

There are also implications for what a heuristic-friendly culture could achieve for reviewer care quality judgements and quality of care concerns. Some of the possible actions in responding to this include discussing, promoting, and funding research on heuristics and its application in the medical context. This includes a common understanding and conceptualization of what heuristics are, how they operate, and which methodological tools we need to develop to study and critically evaluate them. Formal rules, often neglected by practitioners due to the lack of transparency, can be translated into clear and handy decision tools during case-note review that not only facilitate applicability, but also convey the underlying rationale of the tool and help the practitioner to reflect on the existing tools they use.(Bodemer et al., 2015) For instance, the case-note review proforma itself, which is used in conjunction with the case-note review process, can be subjected to re-design for this

heuristic-friendly culture. This is a possible action of the work; the design of the review proforma is typically not scrutinised in this way to minimise the influence of the action of heuristic/biases.

6.1.1.3. Objective 3: Identify the influence of case-note reviewer demographics and select attitudinal measures and case-note features upon case-note reviewer care quality judgements and their care quality confidence judgements (on the validity of case-note reviews)

This objective directly addresses the validity of case-note reviews. Chapter 3 extended the remit beyond Chapter 2's focus on cognitive biases and heuristics by investigating the influence of reviewer attitudes, their demographics, and the case-note characteristics. Therefore, Chapter 2 and 3 are related and both address the variables specific to case-notes and reviewers that plausibly influence the global care quality judgement. A multi-level model was employed to account for the clustering within case-note reviewer care quality judgements and hierarchy by splitting up the case note and reviewer variance components. From Chapter 3, the analysis found that 21% (0.205) of the variation around the global care quality judgement is attributable to the reviewer (Table 23).

Table 23. Reviewer level fixed effects in null model

Reviewer mean care quality judgement (i)			Reviewer mean confidence in care quality judgement (ii)		
CQJ	SD	Intra class correlation	CCQJ	SD	Intra class correlation
4.17	0.82	0.205 [15.3-26.6]	82.6	18.6	0.19[0.14-0.26]

Attitudinal measures do not explain any of this reviewer contribution. Despite the real-world non-significance of this finding between the independent variables and the care quality judgement score, other attitudinal measures could have a significant statistical and practical relationship with the care quality judgement score.

The implications of this research are broad and wide-ranging as the source of this reviewer variation has not been defined from our findings. However, there are still implications for a stakeholder of case-note reviews. Clinicians would like to understand better the sources of this variation and would desire to collaborate with academics to understand some of the potential sources for this variation. In the large scheme of things, 20% is not high, however, in real-world terms it is the contribution of a (maximum) possible one point change on a 5-point Likert scale. In actual real-world terms, this is significant.

For hospital Trusts, there is the burden of undertaking MCNR and then knowing that there is variability across these reviewer judgements. In the Trust board member's mind and the Trustees confidence lies in the fidelity of this process. However, the implications may be less far-reaching if a particular Trust, despite the mandate, does not place a strong emphasis for QI from MCNR but their other QI interventions. Given the mandate to review scopes across all acute NHS foundation Trusts, it is in the interests of the Trust board, Trustees, staff and all the patients to know that the MCNR is as reliable as it can be.

For policymakers who endorse and /or mandate MCNR, it will improve the fidelity of MCNR judgements, which then produce care quality learning that are themselves lessons for quality improvement. The policymakers must elicit reasons for the mandate to review on empirical grounds first, and any other reasons later. If there are indeed empirical research in support or highlight the need for MCNR to improve, then it is necessary for the policymaker to consider the significance of the research and its impact for those affected by this policy. In effect, it is every person who will die at some point, which is everyone. But naturally, the

policymaker would have to consider what the consequences are around this reviewer-specific variability. It needs to be communicated to policymakers that this variability does affect case-note reviewer judgements and so they need to show effort to address this gap in the knowledge about the variability in the MCNR judgements. More research needs to be gathered to understand what these factors are, and what the real-world impact of these are. Presently, it is an unknown quantity, however, its consequences are very likely non-trivial. In sum, the role of the policymaker is to endorse policies based on evidence, and we find an unexplained variance contribution from the reviewer for the care quality judgement score.

6.1.1.4. Objective 4: Examine the potential quality improvement potential and any mechanisms from the use of case-note reviews in hospitals (on the use of case-note reviews)

This objective directly addresses the use of case-note reviews. Chapter 4 found case-note reviewing to be well-embedded in our UK National Health Service Trust case study. There were four things we learnt from this case-study. First, the case-note reviews are well-embedded with the entire Trust using, if not, aware of case-note reviews being used. Second, information flows well from the board-to-ward but less so from the ward-to-board. The specialties tended to hold more MCNR information than they released to the board-level; much of this was attributable to internally derived learning from their reviews. Third, individual, system and cultural factors affect QI from case-note reviews. For instance, certain specialties expressed a greater utility from case-note reviews than other specialties. Fourth, there are different perceived purposes for case-note review which depend on the specialty, the reviewer, and their desires for case-note reviews.

There are several implications for stakeholders. Firstly, healthcare organisations[61] and their staff need to be made aware of factors affecting the use of MCNR for QI and we address this in our case study. Undoubtedly MCNR will have different expressions in each healthcare organisation, but that it will be because of their bespoke purpose and local-specialty expression, their perceived utility of the tool, the existing specialty culture for QI and the complexity of their patients. This is knowledge for the clinical reviewer; however, it also provides understanding for these reviewers and the Trust to perceive better the factors operating in their specialty, and thus come to understand how to maximise the utility of MCNR in their setting. Though MCNR is prescriptive, there is no definite expression of it in any Trust, neither have the Royal College of Physicians given bespoke guidelines in their structured-implicit review training for the specialty-specific use of MCNR.(The Royal College of Physicians, 2016b) Our case study goes a step further than the RCP training to provide a context-enriched case-study of how MCNR was implemented in a Trust with all the barrier and facilitators mentioned. Secondly, the Royal College of Physicians[62] needs to take note of these findings so that they can modify their training, if required, in future to help Trusts to adapt this tool and have it well embedded against their existing QI and technological milieu. The findings of our case study will prove instructive for reviewer

[61] Though this was undertaken in an acute NHS foundation Trust, the principles certainly extend to other healthcare organisations who must review and disseminate these findings (or not) throughout the organisation.

[62] It is not strictly limited to the RCP, but can include any case-note review trainers and/or methodologists who will upskill clinical reviewers to a new review methodology

methodologists and instructors as we cover the whole implementation cycle from its embedding to QI outputs.

In retrospect, the implications can be summarised as follows. The reviewer-level variation and unreliability can be characterised, understood, and possibly mitigated, quality improvement certainly involves the reviewer and other domain-specific areas. With further consideration of the case study's context, which include the facilitators and barriers for embedding and information flow, this could help healthcare organisations to better adapt the review intervention to maximise the utility of MCNR for quality improvement.

6.1.1.5. Objective 5: Discuss the research process and limitations

This objective is broken down into two parts on first their contributions to case-note review validity and second, their contributions to case-note review use.

Implications for the validity of case-note reviews

Chapter 1 is concerned about the measurement properties of using hospital mortality statistics as a proxy for hospital care quality via the detection of preventable deaths. The case-note reviews as a proxy for quality of care. The limitations were concerned with the systematic review methodology, however, all the information pertaining to the number of reviewers required for reliable review (kappa = 0.8-0.9) is robust. The implication for the reviewer is that this equation (Spearman-Brown formula) needs to be applied by reviewers in their own context to understand how many reviewers are required to have a reliable review.

Chapter 2 is the first chapter concerned with the psychology of the case-note reviewer. Biases and heuristics have been elicited in populations, in some studies, where

there are questions raised about the generalisability of that study population to a wider audience. For instance, many of the paradigmatic studies in the cognitive biases and heuristics field in the 1970-80s were undertaken in undergraduate psychology students and other non-representative groups.(Kahneman et al., 1982, Gilovich et al., 2002) However, different sub-fields and disciplines have continued to elicit these biases and heuristics in their own experiments, and the medical field, for instance, which has been a field well-studied and has involved clinicians.(Dobler et al., 2019, Richie and Josephson, 2018, Connolly and Jochen, 2003, Flach et al., 2002, Foster and Harrison, 2008, Mamede et al., 2010, Mendel et al., 2011, Blumenthal-Barby and Krieger, 2015b) The cognitive biases paradigm has been discussed further revealing more nuance from the different schools of cognitive psychology, which see cognitive biases as being part of an adaptive process(Gigerenzer and Brighton, 2009) or some others indicating their irrelevance to decision-making in favour of contextualised holistic decisions.(Klein, 1999, Klein et al., 1986) A behavioural framework is employed to yield possible nudges, which is through indirect influence upon subject's environment. A list of these is provided in 4.

Chapter 3 is the second chapter concerned with the psychology of the case-note reviewer. In psychological terms, the grand assumption in psychology is that the measurement of psychological attributes is scientific and quantitative. The current psychological research process has a weakness as it assumes that hypotheses are scientific without demonstrating the additivity of the different magnitudes for one psychological attribute. For a quantitative measure to be obtained, the must believe in two statements (Table 24).

Table 24. Assumptions required for quantitative psychological measurement

Key assumptions of the quantitative psychological paradigm
i) measurement consists entirely in making numerical assignments to things according to some rule
ii) attribute possesses an additive structure

Without some understanding of the nature and foundation of these entities, it is unlikely we can mitigate these biases and lower the variability between reviewer care quality judgements on the same set of case-notes. It is important to have the correct assumptions concerning psychological processes and a realistic view of the nature of cognitive biases and heuristics. However, I am unsure precisely what empirical studies can be conducted which consider the additivity of psychological constructs, just like with physical quantities and constructs.

Implications for the use of case-note reviews

Chapter 4 is the qualitative case-study service evaluation into the embedding, information flow and QI contribution from MCNR. The research process offers more insights than any quantitative method could have offered on the implementation of MCNR, the information flow from it and the QI contribution. Specifically, the qualitative criteria of credibility, transferability, dependability and confirmability were reasonably met.(Guba, 1981) One limitation of this study could have been that I did not employ ethnographic methods, however, given the medical nature of the reviewer's daily clinical work. It was a difficult to overcome this ethical quandary. In-depth structured interviews and the collection of key documents provided sufficiently rich data to triangulate with.

6.1.1.6. Objective 6: Implications for stakeholder groups (Chapter 6)

Given the findings of this thesis, I discuss the findings and implications for those directly involved in reviewing and also those stakeholders indirectly involved and/or influenced by judgements from the retrospective case-note review process. (Cleary and McNeil, 1988) This broadly fits into the two following categories. First, whether the case-note reviews are valid, and the extent to which they are. Secondly, the use of case-note reviews.

6.2. Implications for the validity of case-note reviews

6.2.1. Role of variations and biases in care quality judgements in QI: Implications for case-note reviewers

Our findings imply that **case-note reviewers** know about cognitive biases and heuristics and be more aware of their potential influence on mortality case-note review care quality judgements. These are pertinent for case-note reviewers and review methodologists. The implication this PhD has case-note reviewer regarding the validity of case-note reviews is as follows:

- Heed the plausible influence of these biases and heuristics during reviewer care quality judgements. Given biases and heuristics influence all types of decision-making, then reviewer care quality judgements are potentially influenced.
- Explore the significant amount of literature exhibiting moderate levels of disagreement between replicate reviews by different reviewers for the same batch of case-notes.(Lilford et al., 2007, Goldman, 1994, Hutchinson et al., 2010c, Benning et al., 2011, Goldman, 1992, MacKenzie et al., 1992) Further

focussed research is also offered by us that there are, as yet, unidentified reviewer-specific factors contributing to the variability in the reviewer care quality judgement scores which warrants further exploration.

- Consider the influence of cognitive biases and heuristics factors, and work to identify other factors, that could contribute to this variability in the reviewer care quality judgement. The methodologist must consider these factors to ensure that the case-note review process is as robust and fit-for-purpose to retrospectively appraise care quality in an unbiased and reliable manner (see Chapter 2 and Chapter 3 Results and Discussion sections)

6.2.2. Role of case-note reviews for QI and the variation between reviewers still poorly understood - implications for policymaking

Policymakers always need to consider the (positive or negative) effects their policies have on their intended setting. The finding from Chapter 1 will be of relevance to reviewers as the care quality judgements are not final and do not have high levels of agreement and therefore must have research commissioned to improve methodologies and not put undue emphasis on their results. For instance, they must not use these reviews for ranking hospitals or provider's use. The findings from Chapter 4 's case study will be of the most relevance to policymakers; the case study has implications which the policymaker should be concerned about. As the UK's National Quality Board (NQB) has mandated the undertaking of mortality case-note reviews in all hospitals, there needs to be a better understanding of the factors that influence this programme of work. If there are factors, arising from research, which improve the effectiveness of the programme for QI purposes, ought to be investigated by the NQB (or any

other country's equivalent) to see if there could be further benefits reaped for the MCNRs. Also, if there are factors which inhibit the embedding of MCNR, which is contrary to the NQBs goal, then it is in their best interests to mitigate, and if possible, nullify these effects. The NQB has set the goal of having all acute NHS Foundations Trusts conducting MCNRs and they should heed the findings from the case study which identify key barriers to MCNR embedding (see Chapter 4 Results). Also, in relation to the barriers to the embedding of MCNR, any factors that cause harm or hamper the MCNR programme, it is in the best interests of the NQB to address these issues - on behalf of all the hospitals - and understand these factors and to respond accordingly. The implication this PhD has for policymaking concerned with the validity of case-note reviews is as follows:

- Carefully appraise any potential egregious use of hospital mortality statistics beyond what they can reliably detect (Chapter 1 Results and Discussion)
- Be mindful about the potential action of cognitive biases and heuristics, and where reasonable, have the knowledge and awareness to employ strategies of these biases and heuristics (Chapter 2 Results)
- Be involved in further research to explore the action of these biases and heuristics on case-note reviewers during the process of case-note review judgement (Chapter 2 Discussion)
- Appreciate that biases and heuristics could influence the care quality judgements of case-note reviewers and take steps to learn more about their activity and the way in which they can be mitigated (Chapter 2 Results and discussion)
- Consider the influence of variability in case-note reviewer judgements (Chapter 3 results) and identify possible sources of this from their experience

and their lived practice so that further hypothesis generation and testing can be considered to explore its source.

- Reflect on the viability of using batch review calibration methods to identify the necessary numbers of reviews needed for a highly reliable judgement of a case-note review (Chapter 1 discussion)

6.2.3. For researchers

Research is needed to help advance any field of inquiry. Research is needed in these areas because there do remain several unknown effects upon case-note reviews as presented in this thesis. There is also the possibility of new fecund research between case-note reviews and research and other disciplinary areas such as psychology and healthcare implementation science. In terms of researcher groupings, I refer specifically to psychologists, behavioural scientists, qualitative researchers, and healthcare data scientists, health informatics and information system experts.

The implication this PhD research has for these researcher groups studying the validity of case-note reviews is as follows:

- Psychologists need to:
  - Identify the relevance of other attitudinal measures for case-note reviewer care quality judgements and consider if the sources of variability in the reviewer judgement are found in attitudes or another psychological (or even non-psychological) component (Chapter 3 Results and Discussion)
  - Understand better the reviewer decision-making within the case-note review process and both how and why they make the judgements they do (i.e., case-note reviewers). This is to build a better mental model of how reviewer's think

when handling the case-notes and make the judgements they do which would help situate any plausible influence of cognitive biases and heuristics (Chapter 2 Discussion)

    - Understand better the reviewer decision-making around the case-note review process through the collaboration of those who support during the case-note review process (i.e., patient safety officers, administrators, board members)
    - Consider adopting a behavioural framework (e.g., MINDSPACE behavioural science framework) to make sense of the 'choice architecture' of case-note reviewers and their case-note review judgement

- Behavioural scientists need to:
    - Consider from the thesis how case-note reviews are operationalised in a hospital setting and seek to develop "nudge" strategies to help better embed MCNR programmes in hospitals
        - Consider the barriers to all aspects of the MCNR implementation and be able to explain and then develop strategies to tackle these barriers for its embedding, information flow and contribution to QI.
        - Consider the facilitators to all aspects of MCNR implementation and be able to explain how and why they work.

6.3. Implications for the use of case-note reviews

6.3.1. Role of the use of Case-note reviews in QI: Implications of findings for healthcare leaders and managers

The leaders and board members of healthcare organizations are ultimately responsible for the care provided from their organization. Healthcare managers are ultimately responsible for care delivery within their teams. As mortality case-note reviews are mandated for use in all hospitals, it is the responsibility of the healthcare leaders, which include board members, and healthcare managers to ensure this mandate is actioned well and effectively. The implication this PhD has for healthcare leadership on the topic of the use of case-note reviews is as follows:

- Healthcare leaders and managers should then seek to understand how and why factors can help or jeopardize the implementation of a fully functional MCNR process capable of identifying learning for QI purposes
- Leaders and managers should consider any information available which can inform them concerning the barriers and facilitators to MCNR embedding, information flow and QI contribution
- Healthcare leaders and managers would do well to consider results and discussion from our case study which address the barriers and facilitators to MCNR embedding, information flow and its potential QI contribution (see Chapter 4 Results and Discussion sections)

The implication this PhD has for policymaking concerned with the use of case-note reviews is as follows:

- exploring how and why certain specialties, with their micro-cultures, are more inclined to first, use MCNR and secondly, to have QI benefit from its use (Chapter 4 Results and Discussion)
- develop guidance to support specialties known to benefit from MCNR

- develop guidance and an infrastructure to support a specialty micro-culture conducive to QI
- exploring how and why certain types of more complex and frail patients are a consistent source of MCNR learning (Chapter 4 Results and Discussion)
- develop some guidance around its use for these types of patients
- Consider, appreciate, and reflect on how these identified barriers and facilitators to MCNR embedding and information flow influence clinical practice (Chapter 4 Results)
- Consider some changes that can be made to help MCNR embed and flow better in their setting (Chapter 4 Results)
    - Consider, appreciate, and reflect on how the barriers and facilitators to MCNR embedding and information flow identified in our case study influence the wider practice (Chapter 4 Results)

The implication this PhD has for these researcher groups studying the use of case-note reviews is as follows:

- Qualitative health researchers need to:
    - Understand better the case-note reviewer behaviours around MCNR and any work especially when overlapping with other QI interventions
    - Undertake ethnographic work of case-note reviewers to follow the process of a review from beginning to the end
        - This would help to us to appreciate the temporal flow of case-note reviews and so help arrive at a better understanding of how the barriers, facilitators, and workarounds at operating in the daily MCNR process

    - Explore how the use of data collection tools could influence MCNR embedding, information flow and its contribution to QI
    - Examine and build upon Chapter 4's findings on how and why training is (not) taken up within the specific setting
        - What are the main barriers and facilitators to the uptake of MCNR related training?
- Health informaticians need to:
    - Establish the influence of the modality of the case-note record most well suited to MCNR purposes (i.e., electronic, paper-based or hybrid case-note format for MCNR)
    - Consider the influence of the electronic and the paper-based case-note on the MCNR embedding, information flow and contribution to QI[63]
        - Research shows that the poor state and formatting of case-notes has an impact upon the perceived care quality by clinicians.(Francois et al., 2014)

In having developed a comprehensive set of implications for a wide-range of stakeholders, I will next turn to the likely steps mortality case-note reviews can move in to ensure that it is used as effectively as possible.

[63] It was not possible to differentiate between paper-based and electronic case-notes in the case-study, and so a study is needed to explore whether the case-note modality has a significant role to play for MCNR embedding, information flow and contribution to QI.

6.4. What does the future look like for mortality case-note reviews?

Since the advent of modern case-note review, there have been several ways it has been employed over the decades. These have not been without their own challenges. The main approaches can be grouped into summative and formative approaches. The summative approaches have their aim to benchmark using mortality case-note reviews. The formative approaches are for the internal hospital use and local quality improvement purposes. I will now set out some of the main challenges using these approaches.

6.4.1. Summative approaches: challenges

The use for summative means is for the benchmarking of different hospitals. They are compared in a league table to help present any disparities in care. To ascertain whether one hospital is performing better than any other, one needs to have two objects in mind. Firstly, that the actual care quality is being correctly measured and reported. This is important because, as has been seen with hospital mortality statistics, they make assumptions which are unwarranted. Now that it is known that the reviewer contributes about a fifth of the variance to the total variance, it must be then explored to what extent this knowledge affects the underlying integrity of the care quality judgement. This is important because 20% is not an insignificant proportion and it remains unexplained, as it stands. There needs to be a better understanding of what exactly is contributing to the care quality judgements before one can make a secure assessment of its validity for any type of benchmarking process. In real-world

terms, if the variance is left unchecked, this added noise[64] to the judgement could lead to incommensurable comparisons. Any comparison between estimations of care quality based on the batch of case-notes would not be possible because factors influence each reviewer's judgement of the case-note review. This is the corollary of Chapter 3 . This is one challenge to overcome. For MCNR, there is a need to clarify what is this contribution to the noise of care quality judgements.

Secondly, the local care quality must be considered in the evaluation of care. And these require that mortality statistics are employed. Do note, that there is the low-to-moderate level of disagreement between reviewers undertaking replicate reviews on the same batch of case-notes (Goldman, 1992, MacKenzie et al., 1992, Goldman, 1994, Lilford et al., 2007, Hutchinson et al., 2010c, Benning et al., 2011) One way to overcome this is to have a batch of case-notes notes which are reviewed a replicate number of times to establish estimate statistical parameters (i.e. calibration review sets to obtain an accurate and reliable variance around the preventability judgement in any given healthcare context (e.g. hospital or Trust, or healthcare organization)). So, for case-note reviews to be fairly represented, there is the need to understand the degree of variation in the judgement of care quality to factor this into summative evaluations about the care in a hospital or Trust. Without the knowledge of this local variation, the preventability estimates would not be comparable and will certainly lead to erroneous classification through misleading under-performing and hyper-performing metrics. The underlying reason is that the care quality and contextual factors will vary from

[64] It is another term for variance, which has gained popularity as explaining the unpredictability of human predictions and judgements. The magnitude may vary depending on which unknown factors are at play influencing the care quality judgement

Trust to Trust, and it is this statistical parameter called the variance that can be used to equalise for any local differences in this variance. Obtaining this figure for all trusts is a great endeavour and costly and still does not eliminate the variation due to reviewers. It is therefore best not to do such summative reviews as alluded to in Chapter 3 .This then makes it possible to then compare across the preventability figures of healthcare organizations.

On balance having considered some of the challenges employing MCNR, the summative approach is beset by technical measurement concerns over fidelity whilst the formative approach is beset by specialty and cultural challenges. Though summative approaches are seen as informative by both clinicians and the public because they have powerful face validity. (Normand et al., 2016) But one of the dangers of employing a summative approach for MCNR is that this measurement procedure first needs to consider the influence of local preventable death variability within healthcare organisations. As the situation stands, appropriate measurement procedures for calculating this inter-hospital variance are not widespread with only two studies having obtained such outputs.(Baines et al., 2013, Baines et al., 2015, Manaseki-Holland et al., 2016) Thus, it does not seem promising for the practical use of a summative method.

6.4.2. Formative approaches: the challenges

The formative approaches are for local use. This is either within the hospital or the specialty in the hospital for quality improvement purposes. Firstly, to learn consistently from case-note reviews, there needs to be a structure in place for MCNR to be embedded. The culture, especially of the board and senior clinicians, is responsible for a supportive culture for MCNR to flourish. If the board and senior clinicians are not supportive of MCNR then it will be an uphill struggle for the hospital to obtain any quality improvement from MCNR. It

cannot be known the exact culture MCNR will be used in, however, there need to be pro-active engagement and ownership for MCNR to be embedded and utilised. This is the first challenge to overcome. Secondly, the training needs to be concise and effective considering the range of demands placed on MCNR in different specialties. Given the lack of guidance on the application of MCNR in each specialty, there needs to be greater re-assurance that the method is being implemented effectively across these different specialties and is fit-for-purpose. As observed in our case study, simply giving different specialties the license to undertake MCNR does not necessarily entail that it is embedded well nor used effectively. Thirdly, it is less a challenge than it is an opportunity, but there would be benefit for patients and staff to understand how and why some specialties benefit from MCNR more (i.e., more valuable QI contributions from MCNR). Policymakers should seek to address this point and make guidance more specific to ensure MCNR is used in a cost-effective manner. It would be an opportunity to explore MCNR process in different specialties.

On the other hand, the formative approach is less inclined to measurement assumptions but requires more buy-in from healthcare organisations. With the policy in place to review a proportion of deaths, all that is required is reasonably required for MCNR's formative use is that specialty-specific training is developed to ensure MCNR is efficiently used and the support and buy-in from senior clinicians and board members. This is more reasonable to achieve than the summative approach, however, there is certainly more organisational case study research needed to understand the best use for MCNR in the Trust and what the best way is to encourage its adoption and use across the entire organisation. These challenges are not small, but our case-study has helped provide a working example of how barriers and facilitators affect all levels of the MCNR process. This is a topic that qualitative researchers ought to study more closely and build upon. In the next mini-section, I

shall discuss how the gradual transition of case-notes from a paper-to-electronic format offer fresh opportunities for case-note review use for hospital quality improvement.

6.4.3. The Digital Revolution for Case-notes

The electronic medical record (EMR) is an enabling technology that allows hospital clinicians to pursue more powerful quality improvement programs than is possible with paper-based records.(Miller and Sim, 2004) This EMR databases contain the history of hospital encounters, records of diagnoses and interventions, lab test, medical images, and clinical narratives.(Bohr and Memarzadeh, 2020) There are two important considerations to consider around the use of electronic medical records for MCNR use. The key is the first, how the data is entered into the electronic record and second, how the information contained within extant paper-based records is imported into the electronic record (or an electronic format).

The prospect of digitisation provides a wealth of challenges. The UK has previously implemented a national transition from paper-based to electronic medical records transition for general practice in the early 2000's.(Wilson and Khansa, 2018) The programme was a failure and for several reasons including, from which we can glean lessons for any major transitions from paper records to EMR in hospitals. Achieving quality improvement through EMR use is neither low-cost nor straight-forward. There are two main limitations for the entry and the systematic extraction of usable information, as useful insights, from EMRs. The digitisation is the first step, but it is also required for protocols to be in place of good practise to address some of the pertinent suggestion from a recent Professor Ben Goldacre's review into the use of healthcare data.(Goldacre, 2022)

The first limitation is the variability of the input of the data, such as differences in where patient information is entered into the EMR (for example, smoking history may be entered as social history, a risk factor or as part of the encounter note). Another example is the myriad ways for practitioners to characterize whether a patient smokes tobacco or not. To deal with this, there needs to be a standardised code to indicate whether the patient currently smokes', 'was a previous smoker' or 'never smoked.' Further, some electronic record systems have better background disease coding than others. There is no expectation by the network that the practitioner codes any data so we must do that in the background. If a practitioner uses short forms such as "DM", "T2DM" or "Diab" for the diagnosis of diabetes mellitus, data managers have developed synonym lists that are all coded to ICD coded of "250" for diabetes mellitus.

The second limitation is data is difficult to code or analyse, that of which is found in scanned documents or case notes. Solutions to these problems include providing a more structured data entry that has less free text and choosing an EMR capable of communicating with other electronic systems (for example, in hospitals, laboratories and imaging) to allow direct storage of as much clinical information as possible in the EMR. With these considerations, and with narrative information more prevalent than constrained closely defined fields, I now turn to how this information can be narratively mined.

The third limitation is the lack of trusted research environments that allows for the open sharing of NHS data to those with the expertise to produce analytical insights into the data. Developing these environments requires a lot of technical expertise and deep knowledge about the technical infrastructure, which has been alluded to in an earlier section in 4.7.9. Information infrastructure. In other words, strong data engineering skills are needed 4.7.9. Information infrastructure

to meet the needs of each of the users to securely access the NHS data.(Goldacre, 2022) This would provide more possibilities for Data science

learning from datasets, around these case-note (and their content) as already alluded to earlier in the methods outline of this thesis.

6.4.4. What is narrative mining?

As alluded to Narrative mining

It typically involves digital computers applying algorithms to discover these data insights and this is important for case-note reviews as much of the data entry for case-note are in free-text form. Narrative mining (or natural language processing (NLP)) has been used to extract clinical information from electronic medical records of patients with hepatocellular carcinoma (HCC) to successfully grade the stage of HCC with high degree of concordance (75%) with the manual case-note reviews. Furthermore, in another study, there have been attempts for EMR to detect cases using the following case-detection methods (CDM) - that range from manual searches to machine learning algorithms - with median accuracy highest for those which adopted code/text combined algorithms compared to codes-only. A wide range of studies showed that information extracted from EMR text were used to identify varied conditions with variable degrees of success (Table 25). The CDMs reviewed are

adequate for recruitment into exploratory, hypothesis-generating research. However, for any use for recruitment into clinical trials then higher sensitivities and specificities (>99%) are needed to reduce the likelihood of "troublesome" false positives.(Chen et al., 2019) There is a viable process for narrative mining from medical records for diagnostic and predictive purposes; I shall discuss how it could be implemented for case-notes and to help reviewers review electronic case-notes.

Table 25. Types of Case-Detection Algorithms

Type of case-detection
No additional algorithm (manual review of information)
Single keyword or code sufficient to define case
Same NLP algorithm as extracted info also detected cases (text only)
New rule based CDA (text only)
Logistic regression or machine learning CDA (text only)
New rule based CDA (combining text with codes, labs, or medication)
Logistic regression CDA (combining text with codes, labs, or medication)
Machine learning algorithm (combining text with codes, labs, or medication)

6.4.5. How can narrative mining be applied for case-note reviews?

The rapid progress and developments within the IT field makes the problems of paper-based patient documentation all the clearer. It can be available only in one physical location at a time and it is often both poorly organized and collated, while documents may be incomplete and/or illegible, data may be acquired that is redundant which wastes time and resource. The data is usually physically stored at different sites which makes accessing them problematic in

terms of access logistics. Moreover, there is the requirement for high-resource, via personnel and space requirement for the routing, archiving and the maintenance of the paper-based patient documents. The conventional paper-based patient record is fast reaching its limits for practicability and efficiency in today's modern fast-paced and increasingly data-driven world. Chaudhry and colleagues conducted a systematic review into the impact of health information technology (HIT) on quality, efficient and costs of medical care and found that across four benchmarked organisations, HIT led to better quality, specifically through increased adherence to guideline-based care, enhanced surveillance and monitoring, and decreased medication errors and efficiency, specifically through the decreased utilization of care. The author comment on the inevitable "death knell" for paper-based records:

> *"Given the fragmented nature of health care, the large volume of transactions in the system, the need to integrate new scientific evidence into practice, and other complex information management activities, the limitations of paper-based information management are intuitively apparent"(Chaudhry et al., 2006)*

It has the potential to extend to case-notes because of the high preponderance of free-text o structured text entries compared to the other parts of the note such as the clinical information. Narrative mining techniques can be employed to both lighten the cognitive load on clinicians during their reviews and to expand the digitised data repository of case-note for further data insights.

6.4.5. Issues with electronic medical records

Despite the promise of HIT for electronic medical records, there are several "stubborn" challenges facing EMR advocates. In no specific order, these different versions of

EMR are often incompatible with each other. And if the training is conducted, then its needs to be delivered comprehensively and efficiently to all clinicians and administrators.

Presently, medical records defy simple organization or searchability and do appear to counter (and somewhat incompatible) to a human's ability to grasp high-level generalisations from the data about the person waiting to be seen in person or have their case-note reviewed. Just the fact that it takes many sessions across many hours to be trained to use EMR indicates that the complexity of working with them may not be cost-effective nor efficient for healthcare staff, certainly not at the outset. This sympathy was also reflected in the case-study also, with interviewees who review preferring paper case-notes over EMR-type notes because of their comparative ease to navigate.

Electronic health records can be used to harness the power of Artificial Intelligence (AI) tools. AI is loosely defined as "the theory and development of computer systems able to perform tasks normally requiring human intelligence." I state "loosely" because AI itself owes its success to its flexibility to accommodate for a wide-range of unstructured data types, which led the Stanford University's One Hundred Year Study of Artificial Intelligence (AI100) to study and anticipate, on an annual basis, how the effects of artificial intelligence will ripple through every aspect of how people work, live and play.(Ajunwa; et al., 2021) They opine that:

> *"The lack of a precise, universally accepted definition of AI probably has helped the field to grow, blossom, and advance at an ever-accelerating pace."*

As with the relevant medical literature, ideally AI could mine and integrate all the data about a patient, if they were structured comprehensively, neatly, and compactly. There has been no such streamlined product, but it would bring much benefit to unify healthcare information under the National Health System (NHS) in the United Kingdom. If this ever does happen, it will represent not only a means to promote the efficiency of the doctor's workflow but also a tool for a more meaningful and thorough prospective and retrospective assessment of each patient. And it will ultimately prove to be immensely valuable and also cost-effective - in not duplicating work unnecessarily and having it available at all time and open to clinician analysis or algorithmic analysis - to everyone in his or her life journey through health and illness.(Topol, 2019) In sum, there are primarily two challenges. First, to get the physical information into an electronic form. The second is to have the ability to generate insights from this data. The first issue is more of a logistical issue whilst the second issue is an AI issue. However, there is the possibility of combining these two steps into one step; that is to have notes scanned and simultaneously mined for text-based analysis. These options will be discussed in turn.

6.4.6. Importing paper-based records through e-scanning

First, the information needs to be recognised. If the information is captured as free script, that is formally recognised symbols a computer can recognise, that is not unique to an individual, then a method call Optical Character Recognition can be employed to extract the text information into an electronic form. Optical character recognition (OCR) is software that used a more advanced usage of matrix method. It is also known as pattern matching. It is a procedure for the interpretation of intelligible characters (i.e. print script) to machine readable characters in an optically scanned and digitized text.(Dhande and Kharat, 2017) One can then

search within this document for keywords. Also, scanning the physical case-notes into the electronic system may appear straightforward, however, there are some challenges posed by the order in which administrators scan the case-notes and the inter-operability of the EMR interface to allow case-note reviewers to navigate the document easily to confirm or disconfirm their initial convictions about care quality. These were all ideas raised from Chapter 4 ’s case-study which need to be addressed for any EMR system informed by the scanning of the original physical notes. There are reasonable strategies to tackle this: the pages could be numbered, or the notes could be scanned in a certain ‘sectioned’ order according to a system of priority. The exact barriers differ to the setting and require bespoke strategies to overcome them.

6.4.7. Natural Language Processing (NLP)

Once the notes are imported as EMR, then the EMR data can be subjected to various algorithms. An algorithm is set of rules followed for the purpose of solving a problem. This problem is first to extract high-level information from the scanned documents. Natural language processing can be employed to extract information from unstructured data such as *printed* free text entries. It must be noted that Natural language processing (NLP) relates to the interaction between computers and humans who use natural language. The aim of NLP for the computers to find meaningful patterns from the human language found in free text data. NLP is crucial for many applications of big data analysis within healthcare, particularly for EMRs and the translation of narratives provided by clinicians. It is typically used in operations such as extraction of information, conversion of unstructured data into structured data, and categorization of data and documents. NLP makes use of various classifications to infer meaning from unstructured textual data and allows clinicians to work more freely using

language in a “natural way” as opposed to fitting sequences of text into input options to serve the computer. NLP has been employed to generate meaningful information from unstructured data in the specific domain of incident reports and adverse events. It is well known that the analysis of incident reports and adverse events in healthcare is considered an important part of quality improvement and patient safety. NLP has been shown to perform favourably compared to manual annotation across a range of classification tasks. However, these are across binary classifications (i.e. comes to harm or no harm, fall or no fall, drug error or no drug error) which may lead to positive chance findings from multiple hypothesis testing alone and the higher likelihood of generating false-positives.(Benjamini and Hochberg, 1995) There is a future for investigating the more complex relations between many components of care, not just for binary classifications, during the courses and development of medical errors from patient safety reports; this study shows the promise of the automatic classification of clinical reports in the clinical setting. These are directly translatable to the hospital setting for use of classifying case-notes. Due to the sheer number of case-notes available to review, this method has to help screening across all case-notes, for certain features, which frees up clinical reviewer time to focus on the case-notes most likely with care quality concerns.(Liang and Gong, 2017)

6.4.8. Unsupervised deep learning from handwritten medical records

It is a different study entirely for text entries that are not printed. A significant proportion of case-note can be handwritten – cursive and/or block letters – and so different NLP methods need to be employed to extract meaningful information from the case-notes. A state-of-the-art AI technique called “deep learning” has been employed methods recognise doctor’s cursive

handwriting.[65] Deep learning computers can “learn” from experiences just like the human brain learns from stimuli. It is through this hierarchical layering that it can then comprehend the world. Since computers can acquire knowledge from experiences, there is no need to have a human computer operator to specify all the knowledge that the computer needs. Through hierarchy of concepts, computers will be able to learn complicated concepts by building them out of more basic ones. In numerous software disciplines such as computer vision, speech recognition, language processing, robotics, bioinformatics, video games, search engines, online advertising, and finance, it is now proven that deep learning is widely employed in these aforementioned areas.(Goodfellow et al., 2016)

In more detail, deep learning employs artificial neural networks to classify objects as data points. A neural network (also called an artificial neural network) is an adaptive system that learns by using interconnected nodes or neurons in a layered structure that resembles a human brain. A neural network can learn from data—so it can be trained to recognize patterns, classify data, and forecast future events. Of all deep neural network types, the most popular are convolutional neural networks (CNN). A CNN convolves - convolutional means combining all into one - learned features with input data, and uses 2-dimensional

---

[65] The difference between conventional machine learning (ML) and deep learning is that the former is supervised and needs to have the model trained and guided by a human operator and the latter is not. The latter method is not supervised because of the use of many hidden layers which can be re-configured and re-weighted according to the data. The only caveat is that deep learning requires a lot of data for the model to converge unsupervised in producing good outputs.

convolutional layers, making this architecture well suited to processing 2-dimensional data, such as images. Thus, imagine a computer trying to represent some handwriting; the computer would take hold of this handwritten image and represent it using a pixel grid. The more highly resolved the grid, the more data is then convolved into the CNN. This process has been utilised with success with doctor's cursive handwriting. For the training stage for the model,[66] the training accuracy was 76%, which were undertaken on 70% of the dataset. The test accuracy was 72% based on the remaining 30% of the data.(Fajardo et al., 2019) The accuracy is modest, however, there realistically needs to be an accuracy above 95% across all hospital case-notes for it to be employable for care quality issue detection. This is because the false positive finding for a care quality issue may still lead case-note reviewers to review more notes than as they did before. However, any improvement in the case-detection stage will help alleviate some of the time and effort to review the case-notes. If there was a high sensitivity - that is the true positive rate of detection on a certain care quality feature – then this would have a desirable outcome which could reduce the workload on clinicians whilst maintaining high sensitivity to a said care quality construct of their choosing.

Case-notes, however, feature both print and handwritten formats which themselves are either structured or unstructured. To help overcome the variability of handwriting types observed in a case-note, there are methods that can help set pre-defined these field types and optimise the recognition of these different writing types. In a study by Tomoiaga and colleagues who sought to train and test a neural network to identify real-life handwritten text within different

[66] In machine learning and deep learning, the training data, which is the initial reserve of data used to develop the model.

field types, they applied a convoluted recurrent neural network (CRNN) approach. With this method, they generated their own cleverly devised training data and lowered the character error rate from 73% - on a conventional CNN to – a modest 23%. The synthetic training data they amassed is able to generate a greater variability than is otherwise observed in the public source materials.(Tomoiaga et al., 2019) The method was optimised by training on a synthetic data to overcome the problem of previous public datasets poorly generalizing to real-life structured forms. They employed the state-of-the-art CRNN with VGG – a particular configuration with a small filter size (3x3) with greater depth of weighted layers that breaks up the data into even smaller discrete packets than is typically used.(Simonyan and Zisserman, 2014) The field types were also accounted for through these synthetic training data. The handwriting of individual contributors to the handwriting data, that is the variance in handwriting naturally observed within individuals, will also vary considerably from one person to another. This is well accounted for within the synthetic training dataset; this bodes well for the consideration of clinician.

This method can better resolve real-world handwritten forms with very good accuracy. However, there still need to be further improvements for the accuracy of detecting information from real-life documents such as insurance forms and medical records, but the groundwork is in place with CRNN to be a prospective approach for the digital archiving of case-notes and its content classification. What next needs to be done is for the reviewers to agree amongst themselves precisely what the standards of good care and bad care are to help the CRNN train adequately. Once this is resolved, then the model will be able to train autonomously. But these are deep, ethical, and human judgements are a matter that machines have long since struggled to grasp. It must be noted that there has been no machine which can

think on such a deep and ethical level, and this is requisite for any sensible care quality rating. And so, next we turn to the limitations of this whole approach.

6.4.9. Limitations of Deep Learning

Deep learning is very data intensive as it is guided not by a human operator but sufficient information from the data. Assuming there is sufficient data to train the model effectively, there is the need for the computing hardware to match the demands of the data-intensive process. Specifically, deep learning requires training on graphical processing units (GPU) which have an abundance of computing cores over and beyond that of computing processing units (CPUs), which are now more commonly found in healthcare use. GPUs are more suited to graphically render intensive processes because of their high throughput (data flow, speed, and refresh rates) which makes them well suited for graphical representations in highly intensive activities like in computer games. The strength of the GPU is that it breaks complex problems into thousands or millions of separate tasks and works on these in parallel, thus the term parallel processing. The disadvantage of this financial cost of this necessary computing power. Next, there is the time needed to train the model. The time taken to train can take days if not months to train as a function of the amount of data and the number of hidden layers (essential for CNNs) and the efficiency of the code itself. The high-level of computing power, cost, time to compute required for this deep learning are the main limitations to consider. This assumes that data scientists and machine learning practitioners have the case-notes in a good order electronically for deep learning processing. And lastly, there is the issue of ethical algorithms. These algorithms are shaped by the developers who programmed it, and it must also be responsive to human values in the data it is being trained and tested on. These are important considerations for the data engineers and scientist who will be responsible for

developing the models with clinical input to ensure the model is developed as ethically as possible. Human supervision is needed to ensure that the good care quality remains just that if it is allowed to judge the care quality of case-note directly or as an aide to reviewers.(Chen et al., 2020)

6.5. A promising future

Due to the power of the open-source data in fields such as machine learning and artificial intelligence, the field is likely to progress much over the coming years to the point where the technology can be implemented at scale on an organisational scale such as the NHS Trust or even the whole UK National Health Service. However, the greatest barrier to overcome would be for a system to become the technological "*lingua franca*" across the organisation. It is no mean feat to achieve this, but the technology is as well placed as it ever has been to allow case-notes to be scanned with good accuracy and for machine learning algorithms to be able to detect the text and/handwritten content. The aim is for these techniques to help case-note reviewers screen case-notes and ensure the most problematic case-notes – in care quality terms and the greatest safety concerns - are not missed and are making their way into human hands. They are not a replacement but an aide for the often-beleaguered doctor of modern times who is short on time and pressed to complete their clinical duties without consideration on more efficient data-driven approaches to their practise. There is the promise of a technological revolution for clinical evaluation of case-note reviews. However, decision-making on the case-notes themselves must still be guided by an experienced clinical reviewer. Time to tell what this interaction and its practise will look like, and what the implications will be for retrospective care quality judgements.

6.5.1. Translation into Quality Improvement

How would this all translate into the organisational setting of a hospital and its case-note process? Taken from the findings in our case study, the acute medical wards and their reviewers found more benefit, in terms of QI, than other specialties.[67] There is the intention of having a pilot of this infrastructure conducted with members of this speciality to detail how this happens. This can be explored further with focus groups and semi-structured interviews to understand how such a system has utility for them. There can be much insight gleaned from qualitative research to obtain a greater understanding of what this care quality construct is in the mind of the case-note reviewers. After all, having good data is the first requirement, but then the second requirement is to extract information relating to a certain property of interest. In this thesis, the property is "care quality" and all its related constructs. Qualitative research can help reviewers understand what they are looking for and help data scientists to help them mine such information from the electronic medical records. For instance, the discussion could range from the logistical side of case-note reviewing (which includes scanning the notes and having them quickly available to review on the same day or near enough, rather than the protracted 8 weeks delay from case-note generation to case-note review). There could be a desirable use for reviewers to be able to quickly case-note content using keywords and filters. Lastly, the AI algorithm can guide using the selection of a "feature". The case-note contains a lot of data, most of it redundant, and one way of helping

[67] For specialties which found little QI benefit from MCNR (e.g., emergency medicine), this idea can still be discussed with reviewers from their department, however, it is unlikely to yield great promise as the machine still requires discernible "features" selected to inform the data-driven process of which are likely wanting in these specialties.

reviewers make sense of the signal from the noise is for them to select variables they know that are proxies - to some degree or other – for care quality in their clinical opinion (e.g., length of stay, co-morbidities, disease(s) etc.). The reviewers and data scientists/ML engineers can work closely to develop features that can reasonably detect care quality from the case-notes. This is called feature engineering. This would be the first step towards better characterising what good care could look like and help a machine to understand this through text, via NLP. There are modest uses for this which include screening case-notes for certain features (e.g., keywords, features associated with care quality detected in the case-note text content) that case-note reviewers feel are strongly linked to poor care quality. A more ambitious use-case would be for the machine algorithms to help the case-note reviewer in eliciting care quality judgements; the word "helped" is stressed because machines cannot have intuitions or their own understanding about abstract concepts like "care quality" but only data-driven generalisations. A reviewer is always needed to make the final judgement, but the machine can provide some degree of decision-assistance from its generalisations across the data. In this way, the reviewers can take stock of the innumerable factors distilled by the machine and consider how it is to inform their care quality judgements. There could be a positive influence on case-note reviewer agreement levels within case-note reviewer care quality judgements as mentioned at the thesis outset. Biases could be weeded out – at least the identifiable ones - and human sentiment with its 'noise-filled' judgements reduced to a lower level.(Kahneman et al., 2021) That would be on the more optimistic side, but it would be something ambitious to aim for. For this thesis specifically, this method has the prospect of reducing the variability of care quality judgements within reviewer pairs by reducing the extraneous noise permeating reviewer intuitions. This has the prospect of helping reduce the noise in the reviewer's intuitive judgement and thus allow case-note reviews as a method and

hospital use-care the greatest chance to maximise its potential. More significantly, this could contribute to quality improvement than it does today at the local, national, and global levels.

6.6. Conclusion

Firstly, case-note reviews have been largely assumed by healthcare practitioners as a valid tool for quality improvement purposes. However, our studies show that the care quality judgement variability found between reviewers for case-notes varies are attributable to a wide range of unknown factors. Areas to explore this source of variance include cognitive biases, heuristics, additional individual difference measures and multi-faceted, irreducible organisational features of the healthcare providers themselves. Case-note review use in hospital settings has received very little research. This case study studied its use and found that specialty culture, existing national QI frameworks, reviewer patient ownership, clinician workload with their awareness of policies and training influenced the quality improvement contribution from case-note reviews. These findings are likely to be of interest to patient groups, academic healthcare practitioners, behavioural scientists, psychologists, hospital healthcare policy makers and quality improvement specialists. The work is diverse and the methodological challenges considerable, but this research area is as important as any other hospital improvement approach and method – through the case-note reviewer biases, attitudinal measures, practicability of reviewing and electronic medical records - could reduce the noise in clinician judgements with likely repercussions for the enhanced validity and practise of case-note reviews for hospital care quality improvement.

References

ACEVEDO, M. & KRUEGER, J. I. 2004. Two egocentric sources of the decision to vote: The voter's illusion and the belief in personal relevance. *Political Psychology,* 25**,** 115-134.

AJUNWA;, M. L. L. I., BERGER;, G., BOUTILIER;, C., CURRIE;, M., DOSHI-VELEZ;, F., HADFIELD;, G., HOROWITZ;, M. C., ISBELL;, C., KITANO;, H., LEVY;, K., LYONS;, T., MITCHELL;, M., SHAH;, J., SLOMAN;, S., VALLOR;, S. & WALSH, T. 2021. Gathering Strength, Gathering Storms: The One Hundred Year Study on Artificial Intelligence (AI100) 2021 Study Panel Report. Stanford University, Stanford.

AL-ABRI, R. & AL-BALUSHI, A. 2014. Patient satisfaction survey as a tool towards quality improvement. *Oman medical journal,* 29**,** 3-7.

ALBARRACÍN, D. & WYER, R. S., JR. 2000. The cognitive impact of past behavior: influences on beliefs, attitudes, and future behavioral decisions. *Journal of Personality and Social Psychology,* 79**,** 5-22.

ALHAKAMI, A. S. & SLOVIC, P. 1994. A psychological study of the inverse relationship between perceived risk and perceived benefit. *Risk analysis,* 14**,** 1085-1096.

ALVESSON, M. Skoldberg,(2000) Reflexive methodology: New vistas for qualitative research. London: Sage.

ALVESSON, M. & SKÖLDBERG, K. 2008. *Tolkning och reflektion: vetenskapsfilosofi och kvalitativ metod,* Studentlitteratur.

AMIN, A., KESHISHIAN, A., TROCIO, J., DINA, O., LE, H., ROSENBLATT, L., LIU, X., MARDEKIAN, J., ZHANG, Q. & BASER, O. 2017. Risk of stroke/systemic embolism, major bleeding and associated costs in non-valvular atrial fibrillation patients who initiated apixaban, dabigatran or rivaroxaban compared with warfarin in the United States Medicare population. *Current medical research and opinion,* 33**,** 1595-1604.

AMOS, T., APPLEBY, L. & SHAW, J. 1997. National Confidential Inquiry into Suicide and Homicide by People with Mental Illness: recent developments. *J International journal of psychiatry in clinical practice,* 1**,** 69-71.

ANDERSON, J. 1962. Studies in empirical philosophy.

ANDERSON, J. R., BOTHELL, D., BYRNE, M. D., DOUGLASS, S., LEBIERE, C. & QIN, Y. 2004. An integrated theory of the mind. *Psychological review,* 111**,** 1036.

ARAH, O. A., WESTERT, G. P., HURST, J. & KLAZINGA, N. S. 2006. A conceptual framework for the OECD Health Care Quality Indicators Project. *International Journal for Quality in Health Care,* 18**,** 5-13.

ARCHER, M., BHASKAR, R., COLLIER, A., LAWSON, T. & NORRIE, A. 2013. *Critical realism: Essential readings*, Routledge.

ARCHER, M. S. & ARCHER, M. S. 1995. *Realist social theory: The morphogenetic approach*, Cambridge university press.

ARMSTRONG, D., LILFORD, R., OGDEN, J. & WESSELY, S. 2007. Health-related quality of life and the transformation of symptoms. *Sociology of Health and Illness,* 29**,** 570-583.

ASHTON, C. M., KUYKENDALL, D. H., JOHNSON, M. L. & WRAY, N. P. 1999. An empirical assessment of the validity of explicit and implicit process-of-care criteria for quality assessment. *Medical Care,* 37**,** 798-808.

ATKINS, S., LEWIN, S., RINGSBERG, K. C. & THORSON, A. J. B. H. S. R. 2011. Provider experiences of the implementation of a new tuberculosis treatment programme: a qualitative study using the normalisation process model. 11**,** 275.

AUSTRALIAN COUNCIL FOR SAFETY QUALITY IN HEALTH CARE 2005. Clinical handover and patient safety: Literature review report.

AYLIN, P., BOTTLE, A., JEN, M. H. & MIDDLETON, S. HSMR mortality indicators.

AYLIN, P., BOTTLE, A. & MAJEED, A. 2007. Use of administrative data or clinical databases as predictors of risk of death in hospital: comparison of models. *Bmj,* 334**,** 1044.

BABCOCK, L. & LOEWENSTEIN, G. 1997. Explaining bargaining impasse: The role of self-serving biases. *Journal of Economic perspectives,* 11**,** 109-126.

BACON, F. 1858. *The physical and metaphysical works of Lord Bacon,* HG Bohn.

BAINES, R. J., LANGELAAN, M., DE BRUIJNE, M. C., ASSCHEMAN, H., SPREEUWENBERG, P., VAN DE STEEG, L., SIEMERINK, K. M., VAN ROSSE, F., BROEKENS, M. & WAGNER, C. 2013. Changes in adverse event rates in hospitals over time: A longitudinal retrospective patient record review study. *BMJ Quality and Safety,* 22**,** 290-298.

BAINES, R. J., LANGELAAN, M., DE BRUIJNE, M. C. & WAGNER, C. 2015. Is researching adverse events in hospital deaths a good way to describe patient safety in hospitals: A retrospective patient record review study. *BMJ Open,* 5.

BAKER, G. R. 2011a. The contribution of case study research to knowledge of how to improve quality of care. *BMJ Quality & Safety,* 20**,** i30.

BAKER, G. R. 2011b. The contribution of case study research to knowledge of how to improve quality of care. *BMJ quality & safety,* 20**,** i30-i35.

BAKER, G. R., NORTON, P. G., FLINTOFT, V., BLAIS, R., BROWN, A., COX, J., ETCHELLS, E., GHALI, W. A., HEBERT, P., MAJUMDAR, S. R., O'BEIRNE, M., PALACIOS-DERFLINGHER, L., REID, R. J., SHEPS, S. & TAMBLYN, R. 2004. The Canadian Adverse Events Study: the incidence of adverse events among hospital patients in Canada. *Cmaj,* 170**,** 1678-86.

BANBURY, S. P. & BERRY, D. C. 2005. Office noise and employee concentration: Identifying causes of disruption and potential improvements. *Ergonomics,* 48**,** 25-37.

BARR, J. K., GIANNOTTI, T. E., SOFAER, S., DUQUETTE, C. E., WATERS, W. J. & PETRILLO, M. K. 2006. Using public reports of patient satisfaction for hospital quality improvement. *Health services research,* 41**,** 663-682.

BARRY, C. A., BRITTEN, N., BARBER, N., BRADLEY, C. & STEVENSON, F. 1999. Using reflexivity to optimize teamwork in qualitative research. *Qualitative health research,* 9**,** 26-44.

BATALDEN, P. B. & DAVIDOFF, F. 2007. What is "quality improvement" and how can it transform healthcare? : BMJ Publishing Group Ltd.

BATE, S., ROBERT, G. & MCLEOD, H. 2002. Report on the'Breakthrough'Collaborative approach to quality and service improvement within four regions of the NHS. A research based investigation of the Orthopaedic Services Collaborative within the Eastern, South & West, South East and Trent regions.

BAUMEISTER, R. F. & VOHS, K. D. 2007. *Encyclopedia of social psychology,* Sage.

BENJAMINI, Y. & HOCHBERG, Y. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal statistical society: series B (Methodological),* 57**,** 289-300.

BENNING, A., GHALEB, M., SUOKAS, A., DIXON-WOODS, M., DAWSON, J., BARBER, N., FRANKLIN, B. D., GIRLING, A., HEMMING, K. & CARMALT, M.

2011. Large scale organisational intervention to improve patient safety in four UK hospitals: mixed method evaluation. *Bmj,* 342**,** d195.

BENSON, B. 2016. *Cognitive bias cheat sheet* [Online]. Available: https://betterhumans.pub/cognitive-bias-cheat-sheet-55a472476b18#.f6daic2o3 [Accessed 13/08/2018].

BENSON, R. & HARP, N. 1994. Using systems thinking to extend continuous quality improvement. *The Quality letter for healthcare leaders,* 6**,** 17-24.

BERWICK, D. 2013. An independent report to the Department of Health. A promise to learn–a commitment to act. Improving the safety of patients in England.

BERWICK, D. M., NOLAN, T. W. & WHITTINGTON, J. 2008. The triple aim: care, health, and cost. *Health affairs,* 27**,** 759-769.

BION, J., ALDRIDGE, C., GIRLING, A. J., RUDGE, G., SUN, J., TARRANT, C., SUTTON, E., WILLARS, J., BEET, C. & BOYAL, A. 2020. Changes in weekend and weekday care quality of emergency medical admissions to 20 hospitals in England during implementation of the 7-day services national health policy. *BMJ Quality & Safety*.

BION, J., ALDRIDGE, C. P., GIRLING, A., RUDGE, G., BEET, C., EVANS, T., TEMPLE, R. M., ROSEVEARE, C., CLANCY, M. & BOYAL, A. 2017. Two-epoch cross-sectional case record review protocol comparing quality of care of hospital emergency admissions at weekends versus weekdays. *Bmj open,* 7**,** e018747.

BJÖRKLUND, F. & BÄCKSTRÖM, M. 2008. Individual differences in processing styles: validity of the Rational–Experiential Inventory. *Scandinavian journal of psychology,* 49**,** 439-446.

BLAND, J. M. & ALTMAN, D. G. 1997. Statistics notes: Cronbach's alpha. *BMJ,* 314**,** 572.

BLAWATT KEN, R. 2016. Appendix A: List of Cognitive Biases. *Marconomics.* Emerald Group Publishing Limited.

BLUMENTHAL-BARBY, J. & KRIEGER, H. 2014. Cognitive biases and heuristics in medical decision making a critical review using a systematic search strategy. *Medical Decision Making***,** 0272989X14547740.

BLUMENTHAL-BARBY, J. S. & KRIEGER, H. 2015a. Cognitive biases and heuristics in medical decision making: A critical review using a systematic search strategy. *Medical Decision Making,* 35.

BLUMENTHAL-BARBY, J. S. & KRIEGER, H. 2015b. Cognitive biases and heuristics in medical decision making: A critical review using a systematic search strategy. *Medical Decision Making,* 35**,** 539-557.

BMJ BEST PRACTISE. 2002. *Understanding risk* [Online]. Available: https://bestpractice.bmj.com/info/toolkit/practise-ebm/understanding-risk/ [Accessed 19/02/2021].

BODEMER, N., HANOCH, Y. & KATSIKOPOULOS, K. V. 2015. Heuristics: foundations for a novel approach to medical decision making. *Internal and emergency medicine,* 10**,** 195-203.

BOHMER, R. M. 2011. The four habits of high-value health care organizations. *N Engl J Med,* 365**,** 2045-2047.

BOHR, A. & MEMARZADEH, K. 2020. The rise of artificial intelligence in healthcare applications. *Artificial Intelligence in healthcare.* Elsevier.

BOONYASAI, R. T., WINDISH, D. M., CHAKRABORTI, C., FELDMAN, L. S., RUBIN, H. R. & BASS, E. B. 2007. Effectiveness of teaching quality improvement to clinicians: a systematic review. *Jama,* 298**,** 1023-1037.

BOURKE, B. 2014. Positionality: Reflecting on the research process. *Qualitative Report,* 19.

BRADLEY, E., HOLMBOE, E., MATTERA, J., ROUMANIS, S., RADFORD, M. & KRUMHOLZ, H. 2004. Data feedback efforts in quality improvement: lessons learned from US hospitals. *BMJ Quality & Safety,* 13**,** 26-31.

BRAITHWAITE, J., HERKES, J., LUDLOW, K., TESTA, L. & LAMPRELL, G. 2017. Association between organisational and workplace cultures, and patient outcomes: systematic review. *BMJ open,* 7.

BRENNAN, T. A., LEAPE, L. L., LAIRD, N. M., HEBERT, L., LOCALIO, A. R., LAWTHERS, A. G., NEWHOUSE, J. P., WEILER, P. C. & HIATT, H. H. 1991. Incidence of adverse events and negligence in hospitalized patients: Results of the Harvard medical practice study I. *New England Journal of Medicine,* 324**,** 370-376.

BRICKER, P. 2014. Ontological commitment.

BRITTEN, N. 1995. Qualitative research: qualitative interviews in medical research. *Bmj,* 311**,** 251-253.

BROOK, R. H., MCGLYNN, E. A. & SHEKELLE, P. G. 2000. Defining and measuring quality of care: a perspective from US researchers. *International Journal for Quality in Health Care,* 12**,** 281-295.

BUCK, N., DEVLIN, H. B., LUNN, J. N. & VICKERS, M. D. 1987. *The report of a confidential enquiry into perioperative deaths,* King's Fund Publishing Office.

BUI, T. C., KRIEGER, H. A. & BLUMENTHAL-BARBY, J. S. 2015. Framing effects on physicians' judgment and decision making. *Psychological Reports,* 117**,** 508-522.

BURNS, R. B., FREUND, K. M., MOSKOWITZ, M. A., KASTEN, L., FELDMAN, H. & MCKINLAY, J. B. 1997. Physician characteristics: Do they influence the evaluation and treatment of breast cancer in older women? *The American Journal of Medicine,* 103**,** 263-269.

BURRUS, J. & CARNEY, L. 2015. Social Beliefs and Attitudes: Personality and the Isms. *In:* WRIGHT, J. D. (ed.) *International Encyclopedia of the Social & Behavioral Sciences (Second Edition).* Oxford: Elsevier.

BUSENITZ, L. W. & BARNEY, J. B. Biases and heuristics in strategic decision making: Differences between entrepreneurs and managers in large organizations. Academy of management proceedings, 1994. Academy of Management Briarcliff Manor, NY 10510, 85-89.

BYRNE, A. 2013. Mental workload as a key factor in clinical decision making. *Advances in Health Sciences Education,* 18**,** 537-545.

CABITZA, F., ELLINGSEN, G., LOCORO, A. & SIMONE, C. 2019. Repetita Iuvant: Exploring and Supporting Redundancy in Hospital Practices. *Computer Supported Cooperative Work (CSCW),* 28**,** 61-94.

CACIOPPO, J. T., PETTY, R. E., FEINSTEIN, J. A. & JARVIS, W. B. G. 1996. Dispositional differences in cognitive motivation: The life and times of individuals varying in need for cognition. *Psychological Bulletin,* 119**,** 197.

CALDER, L. A., FORSTER, A. J., STIELL, I. G., CARR, L. K., BREHAUT, J. C., PERRY, J. J., VAILLANCOURT, C. & CROSKERRY, P. 2012. Experiential and rational decision making: a survey to determine how emergency physicians make clinical decisions. *Emergency Medicine Journal,* 29**,** 811-816.

CAMACHO, L. A. & RUBIN, H. R. 1998. Assessment of the validity and reliability of three systems of medical record screening for quality of care assessment. *Medical care,* 36**,** 748-751.

CAMPBELL, M., STYLIANOU, A. C. & SHROPSHIRE, J. 2016. The impact of attitudinal factors on intention to report workplace Internet abuse. *Journal of Information Privacy and Security,* 12**,** 68-83.

CAMPBELL, M. J., JACQUES, R. M., FOTHERINGHAM, J., MAHESWARAN, R. & NICHOLL, J. 2012. Developing a summary hospital mortality index: retrospective analysis in English hospitals over five years. *Bmj,* 344**,** e1001.

CARAYON, P., WETTERNECK, T. B., RIVERA-RODRIGUEZ, A. J., HUNDT, A. S., HOONAKKER, P., HOLDEN, R. & GURSES, A. P. 2014. Human factors systems approach to healthcare quality and patient safety. *Applied ergonomics,* 45**,** 14-25.

CARE QUALITY COMMISSION, U. 2016. Learning, candour and accountability: A review of the way NHS trusts review and investigate the deaths of patients in England.

CARLING, C. L. L., KRISTOFFERSEN, D. T., OXMAN, A. D., FLOTTORP, S., FRETHEIM, A., SCHÜNEMANN, H. J., AKL, E. A., HERRIN, J., MACKENZIE, T. D. & MONTORI, V. M. 2010. The effect of how outcomes are framed on decisions about whether to take antihypertensive medication: a randomized trial. *PLoS ONE,* 5.

CARNEY, P. A., ELMORE, J. G., ABRAHAM, L. A., GERRITY, M. S., HENDRICK, R. E., TAPLIN, S. H., BARLOW, W. E., CUTTER, G. R., POPLACK, S. P. & D'ORSI, C. J. 2004. Radiologist uncertainty and the interpretation of screening. *Medical Decision Making,* 24**,** 255-264.

CARTWRIGHT, N. & BRADBURN, N. 2011. A theory of measurement. National Academies Press.

CARTWRIGHT, N. & CARTWRIGHT, N. 1983. How the laws of physics lie.

CASTI, J. L. 2010. *Mood matters: from rising skirt lengths to the collapse of world powers,* Springer Science & Business Media.

CECCHINI, M., SASSI, F., LAUER, J. A., LEE, Y. Y., GUAJARDO-BARRON, V. & CHISHOLM, D. 2010. Tackling of unhealthy diets, physical inactivity, and obesity: health effects and cost-effectiveness. *The Lancet,* 376**,** 1775-1784.

CERVERO, R. M. & GAINES, J. K. 2015. The impact of CME on physician performance and patient health outcomes: an updated synthesis of systematic reviews. *Journal of Continuing Education in the Health Professions,* 35**,** 131-138.

CHAMBERS, D., CANTRELL, A. & BOOTH, A. 2019. Factors that facilitate the implementation of interventions to reduce preventable hospital admissions with a focus on cardiovascular or respiratory conditions: an evidence map and realist synthesis. *Health Services and Delivery Research*.

CHAPMAN, G. B. 2003. Time discounting of health outcomes. *Time and Decision: Economic and Psychological Perspectives on Intertemporal Choice.* Russell Sage Foundation.

CHAUDHRY, B., WANG, J., WU, S., MAGLIONE, M., MOJICA, W., ROTH, E., MORTON, S. C. & SHEKELLE, P. G. 2006. Systematic review: impact of health information technology on quality, efficiency, and costs of medical care. *Annals of internal medicine,* 144**,** 742-752.

CHEN, I. Y., PIERSON, E., ROSE, S., JOSHI, S., FERRYMAN, K. & GHASSEMI, M. 2020. Ethical Machine Learning in Healthcare. *Annual Review of Biomedical Data Science,* 4.

CHEN, L., SONG, L., SHAO, Y., LI, D. & DING, K. 2019. Using natural language processing to extract clinically useful information from Chinese electronic medical records. *International journal of medical informatics,* 124**,** 6-12.

CHEN, Y. F., BOYAL, A., SUTTON, E., ARMOIRY, X., WATSON, S., BION, J. & TARRANT, C. 2016. The magnitude and mechanisms of the weekend effect in hospital admissions: A protocol for a mixed methods review incorporating a systematic review and framework synthesis. *Systematic Reviews,* 5**,** 84.

CHER, D. J., MIYAMOTO, J. & LENERT, L. A. 1997. Incorporating risk attitude into markov-process decision models:: importance for individual decision making. *Medical Decision Making,* 17**,** 340-350.

CLEARY, P. D. & MCNEIL, B. J. 1988. Patient satisfaction as an indicator of quality care. *Inquiry***,** 25-36.

COCHRANE COLLABORATION 2019. Data Extraction and Assessment Form - Cochrane EPOC for Randomised and Non-randomised controlled trials.

COHEN, H., HUNT, B. J., EFTHYMIOU, M., ARACHCHILLAGE, D. R., MACKIE, I. J., CLAWSON, S., SYLVESTRE, Y., MACHIN, S. J., BERTOLACCINI, M. L. & RUIZ-CASTELLANO, M. 2016. Rivaroxaban versus warfarin to treat patients with thrombotic antiphospholipid syndrome, with or without systemic lupus erythematosus (RAPS): a randomised, controlled, open-label, phase 2/3, non-inferiority trial. *The Lancet Haematology,* 3**,** e426-e436.

COHEN, J. 2013. *Statistical power analysis for the behavioral sciences*, Academic press.

COHEN, S. M. 2000. Aristotle's metaphysics.

COLEMAN, J. S. 1958. Relational analysis: the study of social organizations with survey methods. *Human organization,* 17**,** 28.

COLLINGWOOD, R. G. & COLLINGWOOD, R. G. 1994. *The idea of history,* Oxford University Press on Demand.

CONNOLLY, T. & JOCHEN, R. 2003. Omission bias in vaccination decisions: Where's the "omission"? Where's the "bias"? *Organizational Behavior and Human Decision Processes,* 91**,** 186-202.

COOK, M. J., NOYES, J. M. & MASAKOWSKI, Y. 2007. *Decision making in complex environments*, Ashgate Publishing, Ltd.

COOK, R. I. & WOODS, D. D. 1994. Operating at the sharp end: the complexity of human error. *Human error in medicine,* 13**,** 225-310.

COOKE, C. R. & IWASHYNA, T. J. 2014. Sepsis mandates: improving inpatient care while advancing quality improvement. *Jama,* 312**,** 1397-1398.

CRONSHAW, H. L., DANIELS, R., BLEETMAN, A., JOYNES, E. & SHEILS, M. 2011. Impact of the Surviving Sepsis Campaign on the recognition and management of severe sepsis in the emergency department: are we failing? *Emergency Medicine Journal,* 28**,** 670-675.

CROSKERRY, P. 2002. Achieving quality in clinical decision making: Cognitive strategies and detection of bias. *Academic Emergency Medicine,* 9**,** 1184-1204.

CROSKERRY, P. & CHISHOLM, C. What does human factors ergonomics need to know about front-line medicine? International Conference on Healthcare Systems Ergonomics and Patient Safety: Human Factor, a Bridge Between Care and Cure, HEPS 2005, 2005 Florence. 77-80.

CROSSAN, F. 2003. Research philosophy: towards an understanding. *Nurse researcher,* 11**,** 46-56.

DANERMARK, B. 2002. Interdisciplinary research and critical realism the example of disability research. *Alethia,* 5**,** 56-64.

DANZIGER, S., LEVAV, J. & AVNAIM-PESSO, L. 2011. Extraneous factors in judicial decisions. *Proceedings of the National Academy of Sciences,* 108**,** 6889-6892.

DASGUPTA, N. 2013. Chapter Five - Implicit Attitudes and Beliefs Adapt to Situations: A Decade of Research on the Malleability of Implicit Prejudice, Stereotypes, and the Self-Concept. *In:* DEVINE, P. & PLANT, A. (eds.) *Advances in Experimental Social Psychology.* Academic Press.

DAVIES, H. T., NUTLEY, S. M. & MANNION, R. 2000. Organisational culture and quality of health care. *BMJ Quality & Safety,* 9**,** 111-119.

DAVIS, D. P., JANDRISEVITS, M. D., ILES, S., WEBER, T. R. & GALLO, L. C. 2012. Demographic, Socioeconomic, and Psychological Factors Related to Medication Non-adherence Among Emergency Department Patients. *The Journal of Emergency Medicine,* 43**,** 773-785.

DAVIS, P., LAY-YEE, R., BRIANT, R., ALI, W., SCOTT, A. & SCHUG, S. 2003. Adverse events in New Zealand public hospitals II: Preventability and clinical context. *New Zealand Medical Journal,* 116.

DE BECKER, G. 1997. *The Gift of Fear: Listening to the Intuition that Protects Us from Danger,* Little, Brown & Company.

DE BONO, E. 2017. *Six thinking hats*, Penguin uk.

DE BRÚN, T., O'REILLY-DE BRÚN, M., O'DONNELL, C. A. & MACFARLANE, A. 2016a. Learning from doing: the case for combining normalisation process theory and participatory learning and action research methodology for primary healthcare implementation research. *BMC health services research,* 16**,** 346-346.

DE BRÚN, T., O'REILLY-DE BRÚN, M., O'DONNELL, C. A. & MACFARLANE, A. 2016b. Learning from doing: the case for combining normalisation process theory and participatory learning and action research methodology for primary healthcare implementation research. *BMC health services research,* 16**,** 346.

DE SAVIGNY, D. & ADAM, T. 2009. *Systems thinking for health systems strengthening*, World Health Organization.

DE VRIES, E. N., RAMRATTAN, M. A., SMORENBURG, S. M., GOUMA, D. J. & BOERMEESTER, M. A. 2008. The incidence and nature of in-hospital adverse events: a systematic review. *Quality and Safety in Health Care,* 17**,** 216.

DELLINGER, R. P., CARLET, J. M., MASUR, H., GERLACH, H., CALANDRA, T., COHEN, J., GEA-BANACLOCHE, J., KEH, D., MARSHALL, J. C. & PARKER, M. M. 2004. Surviving Sepsis Campaign guidelines for management of severe sepsis and septic shock. *Intensive care medicine,* 30**,** 536-555.

DELLINGER, R. P., LEVY, M. M., CARLET, J. M., BION, J., PARKER, M. M., JAESCHKE, R., REINHART, K., ANGUS, D. C., BRUN-BUISSON, C. & BEALE, R. 2008. Surviving Sepsis Campaign: international guidelines for management of severe sepsis and septic shock: 2008. *Intensive care medicine,* 34**,** 17-60.

DELLINGER, R. P., LEVY, M. M., RHODES, A., ANNANE, D., GERLACH, H., OPAL, S. M., SEVRANSKY, J. E., SPRUNG, C. L., DOUGLAS, I. S. & JAESCHKE, R. 2013. Surviving Sepsis Campaign: international guidelines for management of severe sepsis and septic shock, 2012. *Intensive care medicine,* 39**,** 165-228.

DENZIN, N. K. 1978. The research act: An introduction to triangulation–Sociological methods. New York: McGraw-Hill.

DENZIN, N. K. 2017. *Sociological methods: A sourcebook*, Routledge.

DENZIN, N. K. & LINCOLN, Y. S. 2005. Introduction: The discipline and practice of qualitative research.

DEPARTMENT OF HEALTH EXPERT GROUP (CHAIRMAN, C. 2000. An organisation with a memory. 92.

DEPARTMENT OF HEALTH, U. K. 2013. Quality in the new health system - Maintaining and improving quality from April 2013.

DESHARNAIS, S. I., CHESNEY, J. D., WROBLEWSKI, R. T., FLEMING, S. T. & MCMAHON JR, L. F. 1988. The Risk-Adjusted Mortality Index: a new measure of hospital performance. *Medical care***,** 1129-1148.

DHANDE, P. S. & KHARAT, R. Character Recognition for Cursive English Handwriting to Recognize Medicine Name from Doctor's Prescription. 2017 International Conference on Computing, Communication, Control and Automation (ICCUBEA), 2017. IEEE, 1-5.

DISABILITIES;, U. O. B. N. F. C. F. L. 2016. The Learning Disabilities Mortality Review (LeDeR) Programme Annual Report 2016–2017.

DIXON, J., SANDERSON, C., ELLIOTT, P., WALLS, P., JONES, J. & PETTICREW, M. 1998. Assessment of the reproducibility of clinical coding in routinely collected hospital activitydata: a study in two hospitals. *Journal of Public Health,* 20**,** 63-69.

DIXON-WOODS, M. 2019. How to improve healthcare improvement—an essay by Mary Dixon-Woods. *bmj,* 367.

DOBLER, C. C., MORROW, A. S. & KAMATH, C. C. 2019. Clinicians' cognitive biases: a potential barrier to implementation of evidence-based clinical practice. *BMJ Evidence-Based Medicine,* 24**,** 137.

DOLAN, P., HALLSWORTH, M., HALPERN, D., KING, D., METCALFE, R. & VLAEV, I. 2012. Influencing behaviour: The mindspace way. *Journal of Economic Psychology,* 33**,** 264-277.

DONABEDIAN, A. 1966. Evaluating the quality of medical care. *The Milbank memorial fund quarterly,* 44**,** 166-206.

DONABEDIAN, A. 1988. The Quality of Care: How Can It Be Assessed? *JAMA,* 260**,** 1743-1748.

DONABEDIAN, A. 2002. *An introduction to quality assurance in health care,* Oxford University Press.

DONALDSON, L. 2002. An organisation with a memory. *Clinical Medicine,* 2**,** 452-457.

DONOGHUE, J. O., KANE, T. O., GALLAGHER, J., COURTNEY, G., AFTAB, A., CASEY, A., TORRES, J. & ANGOVE, P. 2011. Modified early warning scorecard: the role of data/information quality within the decision making process. *Electronic Journal of Information Systems Evaluation,* 14**,** pp100-109-pp100-109.

EDMONDSON, A. C., BOHMER, R. M. & PISANO, G. P. 2001. Disrupted routines: Team learning and new technology implementation in hospitals. *Administrative science quarterly,* 46**,** 685-716.

EDWARDS, M. T. 2013. A longitudinal study of clinical peer review's impact on quality and safety in US hospitals. *Journal of Healthcare Management,* 58**,** 369-384.

EDWARDS, P. K., O'MAHONEY, J. & VINCENT, S. 2014. *Studying organizations using critical realism: A practical guide*, OUP Oxford.

EFFECTIVE PUBLIC HEALTH PRACTISE PROJECT (EPHPP) 2018. Quality assessment tool.

EISENBERG, J. M. 1979. Sociologic influences on decision-making by clinicians. *Annals of internal medicine,* 90**,** 957-964.

EISENBERG, J. M., KITZ, D. S. & WEBBER, R. A. 1983. Development of attitudes about sharing decision-making: a comparison of medical and surgical residents. *Journal of Health and Social Behavior,* 24**,** 85-90.

EISNER, E. W. 2017. *The enlightened eye: Qualitative inquiry and the enhancement of educational practice*, Teachers College Press.

ENDSLEY, M. R. 1995. Toward a theory of situation awareness in dynamic systems. *Human factors,* 37**,** 32-64.

EPLEY, N. & GILOVICH, T. 2006. The anchoring-and-adjustment heuristic: Why the adjustments are insufficient. *Psychological science,* 17**,** 311-318.

EPSTEIN, B. 2018. Social Ontology.

ERBER, M. W., HODGES, S. D. & WILSON, T. D. 1995. Attitude strength, attitude stability, and the effects of analyzing reasons. *Attitude strength: Antecedents and consequences***,** 433-454.

ETIKAN, I., MUSA, S. A. & ALKASSIM, R. S. 2016. Comparison of convenience sampling and purposive sampling. *American journal of theoretical and applied statistics,* 5**,** 1-4.

EVANS, D., BOOKER, S., DAYKIN, N., POWELL, J., TAYLOR, P. & WEIL, S. 2008. National Evaluation of Patient Advice and Liaison Services (PALS): Final Report. *University of the West of england, Bristol*.

EVANS, J. 2008. Dual-processing accounts of reasoning, judgment, and social cognition. *Annu. Rev. Psychol.,* 59**,** 255-278.

FAJARDO, L. J., SORILLO, N. J., GARLIT, J., TOMINES, C. D., ABISADO, M. B., IMPERIAL, J. M. R., RODRIGUEZ, R. L. & FABITO, B. S. Doctor's Cursive Handwriting Recognition System Using Deep Learning. 2019 IEEE 11th international conference on humanoid, nanotechnology, information technology, communication and control, environment, and management (HNICEM), 2019. IEEE, 1-6.

FARMER, T., ROBINSON, K., ELLIOTT, S. J. & EYLES, J. J. Q. H. R. 2006. Developing and implementing a triangulation protocol for qualitative health research. 16**,** 377-394.

FAULKNER, S. L. & TROTTER, S. P. 2017. Data saturation. *The international encyclopedia of communication research methods***,** 1-2.

FEREDAY, J. & MUIR-COCHRANE, E. 2006. Demonstrating rigor using thematic analysis: A hybrid approach of inductive and deductive coding and theme development. *International journal of qualitative methods,* 5**,** 80-92.

FERLIE, E., FITZGERALD, L., WOOD, M. & HAWKINS, C. 2005. The nonspread of innovations: the mediating role of professionals. *Academy of management journal,* 48**,** 117-134.

FERLIE, E. B. & SHORTELL, S. M. 2001. Improving the quality of health care in the United Kingdom and the United States: a framework for change. *The Milbank Quarterly,* 79**,** 281-315.

FETHERSTONHAUGH, D., SLOVIC, P., JOHNSON, S. & FRIEDRICH, J. 1997. Insensitivity to the value of human life: A study of psychophysical numbing. *Journal of Risk and uncertainty,* 14**,** 283-300.

FEYERABEND, P. 1993. *Against method,* Verso.

FINCH, T. L., MAIR, F. S., O'DONNELL, C., MURRAY, E. & MAY, C. R. J. B. M. R. M. 2012. From theory to'measurement'in complex interventions: methodological lessons from the development of an e-health normalisation instrument. 12**,** 69.

FINKEL, M. 2020. *On flexibility,* Stanford University Press.

FINLAY, L. 1998. Reflexivity: an essential component for all research? *British Journal of Occupational Therapy,* 61**,** 453-456.

FINUCANE, M. L., ALHAKAMI, A., SLOVIC, P. & JOHNSON, S. M. 2000. The affect heuristic in judgments of risks and benefits. *Journal of behavioral decision making,* 13**,** 1-17.

FINUCANE, M. L., MERTZ, C., SLOVIC, P. & SCHMIDT, E. S. 2005. Task complexity and older adults' decision-making competence. *Psychology and aging,* 20**,** 71.

FISCHHOFF, B. 1977. Perceived informativeness of facts. *Journal of Experimental Psychology: Human Perception and Performance,* 3**,** 349.

FITZGERALD, C. & HURST, S. 2017. Implicit bias in healthcare professionals: a systematic review. *BMC medical ethics,* 18**,** 19.

FLACH, S. D., CANARIS, G. J., TAPE, T. G., HUNTLEY, K. M. & WIGTON, R. S. 2002. The effects of proficiency and bias on residents' interpretation of the microscopic urinalysis. *Medical Decision Making,* 22**,** 318-325.

FLETCHER, A. 2017. Applying critical realism in qualitative research: methodology meets method. *International Journal of Social Research Methodology,* 20**,** 181-194.

FLINDERS, D. J. 1997. InterViews: An introduction to qualitative research interviewing: Steinar Kvale. Thousand Oaks, CA: Sage Publications, 1996. Pergamon.

FLYVBJERG, B. 2006. Five misunderstandings about case-study research. *Qualitative inquiry,* 12**,** 219-245.

FORREST, C. B., NUTTING, P. A., VON SCHRADER, S., ROHDE, C. & STARFIELD, B. 2006. Primary care physician specialty referral decision making: patient, physician, and health care system determinants. *Medical Decision Making,* 26**,** 76-85.

FORROW, L., TAYLOR, W. C. & ARNOLD, R. M. 1992. Absolutely relative: How research results are summarized can affect treatment decisions. *The American Journal of Medicine,* 92**,** 121-124.

FORSTER, A. J., TALJAARD, M., BENNETT, C. & VAN WALRAVEN, C. 2012. Reliability of the peer-review process for adverse event rating. *PloS one,* 7**,** e41239.

FOSTER, K. H. & HARRISON, E. 2008. Effect of presentation bias on selection of treatment option for failed endodontic therapy. *Oral Surgery, Oral Medicine, Oral Pathology, Oral Radiology and Endodontology,* 106**,** e36-e39.

FOX, C. R. & TVERSKY, A. 1995. Ambiguity aversion and comparative ignorance. *The quarterly journal of economics,* 110**,** 585-603.

FOX, K. A., PICCINI, J. P., WOJDYLA, D., BECKER, R. C., HALPERIN, J. L., NESSEL, C. C., PAOLINI, J. F., HANKEY, G. J., MAHAFFEY, K. W. & PATEL, M. R. 2011. Prevention of stroke and systemic embolism with rivaroxaban compared with

warfarin in patients with non-valvular atrial fibrillation and moderate renal impairment. *European heart journal,* 32**,** 2387-2394.

FOX, R. C. 1989. The sociology of medicine: a participant observer's view.

FRANCIS, R. 2010. *Independent Inquiry into care provided by Mid Staffordshire NHS Foundation Trust: January 2005-March 2009*, The Stationery Office.

FRANCOIS, P., BATAILLER, P., SEIGNEURIN, A., VITTOZ, J.-P., SELLIER, E. & LABARERE, J. 2014. Medical record-keeping and patient perception of hospital care quality. *International journal of health care quality assurance*.

FRANKEL, R. M., FLANAGAN, M., EBRIGHT, P., BERGMAN, A., O'BRIEN, C. M., FRANKS, Z., ALLEN, A., HARRIS, A. & SALEEM, J. J. 2012. Context, culture and (non-verbal) communication affect handover quality. *BMJ quality & safety,* 21**,** i121-i128.

FREEMAN, M. 2016. *Modes of thinking for qualitative data analysis*, Taylor & Francis.

FUSCH, P. I. & NESS, L. R. 2015. Are we there yet? Data saturation in qualitative research. *The qualitative report,* 20**,** 1408.

GALE, N. K., HEATH, G., CAMERON, E., RASHID, S. & REDWOOD, S. 2013. Using the framework method for the analysis of qualitative data in multi-disciplinary health research. *BMC Medical Research Methodology,* 13**,** 117.

GEERTZ, C. 1973. Thick description: Toward an interpretive theory of culture. *Turning points in qualitative research: Tying knots in a handkerchief,* 3**,** 143-168.

GERRITY, M. S., WHITE, K. P., DEVELLIS, R. F. & DITTUS, R. S. 1995. Physicians' reactions to uncertainty: refining the constructs and scales. *Motivation and Emotion,* 19**,** 175-191.

GIBBS, J., CLARK, K., KHURI, S., HENDERSON, W., HUR, K. & DALEY, J. 2001. Validating risk-adjusted surgical outcomes: chart review of process of care. *International Journal for Quality in Health Care,* 13**,** 187-196.

GIGERENZER, G. 2008a. Bounded and rational. *Philosophie: Grundlagen und Anwendungen/Philosophy: Foundations and Applications.* mentis.

GIGERENZER, G. 2008b. *Gut feelings: Short cuts to better decision making*, Penguin Uk.

GIGERENZER, G. 2018. The Bias Bias in Behavioral Economics. *Review of Behavioral Economics,* 5**,** 303-336.

GIGERENZER, G. & BRIGHTON, H. 2009. Homo Heuristicus: Why Biased Minds Make Better Inferences. *Topics in Cognitive Science,* 1**,** 107-143.

GIGERENZER, G. & GAISSMAIER, W. 2011a. Heuristic decision making. *Annu Rev Psychol,* 62.

GIGERENZER, G. & GAISSMAIER, W. 2011b. Heuristic decision making. *Annu Rev Psychol,* 62**,** 451-82.

GILOVICH, T. 2008. *How we know what isn't so*, Simon and Schuster.

GILOVICH, T., GRIFFIN, D. & KAHNEMAN, D. 2002. *Heuristics and biases: The psychology of intuitive judgment*, Cambridge university press.

GIRLING, A. J., HOFER, T. P., WU, J., CHILTON, P. J., NICHOLL, J. P., MOHAMMED, M. A. & LILFORD, R. J. 2012a. Case-mix adjusted hospital mortality is a poor proxy for preventable mortality: a modelling study. *Bmj Quality & Safety,* 21**,** 1052-1056.

GIRLING, A. J., HOFER, T. P., WU, J., CHILTON, P. J., NICHOLL, J. P., MOHAMMED, M. A. & LILFORD, R. J. 2012b. Case-mix adjusted hospital mortality is a poor proxy for preventable mortality: A modelling study. *BMJ Quality and Safety,* 21**,** 1052-1056.

GLADWELL, M. 2005. Blink. London: Allen Lane. Penguin.

GLUYAS, H. 2018. Understanding the human and system factors involved in medication errors. *Nurs Stand*.

GODIN, M., MURRAY, P. & MEHTA, R. L. Clinical approach to the patient with AKI and sepsis. Seminars in nephrology, 2015. Elsevier, 12-22.

GOLDACRE, B. 2022. Better, broader, safer: using health data for research and analysis. Department of Health and Social Care, UK.

GOLDMAN, A. I. 1999. *Knowledge in a social world*, Clarendon Press Oxford.

GOLDMAN, R. L. 1992. The reliability of peer assessments of quality of care. *Journal of American Medical Association,* 267**,** 958-960.

GOLDMAN, R. L. 1994. The reliability of peer assessments of quality of care. *Journal of American Medical Association,* 17**,** 3-21.

GOODFELLOW, I., BENGIO, Y. & COURVILLE, A. 2016. *Deep learning*, MIT press.

GOWENSMITH, W., MURRIE, D. & BOCCACCINI, M. 2012. Field Reliability of Competence to Stand Trial Opinions: How Often Do Evaluators Agree, and What Do Judges Decide When Evaluators Disagree? *Law and human behavior,* 36**,** 130-9.

GRAZ, B., WIETLISBACH, V., PORCHET, F. & VADER, J. P. 2005. Prognosis or "curabo effect?" Physician prediction and patient outcome of surgery for low back pain and sciatica. *Spine,* 30**,** 1448-1452.

GREENGLASS, E., SCHWARZER, R., JAKUBIEC, S. D., FIKSENBAUM, L. & TAUBERT, S. 2014. The Proactive Coping Inventory (PCI): A Multidimensional Research Instrument. Paper Presented at the 20th International Cracow, Poland, 1999. *Conference of the STAR (Stress and Anxiety Research Society).*

GREENHALGH, T., ROBERT, G., BATE, P., MACFARLANE, F. & KYRIAKIDOU, O. 2008. *Diffusion of innovations in health service organisations: a systematic literature review,* John Wiley & Sons.

GROOPMAN, J. E. & PRICHARD, M. 2007. *How doctors think,* Houghton Mifflin Boston.

GUBA, E. G. & LINCOLN, Y. S. 1994. Competing paradigms in qualitative research. *Handbook of qualitative research,* 2**,** 105.

GUBA, E. G. J. E. 1981. Criteria for assessing the trustworthiness of naturalistic inquiries. 29**,** 75.

GUPTA, M., SCHRIGER, D. L. & TABAS, J. A. 2011. The presence of outcome bias in emergency physician retrospective judgments of the quality of care. *Annals of Emergency Medicine,* 57**,** 323-328.e9.

GUTHRIE, C., RACHLINSKI, J. J. & WISTRICH, A. J. 2007. Blinking on the bench: How judges decide cases. *Cornell L. Rev.,* 93**,** 1.

HAJJAJ, F. M., SALEK, M. S., BASRA, M. K. A. & FINLAY, A. Y. 2010. Non-clinical influences on clinical decision-making: a major challenge to evidence-based practice. *Journal of the Royal Society of Medicine,* 103**,** 178-187.

HAMMOND, M. E. H., STEHLIK, J., DRAKOS, S. G. & KFOURY, A. G. 2021. Bias in medicine: lessons learned and mitigation strategies. *Basic to Translational Science,* 6**,** 78-85.

HANNA, P. & HARRISON, B. 2004. *Word and world: Practice and the foundations of language,* Cambridge University Press.

HARRIS, P. A., TAYLOR, R., THIELKE, R., PAYNE, J., GONZALEZ, N. & CONDE, J. G. J. J. O. B. I. 2009. Research electronic data capture (REDCap)—a metadata-driven methodology and workflow process for providing translational research informatics support. 42**,** 377-381.

HARTELOH, P. P. 2003. The meaning of quality in health care: a conceptual analysis. *Health Care Analysis,* 11**,** 259-267.

HARVEY, N. 1994. Relations between confidence and skilled performance.

HAYES, R. P., STOUDEMIRE, A. S., KINLAW, K., LYNN DELL, M. & LOOMIS, A. 1999. Changing attitudes about end-of-life decision making of medical students during third-year clinical clerkships. *Psychosomatics,* 40**,** 205-211.

HAYWARD, R. A. & HOFER, T. P. 2001. Estimating hospital deaths due to medical errors: preventability is in the eye of the reviewer. *Jama,* 286**,** 415-420.

HAYWARD, R. S. A., GUYATT, G. H., MOORE, K. A., MCKIBBON, A. & CARTER, A. O. 1997. Canadian physicians attitudes about and preferences regarding clinical practice guidelines. *Canadian Medical Association Journal,* 156**,** 1715.

HEALTH QUALITY IMPROVEMENT PARTNERSHIP (HQIP) 2015a. Tender for Retrospective case record review programme - Lot 1 Tender Documentation for a National Retrospective Case Record Review Programme - Establishment and roll-out a standardised methodology and process for retrospective case record review (RCRR) for adult, acute care deaths in England and Scotland to improve understanding and learning about problems in care that may have contributed to a patient's death

.

HEALTH QUALITY IMPROVEMENT PARTNERSHIP (HQIP) 2015b. Tender for Retrospective case record review programme - Lot 2 Tender Documentation for a National Programme to use retrospective record review to estimate the proportion of hospital deaths due to problems in care in English acute NHS trusts.

HEALTH, S. O. S. F. 2008. *High quality care for all: NHS next stage review final report,* The Stationery Office.

HEALTHCARE AND WORKFORCE IMPROVEMENT. 2017. *Dimensions of Quality* [Online]. Available: http://www.qaproject.org/methods/resdimension.html [Accessed].

HELLER, R. F. & SALTZSTEIN, H. D. 1992. Heuristics in Medical and Non-medical Decision-making. *The Quarterly Journal of Experimental Psychology Section A,* 44**,** 211-235.

HEMMERICH, J. A., ELSTEIN, A. S., SCHWARZE, M. L., MOLISKI, E. G. & DALE, W. 2012. Risk as feelings in the effect of patient outcomes on physicians' future treatment decisions: A randomized trial and manipulation validation. *Social Science and Medicine,* 75**,** 367-376.

HENSHER, M., EDWARDS, N. & STOKES, R. 1999. International trends in the provision and utilisation of hospital care. *Bmj,* 319**,** 845-848.

HESS, T. M. 2001. Ageing-related influences on personal need for structure. *International Journal of Behavioral Development,* 25**,** 482-490.

HEYWORTH, J., KEEP, J. & DANIELS, R. 2009. On behalf of the Clinical Effectiveness Committee of the College of Emergency Medicine. Sepsis pack 2009.

HOFER, T. P., ASCH, S. M., HAYWARD, R. A., RUBENSTEIN, L. V., HOGAN, M. M., ADAMS, J. & KERR, E. A. 2004. Profiling quality of care: Is there a role for peer review? *BMC health services research,* 4**,** 9.

HOFER, T. P., BERNSTEIN, S. J., DEMONNER, S. & HAYWARD, R. A. 2000. Discussion between reviewers does not improve reliability of peer review of hospital quality. *Medical Care,* 38**,** 152-161.

HOFMANN, W., FRIESE, M. & WIERS, R. W. 2008. Impulsive versus reflective influences on health behavior: A theoretical framework and empirical review. *Health Psychology Review,* 2**,** 111-137.

HOGAN, H. 2016. The problem with preventable deaths. *BMJ Quality and Safety,* 25**,** 320-323.

HOGAN, H., HEALEY, F., NEALE, G., THOMSON, R., BLACK, N. & VINCENT, C. 2014. Learning from preventable deaths: Exploring case record reviewers' narratives using change analysis. *Journal of the Royal Society of Medicine,* 107**,** 365-375.

HOGAN, H., HEALEY, F., NEALE, G., THOMSON, R., VINCENT, C. & BLACK, N. 2012a. Preventable deaths due to problems in care in English acute hospitals: a retrospective case record review study. *BMJ quality & safety***,** bmjqs-2012-001159.

HOGAN, H., HEALEY, F., NEALE, G., THOMSON, R., VINCENT, C. & BLACK, N. 2012b. Preventable deaths due to problems in care in English acute hospitals: A retrospective case record review study. *BMJ Quality and Safety,* 21**,** 737-745.

HOGAN, H., ZIPFEL, R., NEUBURGER, J., HUTCHINGS, A., DARZI, A. & BLACK, N. 2015. Avoidability of hospital deaths and association with hospital-wide mortality ratios: retrospective case record review and regression analysis. *Bmj-British Medical Journal,* 351.

HOLLNAGEL, E. 2018. *Safety–I and safety–II: the past and future of safety management*, CRC press.

HOLLNAGEL, E., WEARS, R. L. & BRAITHWAITE, J. From Safety-I to Safety-II: a white paper.

HORNE, J. A. & ÖSTBERG, O. 1976. A self-assessment questionnaire to determine morningness-eveningness in human circadian rhythms. *International journal of chronobiology*.

HSEE, C. K. 1998. Less is better: When low-value options are valued more highly than high-value options. *Journal of Behavioral Decision Making,* 11**,** 107-121.

HU, J., JORDAN, J., RUBINFELD, I., SCHREIBER, M., WATERMAN, B. & NERENZ, D. 2017. Correlations among hospital quality measures: what “Hospital Compare” data tell us. *American Journal of Medical Quality,* 32**,** 605-610.

HUBER, J., PAYNE, J. W. & PUTO, C. 1982. Adding asymmetrically dominated alternatives: Violations of regularity and the similarity hypothesis. *Journal of consumer research,* 9**,** 90-98.

HUBERMAN, M. & MILES, M. B. 2002. *The qualitative researcher's companion,* Sage.

HUSSEY, P. S., WERTHEIMER, S. & MEHROTRA, A. 2013. The association between health care quality and cost: a systematic review. *Annals of internal medicine,* 158**,** 27-34.

HUTCHINSON, A., COSTER, J., COOPER, K., MCINTOSH, A., WALTERS, S., BATH, P., PEARSON, M., RANTELL, K., CAMPBELL, M. & NICHOLL, J. 2010a. Assessing quality of care from hospital case notes: comparison of reliability of two methods. *Quality and Safety in Health Care***,** qshc. 2007.023911.

HUTCHINSON, A., COSTER, J., COOPER, K., MCINTOSH, A., WALTERS, S., BATH, P., PEARSON, M., YOUNG, T., RANTELL, K. & CAMPBELL, M. 2010b. Comparison of case note review methods for evaluating quality and safety in health care. *Health Technology Assessment,* 14**,** 1-170.

HUTCHINSON, A., COSTER J, E., COOPER K, L., MCINTOSH, A. & WALTERS S, J. 2010c. Comparison of case note review methods for evaluating quality and safety in health care. *Health Technology Assessment,* 14**,** 170.

HUTCHINSON, A., COSTER, J. E., COOPER, K. L., MCINTOSH, A., WALTERS, S. J., BATH, P. A., PEARSON, M., YOUNG, T. A., RANTELL, K., CAMPBELL, M. J. & RATCLIFFE, J. 2010d. Comparison of case note review methods for evaluating quality and safety in health care. *Health Technology Assessment,* 14**,** 1-170.

HUTCHINSON, A., COSTER, J. E., COOPERK.L, MCINTOSH, A., WALTERS, S. J., BATH, P. A., PEARSON, M., RANTELL, K., CAMPBELL, M. J., NICHOLL, J. &

IRWIN, P. 2010e. Assessing quality of care from hospital case notes: Comparison of reliability of two methods. *Quality and Safety in Health Care,* 19.

IBRAHIM, H., MUNKHBAYAR, U., TOIVGOO, A., HUMPHRIES, C., OCHIR, C., NARULA, I. S., LILFORD, R. & MANASEKI-HOLLAND, S. 2019. Can universal patient-held health booklets promote continuity of care and patient-centred care in low-resource countries? The case of Mongolia. *BMJ quality & safety,* 28**,** 729-740.

IEZZONI, L. I. 1996. 100 apples divided by 15 red herrings: a cautionary tale from the mid-19th century on comparing hospital mortality rates. *Annals of Internal Medicine,* 124**,** 1079-1085.

IEZZONI, L. I. 2003. *Risk adjustment for measuring health care outcomes*, Health Administration Press Chicago.

INQUIRY, M. S. N. F. T. P. 2013. *Report of the Mid Staffordshire NHS Foundation Trust Public Inquiry: Executive Summary*, The Stationery Office.

INSTITUTE FOR HEALTH IMPROVEMENT (IHI) 2009. IHI Global Trigger Tool for Measuring Adverse Events: Second Edition. http://www.ihi.org/_layouts/15/ihi/login/login.aspx?ReturnURL=%2fresources%2fPages%2fIHIWhitePapers%2fIHIGlobalTriggerToolWhitePaper.aspx: Institute for Health Improvement (IHI).

INSTITUTE OF HEALTHCARE IMPROVEMENT (IHI) 2008. IHI Global Trigger Tool for Measuring Adverse Events (UK version). http://www.ihi.org/resources/_layouts/download.aspx?SourceURL=%2fresources%2fKnowledge+Center+Assets%2fTools+-+IHIGlobalTriggerToolforMeasuringAdverseEvents_df8a18b6-52cc-4674-8258-030941832115%2fIHIGlobalTriggerTool_UKVersion2008.pdf: Institute of Healtcare Improvement (IHI).

INTELLIGENCE, D. F. 2009. The Dr Foster Hospital Guide 2009: How Safe is Your Hospital. *London: Dr Foster Intelligence*.

JAMES, B. C. & SAVITZ, L. A. 2011. How Intermountain trimmed health care costs through robust quality improvement efforts. *Health Affairs,* 30**,** 1185-1191.

JANSEN, A. C. M., VAN AALST-COHEN, E. S., HUTTEN, B. A., BÜLLER, H. R., KASTELEIN, J. J. P. & PRINS, M. H. 2005. Guidelines were developed for data collection from medical records for use in retrospective analyses. *Journal of Clinical Epidemiology,* 58**,** 269-274.

JARMAN, B., BOTTLE, A., AYLIN, P. & BROWNE, M. 2005. Monitoring changes in hospital standardised mortality ratios. *Bmj,* 330**,** 329.

JARMAN, B., PIETER, D., VAN DER VEEN, A. A., KOOL, R. B., AYLIN, P., BOTTLE, A., WESTERT, G. P. & JONES, S. 2010. The hospital standardised mortality ratio: a powerful tool for Dutch hospitals to assess their quality of care? *Quality and Safety in Health Care,* 19**,** 9.

JERRARD, H. & MCNEILL, D. 1992. Systems of units. *Dictionary of Scientific Units.* Springer.

JONES, E. E. & HARRIS, V. A. 1967. The attribution of attitudes. *Journal of Experimental Social Psychology,* 3**,** 1-24.

JONES, K., WRIGHT, C. & BELL, A. 2012. Do multilevel models ever give different results. *Or why bother with multilevel models*.

JUDA, M., VETTER, C. & ROENNEBERG, T. 2013. The Munich chronotype questionnaire for shift-workers (MCTQShift). *Journal of biological rhythms,* 28**,** 130-140.

JULIUSSON, E. Á., KARLSSON, N. & GÄRLING, T. 2005. Weighing the past and the future in decision making. *European Journal of Cognitive Psychology,* 17**,** 561-575.

KAHN, K. L., RUBENSTEIN, L. V., SHERWOOD, M. J. & BROOK, R. H. 1989. Structured Implicit Review for Physician Implicit Measurement of Quality of Care.

KAHNEMAN, D. 2011. *Thinking, fast and slow*, Macmillan.

KAHNEMAN, D. & KLEIN, G. 2009. Conditions for Intuitive Expertise: A Failure to Disagree. *American Psychologist,* 64**,** 515-526.

KAHNEMAN, D., KNETSCH, J. L. & THALER, R. H. 1991. Anomalies: The endowment effect, loss aversion, and status quo bias. *The journal of economic perspectives,* 5**,** 193-206.

KAHNEMAN, D., SIBONY, O. & SUNSTEIN, C. R. 2021. *Noise: a flaw in human judgment*, Little, Brown.

KAHNEMAN, D., SLOVIC, P. & TVERSKY, A. 1982. Judgement under uncertainty: Heuristics and biases. Cambridge University Press, Cambridge.

KAHNEMAN, D. & TVERSKY, A. 1984. Choices, values, and frames. *American Psychologist,* 39**,** 341.

KAMAL, R. N., LINDSAY, S. E. & EPPLER, S. L. 2018. Patients should define value in health care: a conceptual framework. *The journal of hand surgery*, 43**,** 1030-1034.

KAPLAN, H. C., PROVOST, L. P., FROEHLE, C. M. & MARGOLIS, P. A. 2012. The Model for Understanding Success in Quality (MUSIQ): building a theory of context in healthcare quality improvement. *BMJ quality & safety,* 21**,** 13-20.

KARASSAVIDOU, E., GLAVELI, N. & ZAFIROPOULOS, K. 2011. Assessing hospitals' readiness for clinical governance quality initiatives through organisational climate. *Journal of health Organization and Management*.

KARNIELI-MILLER, O., STRIER, R. & PESSACH, L. 2009. Power relations in qualitative research. *Qualitative health research,* 19**,** 279-289.

KASHANI, K., ROSNER, M. H., HAASE, M., LEWINGTON, A. J., O'DONOGHUE, D. J., WILSON, F. P., NADIM, M. K., SILVER, S. A., ZARBOCK, A. & OSTERMANN, M. 2019. Quality improvement goals for acute kidney injury. *Clinical Journal of the American Society of Nephrology,* 14**,** 941-953.

KEIJZERS, G., FATOVICH, D. M., EGERTON-WARBURTON, D., CULLEN, L., SCOTT, I. A., GLASZIOU, P. & CROSKERRY, P. 2018. Deliberate clinical inertia: using meta-cognition to improve decision-making. *Emergency Medicine Australasia,* 30**,** 585-590.

KELLER, N., COKELY, E. T., KATSIKOPOULOS, K. V. & WEGWARTH, O. 2010. Naturalistic Heuristics for Decision Making. *Journal of Cognitive Engineering and Decision Making,* 4**,** 256-274.

KELLEY, E. & HURST, J. 2006. Health care quality indicators project: conceptual framework paper.

KELLY, T. 2005. The epistemic significance of disagreement. *Oxford studies in epistemology,* 1.

KEOGH, B. 2013. *Review into the quality of care and treatment provided by 14 hospital trusts in England: overview report,* NHS.

KHATRI, N. 2009. Consequences of power distance orientation in organisations. *Vision,* 13**,** 1-9.

KIRKUP, B. & INVESTIGATION, M. B. 2015. *The Report of the Morecambe Bay Investigation: An independent investigation into the management, delivery and outcomes of care provided by the maternity and neonatal services at the University Hospitals of Morecambe Bay NHS Foundation Trust from January 2004 to June 2013,* Stationery Office.

KISH, L. 1965. Survey sampling.

KLEIN, G. A. 1999. *Sources of power: How people make decisions*, MIT press.

KLEIN, G. A., CALDERWOOD, R. & CLINTON-CIROCCO, A. Rapid decision making on the fire ground. Proceedings of the Human Factors and Ergonomics Society annual meeting, 1986. SAGE Publications, 576-580.

KLEIN, M. W., MALONE, M. F., BENNIS, W. G. & BERKOWITZ, N. H. 1961. Problems of measuring patient care in the out-patient department. *Journal of Health and Human Behavior,* 2**,** 138-144.

KLEINMAN, L. C. & DOUGHERTY, D. 2013. Assessing Quality Improvement in Health Care: Theory for Practice. *Pediatrics,* 131**,** S110.

KOBEWKA, D. M., VAN WALRAVEN, C., TURNBULL, J., WORTHINGTON, J., CALDER, L. & FORSTER, A. 2017. Quality gaps identified through mortality review. *BMJ quality & safety,* 26**,** 141-149.

KOBUSINGYE, O. C., HYDER, A. A., BISHAI, D., HICKS, E. R., MOCK, C. & JOSHIPURA, M. 2005. Emergency medical systems in low-and middle-income countries: recommendations for action. *Bulletin of the World Health Organization,* 83**,** 626-631.

KOCH, K.-R. 2013. *Parameter estimation and hypothesis testing in linear models*, Springer Science & Business Media.

KOEHLER, D. J. & HARVEY, N. 2008. *Blackwell handbook of judgment and decision making*, John Wiley & Sons.

KOHN, L. T., CORRIGAN, J. M. & DONALDSON, M. S. 2000. *To err is human:: building a Safer Health System*, National Academies Press.

KOO, T. K. & LI, M. Y. 2016. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *Journal of chiropractic medicine,* 15**,** 155-163.

KORACH, Z. T., YANG, J., ROSSETTI, S. C., CATO, K. D., KANG, M.-J., KNAPLUND, C., SCHNOCK, K. O., GARCIA, J. P., JIA, H. & SCHWARTZ, J. M. 2020. Mining clinical phrases from nursing notes to discover risk factors of patient deterioration. *International journal of medical informatics,* 135**,** 104053.

KRANTZ, D. H. & TVERSKY, A. 1971. Conjoint-measurement analysis of composition rules in psychology. *Psychological review,* 78**,** 151.

KREFTING, L. 1991. Rigor in qualitative research: The assessment of trustworthiness. *J The American journal of occupational therapy,* 45**,** 214-222.

KUHN, T. S. 2012. *The structure of scientific revolutions*, University of Chicago press.

KUMAH, E. 2017. Patient experience and satisfaction with a healthcare system: connecting the dots. *International Journal of Healthcare Management*.

LALONDE, M. 1974. A new perspective on the health of Canadians. *www. phac-aspc. gc. ca/ph-sp/phdd/pdf/perspective. pdf*.

LANGLEY, G. J., NOLAN, K., NORMAN, C., PROVOST, L. & NOLAN, T. 1996. A practical approach to enhancing organizational performance. New York: Jossey-Bass.

LARRICK, R. P. 2004. Debiasing.

LAU, H. & LITMAN, K. C. 2011. Saving lives by studying deaths: using standardized mortality reviews to improve inpatient safety. *The Joint Commission Journal on Quality and Patient Safety,* 37**,** 400-AP2.

LEAPE, L. L., BRENNAN, T. A., LAIRD, N., LAWTHERS, A. G., LOCALIO, A. R., BARNES, B. A., HEBERT, L., NEWHOUSE, J. P., WEILER, P. C. & HIATT, H. 1991. The nature of adverse events in hospitalized patients: Results of the Harvard medical practice study II. *New England Journal of Medicine,* 324**,** 377-384.

LEDOUX, J. 1998. *The emotional brain: The mysterious underpinnings of emotional life*, Simon and Schuster.

LEE, R. I. 1933. *The fundamentals of good medical care: an outline of the fundamentals of good medical care and an estimate of the service required to supply the medical needs of the United States*, University of Chicago Press.

LEGAULT, M. R. 2006. *Think!: Why Crucial Decisions Can't be Made in the Blink of an Eye*, Simon and Schuster.

LEONARD, M., GRAHAM, S. & BONACUM, D. 2004. The human factor: the critical importance of effective teamwork and communication in providing safe care. *BMJ Quality & Safety,* 13**,** i85-i90.

LEVY, M. M., DELLINGER, R. P., TOWNSEND, S. R., LINDE-ZWIRBLE, W. T., MARSHALL, J. C., BION, J., SCHORR, C., ARTIGAS, A., RAMSAY, G. & BEALE, R. 2010. The Surviving Sepsis Campaign: results of an international guideline-based performance improvement program targeting severe sepsis. *Intensive care medicine,* 36**,** 222-231.

LIANG, C. & GONG, Y. 2017. Automated classification of multi-labeled patient safety reports: a shift from quantity to quality measure. *MEDINFO 2017: Precision Healthcare through Informatics.* IOS Press.

LILFORD, R., EDWARDS, A., GIRLING, A., HOFER, T., DI TANNA, G. L., PETTY, J. & NICHOLL, J. 2007. Inter-rater reliability of case-note audit: A systematic review. *Journal of Health Services Research and Policy,* 12**,** 173-180.

LILFORD, R. & PRONOVOST, P. 2010a. Using hospital mortality rates to judge hospital performance: A bad idea that just won't go away. *BMJ (Online),* 340**,** 955-957.

LILFORD, R. & PRONOVOST, P. 2010b. Using hospital mortality rates to judge hospital performance: a bad idea that just won't go away. *Bmj,* 340**,** c2016.

LINCOLN, Y. S. & GUBA, E. G. 1985. Naturalistic inquiry. CA: Sage.

LINS DE HOLANDA COELHO, G., H. P. HANEL, P. & J. WOLF, L. 2018. The Very Efficient Assessment of Need for Cognition: Developing a Six-Item Version. *Assessment***,** 1073191118793208.

LIPITZ-SNYDERMAN, A., STEINWACHS, D., NEEDHAM, D. M., COLANTUONI, E., MORLOCK, L. L. & PRONOVOST, P. J. 2011. Impact of a statewide intensive care unit quality improvement initiative on hospital mortality and length of stay: retrospective comparative analysis. *Bmj,* 342.

LOEWENSTEIN, G. F. & PRELEC, D. 2000. Preferences for sequences of outcomes. *Choices, Values, and Frames***,** 565-577.

LOHR, K. 1990. Committee to design a strategy for quality review and assurance. *Medicare: a strategy for quality assurance,* 1.

LOHR, K. N. 1991. Medicare: a strategy for quality assurance. *Journal of quality assurance: a publication of the National Association of Quality Assurance Professionals,* 13**,** 10-13.

LOUW, S., TODD, R. W. & JIMARKON, P. Active listening in qualitative research interviews. Proceedings of the International Conference: Research in Applied Linguistics, April, 2011.

LUDOLPH, R. & SCHULZ, P. J. 2018. Debiasing Health-Related Judgments and Decision Making: A Systematic Review. *Medical Decision Making,* 38**,** 3-13.

MACKENZIE, E. J., STEINWACHS, D. M., BONE, L. R., FLOCCARE, D. J. & RAMZY, A. I. 1992. Inter-rater reliability of preventable death judgments. The Preventable Death Study Group. *The Journal of Trauma,* 33**,** 292-302, 302-3.

MAINZ, J. 2003. Defining and classifying clinical indicators for quality improvement. *International journal for quality in health care,* 15**,** 523-530.

MAKARY, M. A., WICK, E. & FREISCHLAG, J. A. 2011. PPE, OPPE, and FPPE: complying with the new alphabet soup of credentialing. *Archives of Surgery,* 146**,** 642-644.

MALINA, M. A., NØRREKLIT, H. S. & SELTO, F. H. 2011. Lessons learned: advantages and disadvantages of mixed method research. *Qualitative Research in Accounting & Management*.

MAMEDE, S., DE CARVALHO-FILHO, M. A., DE FARIA, R. M. D., FRANCI, D., NUNES, M. D. P. T., RIBEIRO, L. M. C., BIEGELMEYER, J., ZWAAN, L. & SCHMIDT, H. G. 2020. 'Immunising' physicians against availability bias in diagnostic reasoning: a randomised controlled experiment. *BMJ Quality & Safety*.

MAMEDE, S., VAN GOG, T., VAN DEN BERGE, K., RIKERS, R. M. J. P., VAN SAASE, J. L. C. M., VAN GULDENER, C. & SCHMIDT, H. G. 2010. Effect of availability bias and reflective reasoning on diagnostic accuracy among internal medicine residents. *JAMA - Journal of the American Medical Association,* 304**,** 1198-1203.

MANARY, M. P., BOULDING, W., STAELIN, R. & GLICKMAN, S. W. 2013. The patient experience and health outcomes. *The New England journal of medicine*.

MANASEKI-HOLLAND, S., LILFORD, R. J., BISHOP, J. R., GIRLING, A. J., CHEN, Y.-F., CHILTON, P. J., HOFER, T. P., ALNER, V., BLEASDALE, J. & BOSANKO, N. 2016. Reviewing deaths in British and US hospitals: a study of two scales for assessing preventability. *BMJ Quality & Safety,* 26**,** 408-416.

MANASEKI-HOLLAND, S., LILFORD, R. J., TE, A. P., CHEN, Y., GUPTA, K. K., CHILTON, P. J. & HOFER, T. P. 2019a. Ranking Hospitals Based on Preventable Hospital Death Rates: A Systematic Review With Implications for Both Direct Measurement and Indirect Measurement Through Standardized Mortality Rates. *Milbank Quarterly,* 97**,** 228-284.

MANASEKI-HOLLAND, S., LILFORD, R. J., TE, A. P., CHEN, Y.-F., GUPTA, K. K., CHILTON, P. J. & HOFER, T. P. 2019b. Ranking hospitals based on preventable hospital death rates: A systematic review with implications for both direct measurement and indirect measurement through standardized mortality rates. *The Milbank Quarterly,* 97**,** 228-284.

MANDOLESI, L., POLVERINO, A., MONTUORI, S., FOTI, F., FERRAIOLI, G., SORRENTINO, P. & SORRENTINO, G. 2018. Effects of physical exercise on cognitive functioning and wellbeing: biological and psychological benefits. *Frontiers in psychology,* 9**,** 509.

MANNION, R. & BRAITHWAITE, J. 2012. Unintended consequences of performance measurement in healthcare: 20 salutary lessons from the English National Health Service. *Internal Medicine Journal,* 42**,** 569-574.

MANNION, R. & DAVIES, H. 2018a. Understanding organisational culture for healthcare quality improvement. *BMJ (Online),* 363.

MANNION, R. & DAVIES, H. 2018b. Understanding organisational culture for healthcare quality improvement. *Bmj,* 363.

MANNION, R. & DAVIES, H. T. 2015. Cultures of silence and cultures of voice: the role of whistleblowing in healthcare organisations. *International journal of health policy and management,* 4**,** 503.

MARCIN, J. P., ROMANO, P. S., DHARMAR, M., CHAMBERLAIN, J. M., DUDLEY, N., MACIAS, C. G., NIGROVIC, L. E., POWELL, E. C., ROGERS, A. J. & SONNETT, M. 2018. Implicit review instrument to evaluate quality of care delivered by physicians to children in emergency departments. *Health services research,* 53**,** 1316-1334.

MAREWSKI, J. N. & GIGERENZER, G. 2012a. Heuristic decision making in medicine. *Dialogues in clinical neuroscience,* 14**,** 77.

MAREWSKI, J. N. & GIGERENZER, G. 2012b. Heuristic decision making in medicine. *Dialogues in Clinical Neuroscience,* 14**,** 77-89.

MARKS, A. D., HINE, D. W., BLORE, R. L. & PHILLIPS, W. J. 2008. Assessing individual differences in adolescents' preference for rational and experiential cognition. *Personality and Individual Differences,* 44**,** 42-52.

MARSHALL, M. 2009. Applying quality improvement approaches to health care. British Medical Journal Publishing Group.

MARTEAU, T. M., HOLLANDS, G. J. & FLETCHER, P. C. 2012. Changing human behavior to prevent disease: the importance of targeting automatic processes. *science,* 337**,** 1492-1495.

MASO, G., MONASTA, L., PICCOLI, M., RONFANI, L., MONTICO, M., DE SETA, F., PAROLIN, S., BUSINELLI, C., TRAVAN, L., ALBERICO, S. & MULTICENTER STUDY GROUP ON MODE OF DELIVERY IN FRIULI VENEZIA, G. 2015. Risk-adjusted operative delivery rates and maternal-neonatal outcomes as measures of quality assessment in obstetric care: a multicenter prospective study. *BMC pregnancy and childbirth,* 15**,** 20-20.

MAY, C. & FINCH, T. 2009. Implementing, embedding, and integrating practices: an outline of normalization process theory. *Sociology,* 43**,** 535-554.

MAY, C. R., JOHNSON, M. & FINCH, T. 2016. Implementation, context and complexity. *Implementation Science,* 11**,** 1-12.

MAY, C. R., MAIR, F., FINCH, T., MACFARLANE, A., DOWRICK, C., TREWEEK, S., RAPLEY, T., BALLINI, L., ONG, B. N. & ROGERS, A. 2009. Development of a

theory of implementation and integration: Normalization Process Theory. *Implementation Science,* 4**,** 29.

MAZZOCATO, P., SAVAGE, C., BROMMELS, M., ARONSSON, H. & THOR, J. 2010. Lean thinking in healthcare: a realist review of the literature. *Quality and Safety in Health Care,* 19**,** 376.

MAZZOCCO, K. & CHERUBINI, P. 2010. The effect of outcome information on health professionals' spontaneous learning. *Medical Education,* 44**,** 962-968.

MCDONALD, K. M., SUNDARAM, V., BRAVATA, D. M., LEWIS, R., LIN, N., KRAFT, S. A., MCKINNON, M., PAGUNTALAN, H. & OWENS, D. K. 2007. Conceptual frameworks and their application to evaluating care coordination interventions. *Closing the Quality Gap: A Critical Analysis of Quality Improvement Strategies (Vol. 7: Care Coordination).* Agency for Healthcare Research and Quality (US).

MCGRATH, B., WILKINSON, K. & SHAH, R. K. 2015. Notes from a small island: lessons from the UK NCEPOD Tracheotomy Report. *Otolaryngology–Head and Neck Surgery,* 153**,** 167-169.

MCKEE, M. & HEALY, J. 2002. *Hospitals in a changing Europe*, Citeseer.

MCKINLAY, J. B., LIN, T., FREUND, K. & MOSKOWITZ, M. 2002. The unexpected influence of physician attributes on clinical decisions: Results of an experiment. *Journal of Health and Social Behavior,* 43**,** 92-106.

MCKINLAY, J. B., POTTER, D. A. & FELDMAN, H. A. 1996. Non-medical influences on medical decision-making. *Social Science & Medicine,* 42**,** 769-776.

MCLAIN, D. L. 1993. The MSTAT-I: A new measure of an individual's tolerance for ambiguity. *Educational and Psychological Measurement,* 53**,** 183-189.

MEADOWS, D. H. 2008. *Thinking in systems: A primer*, chelsea green publishing.

MEHRABIAN, A. 1972. *Nonverbal communication*, Transaction Publishers.

MEHTA, R. L., KELLUM, J. A., SHAH, S. V., MOLITORIS, B. A., RONCO, C., WARNOCK, D. G. & LEVIN, A. 2007. Acute Kidney Injury Network: report of an initiative to improve outcomes in acute kidney injury. *Critical care,* 11**,** 1-8.

MENDEL, R., TRAUT-MATTAUSCH, E., JONAS, E., LEUCHT, S., KANE, J. M., MAINO, K., KISSLING, W. & HAMANN, J. 2011. Confirmation bias: Why psychiatrists stick to wrong preliminary diagnoses. *Psychological Medicine,* 41**,** 2651-2659.

MENDELSON, A., KONDO, K., DAMBERG, C., LOW, A., MOTÚAPUAKA, M., FREEMAN, M., O'NEIL, M., RELEVO, R. & KANSAGARA, D. 2017. The effects of pay-for-performance programs on health, health care use, and processes of care: a systematic review. *Annals of internal medicine,* 166**,** 341-353.

MENDU, M. L., LU, Y., PETERSEN, A., TELLEZ, M. G., BELOFF, J., FIUMARA, K. & KACHALIA, A. 2020. Reflections on implementing a hospital-wide provider-based electronic inpatient mortality review system: lessons learnt. *BMJ Quality & Safety,* 29**,** 304.

MEYER, A. N. D., PAYNE, V. L., MEEKS, D. W., RAO, R. & SINGH, H. 2013. Physicians' diagnostic accuracy, confidence, and resource requests: A vignette study. *JAMA Internal Medicine,* 173**,** 1952-1961.

MICHELL, J. 1999. *Measurement in psychology: A critical history of a methodological concept,* Cambridge University Press.

MICHELL, J. 2008. Is psychometrics pathological science? *Measurement,* 6**,** 7-24.

MICHELL, J. 2019. The fashionable scientific fraud: Collingwood's critique of psychometrics. *History of the Human Sciences***,** 0952695119872638.

MICHIE, S., VAN STRALEN, M. M. & WEST, R. 2011. The behaviour change wheel: a new method for characterising and designing behaviour change interventions. *Implementation Science,* 6**,** 42.

MILLER, R. H. & SIM, I. 2004. Physicians' use of electronic medical records: barriers and solutions. *Health affairs,* 23**,** 116-126.

MINISTRY OF HEALTH 1957. Report on Confidential Enquiries into Maternal Deaths in England and Wales, 1952–54.

MIRON, A. M. & BREHM, J. W. 2006. Reactance theory-40 years later. *Zeitschrift für Sozialpsychologie,* 37**,** 9-18.

MITCHELL, C., LAWTON, E., MORTON, C., MCCAIN, C., HOLTBY, S. & MAIN, E. 2014. California pregnancy-associated mortality review: mixed methods approach for improved case identification, cause of death analyses and translation of findings. *Maternal and child health journal,* 18**,** 518-526.

MOHAMMED, M. A., DEEKS, J. J., GIRLING, A., RUDGE, G., CARMALT, M., STEVENS, A. J. & LILFORD, R. J. 2009a. Evidence of methodological bias in hospital standardised mortality ratios: retrospective database study of English hospitals. *Bmj,* 338**,** b780.

MOHAMMED, M. A., DEEKS, J. J., GIRLING, A., RUDGE, G., CARMALT, M., STEVENS, A. J. & LILFORD, R. J. 2009b. Evidence of methodological bias in hospital standardised mortality ratios: retrospective database study of English hospitals. *BMJ (Online),* 338**,** 817-820.

MOHAMMED, M. A., MANT, J., BENTHAM, L. & RAFTERY, J. 2005. Comparing processes of stroke care in high-and low-mortality hospitals in the West Midlands, UK. *International Journal for Quality in Health Care,* 17**,** 31-36.

MOHAMMED, M. A., MANT, J., BENTHAM, L., STEVENS, A. & HUSSAIN, S. 2006. Process of care and mortality of stroke patients with and without a do not resuscitate order in the West Midlands, UK. *International Journal for Quality in Health Care,* 18**,** 102-106.

MONACIS, L., DE PALO, V., DI NUOVO, S. & SINATRA, M. 2016. Validation of the rational and experiential multimodal inventory in the Italian context. *Psychological reports,* 119**,** 242-262.

MORAROS, J., LEMSTRA, M. & NWANKWO, C. 2016. Lean interventions in healthcare: do they actually work? A systematic literature review. *International Journal for Quality in Health Care,* 28**,** 150-165.

MOREWEDGE, C. K., YOON, H., SCOPELLITI, I., SYMBORSKI, C. W., KORRIS, J. H. & KASSAM, K. S. 2015. Debiasing decisions: Improved decision making with a single training intervention. *Policy Insights from the Behavioral and Brain Sciences,* 2**,** 129-140.

MORGAN-TRIMMER, S. & WOOD, F. 2016. Ethnographic methods for process evaluations of complex health behaviour interventions. *Trials,* 17**,** 232.

MOUNTFORD, J. & SHOJANIA, K. G. 2012. Refocusing quality measurement to best support quality improvement: local ownership of quality measurement by clinicians. *BMJ quality & safety,* 21**,** 519-523.

MUMMA, G. H. & WILSON, S. B. 1995. Procedural debiasing of primacy/anchoring effects in clinical-like judgments. *Journal of clinical psychology,* 51**,** 841-853.

MURRAY, E., TREWEEK, S., POPE, C., MACFARLANE, A., BALLINI, L., DOWRICK, C., FINCH, T., KENNEDY, A., MAIR, F. & O'DONNELL, C. 2010. Normalisation process theory: a framework for developing, evaluating and implementing complex interventions. *BMC medicine,* 8**,** 63.

NATIONAL INSTITUTE FOR HEALTH AND CARE EXCELLENCE (NICE). 2007. *Acutely ill adults in hospital: recognising and responding to deterioration* [Online]. Available: https://www.nice.org.uk/guidance/cg50 [Accessed 12/08/2020].

NATIONAL QUALITY BOARD UK 2017. National Guidance on Learning from Deaths: A Framework for NHS Trusts and NHS Foundation Trusts on Identifying, Reporting, Investigating and Learning from Deaths in Care.

NCEPOD 2018. Themes and Recommendations to all hospital specialties.

NEUBERG, S. L. & NEWSOM, J. T. 1993. Personal need for structure: Individual differences in the desire for simpler structure. *Journal of Personality and Social Psychology,* 65**,** 113.

NEWTON, J., GRAHAM, J., MCLOUGHLIN, K. & MOORE, A. 2003. Receptivity to change in a general medical practice. *British journal of management,* 14**,** 143-153.

NEZLEK, J. B. 2008. An Introduction to Multilevel Modeling for Social and Personality Psychology. *Social and Personality Psychology Compass,* 2**,** 842-860.

NHS DIGITAL. 2020. *About the Summary Hospital-level Mortality Indicator (SHMI)* [Online]. Available: https://digital.nhs.uk/data-and-information/publications/ci-hub/summary-hospital-level-mortality-indicator-shmi [Accessed 18/05/2020 2020].

NHS IMPROVEMENT. 2020a. *The Natiional Medical Examiner System* [Online]. National Health Service, UK. Available: https://improvement.nhs.uk/resources/establishing-medical-examiner-system-nhs/ [Accessed 03/07/2020].

NHS IMPROVEMENT. 2020b. *What we do* [Online]. Available: https://improvement.nhs.uk/about-us/what-we-do/ [Accessed].

NICHOLL, J. 2007. Case-mix adjustment in non-randomised observational evaluations: the constant risk fallacy. *Journal of epidemiology and community health,* 61**,** 1010-1013.

NORMAN, G. 2010. Likert scales, levels of measurement and the “laws” of statistics. *Advances in health sciences education,* 15**,** 625-632.

NORMAN, G. R., MONTEIRO, S. D., SHERBINO, J., ILGEN, J. S., SCHMIDT, H. G. & MAMEDE, S. 2017. The causes of errors in clinical reasoning: cognitive biases, knowledge deficits, and dual process thinking. *Academic Medicine,* 92**,** 23-30.

NORMAND, S.-L. T., ASH, A. S., FIENBERG, S. E., STUKEL, T. A., UTTS, J. & LOUIS, T. A. 2016. League tables for hospital comparisons. *Annual Review of Statistics and Its Application,* 3**,** 21-50.

O'NEIL, W. 1944. Factors and faculties. *Australasian Journal of Psychology and Philosophy,* 22**,** 55-69.

O'REILLY, D., MCPHERSON, S., SINCLAIR, M. & SMITH, N. 2017. Lessons from a national audit of acute pancreatitis: A summary of the NCEPOD report ‘Treat the Cause’. *Pancreatology,* 17**,** 329-333.

O’HARA, R., COSTER, J. & GOODACRE, S. 2021. Qualitative exploration of the Medical Examiner role in identifying problems with the quality of patient care. *BMJ open,* 11**,** e048007.

O’REILLY, P., LEE, S. H., O’SULLIVAN, M., CULLEN, W., KENNEDY, C. & MACFARLANE, A. 2017. Assessing the facilitators and barriers of interdisciplinary team working in primary care using normalisation process theory: an integrative review. *PloS one,* 12**,** e0177026.

ORGANIZATION, W. H. 2018. Delivering quality health services: a global imperative for universal health coverage.

ORTEGA EGEA, J. M. & ROMÁN GONZÁLEZ, M. V. 2011. Explaining physicians’ acceptance of EHCR systems: An extension of TAM with trust and risk factors. *Computers in Human Behavior,* 27**,** 319-332.

OSGOOD, C. E., SUCI, G. J. & TANNENBAUM, P. H. 1964. *The measurement of meaning*, University of Illinois Press.

OUTCOME, N. C. E. I. P. & DEATH 2017. Inspiring Change. NCEPOD London.

PALUCK, E. L. 2009. Reducing intergroup prejudice and conflict using the media: a field experiment in Rwanda. *Journal of personality and social psychology,* 96**,** 574.

PARKER, A. M., DE BRUIN, W. B. & FISCHHOFF, B. 2007. Maximizers versus satisficers: Decision-making styles, competence, and outcomes. *Judgment and Decision making,* 2**,** 342.

PATTON, M. Q. 1999. Enhancing the quality and credibility of qualitative analysis. *Health services research,* 34**,** 1189.

PAWSON, R. 2006. *Evidence-based policy: a realist perspective*, sage.

PEARSON, M. L., LEE, J. L., CHANG, B. L., ELLIOTT, M., KAHN, K. L. & RUBENSTEIN, L. V. 2000. Structured implicit review: a new method for monitoring nursing care quality. *Medical care***,** 1074-1091.

PETERS, D. H. 2014. The application of systems thinking in health: why use systems thinking? *Health Research Policy and Systems,* 12**,** 1-6.

PETERSON, D. K. & PITZ, G. F. 1986. Effects of amount of information on predictions of uncertain quantities. *Acta Psychologica,* 61**,** 229-241.

PETTIGREW, A., FERLIE, E. & MCKEE, L. 1992. *Shaping Strategic Change: Making change in large organisations, the case of the NHS,* Sage Publications.

PETTIGREW, L. M. & MAYS, N. 2021. What role is there for 'nudging' clinicians? *British Journal of General Practice,* 71**,** 82.

PETTY, R. E. & CACIOPPO, J. T. 2012. *Communication and persuasion: Central and peripheral routes to attitude change*, Springer Science & Business Media.

PITCHES, D. 2014. Hospital Mortality Rates: Position Statement. *Faculty of Public Health (FPH) Bulletins*.

PLANTINGA, A. 1993. *Warrant and proper function*, Oxford University Press.

POPE, C. & MAYS, N. 2006. Qualitative research in health care.

POULTON, T., MURRAY, D. & TEAM, N. E. L. A. P. 2019. Pre-optimisation of patients undergoing emergency laparotomy: a review of best practice. *Anaesthesia,* 74**,** 100-107.

POUW, M. E., PEELEN, L. M., LINGSMA, H. F., PIETER, D., STEYERBERG, E., KALKMAN, C. J. & MOONS, K. G. 2013. Hospital standardized mortality ratio: consequences of adjusting hospital mortality with indirect standardization. *PloS one,* 8**,** e59160.

QU, S. Q. & DUMAY, J. 2011. The qualitative research interview. *Qualitative research in accounting & management*.

RAJA, S. & THOMAS, C. 2019. Standardising care of patients requiring non-invasive ventilation in response to NCEPOD–a chief registrar quality improvement project. *Clinical Medicine,* 19**,** 79.

RALEIGH, V. S. & FOOT, C. 2010. Getting the measure of quality. *Opportunities and challenges*.

RASBASH, J. 2008. Module 4: Multilevel structures and classifications.

REASON, J. 1990. *Human error*, Cambridge university press.

RENDER, M. L., FREYBERG, R. W., HASSELBECK, R., HOFER, T. P., SALES, A. E., DEDDENS, J., LEVESQUE, O. & ALMENOFF, P. L. 2011. Infrastructure for quality transformation: measurement and reporting in veterans administration intensive care units. *BMJ quality & safety,* 20**,** 498-507.

REUTLINGER, A., SCHURZ, G. & HÜTTEMANN, A. 2011a. Ceteris paribus laws.

REUTLINGER, A., SCHURZ, G. & HÜTTEMANN, A. 2011b. Ceteris paribus laws. *Stanford encyclopedia of philosophy*.

REYNA, C. & ORTIZ, M. V. 2016. Psychometric study of the Rational Experiential Inventory among undergraduate Argentinean students. *Revista de Psicología*, 34.

RHODES, A., EVANS, L. E., ALHAZZANI, W., LEVY, M. M., ANTONELLI, M., FERRER, R., KUMAR, A., SEVRANSKY, J. E., SPRUNG, C. L. & NUNNALLY, M. E. 2017. Surviving sepsis campaign: international guidelines for management of sepsis and septic shock: 2016. *Intensive care medicine,* 43**,** 304-377.

RICHARDSON, L. 2003. Writing: A method of inquiry. *Turning points in qualitative research: Tying knots in a handkerchief,* 2**,** 379.

RICHARDSON, W. 2001. Crossing the quality chasm: A New health system for the 21st century. Institute of Medicine. National Academy press.

RICHIE, M. & JOSEPHSON, S. A. 2018. Quantifying heuristic bias: anchoring, availability, and representativeness. *Teaching and learning in medicine,* 30**,** 67-75.

RITCHIE, J., LEWIS, J., NICHOLLS, C. M. & ORMSTON, R. 2013. *Qualitative research practice: A guide for social science students and researchers*, sage.

RITOV, I. & BARON, J. 1995. Outcome Knowledge, Regret, and Omission Bias. *Organizational Behavior and Human Decision Processes,* 64**,** 119-127.

ROBERT A. CAPLAN, M. K. L. P., PHD; FREDERICK W. CHENEY, MD 1991. Effects of Outcome on Physicians' Judgment of Appropriateness of Care. *JAMA - Journal of the American Medical Association,* 266**,** 794.

ROGERS, B. A. & PLEAT, J. 2010. Is there adequate information on operation notes? The application of the Royal College of Surgeons of England guidelines. *Journal of perioperative practice,* 20**,** 339-342.

ROSS, L. & ANDERSON, C. A. 1982. Shortcomings in the attribution process: On the origins and maintenance of erroneous social assessments.

ROSS, L. & NISBETT, R. E. 2011. *The person and the situation: Perspectives of social psychology*, Pinter & Martin Publishers, London.

RUBENSTEIN, L. V., KAHN, K. L., HARRISON, E., SHERWOOD, M. J., ROGERS, W. H. & BROOK, R. H. 1991. Structured Implicit Review of the Medical Record.

SABIN, J. A., RIVARA, F. P. & GREENWALD, A. G. 2008. Physician implicit attitudes and stereotypes about race and quality of medical care. *Medical Care***,** 678-685.

SADOWSKI, C. J. & GULGOZ, S. 1992. Internal consistency and test-retest reliability of the Need for Cognition Scale. *Perceptual and Motor Skills,* 74**,** 610-610.

SALAS, E. & KLEIN, G. A. 2001. *Linking expertise and naturalistic decision making*, Psychology Press.

SANDARS, J. 2009. The use of reflection in medical education: AMEE Guide No. 44. *Medical teacher,* 31**,** 685-695.

SAPOSNIK, G., REDELMEIER, D., RUFF, C. C. & TOBLER, P. N. 2016. Cognitive biases associated with medical decisions: a systematic review. *BMC Medical Informatics and Decision Making,* 16**,** 138.

SARI, A. B.-A., SHELDON, T. A., CRACKNELL, A. & TURNBULL, A. 2007. Sensitivity of routine system for reporting patient safety incidents in an NHS hospital: retrospective patient case note review. *Bmj*, 334**,** 79.

SAYER, R. A. 1992. *Method in social science: A realist approach*, Psychology Press.

SCHEIN, E. H. 1985. Organisational culture and leadership: A dynamic view. *San Francisco*.

SCHIFFER, S. 1991. Ceteris paribus laws. *Mind,* 100**,** 1-17.

SCHILLER, C. J. 2016. Critical realism in nursing: an emerging approach. *Nursing Philosophy,* 17**,** 88-102.

SCHNEIDER, A., SZECSENYI, J., BARIE, S., JOEST, K. & ROSEMANN, T. 2007. Validation and cultural adaptation of a German version of the Physicians' Reactions to Uncertainty scales. *BMC health services research,* 7**,** 81.

SCHROEDER, S. A. & KABCENELL, A. I. 1991. Do bad outcomes mean substandard care? *Jama,* 265**,** 1995-1995.

SEDGWICK, P. & GREENWOOD, N. 2015. Understanding the Hawthorne effect. *Bmj,* 351.

SELBY, N. M., HILL, R. & FLUCK, R. J. 2015. Standardizing the early identification of acute kidney injury: the NHS England national patient safety alert. *Nephron,* 131**,** 113-117.

SELLTIZ, C., WRIGHTSMAN, L. S., COOK, S. W., BALCH, G. I., HOFSTETTER, R. & BICKMAN, L. 1976. Research methods in social relations.

SENA, L. A. 1972. *Units of physical quantities and their dimensions*, Mir Pub.

SERVICES, R. O. C. & LUCE, T. 2003. *Death Certification and Investigation in England, Wales and Northern Ireland: The Report of a Fundamental Review 2003*, Stationery Office.

SHAH, A. K. & OPPENHEIMER, D. M. 2008. Heuristics made easy: An effort-reduction framework. *Psychological bulletin,* 134**,** 207.

SHAHIAN, D. M., WOLF, R. E., IEZZONI, L. I., KIRLE, L. & NORMAND, S.-L. T. 2010. Variability in the measurement of hospital-wide mortality rates. *New England Journal of Medicine,* 363**,** 2530-2539.

SHEERAN, P., GOLLWITZER, P. M. & BARGH, J. A. 2013. Nonconscious processes and health. *Health Psychology,* 32**,** 460.

SHENTON, A. K. 2004. Strategies for ensuring trustworthiness in qualitative research projects. *Education for information,* 22**,** 63-75.

SHI, Q., MACDERMID, J. C., SANTAGUIDA, P. L. & KYU, H. H. 2011. Predictors of surgical outcomes following anterior transposition of ulnar nerve for cubital tunnel syndrome: A systematic review. *Journal of Hand Surgery,* 36**,** 1996-2001.e6.

SIMBORG, D. W. 1981. DRG creep: a new hospital-acquired disease. *New England Journal of Medicine,* 304**,** 1602-1604.

SIMINOFF, L. A. & FETTING, J. H. 1989. Effects of outcome framing on treatment decisions in the real world: Impact of framing on adjuvant breast cancer decisions. *Medical Decision Making,* 9**,** 262-271.

SIMON, H. A. 1979. Rational decision making in business organizations. *The American economic review,* 69**,** 493-513.

SIMONYAN, K. & ZISSERMAN, A. 2014. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.

SIXMA, H. J., KERSSENS, J. J., CAMPEN, C. V. & PETERS, L. 1998. Quality of care from the patients' perspective: from theoretical concept to a new measuring instrument. *Health expectations,* 1**,** 82-95.

SLADE, E., TAMBER, P. S. & VINCENT, J.-L. 2003a. The Surviving Sepsis Campaign: raising awareness to reduce mortality. *Critical Care,* 7**,** 1.

SLADE, E., TAMBER, P. S. & VINCENT, J.-L. 2003b. The Surviving Sepsis Campaign: raising awareness to reduce mortality. BioMed Central.

SLIJKHUIS, J., RIETZSCHEL, E. F. & VAN YPEREN, N. W. 2013. How evaluation and need for structure affect motivation and creativity. *European Journal of Work and Organizational Psychology,* 22**,** 15-25.

SLOVIC, P., FINUCANE, M., PETERS, E. & MACGREGOR, D. G. 2002. Rational actors or rational fools: Implications of the affect heuristic for behavioral economics. *The Journal of Socio-Economics,* 31**,** 329-342.

SLOVIC, P., MONAHAN, J. & MACGREGOR, D. G. 2000. Violence risk assessment and risk communication: The effects of using actual cases, providing instruction, and employing probability versus frequency formats. *Law and human behavior,* 24**,** 271-296.

SMITH, D. J. 2004. *The Shipman Inquiry*, The Stationery Office.

SMITH, G. B., PRYTHERCH, D. R., MEREDITH, P., SCHMIDT, P. E. & FEATHERSTONE, P. I. 2013. The ability of the National Early Warning Score (NEWS) to discriminate patients at risk of early cardiac arrest, unanticipated intensive care unit admission, and death. *Resuscitation,* 84**,** 465-470.

SMITH, M., HIGGS, J. & ELLIS, E. 2008. Factors influencing clinical decision making. *Clinical reasoning in the health professions,* 3**,** 89-100.

SMITH, M. A., ATHERLY, A. J., KANE, R. L. & PACALA, J. T. 1997a. Peer review of the quality of care: reliability and sources of variability for outcome and process assessments. *Journal of American Medical Association,* 278**,** 1573-1578.

SMITH, M. A., ATHERLY, A. J., KANE, R. L. & PACALA, J. T. 1997b. Peer review of the quality of care: reliability and sources of variability for outcome and process assessments. *Jama,* 278**,** 1573-1578.

SNIJDERS, T. & BOSKER, R. 1999. Multilevel modeling: An introduction to basic and advanced multilevel modeling.

SOCIETY OF JUDGEMENT AND DECISION-MAKING. 2020. *Decision Making Individual Differences Inventory (DMIDI)* [Online]. The Society for Judgment and Decision Making. Available: http://www.sjdm.org/dmidi/ [Accessed 2018].

SOFAER, S. & FIRMINGER, K. 2005. Patient perceptions of the quality of health services. *Annu. Rev. Public Health,* 26**,** 513-559.

SPENCER, L., RITCHIE, J., LEWIS, J. & DILLON, L. 2004. Quality in qualitative evaluation: a framework for assessing research evidence.

SPILLETT, M. A. 2003. Peer debriefing: Who, what, when, why, how. *Academic Exchange Quarterly,* 7**,** 36-41.

SPRANCA, M., MINSK, E. & BARON, J. 1991. Omission and commission in judgment and choice. *Journal of Experimental Social Psychology,* 27**,** 76-105.

STAKE, R. E. 2005. Qualitative case studies.

STANFORD, K. 2009. Underdetermination of scientific theory.

STANOVICH, K. E. & WEST, R. F. 2008. On the relative independence of thinking biases and cognitive ability. *Journal of Personality and Social Psychology,* 94**,** 672-695.

STATISTICS, O. F. N. 2013. Avoidable mortality in England and Wales.

STEVENS, S. S. 1946. On the theory of scales of measurement.

STEWART, K., CHOUDRY, M. I. & BUCKINGHAM, R. 2016. Learning from hospital mortality. *Clinical Medicine,* 16**,** 530.

STEWART, M. A. 1995. Effective physician-patient communication and health outcomes: a review. *CMAJ: Canadian medical association journal,* 152**,** 1423.

STIEGLER, M. P. & RUSKIN, K. J. 2012. Decision-making and safety in anesthesiology. *Current Opinion in Anaesthesiology,* 25**,** 724-729.

SUJAN, M. A., HUANG, H. & BRAITHWAITE, J. 2017. Learning from incidents in health care: Critique from a Safety-II perspective. *Safety Science,* 99**,** 115-121.

SULLIVAN, P. A. & ANDERSON, J. M. 2010. The Health Care Debate: If Lack of Tort Reform is Part of the Problem, Federalized Protection for Peer Review Needs to be Part of the Solution. *Roger Williams UL Rev.,* 15**,** 41.

SUSERUD, B.-O. & BRUCE, K. 2003. Ambulance nursing: Part three: Bjorn-Ove Suserud and Karin Bruce examine the hand over of patients to emergency wards.(feature). *emergency nurse,* 11**,** 16-22.

SWANWICK, T. & VAUX, E. 2020. *ABC of quality improvement in healthcare*, John Wiley & Sons.

SWETS, J. A., DAWES, R. M. & MONAHAN, J. 2000. Psychological Science Can Improve Diagnostic Decisions. *Psychological Science in the Public Interest,* 1**,** 1-26.

TANAKA, Y., FUJINO, J., IDENO, T., OKUBO, S., TAKEMURA, K., MIYATA, J., KAWADA, R., FUJIMOTO, S., KUBOTA, M., SASAMOTO, A., HIROSE, K., TAKEUCHI, H., FUKUYAMA, H., MURAI, T. & TAKAHASHI, H. 2015. Are ambiguity aversion and ambiguity intolerance identical? A neuroeconomics investigation. *Frontiers in Psychology,* 5.

TAVAKOL, M. & DENNICK, R. 2011. Making sense of Cronbach's alpha. *International journal of medical education,* 2**,** 53-55.

TE, A., QUINN, L., MANASEKI-HOLLAND, S., MANNION, R., LILFORD, R. & VLAEV, I. 2017. Systematic review of the cognitive biases, heuristics and human factors influencing the review and assessment of hospital medical records: a narrative synthesis (PROSPERO 2017 CRD42017060142) [UNPUBLISHED].

TE. A, SCHMIDTKHE, K., MANASEKI-HOLLAND, S., LILFORD, R., BION, J., VLAEV, I. & HOFER, T. 2021. The influence of physician reviewer characteristics and attitudes on care quality judgements during case note reviews: a multilevel analysis. *Health services research*.

THALER, R. H. & SUNSTEIN, C. R. 2008. *Nudge: Improving decisions about health, wealth, and happiness*, Yale University Press.

THE HEALTH AND SOCIAL CARE INFORMATION CENTRE 2013. Standards for the clinical structure and content of patient records.

THE ROYAL COLLEGE OF ANAESTHETISTS. 2020. *Risk explained* [Online]. [Accessed 19/02/2021].

THE ROYAL COLLEGE OF OBSTETRICIANS AND GYNAECOLOGISTS. 2015. *Understanding how risk is discussed in healthcare* [Online]. Available: https://www.rcog.org.uk/en/patients/patient-leaflets/understanding-how-risk-is-discussed-in-healthcare/ [Accessed].

THE ROYAL COLLEGE OF PATHOLOGISTS 2016. Medical examiners. The Royal College of Pathologists.

THE ROYAL COLLEGE OF PHYSICIANS. 2016a. *National Mortality Case Record Review (NMCRR) programme resources* [Online]. Available: https://www.rcplondon.ac.uk/projects/outputs/national-mortality-case-record-review-nmcrr-programme-resources [Accessed 17/11/2016 2016].

THE ROYAL COLLEGE OF PHYSICIANS 2016b. National Mortality Case Record Review Programme (NMCRR): Using the structured judgement review method - A clinical governance guide to mortality case record reviews.

THE ROYAL COLLEGE OF PHYSICIANS 2016c. National Mortality Case Record Review Programme (NMCRR): Using the structured judgement review method - A guide for reviewers (England)

.

THE ROYAL COLLEGE OF PHYSICIANS. 2016d. *RCP hosts pioneering national Mortality Case Record Review Programme* [Online]. Available: https://www.rcplondon.ac.uk/news/rcp-hosts-pioneering-national-mortality-case-record-review-programme [Accessed 29/04/2016 2016].

THE ROYAL COLLEGE OF PHYSICIANS 2018. A review of the learning from the national mortality review programmes.

THE WELSH GOVERNMENT 2014. Publication of Professor Stephen Palmer's review of the use of risk adjusted mortality data in NHS Wales.

THOMAS, E. & MAGILVY, J. K. 2011. Qualitative rigor or research validity in qualitative research. *Journal for specialists in pediatric nursing*.

THOMAS, E. J., LIPSITZ, S. R., STUDDERT, D. M. & BRENNAN, T. A. 2002. The reliability of medical record review for estimating adverse event rates. *Annals of internal medicine,* 136**,** 812-816.

THOMAS, E. J. & PETERSEN, L. A. 2003. Measuring errors and adverse events in health care. *Journal of general internal medicine,* 18**,** 61-67.

THOMAS, E. J., STUDDERT, D. M., BURSTIN, H. R., ORAV, E. J., ZEENA, T., WILLIAMS, E. J., HOWARD, K. M., WEILER, P. C. & BRENNAN, T. A. 2000. Incidence and types of adverse events and negligent care in Utah and Colorado. *Medical care***,** 261-271.

THOMAS, J. 2009. Medical records and issues in negligence. *Indian journal of urology: IJU: journal of the Urological Society of India,* 25**,** 384.

THOMAS, M. E., BLAINE, C., DAWNAY, A., DEVONALD, M. A. J., FTOUH, S., LAING, C., LATCHEM, S., LEWINGTON, A., MILFORD, D. V. & OSTERMANN, M. 2015. The definition of acute kidney injury and its use in practice. *Kidney International,* 87**,** 62-73.

THOMPSON, M. M., NACCARATO, M. E., PARKER, K. C. & MOSKOWITZ, G. B. The personal need for structure and personal fear of invalidity measures: Historical perspectives, current applications, and future directions. Cognitive social psychology: The Princeton symposium on the legacy and future of social cognition, 2001. 19-39.

TIWARI, D., DONNELL, A., RENAUT, R., RICHARDSON, T. & ALLEN, S. 2020. Reducing hospital mortality: Incremental change informed by structured mortality review is effective. *Future Healthcare Journal,* 7**,** 143.

TOMOIAGA, C., FENG, P., SALZMANN, M. & JAYET, P. Field typing for improved recognition on heterogeneous handwritten forms. 2019 International Conference on Document Analysis and Recognition (ICDAR), 2019. IEEE, 487-493.

TOPOL, E. 2019. *Deep medicine: how artificial intelligence can make healthcare human again*, Hachette UK.

TROCHIM, W. 2016. *Levels of Measurement 2016* [Online]. Available: https://socialresearchmethods.net/kb/measlevl.php [Accessed].

TRUST, S. W. N. A. C. 2017. Mortality Review Policy.

TURNER, P., WONG, M. C. & YEE, K. C. 2006. Understanding interactions of factors influencing clinical handover: insights for information technology. *HIC 2006 and HINZ 2006: Proceedings***,** 328.

TVERSKY, A. & GRIFFIN, D. 2000. Endowments and Contrast in Judgments of Well-Being. *Choices, Values, and Frames***,** 709-725.

TVERSKY, A. & KAHNEMAN, D. 1971. Belief in the law of small numbers. *Psychological Bulletin,* 76**,** 105-110.

TVERSKY, A. & KAHNEMAN, D. 1974. Judgment under uncertainty: Heuristics and biases. *Science,* 185**,** 1124-1131.

TVERSKY, A. & KAHNEMAN, D. 1975. Judgment under uncertainty: Heuristics and biases. *Utility, Probability, and Human Decision Making.* Springer.

TVERSKY, A. & KAHNEMAN, D. 1981. Judgments of and by Representativeness. STANFORD UNIV CA DEPT OF PSYCHOLOGY.

UK RENAL REGISTRY (UKRR). 2014. *The Acute Kidney Injury (AKI) Programme* [Online]. Available: https://www.renalreg.org/projects/the-acute-kidney-injury-aki-programme/ [Accessed].

VAN CAMPEN, C., SIXMA, H., FRIELE, R. D., KERSSENS, J. J. & PETERS, L. 1995. Quality of care and patient satisfaction: a review of measuring instruments. *Medical Care Research and Review,* 52**,** 109-133.

VAN DEN NOORTGATE, W., OPDENAKKER, M.-C. & ONGHENA, P. 2005. The effects of ignoring a level in multilevel analysis. *School Effectiveness and School Improvement,* 16**,** 281-303.

VAN GESTEL, Y. R. B. M., LEMMENS, V. E. P. P., LINGSMA, H. F., DE HINGH, I. H. J. T., RUTTEN, H. J. T. & COEBERGH, J. W. W. 2012. The Hospital Standardized Mortality Ratio Fallacy: A Narrative Review. *Medical Care,* 50**,** 662-667.

VINCENT, C., NEALE, G. & WOLOSHYNOWYCH, M. 2001. Adverse events in British hospitals: Preliminary retrospective record review. *British Medical Journal,* 322**,** 517-519.

VLAEV, I. & DOLAN, P. 2015. Action change theory: A reinforcement learning perspective on behavior change. *Review of General Psychology,* 19**,** 69-95.

VLAEV, I., KING, D., DOLAN, P. & DARZI, A. 2016. The theory and practice of "nudging": changing health behaviors. *Public Administration Review,* 76**,** 550-561.

VORWERK, C., LORYMAN, B., COATS, T., STEPHENSON, J., GRAY, L., REDDY, G., FLORENCE, L. & BUTLER, N. 2009. Prediction of mortality in adult emergency department patients with sepsis. *Emergency Medicine Journal,* 26**,** 254-258.

WALLSTEN, T. S. 1981. Physician and medical student bias in evaluating diagnostic information. *Medical decision making,* 1**,** 145-164.

WALSHE, K. 2007. Understanding what works—and why—in quality improvement: the need for theory-driven evaluation. *International Journal for Quality in Health Care,* 19**,** 57-59.

WALSHE, K., BENNETT, J. & INGRAM, D. 1995. Using adverse events in health-care quality improvement. *International Journal of Health Care Quality Assurance,* 8**,** 7-14.

WALSHE, K. & OFFEN, N. 2001. A very public failure: lessons for quality improvement in healthcare organisations from the Bristol Royal Infirmary. *Quality in Health Care,* 10**,** 250.

WATSON, S. I., CHEN, Y. F., BION, J. F., ALDRIDGE, C. P., GIRLING, A., LILFORD, R. J. & THE HI, S. C. 2018. Protocol for the health economic evaluation of increasing the weekend specialist to patient ratio in hospitals in England. *BMJ Open,* 8.

WEBB, T. L. & SHEERAN, P. 2006. Does changing behavioral intentions engender behavior change? A meta-analysis of the experimental evidence. *Psychological bulletin,* 132**,** 249.

WEBER, E. U. 2010. Risk attitude and preference. *Wiley Interdisciplinary Reviews: Cognitive Science,* 1**,** 79-88.

WEBER, E. U., BLAIS, A. R. & BETZ, N. E. 2002. A Domain-specific Risk-attitude Scale: Measuring Risk Perceptions and Risk Behaviors. *Journal of Behavioral Decision Making,* 15**,** 263-290.

WEBER, E. U. & STERN, P. C. 2011. Public Understanding of Climate Change in the United States. *American Psychologist,* 66**,** 315-328.

WEBSTER, C. 2002. *The national health service: A political history*, Oxford University Press, USA.

WELLINGTON, H. H. 1982. The nature of judicial review. *The Yale Law Journal,* 91**,** 486-520.

WENNBERG, J. E. 1987. The paradox of appropriate care. *Journal of American Medical Association,* 258**,** 2568-2569.

WIKIPEDIA. 2020. *List of Cognitive Biases* [Online]. Available: https://en.wikipedia.org/wiki/List_of_cognitive_biases [Accessed 26/08/2020].

WILLIAMSON, E. J., AITKEN, Z., LAWRIE, J., DHARMAGE, S. C., BURGESS, J. A. & FORBES, A. B. 2014. Introduction to causal diagrams for confounder selection. *Respirology,* 19**,** 303-311.

WILLSON, A., HOPE, D. & SMITHIES, M. 2008. Welsh critical care improvement programme: the 'how to guide'for improving critical care rapid response to acute illness.

WILSON, J. VII—Internal and external validity in thought experiments. Proceedings of the Aristotelian Society, 2016. Oxford University Press, 127-152.

WILSON, K. & KHANSA, L. 2018. Migrating to electronic health record systems: a comparative study between the United States and the United Kingdom. *Health Policy,* 122**,** 1232-1239.

WILSON MCL, R., RUNCIMAN, W. B., GIBBERD, R. W., HARRISON, B. T., NEWBY, L. & HAMILTON, J. D. 1995. The quality in Australian health care study. *Medical Journal of Australia,* 163**,** 458-471.

WILSON, P. F. 1993. *Root cause analysis: A tool for total quality management,* Quality Press.

WILSON, R. M., RUNCIMAN, W. B., GIBBERD, R. W., HARRISON, B. T., NEWBY, L. & HAMILTON, J. D. 1995. The quality in Australian health care study. *Medical Journal of Australia,* 163**,** 458-471.

WOLFE, A. 2001. Institute of Medicine report: crossing the quality chasm: a new health care system for the 21st century. *Policy, Politics, & Nursing Practice,* 2**,** 233-235.

WOLFSON, A. M., DOCTOR, J. N. & BURNS, S. P. 2000. Clinician judgments of functional outcomes: How bias and perceived accuracy affect rating. *Archives of Physical Medicine and Rehabilitation,* 81**,** 1567-1574.

WOOD, W. 2000. Attitude change: Persuasion and social influence. *Annual review of psychology,* 51.

WRIGHT, J., DUGDALE, B., HAMMOND, I., JARMAN, B., NEARY, M., NEWTON, D., PATTERSON, C., RUSSON, L., STANLEY, P. & STEPHENS, R. 2006. Learning from death: a hospital mortality reduction programme. *Journal of the Royal Society of Medicine,* 99**,** 303-308.

WUNSCH, H., ANGUS, D. C., HARRISON, D. A., COLLANGE, O., FOWLER, R., HOSTE, E. A., DE KEIZER, N. F., KERSTEN, A., LINDE-ZWIRBLE, W. T. & SANDIUMENGE, A. 2008. Variation in critical care services across North America and Western Europe. *Critical care medicine,* 36**,** 2787-e8.

YEE, L. M., LIU, L. Y. & GROBMAN, W. A. 2014. The relationship between obstetricians' cognitive and affective traits and their patients' delivery outcomes. *American journal of obstetrics and gynecology,* 211**,** 692. e1-692. e6.

YIN, R. 2010. Analytic generalization. *Encyclopedia of case study research,* 1**,** 20-22.

YIN, R. K. 1981. The case study as a serious research strategy. *Knowledge,* 3**,** 97-114.

YIN, R. K. 2017. *Case study research and applications: Design and methods,* Sage publications.

ZEGERS, M., DE BRUIJNE, M. C., SPREEUWENBERG, P., WAGNER, C., VAN DER WAL, G. & GROENEWEGEN, P. P. 2011. Variation in the rates of adverse events between hospitals and hospital departments. *Int J Qual Health Care,* 23**,** 126-33.

ZIMMERMANN, C., BALDO, C. & MOLINO, A. 2000. Framing of outcome and probability of recurrence: Breast cancer patients' choice of adjuvant chemotherapy (ACT) in hypothetical patient scenarios. *Breast Cancer Research and Treatment,* 60**,** 9-14.

ZWAAN, L., DE BRUIJNE, M., WAGNER, C., THIJS, A., SMITS, M., VAN DER WAL, G. & TIMMERMANS, D. R. 2010. Patient record review of the incidence, consequences, and causes of diagnostic adverse events. *Arch Intern Med,* 170**,** 1015-21.

Appendices

Appendix 1a: PROSPERO form

Citation

An Te, Lauren Quinn, Semira Manaseki-Holland, Russell Mannion, Richard Lilford, Ivo Vlaev. Systematic review of the cognitive biases, heuristics and human factors influencing the review and assessment of hospital medical records: a narrative synthesis. PROSPERO 2017 CRD42017060142 Available from: https://www.crd.york.ac.uk/prospero/display_record.php?ID=CRD42017060142

Review question

Primary: Amongst physicians and nurses, which cognitive biases, heuristics and human factors influence the reviewing and assessment of hospital medical records?
Secondary: How are these biases, heuristics and human factors influencing the reviewing and interpretation of hospital medical records?

Searches

The search databases are:
• MEDLINE, MEDLINE in process (Ovid)
• EMBASE (OvidSP)
• PubMed
• Health Management Information Consortium (HMIC)
• Social Policy and Practise (SPP)
• PsycINFO
• JSTOR
Additional details concerning the search strategy can be found in the attached PDF.

Search strategy

https://www.crd.york.ac.uk/PROSPEROFILES/60142_STRATEGY_20170230.pdf

Types of study to be included

All study design types reporting primary data are eligible for inclusion.

Condition or domain being studied

Cognitive biases, heuristics and human factors influencing hospital medical records across all medical conditions.

Participants/population

Inclusion Criteria:
Clinicians (physicians or nurses) that have either used, handled and assessed the medical record. Primarily, these will be focussed on hospital doctors and nurses, however, if a preponderance of medical records are found from other clinical and non-clinical groups, we shall carefully inspect their relevance to obtain a bearing of their influence upon the review and assessment of hospital medical records.
Exclusion Criteria:
Papers which do not allude and/or are not situated in the hospital setting.

Intervention(s), exposure(s)

The interventions and exposures are not applicable in this atypical systematic review. The interventions and exposures are likely to vary considerably due to the complex nature of identifying, eliciting and measuring cognitive biases, heuristics and human factors influencing the assessment and review of hospital medical records.

Comparator(s)/control

Not applicable.

Context

Inclusion:
The hospital medical record is discussed at length from the angle of cognitive biases, heuristics and human

factors which can influence its content and information across all populations.
Exclusion:
Where there is no mention (or absence of an inferred allusion to 'hospital medical records' and its compilation and/or derivation for its review or assessment in this particular location) OR no consideration of the cognitive biases, heuristics or human factors (including synonymous, analogous or strongly contiguous) in the article/study.

Main outcome(s)

The objective measurement and/or perceived significance of the influence of cognitive biases, heuristics and human factors on the review and assessment of hospital medical records.

Measures of effect

The timing varies considerably as the accessing, reviewing and assessment of hospital medical records do not occur at discrete points in time. They, rather, are accessed and utilised continuously.
Effect measures of the influence of these factors on the review and assessment of hospital medical records with potential consequences within and without each healthcare system.

Additional outcome(s)

The science and/or purported mechanistic processes of the influence of cognitive biases, heuristics and human factors on the review and assessment of hospital medical records.

Measures of effect

Timing measures are essential in grounding the mechanistic processes for these factors; concrete examples from the hospital setting shall be discussed in outlining the case for each proposed mechanism.
The effect measures for these factors correspond closely with the timing measures and these are likely to be covered simultaneously in concrete examples we come across from our searches and anecdotally, as and when required for explanatory illustrative purposes.

Data extraction (selection and coding)

Search results (minus duplicates) will be retrieved for relevancy screening and downloaded to a bibliographic software programme (EndNote X8) which will be available to the University of Birmingham reviewers.
Phase 1: Using the inclusion criteria, the titles and abstracts of relevant references will be pooled by two independent researchers to exclude any that fail to meet the inclusion criteria. In instances of disagreement, the papers will be taken through to the next stage.
Phase 2: Copies of the full text of papers that meet the criteria in Phase 1 will be obtained and assessed by two independent researchers to exclude the papers that affirmatively do not meet the inclusion criteria. Independent hand-searching of full-text reference lists is undertaken to retrieve any further articles. Any disagreements at this stage will be resolved by discussion either by the two researchers or in consultation with another team member. A table of excluded studies will be produced detailing the reasons for any exclusions.
Phase 3: A structured data abstraction form will guide the extraction of information around: (i) key study characteristics (including bibliographic details, setting/country, intervention type/characteristics (if relevant); (ii) type of methodology and reporting; and (iii) aim of study and summary of findings/conclusions, (iv) cognitive biases and heuristics identified or discussed, (v) human factors identified or discussed.
Data extraction will be performed by the principal reviewer researcher and checked by the second reviewer. Where publications lack key details required for quality assessment or for a full data extraction, the relevant review team member(s) will be contacted in order to obtain the further necessary information.

Risk of bias (quality) assessment

The methodological quality of studies will be assessed independently by two reviewers using the following tools:
Firstly, for the non-randomised quantitative studies, the component-based tool developed by the Effective Public Health Practice Project, Canada, will be used, which evidence suggests possesses a relatively high degree of inter-rater reliability in comparison with alternative tools.
Secondly, any qualitative research studies will be quality assessed using the Critical Review Form for Qualitative Studies 2.0 originally developed by the McMaster University Occupational Therapy Evidence-Based Practice Research Group and revised by Letts et al., 2007. The inter-rater reliability of this tool is superior to another tool due to its more focussed data extraction form and it has been chosen to reduce any

unnecessary heterogeneity in our review. (http://www.ephpp.ca/tools.html)
Thirdly, where none of these tools are appropriate, we shall undertake provide a detailed narrative synthesis of the relevant studies. We shall of each of the numerals as indicated in Data Extraction: Phase 3 tabulating and synthesising this information for a narrative discussion of the barriers, levers, limitations and implications from each of studies.

Strategy for data synthesis

As we foresee the study designs to be highly heterogeneous, a narrative synthesis methodology will be used to pool all the data. An interpretive integration of information from quantitative and qualitative studies shall be undertaken. The synthesis will involve the three following steps:
1. First, the data from the quantitative studies will be summarised, synthesised and presented narratively. Statistically significant findings will be grouped together on a thematic basis, and placed into one of three categories: positive (facilitating); negative (barrier); or Unknown relationship
2. Secondly, the data from eligible papers comprising an identifiable qualitative element (including mixed methods papers) will be entered verbatim into NVivo software designed for qualitative data analysis. A line-by-line coding of the findings of primary studies will be conducted. These codes will be linked and organised into related areas, to allow the construction of relationships between these areas.
3. Finally, synthesis three will explore relationships within and between studies, to formulate a new interpretation in which findings from all eligible studies are integrated into a typology of influencing factors and those factors which are unaware of such a contribution. Typologies shall be provided for both influencing and ‘non-influencing’ factors. A conceptual map, derived from idea webbing, shall be produced from the identified relationships.

Analysis of subgroups or subsets

None planned.

Contact details for further information

Mr An Te

Organisational affiliation of the review

University of Birmingham
http://www.birmingham.ac.uk/research/activity/applied-health/index.aspx

Review team members and their organisational affiliations

Mr An Te. University of Birmingham
Ms Lauren Quinn. University of Birmingham
Dr Semira Manaseki-Holland. University of Birmingham
Professor Russell Mannion. University of Birmingham
Professor Richard Lilford. Warwick University
Professor Ivo Vlaev. Warwick University

Type and method of review

Systematic review

Anticipated or actual start date

31 March 2017

Anticipated completion date

20 February 2018

Funding sources/sponsors

None.

Conflicts of interest

None known

Language

English

Country
England

Stage of review
Review Ongoing

Subject index terms status
Subject indexing assigned by CRD

Subject index terms
Bias (Epidemiology); Cognition; Heuristics; Hospital Records; Hospitals; Humans; Medical Records

Date of registration in PROSPERO
31 March 2017

Date of first submission
11 January 2018

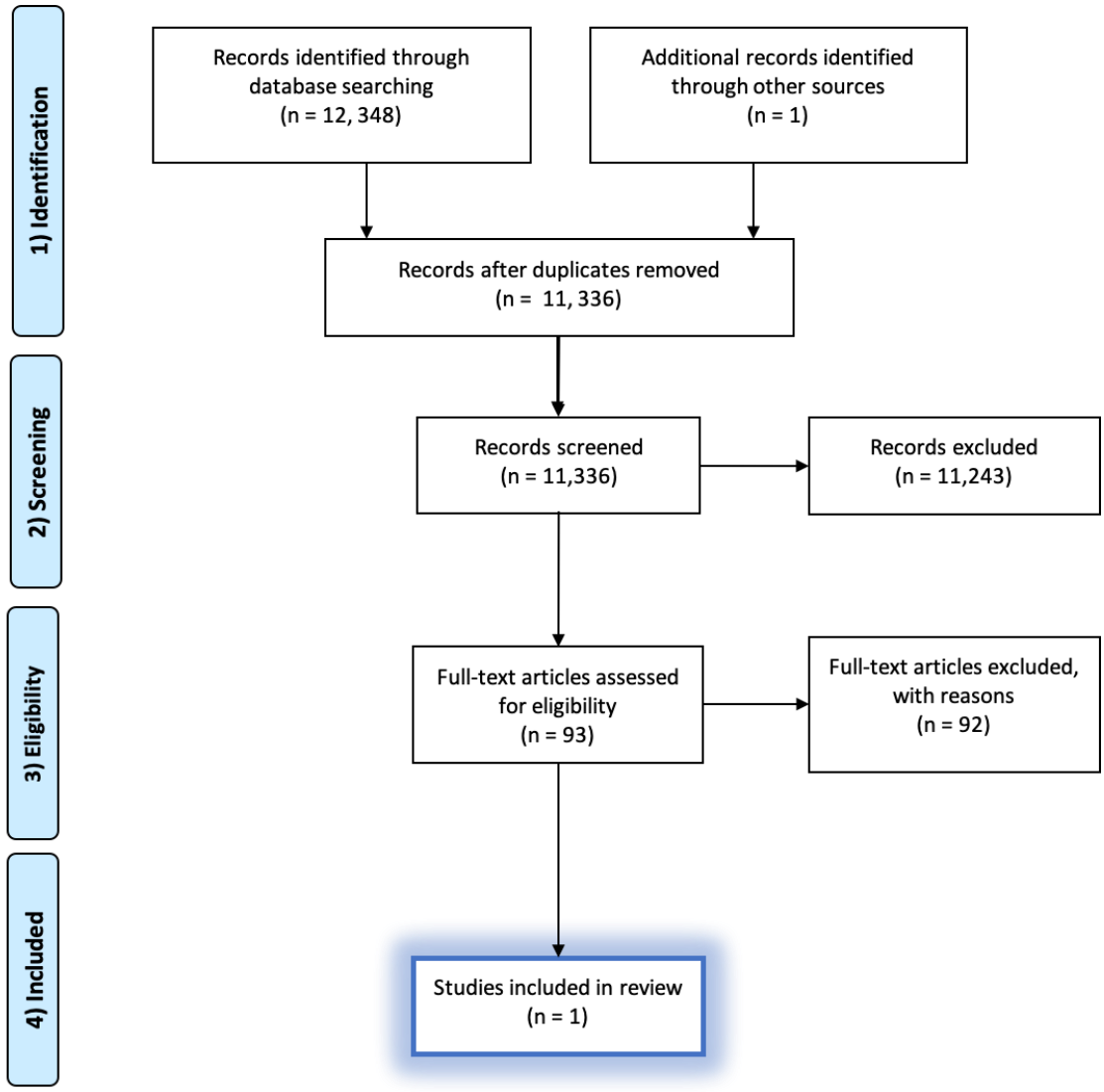
Stage of review at time of this submission

Stage	Started	Completed
Preliminary searches	Yes	Yes
Piloting of the study selection process	Yes	Yes
Formal screening of search results against eligibility criteria	Yes	Yes
Data extraction	Yes	Yes
Risk of bias (quality) assessment	Yes	Yes
Data analysis	No	No

The record owner confirms that the information they have supplied for this submission is accurate and complete and they understand that deliberate provision of inaccurate information or omission of data may be construed as scientific misconduct.
The record owner confirms that they will update the status of the review when it is completed and will add publication details in due course.

Versions
31 March 2017
18 January 2018

Appendix 1b: PRISMA Flow Diagram



Appendix *2*: Consensus Panel Bias Likert form template

Using the Likert scale, please indicate your opinion of the extent to which each heuristic/bias is likely to influence a reviewer's global care quality judgement on the quality of care received by a deceased patient during their admission.[68]

Some of the heuristics/biases have two very different situations in which they can occur. In that case, please score each type of situation/scenario separately.

Score 1 is "very unlikely", 2 "unlikely", 3 "neither likely nor unlikely", 4 "likely" and 5 "very likely."

Cognitive bias or heuristic	**Definition applied to case note reviews**	**Clinical Scenarios**	**Likert scale (1-5)**
			1 2 3 4 5

68

Affect Heuristic	Reviewer care quality judgements may be influenced by incidental (case review unrelated) and integral (case review related) affective states	Clinicians in a state of negative affect can consider information more critically and with higher accuracy or other emotions that may modulate care quality judgements. For example, a heavily sleep deprived reviewer, may have her care quality rating more heavily influenced by affect from the case which may stimulate a higher (or lower) level of scrutiny above that of a routine case review. The care quality ratings may consequently be lower (or higher).[69]					
Ambiguity Aversion	Reviewer care quality judgements may be influenced by how the reviewer perceives therapeutic risks for evidence treatments compared to less or non-evidenced therapeutics	Warfarin and Rivoroxaban are comparably effective anti-coagulants.(Fox et al., 2011, Cohen et al., 2016, Amin et al., 2017) However, the latter has a smaller evidence base. All things being equal, the medical					

[69] For the demonstrative purposes of these scenarios, it is assumed higher ratings equate to better care, lower ratings poorer care. (i.e., a 5-point scale, 1 = very poor, 5 = very good)

		team are accustomed to, and prefer using, warfarin. In the absence of clear evidence, this judgement has been made because of ambiguity over the effectiveness of Rivoroxaban. The reviewer is also inclined to prefer warfarin because there is far more evidence for warfarin's effectiveness than Rivoroxaban. Thus, ambiguity aversion leads to higher (better) care quality judgements when warfarin is used compared to newer anti-coagulating agents.					
Ambiguity Intolerance	Reviewer care quality judgements may be influenced by how the reviewer perceives the completeness of the information presented in the case-notes	A reviewer examines a near complete set of case-notes and grades the care quality received for this case as exemplary. A different reviewer grades the same case-notes as merely 'good' because of ambiguity aversion caused by the absence of					

		information. The first reviewer is tolerant to missing information and gives a higher rating whilst the second reviewer is intolerant to missing information and gives a lower rating.					
Anchoring Bias	Reviewer care quality judgements may be influenced by the past numerical value (i.e., through exposure to past numbers/figures which are quasi-fixed procedural standards)	During clinic work, the reviewer gynaecologist routinely reviews blood test results. She then undertakes a retrospective case-note review of a male patient whose haemoglobin is 125 grams per litre (normal range for males = 138-172 g/L). She does not consider this a concern as she has been 'anchored' to the normal female haemoglobin range of 121-151 g/L in their most recent exposures to clinical practice.					
Availability Bias	Reviewer care quality judgements may be influenced by personal clinical experiences	A retrospective case-note reviewer who has been an expert witness in a coroner's court case because of a					

	which then shape their perceived frequency of future events	patient who died from cardiac arrest secondary to hyperkalaemia, will likely be more attentive to serum potassium levels, throughout a retrospective case-note review process. He may account harm to hyperkalaemia, even though there are other erroneous causes of serum hyperkalaemia e.g., clotted blood sample.					
Bandwagon Effect	Reviewer care quality judgements may be influenced by the widespread, popular adoption of a medical procedure or treatment	A reviewer observes that an acute kidney injury (AKI) treatment protocol is not used. In the reviewer's hospital, the AKI protocol has gained little traction. The reviewer overlooks this omission of the AKI protocol and gives a higher care quality judgement score compared to another reviewer from a hospital which has fully adopted the protocol to his routine practice.					

Commission Bias	Reviewer care quality judgements weight more heavily active treatment over no treatment or *'watch & wait*.'	Immediate antimicrobial treatment is not recommended for asymptomatic pyuria. However, the retrospective case-note reviewer has a general preference for active intervention rather than 'watch & wait' strategies, and therefore gives a higher quality rating to cases in which an antibiotic has been prescribed.					
Confirmation Bias	Reviewers' care quality judgements may be influenced by their tendency to find more support for their initial beliefs and to overlook information that opposes their initial beliefs	A retrospective case-note reviewer is reviewing a case in which the lead consultant for the patient was a close colleague and one whom the retrospective case-note reviewer respected highly. He therefore judges the quality of care as good because he considers this consultant very unlikely to make mistakes.					

Loss/gain framing bias	Reviewer's care quality judgements may be influenced by concerns that their low-quality rating may have adverse personal consequences	The reviewer refrains from escalating his findings from a retrospective case-note review because he fears that doing so will involve additional work, such as the need to undertake a root-cause analysis. He therefore reports no issues with the case and provide a higher care quality judgement than can be justified by the care quality content of this case-note.					
Omission Bias	The converse of commission bias, where reviewer care quality judgements weigh more heavily no action or *'watch & wait'* over action	An elderly woman who is at high risk of surgery, had few viable options available than to undergo a salvage procedure for recurrence of cancer. The retrospective case-note reviewer gives a low rating for care quality judgement because she feels the surgery was inappropriate. Her preference would have been to watch and wait.					

Order Effects: Primacy/recency bias	Reviewers' care quality judgements may be influenced by where different types of information are found in case-notes i.e., at the beginning, middle or end.	A reviewer identifies one error both at the very beginning and very end of a case-notes whilst the almost all parts with the central portion indicating good care content. These errors are more prominent in the reviewer's mind. He ascribes a lower care quality judgement than if the errors had been confined to the central portions (main body) of the record.					
Optimism Bias	Reviewer care quality judgements tend to favour positive judgements of care quality for specialties or areas with which they are unfamiliar.	A renal physician reviews the case-note of a cardiology patient for its care quality. She rates an omission in care more favourably than a cardiologist reviewer familiar with the correct treatment pathways for this patient.					
Outcome Bias	Reviewer care quality judgements may be influenced by the case-note's presentation to the	Having noticed the poor outcome of the case-note, the reviewer is primed to judge the processes of care					

	reviewer concerning therapeutic, diagnostic outcomes, or overall health outcome (e.g., death, adverse outcome, successful procedure, legal case, resolution with relatives	more harshly. She thus incorrectly concludes that any minor deviations from recommended care were responsible for the outcome.					
Representativeness heuristic	Reviewer care quality judgements may be influenced by the resemblance, in the reviewer's mind, between their current retrospective case-note review and previously reviewed case-notes, personal clinical experience or clinical knowledge (anecdotal/evidence-based)	A reviewer spends many hours trying to evaluate the case-note of a patient with a rare respiratory condition with which the reviewer is unfamiliar. She consequently feels compelled to find faults in care to justify her lack of knowledge concerning this case-note's specialty area and rates the care quality lower than a reviewer expert in this condition might have done.					
Sunk-cost effect	Reviewer's care quality judgements may be influenced by the amount of money, effort, or	A gastroenterologist reviews a case-note of a patient treated for a rare respiratory condition. The gastroenterologist spends 20+ hours reviewing this					

	time which the reviewer invested in a retrospective case-note review	case, which is in part attributable a lack of familiarity with the clinical area. She consequently feels compelled to find faults in care to justify the time spent, and rates care quality lower than a competent respiratory physician reviewer might have done.						

Appendix 3: List of human factor and ergonomic studies from the systematic literature search

Under human factors and ergonomics, 125 diverse studies were identified in the final systematic literature search. An n=125 is far too large to extract. Furthermore, after discussion, our human factor criteria were narrower given we did not search all the human databases which could have been pertinent. For instance, a complete search would have considered all studies which were pertinent to medical decision-making; there are very likely pertinent studies which are not found in our search engines/sources focussing on medical decision-making.

Appendix 4. Reasons for full-text exclusion

Article References	**Reasons for Exclusion**
ARTS J, ASH J, BERG M. Extending the understanding of computerized physician order entry: Implications for professional collaboration, workflow, and quality of care. International Journal of Medical Informatics. 2007;76(SUPPL. 1):4-13.	Concerned with the impact of CPOE upon work elements and quality of care but not upon the judgements of the medical record
ARIZA F, KALRA D, POTTS HW. How Do Clinical Information Systems Affect the Cognitive Demands of General Practitioners?: Usability Study with a Focus on Cognitive Workload. Journal of innovation in health informatics. 2015;22(4):379-90.	Concerned with the GP setting. Hospitals systems differ from one another; re-adapt to the system.
Asan O, Tyszka JT, Fletcher K. Patient-centered use of EHRS: Capturing residents' perceptions. Journal of General Internal Medicine. 2016;1):S328-S9.	Viewpoint of patients, not physicians or nurses.
BEN-ASSAIL O, SHABTAI I, LASHON M, HILL S. EHR in emergency rooms: Exploring the effect of key information components on main complaints. Journal of Medical Systems. 2014;38(4).	EMR assignation does not rule out information obtained through other sources

BEN-ASSAIL O, SAGE D, LASHON M, IRONY A, ZIV A. Improving diagnostic accuracy using EHR in emergency departments: A simulation-based study. Journal of Biomedical Informatics. 2015;55:31-40.	Concerned with perceptions and diagnostic accuracy; medical decision-making; retrospective case-note review is a review of all available notes. No extra aid is needed.
BENNETT P, LOWE R. Emotions, and their cognitive precursors: responses to spontaneously identified stressful events among hospital nurses. J Health Psychol. 2008;13(4):537-46.	Not concerned with retrospective case-note review or the review of hospital medical records.
BIRKE LAND S, CHRISTENSEN R, DAMSON N, KASTRUP J. Process-related factors associated with disciplinary board decisions. BMC Health SERV Res. 2013;13:9.	Not concerned with retrospective case-note review or the review of hospital medical records.
BRENNAN TA, LOCALE RJ, LAIRD NL. Reliability and validity of judgements concerning adverse events suffered by hospitalized patients. Med Care. 1989;27(12):1148-58.	No mention of cognitive factors, Cognitive biases, and human factors in the discussion of reliability of the two-phase medical record review.
BROOK RH, APPEL FA. Quality-of-Care Assessment: Choosing a Method for Peer Review. New England Journal of Medicine. 1973;288(25):1323-9.	Full text not available (online access after 1980)

Crayon P, Wetter NECK TB, CARTMILL R, BLOCKY MA, BROWN R, KIM R, ET AL. Characterizing the Complexity of Medication Safety using a Human Factors Approach: An Observational Study in Two Intensive Care Units. BMJ Qual Safe. 2014;23(1):56-65.	No direct impact reported on the judgements during retrospective case-note review.
CARRINGTON JM, GEPHARDT SM, VERANO JA, FINLEY BA. Development of an instrument to measure the unintended consequences of EHRs. Western Journal of Nursing Research. 2015;37(7):842-58.	Concerned with Electronic Health Records and not the judgements of case record review.
CHASE SK. Charting critical thinking: nursing judgements and patient outcomes. DC. 1997;16(2):102-11.	Considers the link between critical thinking and the quality of the medical record.
CHRISTOFIS'S MJ, HILL A, HARPSWELL MS, WATSON MO. Observation chart design features affect the detection of patient deterioration: A systematic experimental evaluation. Journal of Advanced Nursing. 2016;72(1):158-72.	No discussion of the influence on judgements
COHEN T, BLATTER B, PATEL V. Simulating Expert Clinical Comprehension: Adapting Latent Semantic Analysis to Accurately Extract Clinical Concepts from Psychiatric Narrative. J Biomed Inform. 2008;41(6):1070-87.	No direct relevance to factors which influence judgement of retrospective case-note reviews.

COOMBS M. Power and conflict in intensive care clinical decision-making. 2003.	No direct relevance to the factors influencing medical record review judgement
DOLOUR JL, ZUCKERMAN G, WARNER K. Learners' decisions for attending Paediatric Grand Rounds: a qualitative and quantitative study. BMC Med Educ. 2006;6:26.	No direct relevance to medical record review judgement
DUBOIS JM, CARROLL K, GIBB T, KRAUS E, RUBBERLIKE T, VASTER M, ET AL. Environmental Factors Contributing to Wrongdoing in Medicine: A Criterion-Based Review of Studies and Cases. Ethics Behave. 2012;22(3):163-88.	Review paper with no direct relevance to case record review judgements
EARLY WK, TRAN M, DILLON RC, KARPINSKI E. Impact of hindsight bias on interpretation of nonenhanced computed tomographic head scans for acute stroke. Journal of Computer Assisted Tomography. 2010;34(2):229-32.	The MRI is unrealistic. The paper considers hindsight bias amongst radiologists; brain image scans and outcomes will be present in the case-notes; this could influence the judgement of retrospective case-note reviewers.
FARRIS O, PIETKIEWICZ DS, RAHMAN AS, ADAM TJ, POKOMO SV, MELTON GB. A qualitative analysis of EHR clinical document synthesis by clinicians. Amia 2012;Annual Symposium proceedings / AMIA Symposium. AMIA Symposium. 2012:1211-20.	Relevance to the creation of case-notes and medical records; No direct relevance to the factors influencing medical record review judgement

FERNANDO KJ, SIRIWARDENA AK. Standards of documentation of the surgeon-patient consultation in current surgical practice. British Journal of Surgery. 2001;88(2):309-12.	Concerns the integrity of the retrospective case-note review process. There is no mention of judgement elements around the case-note records.
FESSEL WJ, BRUNT EV. Assessing quality of care from the medical record. New England Journal of Medicine. 1972;286(3):134-8.	No direct assessment/review of quality of care; outcomes of those with MI and acute appendicitis. It is concerned with finding correlations between only outcome scores.
FESLER-BIRCH DM. Perioperative nurses' ability to think critically. Qual Manag Health Care. 2010;19(2):137-46.	Concerned with clinical decision-making and not the judgement of case-notes.
GARBEZ RO. Level 2 and level 3 patients in a 5-level triage system: factors related to acuity assignment and trajectory of the emergency department experience: University of California, San Francisco; 2008.	Concerned with clinical decision-making and not the judgement of case-notes.
GHOSH T. Structuration and sensemaking: frameworks for understanding the management of health information systems in the ICU. 2007.	A theoretical paper with no primary data
GILL JM, REESE CLT, DIAMOND JJ. Disagreement among health care professionals about the urgent care needs of emergency department patients. Ann Emerg Med. 1996;28(5):474-9.	No direct relevance to the factors influencing medical record review judgement

GONZALEZ COCINA E, TORRES FP. Electronic medical records. A review and analysis of the current situation. Diraya: Electronic medical records in Andalusia, Spain. [Spanish]. Revista Espanola de Cardiologia Suplementos. 2007;7(C):37C-46C.	Absence of primary data.
GRABER ML, KISSAM S, PAYNE VL, MEYER AND, SORENSEN A, LENFESTEY N, ET AL. Cognitive interventions to reduce diagnostic error: A narrative review. BMJ Quality and Safety. 2012;21(7):535-57.	Review Paper into the cognitive element of clinical decision-making
HIGUCHI KAS, DONALD JG. Thinking processes used by nurses in clinical decision-making. Journal of Nursing Education. 2002;41(4):145-53.	Looks at nurse judgements arounds the chart records, however, there is no comment around factors influencing medical record judgements.
KOOPMAN RJ, STEEGE LM, MOORE JL, CLARKE MA, CANFIELD SM, KIM MS, ET AL. Physician Information Needs and Electronic Health Records (EHRs): Time to Reengineer the Clinic Note. J Am Board Fam Med. 2015;28(3):316-23.	Relevance to the creation of case-notes and medical records; No direct relevance to the factors influencing medical record review judgement
KOSSMAN SP, BONNEY LA, KIM MJ. Electronic health record tools' support of nurses' clinical judgement and team communication. CIN - Computers Informatics Nursing. 2013;31(11):539-44.	No direct relevance to the factors influencing medical record review judgement

LAXMISAN A, HAKIMZADA F, SAYAN OR, GREEN RA, ZHANG J, PATEL VL. The multitasking clinician: decision-making and cognitive demand during and after team handoffs in emergency care. International Journal of Medical Informatics. 2007;76(11-12):801-11.	No direct relevance to the factors influencing medical record review judgement
LEBLANC VR, BROOKS LR, NORMAN GR. Believing is seeing: the influence of a diagnostic hypothesis on the interpretation of clinical features. Acad Med. 2002;77(10 Suppl):S67-9.	Inappropriate study design i.e., head and shoulder photographs are not typically used in medical records.
MCLANE S, TURLEY JP. Taxonomy development and knowledge representation of nurses' personal cognitive artefacts. AMIA Annu Symp Proc. 2009;2009:436-40.	No direct relevance to the factors influencing medical record review judgement
PETERS A, VANSTONE M, MONTEIRO S, NORMAN G, SHERBINO J, SIBBALD M. Examining the Influence of Context and Professional Culture on Clinical Reasoning Through Rhetorical-Narrative Analysis. Qualitative Health Research. 2017;27(6):866-76.	Inclusion of biases and Cognitive biases. No mention of factors influencing medical record review judgement

RESTIVO L, APOSTOLIDIS T, BOUHNIK AD, GARCIAZ S, AURRAN T, JULIAN-REYNIER C. Patients' non-medical characteristics contribute to collective medical decision-making at multidisciplinary oncological team meetings. PLoS ONE. 2016;11 (5) (no pagination)(e0154969).	employs collective decision-making at MDT and not the assessment of the influence of medical records
TILBURT JC, MILLER FG, JENKINS S, KAPTCHUK TJ, CLARRIDGE B, BOLCIC-JANKOVIC D, ET AL. Factors that influence practitioners' interpretations of evidence from alternative medicine trials: a factorial vignette experiment embedded in a national survey. Med Care. 2010;48(4):341-8.	Concerned with judgements of CAM trials and not case record review.
WEINGART SN, DAVIS RB, PALMER RH, BETH HAMEL M, MUKAMAL K, PHILLIPS RS, ET AL. Discrepancies between explicit and implicit review: physician and nurse assessments of complications and quality. Health services research. 2002;37(2):483-98.	Inconsequential discussion with the results presented
WOLKENSTEIN L, BRUCHMULLER K, SCHMID P, MEYER TD. Misdiagnosing bipolar disorder--do clinicians show cognitive biases and heuristics? J Affect Disord. 2011;130(3):405-12.	Inappropriate clinical context for the consideration of cognitive biases and human factors.

ZWAAN L, THIJS A, WAGNER C, VAN DER WAL G, TIMMERMANS DR. Relating faults in diagnostic reasoning with diagnostic errors and patient harm. Acad Med. 2012;87(2):149-56.	Discuss the examples of suboptimal cognitive acts and which factors are correlated with SCA. However, there is no in-depth indication of the causal pathway which may have led to this influence.

Appendix 5: Cochrane Systematic Review Data extraction tool (Caplan et al.)

Data collection form

Intervention review –

Randomised trials and non-randomized trials

This form can be used as a guide for developing your own data extraction form. Sections can be expanded and added, and irrelevant sections can be removed. It is difficult to design a single form that meets the needs of all reviews, so it is important to carefully consider the information you need to collect and design your form accordingly. Information included on this form should be comprehensive and may be used in the text of your review, 'Characteristics of included studies' table, risk of bias assessment, and statistical analysis.

Notes on using a data extraction form:

Be consistent in the order and style you use to describe the information for each included study.

Record any missing information as unclear or not described, to make it clear that the information was not found in the study report(s), not that you forgot to extract it.

Include any instructions and decision rules on the data collection form, or in an accompanying document. It is important to practice using the form and give training to any other authors using the form.

You will need to protect the document to use the form fields (Tools / Protect document)

Review title or ID
Caplan et al. 1991

Study ID (surname of first author and year first full report of study was published e.g., Smith 2001)
Caplan et al. 1991

General Information

Date form completed	24/10/2017
Name/ID of person extracting data	AT
Report title (Title of paper/ abstract/ report that data are extracted from)	Effect of Outcome on Physician Judgements of Appropriateness of Care
Study funding source (Including role of funders)	N/A

Eligibility

Study Characteristics	Review Inclusion Criteria (Insert inclusion criteria for each characteristic as defined in the Protocol)	Yes/ No / Unclear	Location in text (pg & ¶/fig/table)
Type of study	Randomised trial	...	
	Non-randomised trial	...	
Participants	56 Physicians	...	
Types of intervention	'Temporary' and 'Permanent' Outcome Formats	...	
Types of outcome measures	'Appropriate', 'Less than Appropriate' and 'Impossible to Judge'	...	
Decision:	Include		

DO NOT PROCEED IF STUDY EXCLUDED FROM REVIEW

Population and setting

	Description Include comparative information for each group (i.e., intervention and controls) if available	Location in text (pg & ¶/fig/table)
Population description (From which study participants are drawn)	115 anaesthesiologists originally agreed to participate, and 112 (97%) completed their case reviews. Reviewers were in 36 states plus Washington, DC. All 30 US administrative districts of the American Society of Anaesthesiologists were represented. The average (SD) age of reviewers was 51 ± 8 years, and the average number of years in practice was 20 ± 8 years. The distribution of practice settings was 40% private, 31% academic, 22% private practice with teaching responsibilities, and 6% unspecified. All reviewers were currently in practice, except for one reviewer with predominantly administrative duties. Board certification in anaesthesiology or its foreign equivalent was held by 110 reviewers (98%).	Methods

	Description Include comparative information for each group (i.e., intervention and controls) if available	Location in text (pg & ¶/fig/table)
Setting (Including location and social context)	Hospitals	
Inclusion criteria	Case-notes from a database if they could be i) classified as either temporary or permanent outcome, ii) the outcome in each case can be altered to the opposite severity without the necessary alteration case details or compromising plausibility, 3) no evidence of gross errors or negligence.	Methods
Method/s of recruitment of participants	Cases selected from the database of the Closed Claims Project of the American Society of Anethesiologists.	Methods

Methods

	Descriptions as stated in report/paper	Location in text (pg & ¶/fig/table)

Aim of study	Is a permanent injury more likely to elicit a rating of inappropriate care than a temporary injury?	Abstract
Design (e.g., parallel, crossover, non-RCT)	Retrospective Case record review	
Unit of allocation (By individuals, cluster/ groups, or body parts)	Random allocation of case outcomes (i.e., permanent, or temporary harm)	

Quality assessment (Adapted from Effective Public Health Practice Project, Canada's Quality Assessment Tool for Quantitative Studies)

See *Chapter 8* of the Cochrane Handbook. A *dictionary aid* is used to support the quality assessment. Additional domains may be required for non-randomised studies.

*See Assessment Tool for Corresponding Options

Component Ratings	Risk of bias Low to High/Unclear*	Support for judgement	Location in text (pg. & ¶/fig/table)
Selection Bias			
Participants representative of target population	2	Anaesthesiology will not represent all specialties; enable or facilitate therapy.	
Proportion of selected individuals agreeing to participate	1	US-wide sample (36 states)	
SECTION RATING			
Study Design			
Study Design	2		
Randomized? If not, go to C.	Y	Case-note outcomes were randomly allocated.	Methods

Component Ratings	Risk of bias Low to High/Unclear*	Support for judgement	Location in text (pg. & ¶/fig/table)
Selection Bias			
Randomization Description	Y	The original and alternate forms of each case were randomly divided between two sets, and one of these sets was randomly assigned to each reviewer.	
Randomization Method Appropriate	CT		
SECTION RATING			
Confounders			
Important Differences between group	N	Matched case-note pairs were used.	

Component Ratings	Risk of bias Low to High/Unclear*	Support for judgement	Location in text (pg. & ¶/fig/table)
Selection Bias			
Examples of group differences	None	Average age of patient (from case-notes) may have been a factor assuming the clinician took the QALY (QoL and LE) into consideration in the quality-of-care judgements. Quality of care judgements are influenced by the outcomes; however, would this vary depending on age.	

Percentage of confounders controlled	Low	The demographics (i.e., average age, gender balance, length of stay) of the 21 cases were not considered as potentially influencing the magnitude of the observed outcome effects. To control for factors such as a mean age, gender, length of stay and other features which may have contributed to the magnitude of the effect. This is to make a case for extent to which the outcome bias	Methods + Results

Component Ratings	Risk of bias Low to High/Unclear*	Support for judgement	Location in text (pg. & ¶/fig/table)
Selection Bias			
		can be transpired through these factors.	
SECTION RATING			
Blinding			
Assessor awareness of intervention or exposure status	N	Random allocation of notes	Methods
Study participants aware of research question	CT		
SECTION RATING			
Data Collection Methods			
Data collection tools valid	Y		

Component Ratings	Risk of bias Low to High/Unclear*	Support for judgement	Location in text (pg. & ¶/fig/table)
Selection Bias			
Data collection reliable	Y	Those trained in record review used	Methods
SECTION RATING			
Withdrawals and Drop-Outs			
Withdrawals reported with reasons	Y N CT		
Proportion of Participants completing study	Y N CT		
SECTION RATING			
Intervention Integrity			
Participant proportion receiving intervention/exposure	1	Half each	Methods

Component Ratings	Risk of bias Low to High/Unclear*	Support for judgement	Location in text (pg. & ¶/fig/table)
Selection Bias			
Participant proportion receiving intervention/exposure	1	Half	Methods
Intervention Consistent	Y	Use of matched pairs and thorough methodology suggests no inconsistencies	Methods
Unintended intervention	N		
Analyses			
Unit of allocation	Other (Case Notes)		

Component Ratings	Risk of bias Low to High/Unclear*	Support for judgement	Location in text (pg. & ¶/fig/table)
Selection Bias			
Unit of analysis	Community Organisation Practice/Office Individual Other		
Appropriate statistical methods	Y	Chi squared for matched pair data	
Analysis by allocation or intervention	Y	Intervention	
GLOBAL RATING	1		
With both reviewers:			
Reviewer Discrepancy	N		

Component Ratings	Risk of bias Low to High/Unclear*	Support for judgement	Location in text (pg. & ¶/fig/table)
Selection Bias			
Discrepancy Reason	1 2 3		
FINAL DECISION	1		

Participants

Provide overall data and, if available, comparative data for each intervention or comparison group.

	Description as stated in report/paper	Location in text (pg. & ¶/fig/table)
Total no. randomised (Or total pop. at start of study for NRCTs)	42 permutations (21 cases)	
Clusters (If applicable, no., type, no. people per cluster)		

	Description as stated in report/paper	Location in text (pg. & ¶/fig/table)
Baseline imbalances	None due to matched pairs	
Withdrawals and exclusions (if not provided below by outcome)	112 of 119 (97%) completed their reviews. This is a very high completion rate.	
Age	51±8 and practicing 20±8 years	Methods
Sex	N/A	
Race/Ethnicity	N/A	
Severity of illness	N/A	
Co-morbidities	N/A	
Other treatment received (Additional to study intervention)		

	Description as stated in report/paper	Location in text (pg. & ¶/fig/table)
Other relevant socio-demographics	The distribution of practice settings was 40% private, 31% academic, 22% private practice with teaching responsibilities, and 6% unspecified. All reviewers were currently in practice, except for one reviewer with predominantly administrative duties. Board certification in anaesthesiology or its foreign equivalent was held by 110 reviewers (98%).	

For randomised or non-randomised trial - Dichotomous outcome

	Description as stated in report/paper	Location in text (pg. & ¶/fig/table)
Comparison		
Outcome	Appropriate care Inappropriate care Impossible to judge	

	Description as stated in report/paper				Location in text (pg. & ¶/fig/table)
Results Note whether: **...** post-intervention OR **...** change from baseline And whether *... Adjusted OR* *...Unadjusted*	Intervention		Comparison		
	No. events	No. participants	No. events	No. participants	
		112			
Baseline data	Intervention		Comparison		
	No. events	No. participants	No. events	No. participants	
No. missing participants and reasons	7 (112/119)				

For randomised or non-randomised trial - Continuous outcome

<table>
<tr><td colspan="2"></td><td colspan="5">Description as stated in report/paper</td><td>Location in text
(pg. & ¶/fig/table)</td></tr>
<tr><td rowspan="3">Results
Note whether:
... post-intervention
OR
... change from baseline
And whether
... Adjusted OR
...Unadjusted</td><td colspan="3">Intervention</td><td colspan="3">Comparison</td><td rowspan="3"></td></tr>
<tr><td>Mean</td><td>SD (or other variance)</td><td>No. participants</td><td>Mean</td><td>SD (or other variance)</td><td>No. participants</td></tr>
<tr><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td rowspan="3">Baseline data</td><td colspan="3">Intervention</td><td colspan="3">Comparison</td><td rowspan="3"></td></tr>
<tr><td>Mean</td><td>SD (or other variance)</td><td>No. participants</td><td>Mean</td><td>SD (or other variance)</td><td>No. participants</td></tr>
<tr><td></td><td></td><td></td><td></td><td></td><td></td></tr>
</table>

For interrupted time series or repeated measures study

Applicability

Have important populations been excluded from the study? (consider disadvantaged populations, and possible differences in the intervention effect)	... *Yes*	Other specialty groups.
Is the intervention likely to be aimed at disadvantaged groups? (e.g. lower socioeconomic groups)	... *No*	
Does the study directly address the review question?	... *Yes*	Very much so. It is comprehensive in content and its scope in answering the question.

(any issues of partial or indirect applicability)		
Notes:		

Other information

	Description as stated in report/paper	Location in text (pg. & ¶/fig/table)
Key conclusions of study authors	the proportion of ratings for appropriate care decreased by 31 percentage points when the outcome was changed from temporary to permanent and increased by 28 percentage points when the outcome was changed from permanent to temporary. We conclude that knowledge of the severity of outcome can influence a reviewer's judgment of the appropriateness of care.	Abstract
References to other relevant studies	Brook and Appel - Quality-of-Care Assessment: Choosing a Method for Peer Review	

Correspondence required for further study information (what and from whom)	No
Further study information requested (from whom, what and when)	No
Correspondence received (from whom, what and when)	N/A
Notes:	

Appendix 6: Conceptual review results

Cognitive biases and heuristics conceptual review

Cognitive biases and heuristics were sourced from Blumenthal-Barby's systematic review because it was the most comprehensive and inclusive[70] systematic review of the literature pertaining to medical decision-making. This review seeks to identify and appraise the following cognitive biases and heuristics specifically expanding upon these criteria:

Origins (i.e., first formal mention of the term in the published academic literature)

cognitive mechanisms which translate effects to case-note reviewer care quality judgements

study designs and methods translatable to the case-note reviewer setting

study limitations. All references can be sourced from the two systematic reviews from Blumenthal-Barby et al. (2015) and Saposnik et al. (2016).

Each study will be rated on the above criteria using the following traffic light system:

GREEN	Strong indication of effect or study design promise relating to care quality judgements
AMBER	Possible indication of effect or study design promise relating to care quality judgements

[70] We consider that medical decision-making is a particular type of decision-making exclusive to those who think about healthcare concerns. This includes lay people. Blumenthal-Barby included clinicians and laypeople and did not restrict their search to clinicians like Saposnik et al.(2016).

RED	Weak indication of effect or study design promise relating to care quality judgements

For example, a study indicates strong indication of effects if it is highlighted like so:

Stefanie J. Sharman (2011) Current negative mood encourages changes in end-of-life treatment decisions and is associated with false memories

These studies were placed under their respective biases or heuristics. An example of the list of Affect Heuristic studies from these two systematic reviews are provided and my rationale for their salience to discover empirical methods to study biases and heuristics during care quality judgements (CQJ) of case-notes.

Affect Heuristic

Affect Heuristic: Origins

The affect heuristic is defined as a representation of the 'reliance on good or bad feelings experienced in relation to a stimulus. Affect-based evaluations are (typically) quick, automatic, and rooted in experiential thought that is activated prior to reflective judgements. The literature evidences that people are influenced by risks framed in terms of counts (e.g. 'from a pool of 50 patients to Mr Smith, 5 are considered to commit an act of violence') compared with percentages (e.g. Patients similar to Mr Smith are estimated to have a 10% chance of committing an act of violence).(Slovic et al., 2000)

Operation

Paul Slovic and his colleagues elicit a cascade for the operation of the affect heuristic(Slovic et al., 2002):

Events or features to which affect is assigned varies as a function of individual characteristics, the task at hand or an interaction of the two. (Individuals differ in how they react affectively; task differ in their evaluability).

These affective qualities are mapped onto stimuli images i.e., imprinted memories with affect

Affective mapping determines contribution of affect to stimulus images of an individual's 'affect pool.' Each image is marked with some affect.

Individuals regularly consult the 'affect pool' in making judgements that is more easily accessed and more efficient than seeking rational or veridical support.

Risk and benefits

A study by *Alhakami and Slovic* found an inverse relationship between perceived risk and perceived benefit of an activity as being linked to the strength of positive or negative affect associated with that activity.(Alhakami and Slovic, 1994) For example, those who have a dislike for the use of pesticides have formed negative emotions about its utility without further counterweighing any benefits from using pesticides for agricultural purposes.

Evaluability

An evaluability principle thus asserts that the weight of a stimulus attribute in an evaluative judgement or choice is proportional to the ease or precision with which the value of that attribute can be mapped into an affective impression. Affect is imbued upon information(Osgood et al., 1964) and this imprinted affect will come to influence future judgements and decision making. Critically, salient attributes for a decision maker may not be presented as it has not been translated precisely, if at all, into an affective frame of reference.(Slovic et al., 2002) (Relevance: not immediate but it could be to do with local cultures, procedures and protocols.)

Proportion dominance

The use of percentages or proportions is critical for fairly weighing judgements across many judgement tasks. For example, an overfilled ice cream container with 7oz ice cream was valued more highly than an under-filled container with 8oz of ice cream.(Hsee, 1998) (Relevance:

Failures of the experiential system

The affective system is designed to sensitize us to small changes in the environment at the cost of making us less able to appreciate and respond appropriately to larger changes further away from zero. For example, a single death you are personally experiencing will commit your affective system but to the contrary, the affective system, is less committed to deaths described statistically using large numbers.(Fetherstonhaugh et al., 1997) (Relevance: Active or passive voice)

Resource constraints

The affect heuristic is more actively operative in resource-scarce or time-pressured considerations. Consider the example, instead of considering risk and benefits separately, those with negative attitudes to nuclear power perceive its benefits as low and risk as high which leads to a more negative risk-benefit evaluation than would be evident under situations with no time pressures.(Finucane et al., 2000) (Relevance: time pressure is a genuine concern in case-note review. Could the perception of a distasteful intervention/procedure result in a harsher judgement for a case review?)

VERDICT: The affect heuristic plausibly influences reviewer care quality judgements.

Positive and negative affect may influence the reviewer care quality judgements i.e., those who have a dislike for appendectomies, due to their previous negative outcomes post-surgery, comes to influence their care quality judgement over case-notes who've had appendectomies undertaken.

A further example is that the review is mostly a dispassionate review, however, if the reviewer is acquainted with a colleague or aspect of the particular case-note that can be personalised by the reviewer, this may influence their care quality judgement e.g., partiality towards a colleague who is also their long-time friend.

Furthermore, those under time pressure to complete reviews will have a proclivity to express more their personal prejudices e.g., someone who is averse to the prescription of a particular antibiotic (i.e., gentamicin that can cause serious side-effects) will review the care quality lower than if they had carefully considered the risks and benefits of that antibiotic in the scenario.

Affect Heuristic in Medical Decision-making (MDM)

Yee et al. - The relationship between obstetricians' cognitive and affective traits and their patients' delivery outcomes [Example of Studies Included]

In a study investigating the relationship between physician coping skills, need for cognition, tolerance of ambiguity, and anxiety and on their patients' mode of delivery, it was found that

physicians with the most reflective coping (i.e., highest quartile) were significantly less likely (adjusted odds ratio, 0.70; 95% confidence interval, 0.50-0.98) to perform operative vaginal delivery (OVD), an atypical delivery mode. (Higher-than-expected OVD rates have been shown to indicate poorer maternal and neonatal outcomes compared with below and expected OVD rates in an Italian study.) (Maso et al., 2015) However, lower anxiety and higher ambiguity tolerance were associated with an increased risk of chorioamnionitis and postpartum haemorrhage, respectively. There were no identified differences in adverse neonatal outcomes by physician cognitive or affective traits.(Yee et al., 2014) The affect heuristic may be operative in the obstetric setting and it is worth noting that OVD rates may be an overt indicator of egregious harm to either maternal and neonatal health. Thus, more research is warranted to establish any link between affective traits and the rates of the obstetric delivery mode. This does commit the assumption that delivery mode rates are indicative of care quality, which is further contextual issue subject to circumstantial particularities.

CQJ

Obstetric outcomes are influenced by affective traits. Obstetric outcomes involve care quality decisions. Care quality decisions are informed by care quality judgements. Reviewers review notes and there could be a link between affect in clinical obstetrics elements in case-note review.

Note: for reasons of space and its indirect relation to the PhD research questions, the rest of the conceptual review is not included. It can be requested, if needed, from the thesis author.

Appendix 7: Panel Contributions and Qualifications

Expert Panel Member	Review contribution	Qualification as expert consensus panel member	Case-note review credentials	Behavioural science credentials
AT	Chair for meeting, project progenitor and lead; involvement in initial review (PROSPERO ID: CRD42017060142) and involved at all stages Develop and refine applied definitions and scenarios and provide ranking scores	Natural scientist, methodologist who led on conception, direction and each review and write-up stage	Non-reviewer Observed, studied, and became acquainted with case-note reviewing	Became acquainted with behavioural science through literature and expert discussion
KAS	develop and refine applied definitions and scenarios to ensure psychological constructs and plausibility are maintained,	Cognitive psychology lecturer, study methodologist, experimental psychology & philosophy	Non-reviewer	Behavioural scientist

	contributions to theory development and behaviour change interventions and involved at stage Refine applied definitions and scenarios and provide ranking scores	contributions, applied psychology to health and well-being with substantial contributions to manuscript	Became acquainted with case-note reviews through attendance of case-note review training and expert discussion	
CB	Develop and refine applied definitions and scenarios and provide ranking scores	Intensive care consultant (then final year specialty registrar) and	Case-note reviewer trainer	Became acquainted with behavioural science through literature and expert discussion
LQ	Involvement in initial review (PROSPERO ID: CRD42017060142), both develop and refine applied definitions and clinical scenarios	Academic Foundation Programme Doctor with involvement at Stage I data extraction	Non-reviewer Became acquainted with case-note reviews	Became acquainted with behavioural science through

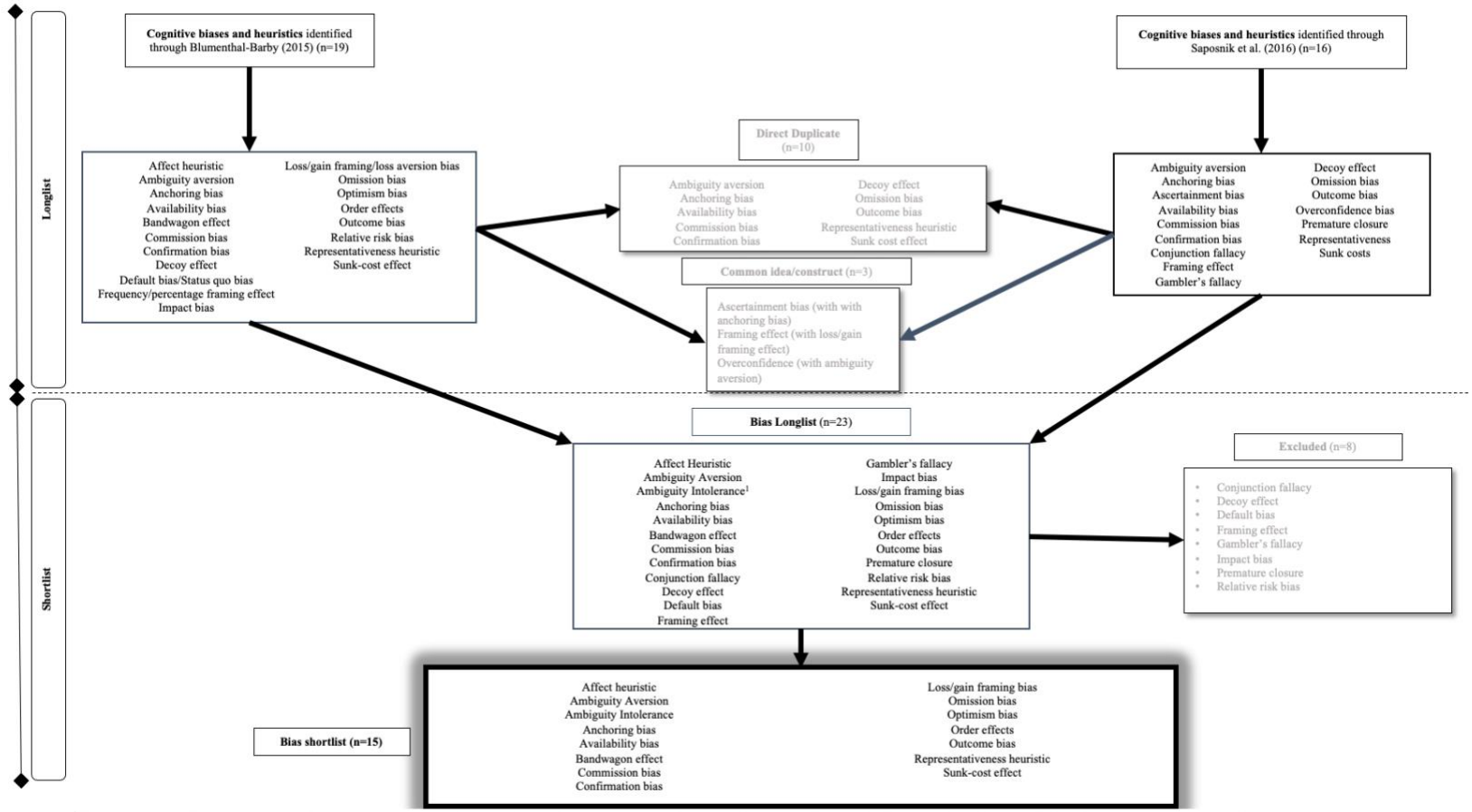
			through expert discussion	literature and expert discussion
JB	Develop applied definitions and scenarios and provide ranking scores	Professor of Intensive Care Medicine, High-Intensity Specialist-Led Acute Care (HiSLAC) Principal Investigator	Case-note reviewer trainer	Became acquainted with behavioural science through literature and expert discussion
TPH	Refine and accept applied definitions and scenarios and provide ranking scores	Associate Director for analytic and information resources, study methodologist, statistician measuring and profiling care	Case-note reviewer trainer	Became acquainted with behavioural science through literature and expert discussion

		quality, internal medicine physician (US)[71]		
RL	Refine and accept applied definitions and scenarios and provide ranking scores	Professor of Public Health, formerly a consultant in Obstetrics and Gynaecology with substantial contributions to manuscript Case-note reviewer trainer	Case-note reviewer trainer	Became acquainted with behavioural science through literature and expert discussion
IV	Involvement in review (PROSPERO ID: CRD42017060142), ensure psychological constructs and plausibility are maintained, contributions to theory development and behaviour change interventions	Professor of Behavioural Science, human decision-making, and behaviour change	Non-reviewer Became acquainted with case-note reviews	Behavioural scientist

[71] Despite TPH not being from the West Midlands, this person has had longstanding involvement with quality improvement and case-note review methodology and was situated in the West Midlands for a sabbatical in 2018.

	Refine and accept applied definitions and scenarios and provide ranking scores		through expert discussion	
SMH	Involvement in initial review (PROSPERO ID: CRD42017060142), Develop and refine applied definitions and scenarios at stage I and II	Health systems (low-middle income countries) expert, retrospective case-note review methodologist and practitioner, previously paediatrician with substantial contributions of the manuscript	Case-note reviewer trainer	Became acquainted with behavioural science through literature and expert discussion

Appendix 8: Shortlisting Process for Biases



[1]We interpreted Saposnik *et al*. as viewing ambiguity intolerance as not distinct from ambiguity aversion as both were considered together in their paper. However, later literature considered these two distinct biases and so they became distinct in our paper.

Appendix 9: Excluded cognitive biases and heuristics

EXCLUDED BIASES	Definition	Reason for Exclusion
Decoy effect	. . . the addition of such [asymmetrically dominated] alternatives increases the share of the item that dominates it… (Huber et al., 1982)	Global care quality judgments are unlikely to be susceptible as no additional discrete alternative is found in case-notes.
Default bias	. . . individuals have a strong tendency to remain at the status quo, because the disadvantages of leaving it loom larger than advantages.(Kahneman et al., 1991)	Global care quality judgments are unlikely to be susceptible to default options, as there is no default position in case-notes.
Conjunction fallacy	…Violations of the conjunction rule, P(A&B) < P(B), are observed in both between-subjects and within-subjects comparisons, with both fictitious and real-	Global care quality judgments are not susceptible to conjunction rules or fallacies as case-notes present

	world events…(Tversky and Kahneman, 1981)	person-specific information, not generalisable information.
Frequency/percentage framing effect	. . . frequency scales generally . . . lead to higher perceived risk. . . .(Slovic et al., 2000)	Global care quality judgments are not plausibly influenced by the quantitative risk presentation format as percentages or rates are not found in case-notes.
Gambler's fallacy	…when subjects act as if every segment of a random sequence/set of events must reflect the true proportion; if the sequence has strayed from the population proportion, a corrective bias in the other direction is expected…(Tversky and Kahneman, 1971)	Global care quality judgments are not influenced by the gambler's corrective bias as case-notes present a one-off, unrepeatable sequence of medical events.
Impact Bias	. . failure to anticipate our remarkable ability to adapt to new states. People tend to	Global care quality judgments are not influenced by the changing state

	overestimate the long-term impact of both positive events . . . and negative events. . . .(Tversky and Griffin, 2000)	of the patient's health as a case note is an immutable care document and is not susceptible to changing its information state.
Premature closure	…accepting the first plausible diagnosis before it has been fully verified.(Stiegler and Ruskin, 2012, Croskerry, 2002)	Global care quality judgments are not influenced by premature closure as case-note reviews are finalised medical documents precluding any prospective diagnostic decision-making.
Relative risk bias	. . . a stronger inclination to [choose treatment] when presented with the relative . . . risk than when presented with the same [information] described in terms of the absolute . . . risk. (Forrow et al., 1992)	Global care quality judgments are not plausibly influenced by the relative risk as comparative therapeutic information is not found in case-notes.

Appendix 9 References

CROSKERRY, P. 2002. Achieving quality in clinical decision making: Cognitive strategies and detection of bias. *Academic Emergency Medicine,* 9**,** 1184-1204.

FORROW, L., TAYLOR, W. C. & ARNOLD, R. M. 1992. Absolutely relative: How research results are summarized can affect treatment decisions. *The American Journal of Medicine,* 92**,** 121-124.

HUBER, J., PAYNE, J. W. & PUTO, C. 1982. Adding asymmetrically dominated alternatives: Violations of regularity and the similarity hypothesis. *Journal of consumer research,* 9**,** 90-98.

KAHNEMAN, D., KNETSCH, J. L. & THALER, R. H. 1991. Anomalies: The endowment effect, loss aversion, and status quo bias. *The journal of economic perspectives,* 5**,** 193-206.

SLOVIC, P., MONAHAN, J. & MACGREGOR, D. G. 2000. Violence risk assessment and risk communication: The effects of using actual cases, providing instruction, and employing probability versus frequency formats. *Law and human behavior,* 24**,** 271-296.

STIEGLER, M. P. & RUSKIN, K. J. 2012. Decision-making and safety in anesthesiology. *Current Opinion in Anaesthesiology,* 25**,** 724-729.

TVERSKY, A. & GRIFFIN, D. 2000. Endowments and Contrast in Judgments of Well-Being. *Choices, Values, and Frames***,** 709-725.

TVERSKY, A. & KAHNEMAN, D. 1971. Belief in the law of small numbers. *Psychological Bulletin,* 76**,** 105-110.

TVERSKY, A. & KAHNEMAN, D. 1981. Judgments of and by Representativeness. STANFORD UNIV CA DEPT OF PSYCHOLOGY.

Journal of Medicine. 1992;92(2):121-4.

Appendix 10: Recruitment Solicitation

Box 2: Recruitment Solicitation

Dear [Reviewer Title & Name]

On behalf of the HiSLAC project, we would like to thank you, for your dedicated, hard work in reviewing the large number of case-notes.

We invite you to participate in our study to understand the influence of attitudes upon care quality judgement over the retrospective case record reviews.

We ask you to kindly complete each questionnaire by 30th June 2019. The survey will take about 12 minutes to complete.

We hope the results will guide the training and development of case-note reviewers and potentially increase agreement between the reviewer care quality judgement judgements over hospital admission retrospective case record reviews. At the end of the study, we will send you a copy of the final publication draft before the formal, public dissemination for your own benefit and professional practise. Your participation was very important for this project.

[HERE: RedCap SURVEY LINK]

In 1 weeks', time, we will send you a reminder email to complete the survey.

If there are any queries, please contact: An Te to clarify any concerns.

Best wishes,

Mr An Te, Institute of Applied Health Research, Level 1

Murray Learning Centre

University of Birmingham

Edgbaston, Birmingham, B15 2TT

Appendix 11: HiSLAC list of case-note and case-note reviewer characteristics

Case-note review specific
Sex (Male or Female)
Age at hospital admission
Length of hospital stay (days)
Primary admitting diagnosis (ICD-10)
Comorbid disease (Charlson)
Total number of errors (ℕ)
Care quality judgement (1-5)
Confidence in care quality judgement (0-100)
Reviewer characteristics
Gender
Grade (Specialist Registrar training years 5-7, or Consultant)
Year of graduation from medical school (year)
Current speciality/specialties (list of options)

Appendix 12. Baseline characteristics of level-one and level-two variables

Variables				
CASE-NOTES (n=4,408)	Mean	SD	Min.	Max.
Age at hospital admission	61.2	22.4	17	107
Gender (male (%))	47%			
REVIEWERS (n=72)	Mean	SD	Min.	Max.
Years since graduation	13.0	4.2	7	28
Gender (male (%))	65%			

Appendix 13. Description of level-one, level-two predictors, and outcome variables

Variable	Description
Demographic variables	
Patient Gender	The gender of the patient
Reviewer Gender	The gender of the case-note reviewer
Length of stay	Length of patient stay during the admission episode
Total number of errors identified	Total number of errors identified across the patient's admission
Patient condition pre-admission	Patient condition immediately before the illness that led to this admission
Training variables (reviewer-only)	
Years since graduation	The year the reviewer graduated from medical school
Grade	The medical status and experience of the reviewer (i.e., consultant, or specialty registrar)
Outcome variables	
Global care quality judgement	a 5-point (1= worst, 5=best) Likert scale rating of care quality of the case-note
Confidence in Global care quality judgement	a continuous (0-100) scale rating of the confidence in their global care quality judgement

Independent Variables	COEFFICIENT (POINTS EFFECT)	95% CI (Lower, Upper)
Need for Cognition	0.06	-0.077,0.20
Reviewer years since graduation	-0.02	-0.04,0.00
Length of stay (>7 days)	-0.10**	-0.15,0.04

Appendix 14: Global care quality judgement and confidence in global care quality judgement tables by the three attitudinal measures

Models by global care quality judgement (*Mean = 4.17)*

Independent Variables	COEFFICIENT (POINTS EFFECT)	95% CI (Lower, Upper)
Anxiety due to Uncertainty	0.06	-0.077,0.20
Reviewer years since graduation	-0.02	-0.04,0.00
Length of stay (>7 days)	-0.10**	-0.15,0.04

Models by Confidence in the Global care quality judgement (*Mean = 82.6)*

Independent Variables	**COEFFICIENT (POINTS EFFECT)**	**95% CI (Lower, Upper)**
Personal Need for Structure	0.10	-0.037,0.24
Reviewer years since graduation	-0.02	-0.04,0.00
Patient Length of stay (>7 days)	-0.10**	-0.15,0.04
Independent Variables	**COEFFICIENT (POINTS EFFECT)**	**95% CI (Lower, Upper)**
Anxiety due to uncertainty	-2.78*	-5.09,0.47

Needs help with some activities	-1.19	-2.49,0.13
Dependent for most activities	0.27	-1.90,1.36
Unable to determine	-19.83***	-21.5,-17.2

Independent Variables	**COEFFICIENT (POINTS EFFECT)**	**95% CI (Lower, Upper)**
Need for Cognition	0.11	-2.79,3.01
Needs help with some activities	-1.18	-2.49,0.13
Dependent for most activities	0.27	-1.90,1.36
Unable to determine	-19.82***	-21.2,-17.5

Independent Variables	**COEFFICIENT (POINTS EFFECT)**	**95% CI (Lower, Upper)**
Personal Need for Structure	-3.29	-6.13,0.45

Needs help with some activities	-1.17	-2.49,0.13
Dependent for most activities	0.28	-1.90,1.36
Unable to determine	-19.80***	-21.2,-17.5

Appendix 15. Cross-classified model vs. Simple multi-level model[72]

Random effects parameters	**Simple multi-level model**	**Cross-classified model**
Mean care quality score	4.17	4.16
Standard Error	0.0963	0.0448
Variance partition coefficients		
Between Reviewer (group-level)	0.254	0.135
Between Patient Record within Reviewer (individual-level)	0.0849	0.0859
% of total variance (%)		

72

Between Reviewer (group-level)	20.4%	20.3%
Between Patient Record within Reviewer (individual-level)	79.6%	12.9%
Between review occasion (and other variables)	n/a	66.8%

Appendix 15 Caption

If the study design was not robust, there will be a significant discrepancy between the reviewer variance contribution to the total variance between the simple multilevel model and simple cross-classified model. Due to the close match between the percentage contributions of the "between reviewer" variance for each model, we can therefore conclude that the case-notes are not clustered by a higher-level cross-classified structure (i.e., there was no nested structure in which case-note reviews were reviewed or distributed to reviewers; this assumes case-note reviews were randomly assigned to case-note reviewers).

As a small note, there is a still a large amount of variance not accounted for at the individual level of the cross-classified level, which leaves a significant proportion of variance allotted to the review occasion within the same case-note, alternatively called "noise", which amounts to 66.8.

Appendix 16: Critical realism – one way to represent how case-note reviews really work in a hospital?

Critical Realism (CR) emerged in the 1970-80s through the work of Roy Bhaskar and was elaborated on by others.(Archer and Archer, 1995, Sayer, 1992) CR originated as an alternative approach to positivism and constructivism(Denzin and Lincoln, 2005), by drawing elements from these methodological strains into its own ontology and epistemology for healthcare practise.(Schiller, 2016, Edwards et al., 2014) CR assumes that knowledge is crafted by humans which is a small fraction of the entire causal process underlying all events. And so, human ways of knowing are expected to complement reality. Critical realists can gain knowledge from theories which are 'more or less truthlike' (Danermark, 2002)p.10. The theories that help us get closer to reality, i.e., that help us identify causal mechanisms driving events (Archer et al., 2013) p. xi.

Firstly, I hold the critical realist position in this thesis. Critical realism as a research approach has three basic assumptions as taken from Andrew Sayer's book(Sayer, 1992):

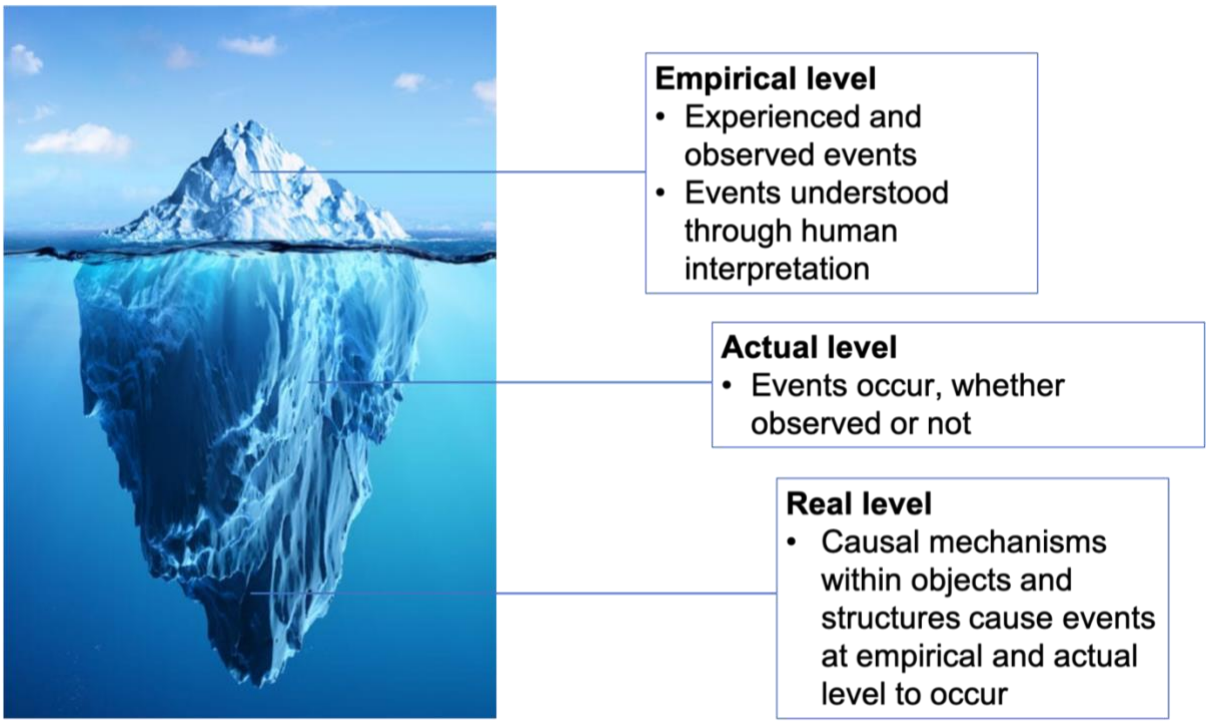
The world exists independently of our knowledge of it.

There is necessity in the world; objects - whether natural or social - necessarily have powers or ways of acting and particular susceptibilities.

The world is differentiated and stratified, consisting not only of events, but objects, including structures, which have powers and liabilities capable of generating events. These structures may be present even when, as in the social world and much of the natural world, they do not generate regular patterns of events.

This is elicited in Figure 16.

Figure 17. Iceberg representation of critical realism's ontology



Applying this three-fold domain could help explain the phenomena behind any quality improvement from hospital case-note reviews. For instance, the *empirical* domain includes the case-note reviewer's experience from reviewing case-notes, engaging with others about the case-note, their first-person centred observations of the case-note review process. Events occur in the *actual* domain and are not necessarily seen by the subject; for instance, and unbeknownst to reviewers, the lessons from the case-note reviews may well not be actioned in the eyes of non-reviewers e.g., healthcare assistants or some other professional. Events arise from mechanisms in the *real* domain. For instance, there is a socio-political mechanism which explains the empirical observation of perceived actions at the board level and the inaction of these measures by healthcare assistants at the ward level. This could be attributable to power dynamics, dislocated organisational cultures, miscommunication, workarounds, and other physical and sociological mechanisms. Chapter 4 will discuss all elements of this ontology, whilst Chapter 2 and Chapter 3 will discuss the empirical domain

concerned with the case-note reviewer's cognitive activity in evaluating the care quality of a case-note.

To produce any care quality judgement, the clinician must be able to access the information content of the case-note. However, if they are not able to do so, this is attributable to a reviewer or case-note specific component. This also naturally includes how and what case-note reviewers consider case-note reviews, as a historical artefact of some kind. As established at the outset of this section, it is important to have the methodology faithfully represent, as best as is possible, the nature of the phenomenon of interest.

The stance of critical realism pervades the entire thesis but will reach its apotheosis in Chapter 4 in relation to the *events* and *reality* of quality improvement from case-note reviews, with a minor discussion regarding the reviewer's cognitive evaluation of the case-note's care quality judgements.

Appendix 17: Normalisation Process Theory

The SSI questions (and themes) were derived using Normalisation Process Theory (NPT), a template commonly used to identify factors promoting or inhibiting the routine incorporation of a complex intervention.(May et al., 2009, Finch et al., 2012, Murray et al., 2010) The data gathering and interpretation were informed by the NPT. As case-note review was neither a simple nor uniform hospital intervention, we consider NPT a useful template for such exploration. The questions were adapted from Murray's table 1. "*NPT in developing a complex intervention*." (Murray et al., 2010) The NPT constructs satisfy the multi-dimensionality of case-note reviews which was used across multiple specialties and both clinical and non-clinical staff. (i.e., constructs captured by coherence, cognitive participation, collective action, reflexive monitoring constructs).

Normalisation Process Theory was founded on three key tenets. First, it proposes practices become embedded in social contexts from people working, individually and collectively, to implement these same practices. Second, implementation was operationalized through four generative mechanisms – coherence, cognitive participation, collective action and reflexive monitoring.(Finch et al., 2012, Murray et al., 2010) If those implementing the case-note reviews can identify coherent reasons for their adoption and use, are also engaged in cascading the reviews, they are more able to adapt their work processes for case-note review use (or dashboards to fit in with practices), and judge them to be valuable once they are in use, then they are more likely to become embedded in routine practice. Third, embedding new ways of working were not a "one-off" process, but requires continuous investment by those involved in its implementation. Diagram 1 present the relations between these constructs.

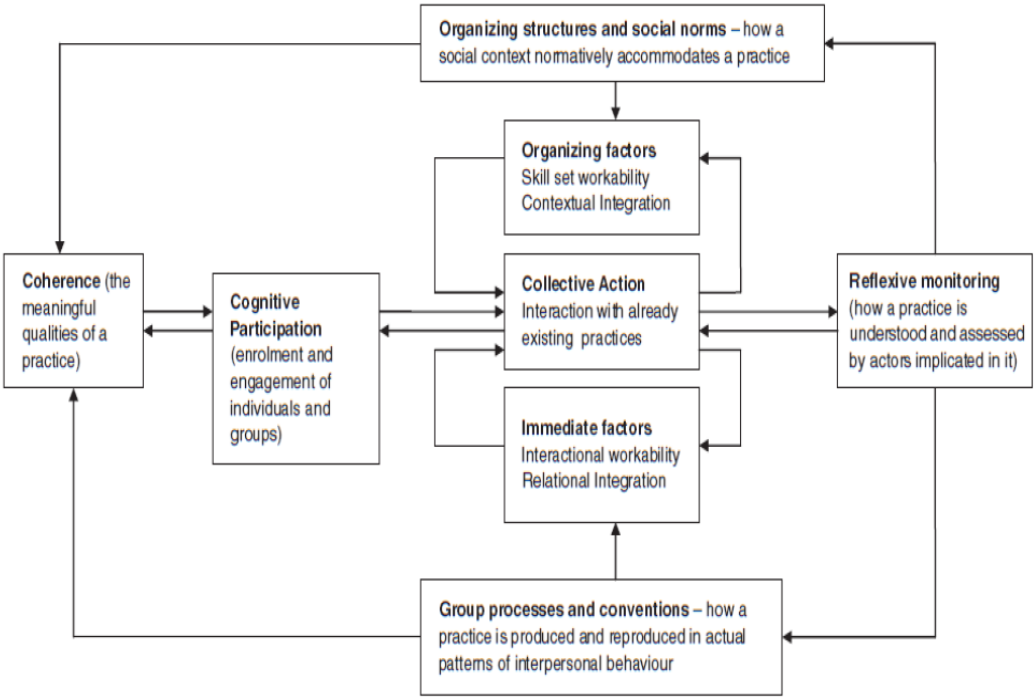


Diagram 1. Model of the construct of normalisation process theory (from May and Finch 2009, p.541)

May and colleagues argue that each of the four generative mechanisms suggests possible research questions(May et al., 2009):

1. Coherence – What is the work? How is a new way of working conceptualized by participants?

2. Cognitive participation – Who does the work? How do they decide whether to engage, and what do they hope to achieve?

3. Collective Action – How does the work get done? How are activities structured, and how are they constrained?

4. Reflexive Monitoring – How are new ways of working – or attempts to introduce them - interpreted by those involved? How do they evaluate new working practices?

As case-note review was an intervention which was gaining advocates, NPT informs the semi-structured interview (SSI) questions.

A significant challenge of implementing the NPT was that many of the constructs overlap, and issues emerging from the interviews and documentary evidence may have too closely related to a separate category. As presented in Diagram 1, it was evident that certain identified themes were placed in more than one construct. Another weakness of the NPT constructs was the lack of focus on the implementation stage particularly for the RCP tool. This concern was nuanced and no moot point. The coding of NPT constructs may favour Coherence and Cognitive Participation with the implementation of an intervention, as these constructs are considered essential to the update and delivery of any new intervention. And the Collective Action and Reflexive Monitoring constructs square better with the actual experiences of user of an intervention. However, this general dichotomy was complicated by implementation work often requiring both sense-making and experience (e.g., using case-note review enables one to make better sense of its implementation and vice-versa). To ameliorate this complexity, we needed to appreciate the local speech context in which a code was considered to best fulfil its function. This caveat needs to be borne in mind for coders not appreciating the sensitivity of the constructs to the coding context.

The constructs and their definitions were ill-adapted for developing questions to elicit the core concerning the uptake of the RCP review methodology and the appreciation of inter-management dynamics.(Atkins et al., 2011) The NPT offers a professional set of constructs for service-providers, planners and policymakers without truly representing the service-users; the case-notes were a simulacrum for the patient's care.(de Brún et al., 2016a) And linked to the management dynamics, NPT makes no consideration for power and inclusivity in its constructs. Due to the complex organisational dynamics foreseen in the Trust environment, we included an organisational component to capture the inter-group Trust dynamics around case-note reviewing.(Davies et al., 2000)

Appendix 18: Framework Method Table

Framework Theme	**Sub-themes**	**Definition**	**Illustrative example**
Over-arching theme to organise themes and categories	Themes generated from the process of inductive coding	The definition of themes	Illustrative examples from the interview and document data
Framework theme 1	Culture	Descriptions of actual experiences with case-note review	*'I think it's their culture. Medical and surgical specialties have more of a culture of sort of constantly reviewing and quality improvement and others have a culture more of…' (Doctor 8)*
Framework theme 2	Workload	Case-note review considered from an organisational Trust perspective	*there is a workload part that will affect the reviewers, the lead reviewers and that's a debate we have quite often is how often are we going to get through this. We manage about a 2 and half month behind the emergency division which has the heaviest workload on learning from death reviews (board member 1)*

Framework theme 3	Lack of time	Descriptions of quality improvement processes other than case-note review	*I think it is. We used to do it when we had time. But the thing is… I used to sit with piles of notes on there [points to filing cabinet] waiting to go through because we never find the time because there is always something else which became a priority (Nurse 1)*
Framework theme 4	Lack of learning from review	Descriptions of attitudes and opinions pertaining to the case-note review process	*It's supposed to be learning from mistakes, mortalities… if there's anything we can do to improve patient care… a lot of the time… there is nothing. But the patient has still died as the patient has multiple co-morbidities. (Doctor 4)*

Appendix 19: Case study selection criteria

Criteria for Single Case NHS Trust selection

Category	**Rationale for selecting case-study**
Learning opportunities	According to the two most recent CQC reports, there has been an overall rating improvement from 'need for improvement' in one year to 'good' two years thereafter. This indicates that there is a dynamic and improving Trust which offers the opportunity of potential lessons arising from case-note reviews.

Hospital/Trust Size	The Trust is a medium sized Trust with ~500 beds (UK hospitals range from 214-1000 beds) with most of the beds situated at x hospital. The Trust exhibits all the key services such as 'main theatres (plus four-day surgery theatres) providing planned and emergency surgical facilities for trauma and orthopaedics, general surgery, urology, and gynaecology as well as a wide range of day procedures.' With a comprehensive range of services offered, this case study will seek to capture the range of approaches to case-note reviews in these services; this will raise the likelihood of transferability of our theory to a wider range of Trusts. In the same vein as E.G. Guba's notion of *transferability*, statistically speaking, studying a medium sized Trust will transferable to significant proportion of hospitals based on minimal specification of services provided at these healthcare centres.(Krefting, 1991)
Receptivity	The Trust were very receptive to our approach which is exhibited by the Trust executive team being willing to interview us.
Ethical procedures	The Trust offered a robust and pro-active and expedited ethical process compared with other Trusts. This ensured that the time was maximised for data collection and analysis.

14th May 2018

Mr An Te
[redacted]

Dear Mr An Te,

Re: Service Evaluation: Preventable Mortality Review Processes for Deceased Inpatient Case Notes in Acute Medical Hospital Trusts [redacted]

Letter of Access for Research/Service Evaluation

In accepting this letter, [redacted] confirms your right of access to conduct research/service evaluation through their organisation for the purpose and on the terms and conditions set out below. This right of access commences on **Monday 14th May 2018** and ends on **Friday 31st May 2019** unless terminated earlier in accordance with the clauses below.

You have a right of access to conduct such research/ service evaluation as confirmed in writing in the letter of permission for research from [redacted]

The information supplied about your role in research at University of Birmingham has been reviewed and you do not require an honorary research contract with this NHS organisation. We are satisfied that such pre-engagement checks as we consider necessary have been carried out. Evidence of checks should be available on request to this NHS organisation.

You are considered to be a legal visitor to [redacted] premises. You are not entitled to any form of payment or access to other benefits provided by this NHS organisation to employees and this letter does not give rise to any other relationship between you and this NHS organisation, in particular that of an employee.

While undertaking research/ service evaluation through [redacted] you will remain accountable to your University/Employer, University of Birmingham but you are required to follow the reasonable instructions of [redacted] Research Manager in this NHS organisation or those instructions given on their behalf in relation to the terms of this right of access.

Appendix 21: Pilot interview question template

The questions are grouped into their respective Normalisation Process Theory construct.

Tell me a little about your role?

Can you tell me about your role in the case note review process?

What is the aim of case note review in your Trust?

Can you describe in your own words what the case note review process is?

Is case note reviewing a team effort or an individual's endeavour?

Is case note review apportioned in your job plan?

Is case note review applied differently across the hospitals?

Coherence

Is case note review and its processes easy to describe?

Can the people easily describe the process of case note review?

Can the people you work with describe the purpose of case note review?

Is case note review clearly distinct from other interventions?

Does case note review have a clear purpose for reviewers, managers, board members and patient safety instructors?

Do reviewers, managers, board, and patient safety instructors have a shared sense of the purpose of case note review?

In your eyes, what do they feel is the purpose of case note review?

Is this shared by others in your group?

Is this shared, as whole, across the hospital?

What benefits will the case note review bring and to whom?

What are the benefits for the review process?

What are the benefits for reviewer learning?

What are the benefits for quality improvement across the Trust?

What are the benefits for participants groups?

Hospital staff

Relatives

Board

Trust

National

What quality improvement examples are there in your specialty?

What about across the hospital?

Are these benefits likely to be valued?

Will case note review fit with the overall goals and activity of the organisation?

Through the lens of each of the values:

Is it safe?

Is it effective?

Is it compassionate?

Is it trustworthy?

Cognitive Participation

Are reviewers, managers, board members and patient safety instructors likely to think it is a good idea?

Are reviewers, managers, board members and patient safety instructors considerate to case note reviews in their job plans/formal job description?

Will reviewers, managers, board, and instructors be prepared to invest time, energy and work in case note reviews?

Collective action

How will the case note reviews affect the work of reviewers, managers, board members and patient safety instructors?

How does this affect the review process?

How does this work affect the quality improvement across the Trust?

Will case note reviewing promote or impede the work of clinical reviewers, managers, board members and patient safety instructors?

Did reviewers, managers, board members and patient safety instructors have training before using case note reviews?

What went into the training?

What was the duration of training?

Who was involved in the training?

Did you have training before/after the Royal College of Physicians conveyed the structured case note review form?

What impact will case note reviews have on wider Trust concerns and responsibility of the reviewers, managers, board members and patient safety instructors?

What is case note review's effect upon the division of labour?

What is case note review's effect upon resource use?

Reflexive Monitoring (or RCP)

How do reviewers, managers, board members and patient safety instructors perceive case note review?

Is case note reviewing likely to be perceived as advantageous for patients and staff?

Can reviewers, managers, board members and patient safety instructors give feedback about the intervention once it is in use?

Can case note review be adapted or improved on the basis of your experience?

Is there anything about reviewer personalities affecting case note review?

Is there anything about the local situation or environment which would affect case note review?

What do you feel is the best way to bring about quality improvement?

Organisational Focus

Have the board changed the organisational strategy of the back of case note review? (Strategy)

Has there been service re-organisation that has led to any specific services changes because of case note review?

Have there been any new incentives (rewards/punishments) because of what case note review has found?

From what case note review has found, have there been any sub-cultures of professionals that have tried to block or facilitate any change?

Appendix 22: Consent and Participant information forms

CONSENT FORM

Organisation:……..

Department: …….

Mortality data and review processes for deceased inpatient case-notes in acute medical hospital Trusts in the West Midlands

An Te, [redacted]; Researcher's position: PhD Student

Semira Manaseki-Holland, [redacted]

Basic Information

Question 1. (Please tick *one* as appropriate)

What is your hospital position?

Practicing clinician (Senior – Consultant level)

Practicing clinician (Junior – below Consultant level)

Administrative staff (Junior)

Administrative staff (Senior)

Junior Managers

Senior manager

Other (please specify: ……)

Question 2. (Please tick one or more, as appropriate)

Which of the below are you affiliated to?

Trust board member

Mortality committee member

Other committee or group member in Trust (please

specify:……..)

No committee or no specific role

Consent to the research

I understand that I may withdraw from the study at any time, and that I will not be identified in the research report. I have been told that my treatment and care will not be affected if I take part in this study.

Please tick below if you agree to participate.

I, …….. agree to take part in this service evaluation.

Signature of participant: …… Date:

Signature of researcher: …… Date:

Appendix: Information sheet

Participant information sheet

Department: …….

Organisation:……..……..

A Service Review of deceased inpatient case-note review processes in a West Midlands NHS Acute Care Trust

Description of the proposed review

We would like to identify factors which inhibit or promote the application and utilisation of deceased case-note review in your Trust. I shall be interviewing healthcare professionals and managers involved in the case-note review process. I shall consider different professional groups within your Trust to better understand some of the dynamics around quality improvement.

Invitation to participate and explanation of what participation entails

You have been selected as you have been identified as an individual who is involved in the mortality case-note review process in your hospital.

We would like to ask you to participate in a semi-structured interview to explore the mortality case-note review processes at your Trust.

Your participation is entirely voluntary, and you are free to withdraw at any point during the interview. If you choose to participate or not, there are no implications for your work or consequences or information given about your participation or information you impart to other Trust staff or management, or anyone outside of the Trust. The information will be used anonymously in a report.

You can expect to speak about your experiences and perceptions of the mortality case-note review process in your hospital. These interviews may provide insights into both the benefits and shortfalls of the mortality case-note review process; we ask that you abstain from identifying persons and job titles in this interview. We shall anonymise all identifiable information, and if it is not possible, we shall not use this data in our analysis.

Please seek further clarification from the reviewer if there is anything which they do not understand prior to participating.

Reimbursements and expenses

There are no reimbursements from these interviews. However, the benefit of participating is that the project will inform your local mortality case-note review processes and help to improve quality of care at your Trust.

Confidentiality, Data Security, Job Security

All information will be treated in a confidential manner and no information will be shared with participants external to this team. During interviews, the reviewer will be unable to commit sufficient information in writing and therefore we would like to record the interviews so that valuable information which can be correctly transcribed and analysed. Voice recordings will not be shared with anybody other than the reviewers involved in the service review. They will be securely digitally stored on encrypted University servers. After initial data transcription, and in the final report, all information will remain anonymous; you will not be recognised as a pseudonym will be used. Your job situation will not be affected in any way whether you take part in the review.

The reviewer will securely dispose of voice-recordings upon project completion.

Results of the Review

While no individual staff information will be disseminated and summary information from this interview will form a part of a PhD thesis for the reviewer and be used to form

recommendations for the trust. Our findings may be published in peer-reviewed journals to help other trusts learn from this review but the name of the Trust will be anonymous.

What if I want to withdraw the data?

Your information, with any identifiable information, will be immediately erased from our local University servers.

Review Funding

There is no external funding for this project. The PhD student is self-funded and supervisors are core funded at the University of Birmingham.

Contact Details

Reviewer's name, position, and contact details: An Te, PhD Student, [redacted]

Primary Supervisor's name & contact details: Dr Semira Manaseki-Holland, Senior Clinical Lecturer, [redacted]

Secondary Supervisor's name & contact details: Professor Russell Mannion, Chair of Health Systems

Appendix 23: Case-study Document Names and Properties of Internal Reports informed by case-note reviews

This table summarises the key properties of the reports generated wholly or partially informed by case-note reviews. The purpose, the creator/publisher, the frequency and who receives and gets to read the report were each identified from participant interviews through further correspondence:

Item	Purpose	Publisher	Frequency	Received by
	What is the purpose of the item?	Person/committee responsible for creating the document	how often is this document produced?	Who receives this document?
Patient Safety Newsletter	Convey patient safety lessons across the Trust	Patient Safety Team	Quarterly	Trust
Mortality Report	re-assurance of mortality process for Doctor 6C	Mortality Surveillance Committee	Quarterly	CGC
Patient Safety Report	Tally numbers of medication errors, serious incidents, their nature, and any incidental learning	Patient Safety Surveillance Committee	Monthly	AOGG, CGC
Mortality Update	inform and re-assure the board of directors on the mortality-specific Trust concerns	Associate Medical Director Governance	Bi-annual	Board of directors

Appendix 24: COREQ Checklist

COREQ (COnsolidated criteria for REporting Qualitative research) Checklist

A checklist of items that should be included in reports of qualitative research. You must report the page number in your manuscript where you consider each of the items listed in this checklist. If you have not included this information, either revise your manuscript accordingly before submitting or note N/A.

Topic	Item No.	Guide Questions/Description	Reported on Page No.
Domain 1: Research team and reflexivity			
Personal characteristics			
Interviewer/facilitator	1	Which author/s conducted the interview or focus group?	3
Credentials	2	What were the researcher's credentials? E.g. PhD, MD	17
Occupation	3	What was their occupation at the time of the study?	17
Gender	4	Was the researcher male or female?	17
Experience and training	5	What experience or training did the researcher have?	17
Relationship with participants			
Relationship established	6	Was a relationship established prior to study commencement?	4-5
Participant knowledge of the interviewer	7	What did the participants know about the researcher? e.g. personal goals, reasons for doing the research	5
Interviewer characteristics	8	What characteristics were reported about the inter viewer/facilitator? e.g. Bias, assumptions, reasons and interests in the research topic	3
Domain 2: Study design			
Theoretical framework			
Methodological orientation and Theory	9	What methodological orientation was stated to underpin the study? e.g. grounded theory, discourse analysis, ethnography, phenomenology, content analysis	2-3
Participant selection			
Sampling	10	How were participants selected? e.g. purposive, convenience, consecutive, snowball	4-5
Method of approach	11	How were participants approached? e.g. face-to-face, telephone, mail, email	3
Sample size	12	How many participants were in the study?	3
Non-participation	13	How many people refused to participate or dropped out? Reasons?	3
Setting			
Setting of data collection	14	Where was the data collected? e.g. home, clinic, workplace	3
Presence of non-participants	15	Was anyone else present besides the participants and researchers?	3
Description of sample	16	What are the important characteristics of the sample? e.g. demographic data, date	3-4
Data collection			
Interview guide	17	Were questions, prompts, guides provided by the authors? Was it pilot tested?	3
Repeat interviews	18	Were repeat inter views carried out? If yes, how many?	3
Audio/visual recording	19	Did the research use audio or visual recording to collect the data?	Appendix II
Field notes	20	Were field notes made during and/or after the inter view or focus group?	3
Duration	21	What was the duration of the inter views or focus group?	Appendix II
Data saturation	22	Was data saturation discussed?	5,14
Transcripts returned	23	Were transcripts returned to participants for comment and/or	Appendix II

Topic	Item No.	Guide Questions/Description	Reported on Page No.
		correction?	
Domain 3: analysis and findings			
Data analysis			
Number of data coders	24	How many data coders coded the data?	5
Description of the coding tree	25	Did authors provide a description of the coding tree?	
Derivation of themes	26	Were themes identified in advance or derived from the data?	5
Software	27	What software, if applicable, was used to manage the data?	5
Participant checking	28	Did participants provide feedback on the findings?	Appendix II
Reporting			
Quotations presented	29	Were participant quotations presented to illustrate the themes/findings? Was each quotation identified? e.g. participant number	6-13
Data and findings consistent	30	Was there consistency between the data presented and the findings?	7,14
Clarity of major themes	31	Were major themes clearly presented in the findings?	14
Clarity of minor themes	32	Is there a description of diverse cases or discussion of minor themes?	14

Developed from: Tong A, Sainsbury P, Craig J. Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. *International Journal for Quality in Health Care*. 2007. Volume 19, Number 6: pp. 349 – 357

Once you have completed this checklist, please save a copy and upload it as part of your submission. DO NOT include this checklist as part of the main manuscript document. It must be uploaded as a separate file.

Appendix 25: The "Duty" theme as expressed under the Framework Theory

MEMO: 'Duty'

Definition

Ideology versus practicality: The duty to review is deeply-rooted in the medical profession

Codes

Duty; professionalism; candour

Summary of data

Duty to review

Participants view the duty to review as intrinsically desirable, a sound theoretical principle for learning from care quality issues from the case-note reviews: ***""they look upon reviews as a professional duty." (D4, Consultant Orthopaedic Surgeon)***

In addition, participants were keen to convey their support for a user-led agenda in which new services incorporate families' perspectives, as well as being responsive to the needs of communities.

This was contrasted with the little learning which was perceived from reviews as possibly questioning the importance of reviewing, even as much as challenging it as a duty: ***"It's supposed to be learning from mistakes, mortalities (…) if there's anything we can do to improve patient care (…) a lot of the time… there is nothing. But the patient has still died as the patient has multiple co-morbidities." (D4, Consultant Orthopaedic Surgeon)***

The ideology of the duty of reviewing and the learning obtained from it was contrasted with the practical difficulties of completing 'case-note reviews' in this Trust. So, although the duty of reviewing case-notes with their mandates were philosophically presented as unproblematic, the process of setting up and maintaining consistent reviewing in terms of internal support, leadership and the meaningful engagement with these reviews were seen as

far more challenging: "***it is a competing demand on people who are very busy, and I think that goes throughout the organisation." (SO3, Consultant radiologist and board member)*** Does this mean that participants supported MCNR in theory, but not in practice? ***"I know everybody's really busy and everyone's got targets and the rest of it but we need to learn from what we do. It might make it easier to do what we do instead of saying 'I haven't got time for that.' In fact, we have to make time and learn and make it better through that process." (N3, Nurse reviewer)*** Does this call into question the interviewee's commitment to conducting case-note reviews? Perhaps the reluctance of participants to fully commit to implementing this policy relates to uncertainty about whether this policy is a longstanding one intended to support other initiatives (i.e., medical examiners programme) or is it intended for standalone quality improvement? Given that interviews spanned the introduction of a new mandate to undertake mortality case-note reviews, the political and economic context was clearly present in participants' views, leading many to examine the worth of undertaking MCNR:

"Deaths are much more likely to be unexpected, you know, it's not the usual outcome. So, surgeons have always been interested in their deaths. If you look at expected deaths in elderly people, then that is pushing water uphill, because there will be a view that essentially it will be a chore in a lot of cases because, erm…, we've been doing it a long time and the learning, new learning, from the cases is rare. So, if you're not gaining anything from doing something it becomes a clerical chore." (D1B2, Senior Clinician and Board Member)

Deviant cases

A clinician did not think that MCNR should be mandatory. However, the consultant is in the emergency department, and has obtained little utility from these reviews which could explain this view?

"I can only comment on the ED [emergency department] side of it. The juniors I think do a very efficient thorough job of investigating a lot of cases that we learn nothing form in terms of the amount of depth that they go in for out of hospitals or cardiac arrests. I would say it's not particularly useful but actually they do do a very good job and they write lots of stuff about all of these patients who came in and probably had a cardiac arrest. I don't think the group of patients' we are looking at is useful for ED. I'm not saying the whole thing isn't effective. The way it's done here and the group of patients we focus on in ED is not useful." (D2, Consultant Emergency Medicine Doctor)

<u>Points for further consideration</u>

What are participants' motivations for putting this policy mandate into practice?

What is the MCNR intended to achieve (e.g., improve care quality, emphasise good aspects of learning, enhanced public assurance on the Trust's behalf)?

Are professional ideals (e.g., duty to the care of their patients to review case-notes) compatible with the practical sustainability and efficiency of ensuring care is delivered to all in the NHS at the expense of undertaking fewer case-note reviews?

Appendix 26. Inductive Labels

Mega-themes	**Themes**	**Definition**	**Illustrative example**
Mega-themes to organise themes and categories	Themes generated from the process of inductive coding	The definition of themes	Illustrative examples from the interview and document data
Experience		Descriptions of actual experiences with case note review	
	Case Note Review Process	Descriptions pertaining to the Trust case note review process	so, we initially have a coroner incident where a patient's care has been difficult on the ward and then 6 months later, they then passed away. So, I don't think anything about what they did was responsible for their deterioration.

	Quality Improvement Examples	Descriptions pertaining to quality improvement examples purportedly derived from case note review	So, we initially have a coroner incident where a patient's care has been difficult on the ward and then 6 months later, they then passed away. So, I don't think anything about what they did was responsible for their deterioration. And what we did through case note review, we identified there are certain complex patients which are a nightmare, well for people to look after. And we had a proper meeting with the intensivists, the specialist physicians and we devised a standard operating procedure to ensure that the right people are supervising these patient's care. And I think that was a big improvement. I think doing case notes in terms of endoscopy and we made sure that they are not pressurised to have procedures done quickly. Or hurried. So that's from one extreme to another. I think case note review have been very helpful.

			Erm, we had some time ago, we had a number of deaths that were reviewed on a couple of wards. A couple of the acute medical wards where were seeing a delay in escalation of patients who were having high NEWS scores and led us to looking at electronic observation systems and we didn't have one until that point. And that absolutely drove our commitment to drive the escalation of patient's observations. So that's one really really good example of that.
	RCP Pilot	Descriptions pertaining to the Royal College of Physicians pilot in acute medicine or the RCP methodology	No, no. Not at all. I thought it was really slick and a lot less bulky process to do it that way. I think it was done in timely manner as you're not wading through lots of documents. It was just one sheet where we could capture data on it. It was good.

Organisational		Case note review considered from an organisational Trust perspective	
	Barriers	Discussion of any organisational barriers to case note review	Unfortunately, we don't have great IT here and lots of the systems to get the patient information are two different systems. So that is the most time-consuming bit. Each of the systems doesn't speak to each other so you're going in to get the patient data, but the one stop form is much better. Yeah. they're chasing these notes and all of that in itself is often quite time consuming and frustrating. And that definitely, I know is the feedback that I have. We don't have the notes, or we can't find the notes. That in itself holds it up and then that then puts kind of a sour taste… it's so long-winded to get the notes to do the review and you're drained before you start the review as you're chasing notes.

	Human Resource	Descriptions of the human involvement in case note review	Other specialties have useful input. So, for instance, if you go into intensive care who are anaesthetists, and I am sure they could add to a review of the patient who hadn't quite made it to…there who was otherwise a medical patient. I'm sure that anaesthetists could add to the review of the surgery that went wrong but it is best taken by the specialty within which it is happening. There clearly needs to be other input as well. Erm, not directly involved in the review process but all of the junior doctors are involved in the sense that they're all involved in a 6-month rolling teaching programme which they attend based on the rota and they are working nights or something, they're involved in the case review process and then they give their presentation of the case to a colleague to present on the day.

	Communication	Descriptions relating to the communication during case note review or relating to learning generated	Communication is another issue… that's what you can tell from the note review. And there is nothing worse than having a major complaint that somebody said this, and you look back through the record and there is no record… you have to then accept that what they are saying is likely to be true because… if somebody told me last Friday this, I don't I did, but if I wrote it in the notes or done a contemporaneous letter… I think the quality of the notes of my colleagues on Castle Ward insist on is very high. It is a pleasure to deal with a review. Yeah, the benefits to the board are a consistency and assurance in reporting. I think if you hear the same language being utilised you get a better understanding. Part of the difficulty with any organisation is getting the executive level to understand what is happening on the ground. I think we do that in a consistent way makes it more of an easy two-way process. And the question back with consistency as well.

	Desired Change	Description of a desire for hypothetical or actual change of case note review or processes akin to CNR in the organisation	And incident reporting. I know it's meant to be a no-blame culture and I wouldn't want my name put to this in case it comes back to haunt me… but I personally think there is a blame culture still with incident reporting. And definitely, not for all incidents, but certainly for some I know lots of people who don't bother putting incident forms because well nothing will be done. It will come back on me, so they don't bother. And professionally, everyone gets stressed when they make a mistake so there is a degree of defence which prevents learning.

Non-case note review QI Examples	Quality Improvement Processes	Descriptions of quality improvement processes other than case note review	At the point and at the time, they found out through Datix. But joining the two up through mortality review, that was where it was identified where things could have been done differently. That was then fed back across all divisions through their clinical audit groups and the clinical governance committee without giving the patient information as such. For more of an awareness with mental health issues, learning disabilities and I think that's probably why there are national projects happening around LD. I'll see when the next one is, and it'd be useful to get that perspective of pulling it all together. And it's kind of like pulling it all together. Often the solicitors will talk about inquests which are related to the deaths because the coroner will wanna know how that patient met his death. Often they get the coroner to go to the stand or the people involved. Often there's that linkage there as well.

	Benchmarks	Description of standard/benchmarks/status quo other quality improvement processes	I think that's a difficult one and that's more involved with the governance than within the clinical divisions. There is a little bit of a time lag. Of course, we know of the patients who have died. We have to wait for all of them to be coded. When we send out data to CHKSA, our clinical benchmarker, that's the point where we can get the individual Safety officer 1I scores so we can see what the risk factor was for each patient. So that tends to be about 15 working days after month end, so the August deaths, we'll be getting them through I think tomorrow so we'll have those figures back at the end of the week. We can send that list to governance. Here are all the deaths and here are all the scores and at the moment, the risk score is on those with a risk score of less than 20%.
Perceptions		Descriptions of attitudes and opinions pertaining to the case note review process	

	Benefits	Descriptions of the perceived benefit of case note reviews	Ultimately, I think it will benefit patients. I do think it will benefit patients and clinical teams. And I think if I say that there'd be a cultural benefit, so the wider ramifications will be potentially immense. I think the local changes in the health services locally over the last 25 and 30 years have been divisive.
	Duty	Description of the duty/mandate to undertake case note review	everyone needs to do reviews of death. It's an extra duty that they have to do.
	Privy to consultants	Descriptions of case note reviews as exclusive or involving consultants	At the moment it is very much about getting the consultants to do it. We need to get senior nurses or … it's very much on the consultant's heads. If we can get the senior matrons involved in these reviews would be useful. I'd like to say that it does, but I don't think it reaches its potential to do that. I still think of a person working in siloes. I still think about case note review going about on Wednesday afternoon but where are the nurses in that? I think it's done in siloes.

	Routinisation	Description of the perceived routinisation of case note review as a quotidian tool	Nearly all of my time as [X position], we've done case note review. I think it is well embedded within the Trust.
	Specialty-specific	Description of the specialty-specific process or learning from case note review	What can you standardise and what we already do in the Trust is where we fail on the governance side… it is to identify therapeutic errors, identify delays to treatments… identify any escalation of concerns to the nursing staff, alright… I can't remember what the RCP guidelines were… I did comes across it once at some point and I felt it was OK. I'm sure much of it is covered within the RCP… it's a tricky one… it has to be looked at specialty by specialty really.
	Variability and bias	Descriptions that pertain variability or bias of case note reviewing and its subsequent learning	I think it is for instance, incident reporting is a tool that varies greatly from individual to individual and it is a form process and not everybody follows that form process either. Some boxes will be filled… some won't. Every incident is subjective on two counts. 'How did it impact on me as an individual?' 'God, it was awful, I had such a horrible day!' But the impact on the patient is relatively low but you may have upgraded it because of your perception.

			I think personally it's about personality. I think those who don't like it don't have the right personality. You could ask to do a case note review and because they couldn't do it, they would do a sloppy job and they assume that those people who do do the case note reviews… I'm busy doing case note reviews are actually just doing very little.

Appendix 27: Triangulation Protocol and Consensus Assessment Procedure

Triangulation protocol

Step	**Activity**
Sorting	Sort findings from each data source or method into similarly categorized segments that address the research question(s) of interest to determine the areas of overlapping content or divergence
Convergence coding	Identify themes from each data source. Compare the findings to determine the degree of convergence of a) essence of the meaning and prominence of the themes presented and b) provincial coverage and specific examples provided in relation to each theme. Convergence coding scheme Agreement – there is full agreement between the sets of results on the elements of comparison (e.g., meaning and the prominence are the same, provincial coverage and specific examples are the same). Partial agreement – there is agreement on one but not both components (e.g., the meaning or prominence of themes is the same, provincial coverage or specific examples provided are the same.) Silence – one set of results covers the theme or example, whereas the other set of results is silent on the theme or example.

	Dissonance – There is disagreement between the sets of results on both elements of comparison (e.g., meaning and prominence are different; provincial coverage and specific examples are different).
Convergence assessment	Reviewing all compared segments to provide a global assessment of convergence levels. Indicate when and where researchers have different perspectives on convergence or dissonance findings.
Completeness assessment	Compare the nature and scope of the unique topic areas for each data source or method to enhance the completeness of the united set of findings and identify key differences in scope/or coverage.
Researcher comparison	Compare the assessments of convergence or dissonance and completeness of the united set of findings among multiple researchers to a) clarify interpretations of the findings and b) determine degree of agreement among researchers on triangulated findings. Plan for how disagreements will be handled and how final decisions on interpretations will be made.
Feedback	Feedback of triangulated results to research team and/or stakeholders for review and clarification.

Convergence assessment procedure

Researcher 1		**Researcher 2**		**VERDICT**
Themes	Assessment - Rationale		Assessment - Rationale	0 = no agreement 1 = agreement -1 = disagreement
Barriers/Challenges/Persisting Issues	Agreement - There is strong agreement from the barriers (interviews) and the challenges (documents).	Barriers	**1** Agreement - examples in the documents though when it is mentioned they are consistent with the findings from the interviews	1
End of Life	Agreement - There is a strong agreement across both sources for the End-of-Life aspect. It is one of key foci of case-note reviews.	N/A	N/A	0

Feedback	Agreement - There is strong agreement that feedback is offered from the case-note reviews as indicated from the sources.	Benefits	**1** Agreement - consistency between documents about the benefit of system changes and for future patients	1
Ostensible Learning	Agreement - There is strong agreement for learning generated from the reviews.	Learning	**1** Agreement - positive applications of case-note review as a learning opportunity	1
Other QI Processes	Agreement - There was strong agreement that there were other processes for monitoring and improving care quality.	Other QI Processes	**1** - Agreement	1
Preventable death/Quality improvement examples	Agreement - Quality improvement examples and	Examples of Quality Improvement	**1** Agreement- across sources quality improvement highlighted i.e., use case-note	1

	preventable death is found across both sources.		review findings to identify areas of improvement **1** Agreement - positive examples of translating the learning from case-note reviews into changes	
Teamwork	Agreement - Teamwork with others is contained in both sources.	Communication	Disagreement - interviews suggest that learning and findings are not made accessible to staff and a desire to communicate with relevant families by contrast the documents suggest it is disseminated to staff and there is no mention of communicating; this undermines teamwork with families	-1
Benefits	Partial - Agreement There is a clear use for reviewing from the interviews.	Benefits	**1** Agreement - consistency between documents about the benefit of system changes and	1

	However, there is only a generic normative comment on the use of reviewing (i.e., there is no tangible example given)			
Perceptions	Silence - There is emphasis in the interview but silence from the documents.	Perceptions	Silence - Documents give no evidence of perceptions around reviewing;	1
Case-note review Process or criteria	Silence - There is emphasis in the interview for the case-note review process but there is silence from the documents.	N/A	N/A	0
Culture	Silence - There is emphasis in the interview for the culture of reviewing but	Culture	**2** Partial Agreement - in small sample I have the examples provided are different though generally in agreement with each other,	1 – due to sample discrepancy

	there is only silence from the documents.		focusing on different aspects. Be easier for you to draw a judgement on this one across the data	
Duty	Silence - There is emphasis in the interview for it being a duty (or that be it ought to be done) but there is only silence from the documents.	Desired Change		1
Facilitators or Modulators	Silence - The interviews clearly outline the external drivers or influencers of reviewing. However, this is absent from the documents. It is unclear what influences or drives case-note review from the documents.	N/A	N/A	0

Human Resource	Agreement - There is overlap between the human resource element from the interviews and the documents. The documents exhibit the importance of the team, specialties, however, not to the extent that the interviews offered. Undertaking case-note review is a significant human resource effort!	Training	**2** Partial Agreement - training not undertaken in sample interviews, documents acknowledge that it is being rolled out but that only a limited number have been formally trained (will depend on when this is supposed to start) Board consider training important, from interviews, less so, more non-chalant and more competent.	1
Reflexivity	Silence - The influence of the researcher has affected interview participant perceptions (and possibly their actions). Though the	N/A	N/A	0

	researcher presence was announced on 21st March 2018 at the Mortality Surveillance Meeting, there was no influence. The researcher has yet to attend an MSC meeting.			
Scrutiny	Agreement - Looking for synonyms of 'discuss', it is clear that interviewees and documents suggest scrutiny is present during case-note reviews and/or investigations of care quality.	N/A	N/A	0

Organisational	1 Agreement - Factors beyond the clinical play a role but examples given vary in their presentation Documents indicate external and internal pressures.	Organisational	1 - Agreement	1

Appendix 28: Case-note review embedding using NPT

Embedding (NPT) Construct	**Description**	**Illustrative quote(s)**	**Illustrative quote(s)**	**Findings**
Coherence	How is the work conceptualised by participants?	"It's an extra duty that they have to do. I think it's done to fit in around some of their other priorities. So, if they've got a bit of time at lunchtime or a bit of free time at the end of the day, they might do some. It certainly doesn't get done when you're allocated to, you know, theatre, clinic, ward." (D4, Consultant Orthopaedic Surgeon)	"It's an extra duty that they have to do. I think it's done to fit in around some of their other priorities. So, if they've got a bit of time at lunchtime or a bit of free time at the end of the day, they might do some. It certainly doesn't get done when you're allocated to, you know, theatre, clinic, ward." "they look upon reviews as a professional duty."	Most interviewees had a coherent idea and purpose for all forms of case-note review establishing that comparably little quality improvement was generated for the time and effort invested.

		"they look upon reviews as a professional duty." (D4, Consultant Orthopaedic Surgeon) "It's supposed to be learning from mistakes, mortalities (…) if there's anything we can do to improve patient care (…) a lot of the time… there is nothing. But the patient has still died as the patient has multiple co-morbidities." (D4, Consultant Orthopaedic Surgeon) "I think the process of shadowing would be good. Say for example, you're going to do a case note review. And you've got someone else who can take on the role in future… Going through case note review is a	"It's supposed to be learning from mistakes, mortalities (…) if there's anything we can do to improve patient care (…) a lot of the time… there is nothing. But the patient has still died as the patient has multiple co-morbidities." "I think the process of shadowing would be good. Say for example, you're going to do a case note review. And you've got someone else who can take on the role in future… Going through case note review is a good way of showing them what good quality is. What you don't want to do is have	

		good way of showing them what good quality is. What you don't want to do is have someone panic and miss the actual point of a case note review." (D12, Consultant Gastrointestinal Surgeon) "I think what we've over time really is that we've tried to focus on not only was it unavoidable or death or are there any lapses in the patient which could have contributed to the harm of the patient… from a negative point of view, the quantitative point of view… we kind of get them to look at the qualitative data. Was it a good death? Was it avoidable? So was the care of the relative, was the	someone panic and miss the actual point of a case note review." "I think what we've over time really is that we've tried to focus on not only was it unavoidable or death or are there any lapses in the patient which could have contributed to the harm of the patient… from a negative point of view, the quantitative point of view… we kind of get them to look at the qualitative data. Was it a good death? Was it avoidable? So was the care of the relative, was the care of the patient in line with best practice around the care of the dying as well as, you know, did we	

		care of the patient in line with best practice around the care of the dying as well as, you know, did we manage sepsis or did we treat the diagnosis appropriately. So, we've kind of changed slightly to a more qualitative practise for some of the reviews." (N4, Head Nurse, Board Member) "When I was initially involved or aware of case note review, it was… the purpose was to identify if there were any lapses or any issues with the care… any incidents that occurred which could have contributed to the death and to identify whether the death was… I don't like the term avoidable but that often is the term	manage sepsis or did we treat the diagnosis appropriately. So, we've kind of changed slightly to a more qualitative practise for some of the reviews."	

		that's used. As the same way as you do root cause analysis of when there's an incident, we can learn from the death in terms of any failings in practice." (N4, Head Nurse, Board Member)		
Collective Action	How does the work get done? How are activities structured, and how are they constrained?	"So, we pulled together a deteriorating patient group and we used all that mortality data to drive those changes, to drive that QI across the organisation, in particular around those topics (pneumonia, heart failure, sepsis)." (D12, Consultant Gastrointestinal Surgeon) "When somebody dies their documentation goes to mortuary then to the bereavement office, then to	"So, we pulled together a deteriorating patient group and we used all that mortality data to drive those changes, to drive that QI across the organisation, in particular around those topics (pneumonia, heart failure, sepsis)." "That (time to [mortality] case-note review compiling for review) process takes around eight weeks in total so by the time the deaths	All forms of case-note reviews were delayed by necessary processing (8 weeks) with information requiring creative application to deliver QI.

		coding, and after it's been coded, it goes to the kind of mortality surveillance committee administration, then they get forwarded in a list to each department. That process takes around eight weeks in total so by the time the deaths come to us normally we are in the middle of the months and it's the next month which gives you the two months delay in the review." (D5, Consultant Acute/General Medicine Doctor) "Mortality has its own page there. That's how. It's got a dashboard where we record the incidents. Falls. VTE. Safety thermometer. Then it	comes to us normally we are in the middle of the months and it's the next month which gives you the two months' delay in the review." "What I do for case note review in our department is, all deaths, all the notes come to me, I conduct a review, structured note review, based on my experience of doing note reviews from training but also structured round the global trigger tool, leaders in patient safety, and I go through the notes."	

		drills down into a lot more detail. And there is also measured staffing. And this is our deteriorating patient's group with community acquired pneumonia and AKI care bundles." (SO2, Quality Project Manager) "That (time to [mortality] case-note review compiling for review) process takes around eight weeks in total so by the time the deaths come to us normally we are in the middle of the months and it's the next month which gives you the two months' delay in the review." (D5, Consultant Acute/General Medicine Doctor) "What I do for case note review in our department is, all deaths, all the		

		notes come to me, I conduct a review, structured note review, based on my experience of doing note reviews from training but also structured round the global trigger tool, leaders in patient safety, and I go through the notes." (D7, Consultant Gastroenterologist Surgeon)		
Reflexive Monitoring	How are new ways of working – or attempts to introduce them — interpreted by those involved? How do they evaluate new working practices?	"I think it is (in response to management of review load vs. clinical load). We used to do it when we had time…. There is always something else which became a priority… Time is made and so I think it is really really important that time is set every week rather than or every fortnight than when we do it	"I think it is (in response to management of review load vs. clinical load). We used to do it when we had time…. There is always something else which became a priority… Time is made and so I think it is really really important that time is set every week rather than or every fortnight	All forms of case-note review were conceptually received well; yet, the pragmatics of clinical workload often overwhelmed the desire to do case-note review with those most concerned about reviewing most

		when we have time because we've always got something which becomes more important." (N1, Staff Nurse Reviewer)	than when we do it when we have time because we've always got something which becomes more important."	likely to use it to bring about any possible QI.
Cognitive Participation	Who does the work? How do they decide whether to engage, and what do they hope to achieve?	"deaths are to be reviewed so they will be done by a multi-disciplinary team approach." (B3, Board member and ex-nurse reviewer) "I think, personally, that there is a group of clinicians that are quite energised about mortality reviews. If you take an elective death, it's obviously an unexpected death, they're very keen to review that." (SO1, Patient Safety Officer) "I think everyone, I would hope, understand the (care) process and the	"deaths are to be reviewed so they will be done by a multi-disciplinary team approach." "I think, personally, that there is a group of clinicians that are quite energised about mortality reviews. If you take an elective death, it's obviously an unexpected death, they're very keen to review that." "I think everyone, I would hope, understand the (care) process and the reasons behind it."	All forms of case-note reviews were undertaken with multi-professional input.

		reasons behind it." (D4, Consultant Orthopaedic Surgeon)		
Organisational	The work that defines and organizes itself by units greater than individuals (i.e., team/specialty/other)	"With the case note review, the reason why we started to do the case note reviews here in the Trust as part of case note review was learn from similar problems and about 3 or 4 years ago, I was asked as… before I took up the trialling for the whole hospital, I was responsible for anaesthetic and ITU trainees. Any trainees involved in a critical incident, I had to go with them. The initial management reviews. IMR. But what I found out in that was that there was a common thing running.	"With the case note review, the reason why we started to do the case note reviews here in the Trust as part of case note review was learn from similar problems and about 3 or 4 years ago, I was asked as… before I took up the trailing for the whole hospital, I was responsible for anaesthetic and ITU trainees. Any trainees involved in a critical incident, I had to go with them. The initial management reviews. IMR. But what I found out in that was that there was a	All forms of case-note review were favoured by certain specialties, but not all.

		Communication problem, or lack of knowledge about what to do. Or lack of escalation. We had multiple incidents where patients had poor outcome because they were not escalated to an appropriate… it was a junior doctor talking to another junior doctor when it could have easily been sorted if the junior doctor had spoken to a consultant or a registrar. Someone who probably had a little more insight. So, as part of that, we started the review, the questions we were asking was whether this patient was an appropriate admission to ITU." (D9, Consultant Anaesthetist & Intensivist)	common thing running. Communication problem, or lack of knowledge about what to do. Or lack of escalation. We had multiple incidents where patients had poor outcome because they were not escalated to an appropriate… it was a junior doctor talking to another junior doctor when it could have easily been sorted if the junior doctor had spoken to a consultant or a registrar. Someone who probably had a little more insight. So, as part of that, we started the review, the questions we were asking was whether this patient	

		“Deaths are much more likely to be unexpected, you know, it’s not the usual outcome. So, surgeons have always been interested in their deaths. If you look at expected deaths in elderly people, then that is pushing water uphill, because there will be a view that essentially it will be a chore in a lot of cases because, erm…, we’ve been doing it a long time and the learning, new learning, from the cases is rare. So, if you’re not gaining anything from doing something it becomes a clerical chore.” (D1B2, Senior Clinician and Board Member)	was an appropriate admission to ITU.” “I can only comment on the ED [emergency department] side of it. The juniors I think do a very efficient thorough job of investigating a lot of cases that we learn nothing form in terms of the amount of depth that they go in for out of hospitals or cardiac arrests. I would say it’s not particularly useful but actually they do do a very good job and they write lots of stuff about all of these patients who came in and probably had a cardiac arrest. I don’t think the group of patients’ we are looking at is useful	

		"So specifically in terms of the role of those within the organisation, that is to look at... take an objective as to the circumstances and the journey that led up to the patient dying. And when I say objective, I mean to say with a fresh pair of eyes. So, for me there is something about having a mixed team that may review a set of notes. So, we may have a clinical team that is involved because they will have some insight into the decision-making that they took around that patient's journey." (SO2, Quality Project Manager) "I can only comment on the ED [emergency department] side of it.	for ED. I'm not saying the whole thing isn't effective. The way it's done here and the group of patients we focus on in ED is not useful." "I cannot think of any clinical teams that have tried to block any changes (with information derived from MCNR). I think the biggest issue that we have is ensuring that the appropriate number of MCNR get done because people are busy."	

		The juniors I think do a very efficient thorough job of investigating a lot of cases that we learn nothing form in terms of the amount of depth that they go in for out of hospitals or cardiac arrests. I would say it's not particularly useful but actually they do do a very good job and they write lots of stuff about all of these patients who came in and probably had a cardiac arrest. I don't think the group of patients' we are looking at is useful for ED. I'm not saying the whole thing isn't effective. The way it's done here and the group of patients we focus on in ED is not		

		useful.” (D2, Consultant Emergency Medicine Doctor) “I cannot think of any clinical teams that have tried to block any changes (with information derived from MCNR). I think the biggest issue that we have is ensuring that the appropriate number of MCNR get done because people are busy.” (B3, Board member and ex-nurse reviewer)		

Appendix 29. Barriers and facilitator factors to case-note review information flow

FACILITATORS **Systems** **Themes**	Descriptions	Illustrative quotes	Findings

Guidelines	Examples of guidelines, recommendations or policy which have helped to promote case-note review activity	"There is a Trust policy which is available to the public and everybody else on the public facing page of the Trust, internet page. That sets how we will address and look at mortality across the organisation." (SO2, Quality Safety Manager) "The second things about the mortality review process is that we have changed the way… for instance with the mental capacity assessment (MCA) with patients who have come in with poor mental capacity and we have not been able to prevent it. So, we have incorporated it into our admission sheet. So, every patient that comes into the ITU now gets an MCA done." (D9, Consultant Intensivist) "The SI process is that we've got a policy and we apply it. And it's very robust." (N4, Senior Nurse and Board Member) "I know the surgeons use the format their colleges recommend. They don't have many deaths in surgery, thank goodness, but they do a review and a presentation. There'll be lessons learnt	There are a range of policies and tools within and external to the Trust that help legitimate and promote case-note review.

		and any comments made." (D7, Consultant Gastroenterologist and Expert reviewer) "And there is a Trust policy which is available to the public and everybody else on the public facing page of the Trust, internet page" (SO2, Senior Quality Manager) "Processes are fairly standardised within the Trust. There is mortality review policy which the Trust does follow developed by the mortality review group." (D10B4, Consultant Laparoscopic General & Colorectal Surgeon)	
Format/tools	Examples are given of the tools/forms which help promote case-note review activity	"Well, it's (Royal College of Physician's structured implicit review) obviously quicker and more specific…the one stop (Royal College of Physicians) form is much better." (N6, Advanced Nurse Specialist in General Medicine, Case-note reviewer) "A global trigger tool, you look at certain things in their pathway. Have they had a blood transfusion? … That should	

		be a trigger in your review." (D4, Consultant Orthopaedic Surgeon) "There are a number of streams of work which are happening in this place at the moment. If we start with mortality reviews, per se, for the adult in-patient group and the child in-patient group. The deaths are recorded within the bereavement office and the information team, and the clinical coding team review those notes in terms of coding. Then the quality team then get a list of those deaths from the information team and the associate medical director (AMD) for governance distributes those deaths out across the organisation for peer review." (SO2, Senior Quality Manager) "The tool, the PRISM tool does guide you to the areas where it's known that things can go wrong to contribute to death. It's kind of built on the global trigger tool." "We look through the notes together rather than it be one of us. We then go through it on the online form. Rather than do it	

		direct online... we print the forms off beforehand. We do it by paper. After it's all over one of us writes it up." (D1B2, Senior Clinician and Board Member) "From my knowledge, I haven't been trained in it so… so erm… my knowledge from people in my team that have gone through the training, it's quite an unwieldy (Trust mortality review proforma) tool." (N4, Senior Nurse and Board Member) "I was just going to mention the RCP tool that it gives you the opportunity to talk about what has gone well in the patient. Care as well which is again a much more positive way of looking at the review. So, I think clinicians will value that aspect to it as well." (SO3, Senior Clinician and Board Member) "What I do for case-note review in our department is, all deaths, all the notes come to me, I conduct a review, structured note review, based on my experience of doing note reviews	

		from training but also structured round the global trigger tool" (D7, Consultant Gastroenterologist and Expert reviewer) "He sent the paper; he didn't present it. And I think… Mortality surveillance committee didn't really, erm… I think D5 had a rather negative view on the medical examiner for death's model in that he thought it would be very time consuming and IT consuming. I think that erm… it depends on how thorough that initial review is. And I think that what D5 is proposing is that. D5's recommendation was that 'it's all too onerous really.' But I haven't said that." (D1B2, Senior Clinician and Board Member)	
IT & Administration	Examples where IT or administration help promote case-note review activity	"The deaths are recorded within the bereavement office and the information team, and the clinical coding team review those notes in terms of coding…. It would come to the patient safety team." "Have you seen our patient safety dashboard which comes from deteriorating patient safety? So, every month, they	The administrative and information staff help to facilitate case-note reviews.

		produce a newsletter which goes to the board." (D9, Consultant Intensivist) "Or it's a near miss or whatever but a further review is undertaken and lots of learning comes out, that learning then goes into the patient safety report and in the patient safety newsletter which goes across the whole organisation." (SO2, Patient Safety Officer) "When somebody dies, their documentation goes to mortuary then to the bereavement office, then to coding, and after it's been coded, it goes to the kind of mortality surveillance committee administration, then they get forwarded in a list to each department." (D5, Consultant Acute Medicine Doctor) "When they're scanned, they're not in any sensible order at all. And it is a nightmare, so we hate doing the one's which are on there. Ideally, we try and get them on paper copies if we can." (N1, Nurse Reviewer)	

Existing structures	Example where there is an existing structures or policy that helps to establish and promote reviews as 'normative' practise	"We take the view that it (case-note review) is a collective responsibility. The collection of the information by the juniors… the junior sit with the consultant… they go through the case review and make sure that it's appropriately documented in the PowerPoint. The PowerPoint presented in our M&M is held once a month. It is held on the first Friday of every month." (D10B4, Consultant Laparoscopic General & Colorectal Surgeon) "You have to (national) audit every admission with a definite diagnosis of heart failure, and we've definitely done that year on year shows that the mortality is half, almost half from the cardiology ward compared to the mortality of patients admitted elsewhere even though the patients are sicker in cardiology." (D11, Consultant Cardiologist) …The (mortality) index is higher in cardiology… they're just sicker and that's why they end up here. That's taking into account all those confounding factors. The patients are just	Established modes of working, through collective responsibility or national audits, are conducive for helping case-note review information flow.

		sicker and that's why they end up in a level two area and we look after them. They're less sick when they're not recognised. And year on year, their overall mortality has been coming down a little" (D11, Consultant Cardiologist)	
Committee	Examples are given of committees that serve to promote case-note review activity	"We've had no alerts from the Royal College for about five years now; we know that our SHMIs and our reviews are reasonably robust across the organisation. So, that's the sort of information we get at the mortality surveillance committee." (SO2, Quality Safety Manager)	The mortality surveillance committee plays a central role in governing, monitoring, and receiving feedback from case-note review.
Individual		"The case-note review is how we've been reviewing deaths so here process has been and is those learnings go to mortality surveillance committee." (B1, Board Vice-Chairman, Non-executive director) "I would say the only information you got would the one that went downstream from mortality surveillance committee; the information goes from the department to the MSC and then to the mortality newsletter. It is not something that goes across	

		from department to department. I cannot access [as a consultant] the learning points from other departments but as governance lead I can. One of my peers will not be able to access the mortality reviews, of say, cardiology." (D5, Consultant Acute Medicine Doctor)	
Roles	Examples are given of roles that serve case-note review activity	"Interviewer: Would you say there are any sub-cultures which stand against case-note review? Interviewee: No, I think its part and parcel of our role. It's embedded in our role." (N6, Advanced Nurse Specialist in General Medicine, Case-note reviewer) "So, my role within South Warwickshire is to deal with all the incidents which come into the Trust. We hold initial management review meetings in terms of any incidents that have triggered a moderate, severe, or death for any learning around them." (SO1, Patient Safety Officer)	There are clear roles for reviewers and review-supporting work.
BARRIERS			

Themes	Descriptions	Illustrative quotes	Findings
Systems			
Information technology (IT)	Examples where IT is mentioned in relation to case-note review, particularly to answer the tripartite research questions.	"Unfortunately, we don't have great IT here and lots of the systems to get the patient information are two different systems. So that is the most time-consuming bit. Each of the systems doesn't speak to each other so you're going in to get the patient data, but the one stop form is much better. Yeah." (N6, Advanced Nurse Specialist in General Medicine, Case-note reviewer) "Can you trace a learning opportunity from a particular case-note review back into practise? That's more difficult because you might have changed the practise, you might have had a technology problem, you might have hadn't done the training." (B1, Board Vice-Chairman, Non-executive director)	The degree of optimisation across software packages with the competence of staff in both using and comprehending information plays a role in case-note review information flow.
Patient safety interventions (PSI)	Examples where there are other interventions and options which	"I could do a Datix or clinical incident form (other QI methods) to say you haven't treated this renal failure or I could	Multiple interventions complicate and lead to

	actively compete or overlap with case-note reviews	do a note review of a patient who's died who's maybe had a problem and I could have done that note review six weeks down the line… and it would have the same outcome" (D4, Consultant Orthopaedic Surgeon)	overlap and redundancies in the information flow.
Scarce learning opportunities	Examples where there is little learning from case-note reviews	"Perhaps with geriatricians they have so many that die there doesn't seem to be that many lessons from the case. Maybe they get overwhelmed by the (large number of) case-note reviews." (D8, Academic Specialty Registrar) "It's supposed to be learning from mistakes, mortalities… if there's anything we can do to improve patient care… a lot of the time… there is nothing." (D4, Consultant Orthopaedic Surgeon)	The low preponderance of learning from the case-notes hinders the optimism with which case-note reviews are handled as 'possible' sources for learning.
Culture **Administration**	Examples of poor administration in relation to case-note review	"Often, they're chasing these notes and all of that in itself is often quite time-consuming and frustrating. And that definitely, I know is the feedback that I have. We don't have	The poor administration and handling of reviewed case-notes directly stymie

		the notes, or we can't find the notes." (SO1, Patient Safety Officer)	case-note review information flow.
Specialty silos	Examples where specialties are implicitly or explicitly inhibited from communicating helpful case-note review information with one another	"I think in each specialty it is slightly different. There's no consistency in how they do their reviews which in a way is good because they do what fits for them, for trauma and orthopaedics and endoscopy or A&E." (SO1, Patient Safety Officer) "When somebody dies, their documentation goes to mortuary then to the bereavement office, then to coding, and after it's been coded, it goes to the kind of mortality surveillance committee administration, then they get forwarded in a list to each department. That process takes around eight weeks in total so by the time the deaths come to us normally we are in the middle of the months and it's the next month which gives you the two months delay in the review" (D5, Consultant Acute Medicine Doctor)	The heterogeneity of specialty approaches to case-note review appears to hinder the case-note review information flow.

Individual **Lack of time**	Example where there is a mention of time taken to undertake any aspect of case-note review	"time-consuming and cognitively demanding and can compete with existing... I think it is important, but I think when you have a certain amount of time to deliver, you tend to deliver to the living rather than the dead." (D11, Consultant Cardiologist) "If it's one or two, then that's probably true. But what tends to happen is that people get keen on it, and they spend quite of a lot of time on there." (D1B2, Senior Clinician and Board Member)	The very time taken up by reviewing case-notes was a slight impediment to the case-note review information flow.

Appendix 30. Facilitator and barrier factors to QI from case-note reviews (topic-specific and mandatory MCNR)

FACILITATOR FACTORS	**Descriptions**	**Illustrative quotes**	**Summary**
Systems			
Estimated quality improvement from case-note reviews		"Gut feel out of 100 would be less than 10 that were avoidable. Less than 10, less than 5, I'd say. Learning. From purely on the basis of a medical review which is the model we have now. Knowing my view that it should be multidisciplinary, it should be more qualitative. The learning at the moment probably… well… the learning is the same theme. Like you said earlier about saturation point, we've got to point where we know that inappropriate… well, patients are in hospital, potentially inappropriately, at the end of their life." (N4, Senior Nurse and Board Member)	There is substantially more estimated learning from all forms of case-note than those recorded by the Trust to derive quality improvement.

National Care Quality Board Mandate	Examples of how the national care quality board mandate to review and learn from case-notes are mentioned	"For the last two years I have been on a secondment and one of my jobs was to look at the learning from deaths review document and be involved in writing the mortality review policy in the Trust and visiting all the mortality reviews as they were and looking at how we go forward and actually undertake the training from learning from deaths reviews and learning disability death reviews. That's basically my background." (N3, Nurse case-note reviewer) "There is mortality review policy which the Trust does follow developed by the mortality review group." (D10B4, Laparotomy Consultant Surgeon and Board member) "Currently, we have a mortality policy. That mortality review policy is in the hands of the board. It is public as the national guidance requirements. The board has a dashboard with the number of deaths, avoidability of	The learning from deaths policy has had wide influence and endorsement across all forms of case-note review.

		deaths and as well, any incidents or complaints that may happen linked to a death in the Trust. We have a mortality surveillance committee that meets monthly, which I'm a member of. And they review all the deaths, all the SHMI, RAMI and all the different standards and statistics that are reported nationally."(D5, Consultant General Medicine Doctor)	
Case-note reviews facilitated through Medical Examiners model	Examples of external systems, procedures, and policies such as the 'UK national medical examiner model' which supports the undertaking of learning from case-note review	"Yeah, the thing is because we haven't yet adopted the medical examiner model, this is like a preparation for it. The pilot prepares the possibility of including a medical examiner model in future if that was the way forward and it's using that model to phase the screening." (D5, Consultant General Medicine Doctor) "No. I can't say much. What leads to changes are national drivers. So, with the medical examiner coming up… that's the thing driving the process at the moment. I think	The medical examiner model had helped review staff to co-ordinate their efforts to a government-led initiative indirectly driving forward quality improvement from all forms of case-note review.

		they feel re-assured at the moment." (S1, Healthcare Governance Support Officer)	
Patient Liaison Service (PALS)	Examples from the formal culture that contribute to the QI from case-note review.	"If it doesn't get reported on Datix or if it's… do you remember when I mentioned about deaths that happened outside the hospital, so, there was a case for the death which happened four hours after discharge. So that was picked up after PALS/bereavement, the family have come back to PALS and there was an issue. And we pick up things from PALS, bereavement team as well. Often, we don't know about deaths which have happened outside but it's only when they've contacted PALS to say actually my dad died two hours after discharge and we'll do a review and it comes from that way as well." (SO1, Patient Safety Officer) "We do the patients that have died on our ward. Many of them will be from our specialty. Some won't be. It's the	PALS was found to identify and voice the concerns of the patient's family to the specialty in question and/or the senior clinical management.

		mortality within our department and patients that are under our care. But we would feedback to other specialties if we were involved in that patient's pathway." (D4, Consultant Orthopaedic Surgeon)	
Continued professional development (CPD)/validation requiring case-note reviews	Examples of CPD requiring case-note reviews	"Case-note review timetabled in anybody's job plan, the answer is that anybody has at least 4 hours a week for audit, professional development re-validation and, you know, learning from patients under your care who have died is surely part of that. We allow flexibility… [we] timetable it within people's SPA [supporting professional activities] time." (D1B2, Senior Clinician and Board Member) "They'll be aware of mortality reviews and how it fits into their approach. Appraisals and re-validations and things like that." (SO3, Clinician and board member) "There is a certain amount of time allowed in a job plan for revalidation activities which include some mortality	Case-note reviews were integral to CPD and the revalidation process.

		review for appropriate specialties." (SO3, Clinician and board member)	
National programmes	Examples of national programmes facilitating case-note reviews	"So, we pulled together a deteriorating patient group and we used all that mortality data to drive those changes, to drive that quality improvement across the organisation, in particular around those topics. We've seen a particular drop in sepsis. Pneumonia is a difficult patient group to manage within the organisation, in terms of…" (D4, Consultant Orthopaedic Surgeon) "Well, if you take the surgical specialties if its relatively low volume and they meet together the benefits are obvious because you've got a peer review of practise and erm. you know, just doing the review together, you're sharing practise and educating. The other reviews, have less benefit to the individuals doing them but we do get lessons to come out of it and they come through the deteriorating patient group around sepsis, pneumonia,	National programmes such as the Surviving Sepsis Campaign (SSC) and other specialty-specific programmes (i.e., national emergency laparotomy audit (NELA) and non-specific mortality and morbidity meetings) have, unintentionally or not, helped to facilitate quality improvement from case-note reviews.

		there have been numerous themes that have gone out, resulted in care bundles etc and closed the loop. Are we still? Are we fading out anything new? Not so much, that's again why you may have found some apathy out there because we're not finding so many new things… we do like see any sort of compliance to slide in some areas." (D1B2, Senior Clinician and Board Member) "Early on, we had quite a lot of learning around sepsis and acute kidney injuries, and we've done a lot of work with the deteriorating patient and sepsis since and that's tailed off. We're not seeing those cases any longer. And there's the odd surgical patient; unexpected learning patient where there is real learning from." (N3, Nurse case-note reviewer) "We take the view that it [case-note review] is a collective responsibility. The collection of the information by the juniors… the junior sit with the consultant… they	

		go through the case review and make sure that it's appropriately documented in the PowerPoint. The PowerPoint presented in our M&M [mortality and morbidity] is held once a month. It is held on the first Friday of every month." (D2, Consultant Emergency Medicine Doctor) "We know that patients that get admitted to cardiology are much sicker than patients who get admitted to other parts of the hospital; perhaps the clinicians think they're sicker and they have to come to cardiology. They're much much more poorly. And the patients who have the milder symptoms which sometimes don't get recognised or they have the diagnoses and that was not the reason they got admitted, if they got transferred to any other place. But the national heart failure audit, you have to audit every admission with a definite diagnosis of heart failure, and we've definitely done that year on hear shows that the	

		mortality is half, almost half from the cardiology ward compared to the mortality of patients admitted elsewhere even though the patients are more sick in cardiology." (D11, Consultant Cardiologist)	
National audit	Examples from the culture of audit that support QI through case-note review	"the national heart failure audit, you have to audit every admission with a definite diagnosis." (D11, Consultant Cardiologist) "Where we've failed to act upon NEWS scores, acting upon alerts where patient results have not been explored properly in a perhaps. I think those processes are fairly standardised within the Trust. There is mortality review policy which the Trust does follow developed by the mortality review group." (D10B4, Laparotomy Consultant Surgeon and Board member)	The audit culture lends itself to support the undertaking of case-note reviews. The philosophy of audit and case-note reviews coincide helping to promote quality improvement from all forms of case-note review.
Organizational Culture			

Specialty culture of learning from case-note review	Examples of how the specialty sub-culture influences QI from case-note reviews	"I think the trouble with learning from deaths, like the Shipman report…it doesn't mean that it fits everybody so erm… I think I will come back to what good looks like, and the important things are you have a culture where people can report incidents, they can learn from them, erm… that learning can be disseminated, and you need some sort of mortality review process that's going to give you assurance." (D1B2, Senior Clinician and Board Member) "I think those processes are fairly standardised within the Trust. There is mortality review policy which the Trust does follow developed by the mortality review group." (D10B4, Laparotomy Consultant Surgeon and Board member) "I think the current culture here is good but the consequences we've got the last 10-15 years of high expectations from patients and when you're a patient it's	Within specialties, specialty-specific considerations such as its size, average patient age, mortality rates and established customs and practices are important factors to consider when evaluating quality improvement generated from all forms of case-note reviews.

		a very different position from being here as you're working here. You were under a very different set of emotional boundaries, and it can be very frustrating. That's why people working in healthcare need to be very caring." (B1, Board Vice-Chairman, Non-executive director) "I think some clinical groups by the nature of their work have had more mortality reviews and have had patient safety ingrained in their specialty... I'm thinking of anaesthetics perhaps. General surgery because of the nature of that specialty…." (S02, Consultant Radiologist and Board member) "I think some specialties have come a bit or has come a bit later (to heavily use case-note reviews) to, but I think everyone, I would hope, understand the process and the reasons behind it." (D10B4, Laparotomy Consultant Surgeon and Board member)	

Specialty culture of reviewer ownership of patients	Examples where the ownership of the reviewing process promotes the QI generated from case-note reviews	"The clinical teams need to own the solutions to that because they'll then make it work." "I think its patient ownership. I think nephrologists really own their patients. Surgeons are not quite the same but they're quite… they've got an incentive based on the stats that get put up… they own their patients, and they are aggressive about their patients, and they want… and as a result, if someone dies, they are really keen to learn from it." (B3, Board member and ex-nurse reviewer) "What happened is that with this lady is that there is this constant attempt to communicate with this particular specialty and no senior person at the other end would take ownership of that patient's case." (D11, Consultant Cardiologist) "we've had a couple (of meetings) and it's like we don't have enough here, we can't make any decision… we're	The extent of case-note reviewer ownership of their patients is associated with how much quality improvement is obtained from all forms of case-note reviews.

		non-quorate…" (D5, Consultant General Medicine Doctor)	
Patients' expectations driving quality improvement culture	Example of when patient expectations were perceived to influence quality improvement practise	"I think the current culture here is good but the consequences we've got the last 10-15 years of high expectations from patients and when you're a patient it's a very different position from being here as you're working here." (B1, Board Vice-Chairman, Non-executive director) "The juniors I think do a very efficient thorough job of investigating a lot of cases that we learn nothing from in terms of the amount of depth that they go in for out of hospitals or cardiac arrests. I would say it's not particularly useful but actually they do a very good job, and they write lots of stuff about all of these patients who came in and probably had a cardiac arrest. I don't think the group of patients we are looking at is useful for (our department). I'm not saying the whole thing isn't	Patient expectations were perceived to have increased over recent years which has influenced the expectations around hospital quality improvement.

		effective. The way it's done here and the group of patients we focus on in (our department) is not useful." (D11, Consultant Cardiologist)	
Individual HCP/staff factors			
Policy awareness	Examples where Trust policy around case-notes were discussed	"There is mortality review policy which the Trust does follow developed by the mortality review group." (D10B4, Laparotomy Consultant Surgeon and Board member) "Currently, we have a mortality policy. That mortality review policy is in the hands of the board. It is public as the national guidance requirements. The board has a dashboard with the number of deaths, avoidability of deaths and as well, any incidents or complaints that may happen linked to a death in the Trust." (S02, Consultant Radiologist and Board member)	There was an awareness of the mortality case-note review policy in the Trust, however, topic-specific case-note reviews also generated their own learning and in-Trust policies. This supports case-note review as a policy-making intervention.

		"it's definitely driving policy. So, we've had work streams that have come out of case-note reviews. There could be something around mis-medication. Or medication errors or inappropriate escalation of observations … we had a number of deaths that were reviewed on a couple of wards. A couple of the acute medical wards where were seeing a delay in escalation of patients who were having high NEWS scores and led us to looking at electronic observation systems and we didn't have one until that point. And that absolutely drove our commitment to drive the escalation of patient's observations." (B3, Ex-nurse and board member)	
BARRIER FACTORS	Descriptions	Illustrative quotes	Findings
Systems			
No knowledge or requirement of training	Examples where staff were either unaware or indicated	"I guess if I'd received training, I'd have known that. But I haven't received training. Never had any. The only bit I	Case-note reviewers, on the whole, had no training with the

		have received is when I talked to Dr X, and he told me about how they do things. [What did he tell you?] I didn't even know training was available on it. Is it?" "There hasn't been much formal training. There has been some more formal training on root cause analysis for instance. Many of the people who looks at the deaths where there would have been harm would have had a root cause analysis training. For example, there's been some limited structured training… I think four members of the Trust have been on the Royal College of Physicians training and then that's probably… I go to conference and get training and I go to some of the conferences around learning from deaths for instance…" (N1, Nurse Reviewer)	current case-note reviewer training which appeared to be limited to select individuals.

Competing processes	Examples of QI interventions that overlap or compete with case-note reviews	"I mean it's [case-note review] a distinct process on its own and I think it is connected to a number of the other things." (B3, Board member and ex-nurse reviewer) "I think the outcome could be the same. Say I saw a patient on the ward right now that had gone into renal failure, and no one had done anything about it. I could do a Datix or clinical incident form to say you haven't treated this renal failure or I could do a note review of a patient who's died who's maybe had a problem and I could have do that note review six weeks down the line… and it would have the same outcome that looking at that they hadn't had their renal function treated properly so the outcome might be the same that clinical incident form goes or a change of protocol but I didn't realise it hadn't happened until I did the note review. As opposed to an information change or reporting from Datix." (D4, Consultant Orthopaedic Surgeon)	There are a range of QI processes which overlap with case-note reviews. All forms of case-note reviews were variably used across specialties.

Organizational Culture			
Specialty quality improvement culture	Examples of variations within specialties around case-note review	"I think some clinical groups by the nature of their work have had more mortality reviews and have had patient safety ingrained in their specialty... I'm thinking of anaesthetics perhaps. General surgery because of the nature of that specialty…." (S02, Consultant Radiologist and Board member) "There's no consistency in how they do their reviews which in a way is good because they do what fits for them, for trauma and orthopaedics and endoscopy or A&E. A&E, when they do their deaths, they go through all their deaths and the questions they ask are slightly different to the other specialties. So that's kind of… I look at the proforma side of it whereas the specialties, they do an in-depth review with more of a discussion with their colleagues and see what their colleagues think and	All forms of case-note review find a different use and expression peculiar to each specialty.

		kind of get that peer review… which I think stays within the division." (SO1, Patient Safety Officer) "The junior *really busy* s I think do a very efficient thorough job of investigating a lot of cases that we learn nothing from in terms of the amount of depth that they go in for out of hospitals or cardiac arrests. I would say it's not particularly useful but actually they do do a very good job and they write lots of stuff about all of these patients who came in and probably had a cardiac arrest. I don't think the group of patients' we are looking at is useful for ED. I'm not saying the whole thing isn't effective. The way it's done here and the group of patients we focus on in ED is not useful." (D2, Consultant Emergency Medicine Doctor) "We've noticed quite a lot that other specialties won't bring up the subject of what their plans were… what their plans were if their heart were to stop beating. There's not	

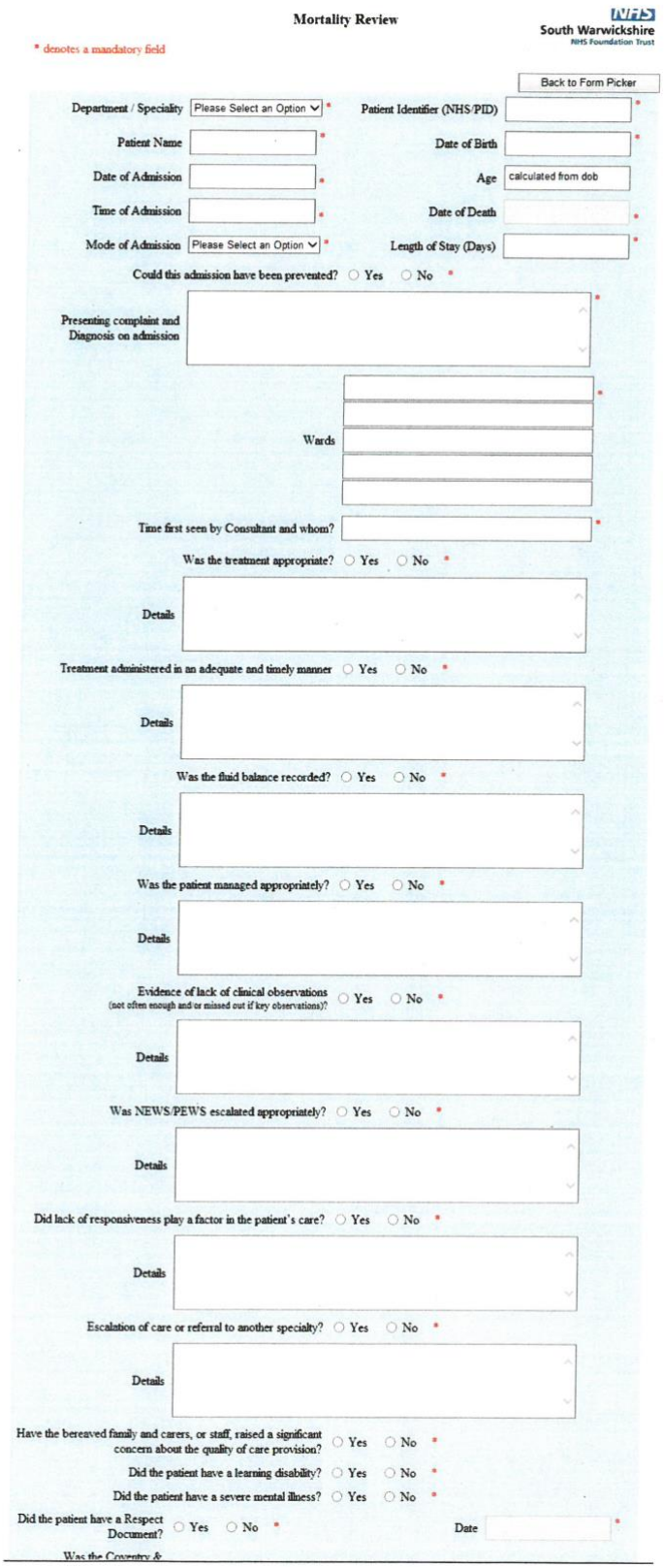
		a lot of input from other specialties. Quite often end up being an acute medical physician that brings up the subject of respect form and plans for the death. That's what we've identified mostly." (N6, Advanced Nurse Specialist in General Medicine, Case-note reviewer)	
Individual/HCP factors			
Busyness	Examples where the interviewees cite business or work overload in relation to their prioritization of case-note reviews	"I think it's not so much about incentive but that they have the time to do it. Yeah, we don't have the time within our busy schedule. We need some time to do it, that's all." (D10B4, Laparotomy Consultant Surgeon and Board member) "It is a competing demand on people who are very busy, and I think that goes throughout the organisation." (SO3, Consultant radiologist and board member) "I know everybody's really busy, and everyone's got targets and the rest of it, but we need to learn from what we do. It might make it easier to do what we do instead of	Being busy is a strong impediment to quality improvement from all forms of case-note review.

		saying ‘I haven’t got time for that.’ In fact, we have to make time and learn and make it better through that process.” (N3, Nurse reviewer)	
Attitudes toward reviewing	Examples of where the interviewee attitudes toward case-note reviews are given	“I think that the feedback you get from the, the doctor doing the high-level decision, it’s quite onerous and erm... we’re not finding anything particularly new. I don’t know if that’s the feedback that you’ve got.” (D1B2, Senior Clinician and Board Member) “It’s supposed to be learning from mistakes, mortalities… if there’s anything we can do to improve patient care… a lot of the time… there is nothing. But the patient has still died as the patient has multiple co-morbidities. It was their end-of-life event that happened to them particularly in orthopaedics where they had a neck of femur fracture	

		but sometimes, we find when they do improve on our practise." (D1B2, Senior Clinician and Board Member) "Perhaps with geriatricians they have so many that die there doesn't seem to be that many lessons from the case. Maybe they get overwhelmed by the (large number of) case-note reviews." (D8, Academic Specialty Registrar) "If you look at the medical specialties, where the majority of deaths occur, just getting, just keep in line with a proportion is difficult as it's another job and as you've probably found, you know, the value, it's not perceived to be very high value because actually a lot of these patients are expected deaths anyway. So, you know, the amount of new learning is relatively small." (D1B2, Senior Clinician and Board Member)	

Attitudes toward training being easy and not important	Examples of where the interviewees give their stance toward case-note review training	"I've read about it. I didn't attend the training… The ministers are desperate to use it as a benchmarking tool and publish it as data. And this is the problem of mortality indicators where they look at them. They're not benchmarking tools. They're smoke detector and learning tools which are not going to be reproducible." (D1B2, Senior Clinician and Board Member) "I guess you could give people training but I guess a lot of it is common sense." (N3, Nurse case-note reviewer) "there's been some limited structured training… I think four members of the Trust have been on the Royal College of Physicians training and then that's probably." (S02, Consultant Radiologist and Board member)	

Appendix 31: Mortality review e-proforma



Mortality Review

NHS South Warwickshire NHS Foundation Trust

* denotes a mandatory field

Back to Form Picker

Department / Speciality: Please Select an Option *

Patient Identifier (NHS/PID): *

Patient Name: *

Date of Birth: *

Date of Admission: *

Age: calculated from dob

Time of Admission: *

Date of Death: *

Mode of Admission: Please Select an Option *

Length of Stay (Days): *

Could this admission have been prevented? ○ Yes ○ No *

Presenting complaint and Diagnosis on admission: *

Wards: *

Time first seen by Consultant and whom? *

Was the treatment appropriate? ○ Yes ○ No *

Details

Treatment administered in an adequate and timely manner ○ Yes ○ No *

Details

Was the fluid balance recorded? ○ Yes ○ No *

Details

Was the patient managed appropriately? ○ Yes ○ No *

Details

Evidence of lack of clinical observations (not often enough and/or missed out if key observations)? ○ Yes ○ No *

Details

Was NEWS/PEWS escalated appropriately? ○ Yes ○ No *

Details

Did lack of responsiveness play a factor in the patient's care? ○ Yes ○ No *

Details

Escalation of care or referral to another specialty? ○ Yes ○ No *

Details

Have the bereaved family and carers, or staff, raised a significant concern about the quality of care provision? ○ Yes ○ No *

Did the patient have a learning disability? ○ Yes ○ No *

Did the patient have a severe mental illness? ○ Yes ○ No *

Did the patient have a Respect Document? ○ Yes ○ No *

Date: *

Was the Coventry &

Appendix 32: Information flow diagram of case-note review derived information for quality improvement purposes