Developing a pathway for remote assessment of surgical wounds with partners in low- and middle-income countries

An approach for efficient trials and resilient perioperative systems

By

James Glasbey
NIHR Doctoral Research Fellow

A thesis submitted to the University of Birmingham
for the degree of Doctor of Philosophy (PhD)

NIHR Global Health Research Unit on Global Surgery

Institute of Cancer & Genomic Sciences
University of Birmingham
July 2022

This report is independent research supported by the National Institute for Health Research through an NIHR Doctoral Fellowship (NIHR300175) and a Royal College of Surgeons (RCSEng) Research Fellowship. The views expressed in this publication are those of the author(s) and not necessarily those of the NHS, the National Institute for Health Research or the Department of Health and Social Care.

Cover art generated using an open-source Artificial Intelligence art generator available at: https://is.gd/artmachine using the following source image.

Abstract

Background: Surgical site infection (SSI) is the most common complication of abdominal surgery, and commonly occurs after hospital discharge. When patients in low- and middle-income countries (LMICs) undergo surgery, they are three times more likely to have a SSI than patients in high-income countries. Returning to hospital for routine face-to-face follow-up is the accepted gold standard for diagnosing a SSI but can be challenging in many environments, and resource intensive for healthcare teams.

Aims: The overall aim of this thesis was to develop a high-quality pathway for remote surgical wound assessment using telemedicine that can be delivered flexibly across diverse healthcare settings. First, I aimed to evaluate the feasibility and accuracy of telemedicine in the detection of SSI in existing data. Second, I aimed to explore the cross-cultural equivalence of a Wound Healing Questionnaire (WHQ) across several LMICs and make recommendations for its adaptation for use in global surgery research and practice. Third, I aimed to test the feasibility and accuracy of the adapted WHQ in diagnosis of SSI.

Methods: The primary outcome of interest in this thesis was SSI reported up to 30-days after surgery using the US Centres for Disease Control criteria. First, I compared the rates of SSI using telemedicine to those with in-person review in two data sources: (A) an international cohort study of adult patients discharged from hospital before 30-days after abdominal surgery; (B) a systematic review with meta-analysis of rates of SSI detection conducted in accordance with PRIMSA guidelines (PROSPERO:192596). Second, to recommend adaptations to the WHQ outcome measure for global implementation, I conducted a mixed-methods study across seven LMICs. Qualitative data were obtained from interviews and focus groups with local researchers with deductive coding aligned to cognitive theory. Quantitative data were collected in a prospective cohort study and Rasch analysis was used to explore measurement properties of the WHQ. I triangulated these data to make recommendations for cross-cultural and cross-language adaptation. Third, I conducted a

validation cohort study within a randomised trial (FALCON, NCT03700749) where consecutive patients undergoing abdominal surgery for a range of indications underwent telephone assessment with the WHQ (index test) up to 72-hours before their face-to-face assessment (reference test). I worked with Community Engagement and Involvement (CEI) partners to optimise the measurement pathway.

Results: The SSI rate reported using telemedicine in the cohort data was lower than with in-person follow-up (11.1% versus 13.4%, p<0.001), which persisted after risk adjustment in a mixed-effects model (adjusted odds ratio: 0.73, 95% confidence interval 0.63-0.84, p<0.001). This was consistent on meta-analysis of all existing data (odds ratio: 0.67, 0.47-0.94, I^2=0.45, p=0.12). Included studies were at a high risk of bias. This indicated the need for a novel, high-quality wound assessment tool. During WHQ adaptation, qualitative data were obtained from 10 structured interviews and 6 focus groups with 47 investigators and quantitative data from 537 patients. Triangulation provided evidence for modification of 9 items, and revision of the response structure. In the validation study, patients were included from three upper-middle (396 patients, 13 hospitals), three lower-middle (746 patients, 19 hospitals), and one low-income country (54 patients, 4 hospitals). Successful telephone contact was achieved in 90.3% (1088/1196) of patients. The WHQ discriminated patients with and without SSI (AUROC 0.869, 95% CI 0.824-0.914). An adapted WHQ cut-off point score of $\geq$4 demonstrated sensitivity of 0.701 (0.610-0.792), specificity of 0.911 (0.878-0.9430), positive predictive value of 0.723 (0.633-0.814) and negative predictive value of 0.901 (0.867-0.935).

Discussion: Current methods for remote detection of SSI are inadequate, missing 1 in 3 patients with infection. This thesis describes the adaptation and validation of the WHQ, demonstrating that a telephone pathway for wound assessment is feasible and moderately accurate. The adapted WHQ is now ready for global implementation in research and routine postoperative surveillance, using the co-designed toolkit to optimise local measurement processes.

Dedication

I dedicate this thesis to some very important people in my life without whom this work would not have been possible. More than that, they continue to inspire me to be a better me every day, and are my main source of inspiration, guidance, and purpose.

Firstly, to my fiancée Steph. Thank you for putting up with my constant stream of crazy new ideas and for always steering me in the right direction. I've loved every minute of our 10-years together and all our adventures along the way. I'm so lucky to have met someone with whom I share such a deep connection and approach to fun, life, and happiness. You are an amazing partner and mum, and I'm forever grateful for your support.

Second, to my son Rupert. You are only 1-year-old, so I assume you won't understand this for a little while! When you have put down *The Squirrels that Squabble* and are ready to read this, I do hope you enjoy it. You are the best thing that's ever happened to me, and I'm unduly proud of you every day. Hopefully some of the work in this thesis will make the world a little better for you and your friends around the world in the future.

Third, to my parents Cathy and Dave. You gave me the most incredible start in life, and taught me the value of hard work, honesty, and integrity. You never pushed me, but instead taught me that inspiration comes from within and to live without fear of the unknown. I hope I can do as good job as you both as a new father to Baby Rupert, and hope you enjoy reading this thesis. I look forward to explaining the finer details over a glass of Norfolk's finest *Ghost Ship* with you soon.

James

Acknowledgements

PhD supervisors

I have been incredibly lucky to be supported by five PhD supervisors and several other advisors and mentors during this programme. First and foremost, I would like to thank Mr Aneel Bhangu for his support and guidance throughout this three-year doctoral programme. Aneel is hugely dedicated to his trainees and has perfected the art of balancing long-term vision with making the most of short-term wins. I've ended my PhD time as a better academic, a better surgeon, and a better person than when I began, which is largely thankfully to his daily input, judgement-free guidance, and inspirational chats. It was an honour to be one of your first PhD students.

Thank you to Prof. Peter Brocklehurst for his invaluable insights on strategy and clinical trial methodology that brought quality and depth into my work and development. Thanks to Dr Jonathan Mathers for his detailed steering in qualitative elements of this thesis; it was a new way of thinking for me, but one I have thoroughly enjoyed and seen huge rewards from. I will continue to adopt mixed methods alongside trials throughout my career and thank you for the invaluable early springboard that you have given me. Many thanks also to Prof. Rupert Pearse, who gave me a totally different view on international research leadership, research quality and governance. It has been an honour to learn from you, both in terms of technical and non-technical skills and I hope to continue working together in global perioperative care long into the future.

Finally, huge thanks to Prof. Dion Morton OBE, who has been my key mentor, guide, and confidant for the past 6-years in Birmingham. Dion has invested a huge amount of time and energy in my growth, and taught me an approach to academic surgery, a passion for

research for patient benefit and a zest for work-life balance that has set me up for the rest of my life – thank you.

PhD advisors

I have benefitted hugely from specific methodological input across the diverse chapters included in this thesis. Much appreciation to Dr Anita Slade and Dr Mike Horton who helped me to design and implement my work in questionnaire adaptation and Rasch analysis, and whose patience and close guidance through these new and often complex methodologies allowed me to grow and learn as a future research leader. Thanks to Dr Rhiannon Macefield, who originally led development of the Bluebelle Wound Healing questionnaire and whose regular input and eye for quality ensured I remained aligned to original objective of the WHQ in improving wound care pathways for our patients. Many thanks to Prof Tom Pinkney who has always provided me with a sounding board for new ideas, top-class methodological expertise and helped me to navigate the clinical-academic interface during my research time. Thanks also to Dr Katie Gillies who provided steering and expertise in behavioural change frameworks in clinical trials methodology. I have learnt so much from you all and hope to continue as lifelong collaborators in these exciting research areas.

Mentors

Next my external mentors. Thanks to Prof Bruce Biccard who is one of the most accomplished, yet most humble, global research leaders I've had the pleasure of meeting. Bruce taught me so much about pragmatic research, grounded in local patient and health system priorities. I feel very lucky to have benefitted from your guidance over the past 3-years. I'm very grateful also to Prof Janet Martin, who taught me so much about how research and policy overlap, and how to close the gap between data and patient benefit. Janet is always the first to provide feedback on a paper draft, and will always do so meticulously, adding huge value and contextual depth at every stage. She also reminded me

its ok to foster a part time love of extreme sports alongside a busy research and clinical career – I hope to see you on the wake soon! Thanks also to Prof Jane Blazeby, whose sage advice and honest, holistic, and insightful mentorship meetings allowed me to reflect on what I truly want from life and my career. Having that opportunity to step back and take an external view was unique, and powerful – I really appreciate it.

International collaborators

This work, whilst presented as part of my thesis, represents the efforts of hundreds of friends and colleagues around the world. From NIHR Global Surgery Unit hub leads, local principal investigators, research team members, and community leaders – it has been a pleasure and an honour to work with you all. A full list of over 500 collaborating surgeons, anaesthetists, critical care doctors, researchers, nurses, and junior doctors that contributed to this work is listed in *Appendix H*. Without each and every one of you this work would not have been possible, and I'm hugely grateful for your efforts towards our shared goal of improving global surgical care. Thanks also to my fellow PhD students Joana Simoes, Elizabeth Li, and Dmitri Nepogodiev – it's been a blast, and thanks for being there for me through the ups and downs of academic life. It's amazing to come out of my PhD years with new friends and collaborating colleagues for life.

Funders

This work was primarily funded by the UK's National Institute of Health Research (NIHR) Doctoral Research Fellowship programme, and I was immensely humbled to receive their support. I am very proud of the incredible, supportive, and holistic clinical academic training infrastructure that the NIHR provides for a diverse and inclusive group of trainees across the UK and hope to continue my career progression in academic surgery in the postdoctoral period. My early pre-doctoral work in September 2019 was supported by a Royal College of Surgeons of England (RCS) Research Fellowship and I held an honorary fellowship in 2019-

2020. Thanks to the college for their training and guidance, and it's been an inspiration to be part of the COVID-19 recovery research group led by Prof. Peter Hutchinson. The work of the college has been transformational in improving patient safety in the UK and around the world during the pandemic. Thank you also to Sara, Kyle, and the team at *Airbnb* (San Francisco, CA) that supported some of the costs of my accommodation and travel during international site work in 2021-22 as part of the '*Live Anywhere'* programme. I have been hugely impressed with your ethos of community and adventure. Thank you for making the world feel like a smaller place - full of friends we just haven't met yet.

Patient and industry partners

Kind thanks to the patients and community members that helped shape the Wound Healing Questionnaire adaptation and optimise the delivery of the TALON-1 and TALON-2 studies. The Wound Healing Questionnaire was licensed from Oxford University Innovation (OUI) Clinical Outcomes group and used with the support of the WHQ development team at the University of Bristol.

Thesis publications

Work from this thesis has been summarised in five manuscripts:

GLASBEY J [First and corresponding author]; NIHR Global Health Research Unit on Global Surgery, *The importance of post-discharge surgical site infection surveillance: An exploration of surrogate outcome validity in an international randomised controlled trial (FALCON).* 2022. ***Under review***

GLASBEY J [First and corresponding author]; NIHR Global Health Research Unit on Global Surgery, *Global use of telemedicine for post-discharge assessment of the surgical wound: international cohort study, and systematic review with meta-analysis*. 2021. Annals of Surgery. ***Accepted ahead of print***

GLASBEY J [First and corresponding author]; NIHR Global Health Research Unit on Global Surgery, *Adaptation of the Wound Healing Questionnaire universal-reporter outcome measure for use in global surgery trials (TALON-1 study): A mixed-methods study and Rasch analysis.* 2022. ***Under review***

GLASBEY J [First and corresponding author]; NIHR Global Health Research Unit on Global Surgery. *Feasibility and diagnostic accuracy of Telephone Administration of an adapted wound heaLing QuestiONnaire for assessment for surgical site infection following abdominal surgery in low and middle-income countries (TALON): protocol for a study within a trial (SWAT).* Trials. 2021 Jul 21;22(1):471. PMID: 34289893

GLASBEY J [First and corresponding author]; NIHR Global Health Research Unit on Global Surgery, *Feasibility and diagnostic accuracy of a telephone Wound Healing*

Questionnaire in detection of surgical site infection following abdominal surgery: A study within a trial in seven low and middle-income countries. 2022. ***Under review***

Other publications

Publications unrelated to this thesis that were published between September 2019 and July 2022 with *Glasbey J* as first, corresponding or senior author are listed below:

1. **GLASBEY JC [First & corresponding author];** COVIDSurg Collaborative. Effect of COVID-19 pandemic lockdowns on planned cancer surgery for 15 tumour types in 61 countries: an international, prospective, cohort study. Lancet Oncol. 2021 Nov;22(11):1507-1517. Epub 2021 Oct 5. PMID: 34624250
2. **GLASBEY JC [First author];** COVIDSurg Collaborative. Elective Cancer Surgery in COVID-19-Free Surgical Pathways During the SARS-CoV-2 Pandemic: An International, Multicenter, Comparative Cohort Study. J Clin Oncol. 2021 Jan 1;39(1):66-78Epub 2020 Oct 6. PMID: 33021869.
3. **GLASBEY JC [Joint first author];** COVIDSurg Collaborative. Mortality and pulmonary complications in patients undergoing surgery with perioperative SARS-CoV-2 infection: an international cohort study. Lancet. 2020 Jul 4;396(10243):27-38Epub 2020 May 29. PMID: 32479829
4. **GLASBEY JC [First & corresponding author];** COVIDSurg Collaborative. Delaying surgery for patients with a previous SARS- CoV-2 infection. Br J Surg. 2020 Nov;107(12):e601-e602. Epub 2020 Sep 25. PMID: 32974904.
5. **GLASBEY JC [Joint first author];** COVIDSurg Collaborative. Preoperative nasopharyngeal swab testing and postoperative pulmonary complications in patients undergoing elective surgery during the SARS-CoV-2 pandemic. Br J Surg. 2021 Jan 27;108(1):88-96. PMID: 33640908.
6. **GLASBEY JC [First author];** NIHR Global Research Health Unit on Global Surgery. Global guidelines for emergency general surgery: systematic review and Delphi prioritization process. BJS Open. 2022 Jan 6;6(1):zrac005. PMID: 35199142; PMCID: PMC8867031.
7. **GLASBEY JC [Senior author];** STARSurg Collaborative and COVIDSurg Collaborative. Death following pulmonary complications of surgery before and during the SARS-CoV-2 pandemic. Br J Surg. 2021 Dec 1;108(12):1448-1464. doi: 10.1093/bjs/znab336. PMID: 34871379.

8. **GLASBEY JC**, Dobbs TD, Abbott TEF. Can patients with asymptomatic SARS-CoV-2 infection safely undergo elective surgery? Br J Anaesth. 2022 Mar 10:S0007-0912(22)00123-4.

9. **GLASBEY JC [First author];** Glasbey J; COVIDSurg and GlobalSurg Collaboratives. Peri-operative outcomes of surgery in children with SARS-CoV-2 infection. Anaesthesia. 2022 Jan;77(1):108-109. doi: 10.1111/anae.15614. Epub 2021 Nov 5. PMID: 34739137.

10. **GLASBEY JC [Joint first author];** ESCP EAGLE Safe Anastomosis Collaborative. ESCP Safe Anastomosis ProGramme in CoLorectal SurgEry (EAGLE): Study protocol for an international cluster randomised trial of a quality improvement intervention to reduce anastomotic leak following right colectomy. Colorectal Dis. 2021 Oct;23(10):2761-2771. doi: 10.1111/codi.15806. Epub 2021 Aug 25. PMID: 34255417.

11. **GLASBEY JC [First author];** EuroSurg Collaborative. Timing of nasogastric tube insertion and the risk of postoperative pneumonia: an international, prospective cohort study. Colorectal Dis. 2020 Dec;22(12):2288-2297. doi: 10.1111/codi.15311. Epub 2020 Sep 18. PMID: 34092023.

12. **GLASBEY JC [Joint first author];** COVIDSurg Collaborative, GlobalSurg Collaborative. SARS-CoV-2 vaccination modelling for safe surgery to save lives: data from an international prospective cohort study. Br J Surg. 2021 Sep 27;108(9):1056-1063. doi: 10.1093/bjs/znab101. PMID: 33761533;

13. **GLASBEY JC [Joint first author];** COVIDSurg Collaborative; GlobalSurg Collaborative. Timing of surgery following SARS-CoV-2 infection: an international prospective cohort study. Anaesthesia. 2021 Jun;76(6):748-758. doi: 10.1111/anae.15458. Epub 2021 Mar 9. PMID: 33690889.

14. **GLASBEY JC [Joint first author];** COVIDSurg Collaborative. COVID-19-related absence among surgeons: development of an international surgical workforce prediction model. BJS Open. 2021 Mar 5;5(2):zraa021. doi: 10.1093/bjsopen/zraa021. PMID: 33688956.

15. **GLASBEY JC [Joint first author];** COVIDSurg Collaborative. Machine learning risk prediction of mortality for patients undergoing surgery with perioperative SARS-CoV-2: the COVIDSurg mortality score. Br J Surg. 2021 Nov 11;108(11):1274-1292.

16. **GLASBEY JC [Joint first author];** COVIDSurg Collaborative. SARS-CoV-2 vaccination to support safe surgery during the pandemic: a modelling study using data from an international prospective cohort study. *BJS*. 2021. Epub ahead of print.

17. **GLASBEY JC [Joint first author];** COVIDSurg Collaborative. Timing of surgery following SARS-CoV-2 infection: an international prospective cohort study. *Anaesthesia*. 2021. Epub ahead of print.

18. **GLASBEY JC**, Almond ML, Ford SJ. The impact of postoperative radiological surveillance intensity on disease free and overall survival from primary retroperitoneal, abdominal and pelvic soft-tissue sarcoma. European Journal of Surgical Oncology. 2021. Epub ahead of print.

19. **GLASBEY JC [Joint first author];** ESCP Safe Anastomosis ProGramme in CoLorectal SurgEry (EAGLE): Study protocol for an international cluster randomised trial of a quality improvement intervention to reduce anastomotic leak following right colectomy. Colorectal Disease. 2021. Epub ahead of print.

20. **GLASBEY JC**. Antibiotic prophylaxis in a global surgical context. Southern African Journal of Anaesthesia and Analgesia. 2020. Epub ahead of print.

21. **GLASBEY J [Joint first author]**; COVIDSurg Collaborative. Outcomes from elective colorectal cancer surgery during the SARS-CoV-2 pandemic. Colorectal Dis. 2020 Nov 15:10.1111/codi.15431. Epub ahead of print. PMID: 33191669

22. **GLASBEY J [Chair & senior author];** STARSurg Collaborative. Challenges of one-year longitudinal follow-up of a prospective, observational cohort study using an anonymised database: recommendations for trainee research collaboratives. BMC Med Res Methodol. 2020 Feb 7;20(1):27. doi: 10.1186/s12874-020-0909-3.

23. **GLASBEY J [1];** STARSurg Collaborative. Perioperative intravenous contrast administration and the incidence of acute kidney injury after major gastrointestinal surgery: prospective, multicentre cohort study. Br J Surg. 2020 Jul;107(8):1023-1032. doi: 10.1002/bjs.11453. Epub 2020 Feb 5. PMID: 32026470.

24. **GLASBEY J [Steering group];** EuroSurg Collaborative. Safety of hospital discharge before return of bowel function after elective colorectal surgery. Br J Surg. 2020 Apr;107(5):552-559. doi: 10.1002/bjs.11422. Epub 2020 Jan 24. PMID: 31976560.

25. **GLASBEY J [Chair & senior author];** STARSurg Collaborative. REspiratory COmplications after abdomiNal surgery (RECON): study protocol for a multi-centre, observational, prospective, international audit of postoperative pulmonary complications after major abdominal surgery. Br J Anaesth. 2020 Jan;124(1):e13-e16. doi: 10.1016/j.bja.2019.10.005. Epub 2019 Nov 26. PMID: 31784036.

26. **GLASBEY J [Chair & senior author];** STARSurg Collaborative. Maximizing opportunities at medical school to support a career in surgery. Br J Hosp Med (Lond). 2019 Nov 2;80(11):670-673. doi: 10.12968/hmed.2019.80.11.670. PMID: 31707879.

Presentations and prizes

Work from this thesis has been presented to the following meetings:

- UK Patient Reported Outcome Measures (UKPROMS) meeting, 17th June 2022 (Virtual, UK)
- Bethune Round Table, 16th June 2022 (Virtual, Canada)
- Trial Methodology Research Partnership (TMRP), 9th February 2022 (Virtual, UK)
- College of Surgeons of East, Central and Southern Africa (COSECSA) Annual Scientific Conference, 1st December 2021 (Lusaka, Zambia)
- Wound Research Network (WReN) 2021 Scientific Meeting, 20th September 2021 (Virtual, UK)
- UK Patient Reported Outcome Measures (UKPROMS) meeting, 17th June 2021 (Virtual, UK)
- Birmingham Clinical Trials Unit (BCTU) Training forum, 2nd June 2021 (Virtual, UK)
- Bethune Round Table, 29th May 2021 (Virtual, Canada)
- Cardiac Surgical Site Infection Network, 6th October 2020 (Virtual, UK)

I'm also honoured to have received the following prizes for work from this thesis:

- RCS(England)/Rosetrees Essay Prize award 2021: *Research most likely to lead to patient benefit*
- Royal Society of Medicine (RSM) Surgery Section: Adrian Tanner Medal 2022
- UK Patient Reported Outcome Measures (UKPROMS): *Early Career Investigator prize 2021*
- UK Patient Reported Outcome Measures (UKPROMS): *Early Career Investigator prize* and *Patient involvement in research prize 2022*
- University of Birmingham 3-minute thesis (3MT) competition: *Runner-up*. Video available at: https://www.youtube.com/watch?v=HNqz_1kLa1A

Abbreviations

AUROC	Area under the receiver operating characteristic curve
BCTU	Birmingham Clinical Trials Unit
CEI	Community Engagement and Involvement
COVID-19	Coronavirus disease 2019
LMIC	Low and middle-income country (defined by World Bank)
NIHR GSU	NIHR Global Health Research Unit on Global Surgery
PPI	Patient and public involvement
POMR	Postoperative mortality rate (within 30-days of surgery)
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PROM	Patient reported outcome measure
RCS	Royal College of Surgeons
RCT	Randomised controlled trial
SARS-CoV-2	Severe acute respiratory syndrome coronavirus 2
SSI	Surgical site infection
STARD	Standards for Reporting Diagnostic accuracy studies
TALON	Telephone Administration of an adapted patient-reported Wound HeaLing QuestiONnaire for assessment of surgical site infection following abdominal surgery in low and middle-income countries
WHO	World Health Organisation
WHQ	Wound Healing Questionnaire

Table of contents

1. Introduction

1.1 The global importance of surgery and anaesthesia systems

1.1.1 Surgery in holistic healthcare systems

Surgery is an essential component of holistic health systems and underpins the treatment of 30% of the global burden of disease [2-5]. Despite this, it has traditionally been viewed as an expensive luxury and neglected from national health policy in resource constrained settings [6-8]. The 2015 Lancet Commission on Global Surgery shone a light on surgery and anaesthesia as neglected components of global health systems, with severe implications on population health, wealth, social cohesion, stability, and communities [2, 7, 9]. Although more than half of the global population lives in low- and middle-income countries (LMICs), less than 20% of the world's surgeons and anaesthetists provide care in these environments [10].

1.1.2 Global Surgery and the Sustainable Development Goals

Patients that require surgery are often of a working age, and poor access to surgical care can lead to prolonged time away from work, temporary or permanent disability or even death with consequential financial ruin for families and communities [9, 11-15]. This has a significant macroeconomic impact on developing economies, and a compelling business case for investment in surgery and anaesthesia has been proposed and ratified by the World Health Organisation [16, 17]. The United Nations Sustainable Development Goal 3 (Good Health) cannot be achieved without improving equity in access to timely, affordable, safe, and high-quality surgery for underserved communities [6, 18]. The three most common operations performed worldwide are abdominal surgery (laparotomy), fixation of a long bone fracture, and caesarean-section. Together these are commonly known as the Bellwether procedures [3, 19, 20]. This thesis focuses on abdominal surgery as a prominent example of a globally important, high-volume procedure performed for a wide variety of non-infectious and infectious diseases across most hospitals in all countries around the world [21].

1.2 Global variation in the safety of surgery

1.2.1 The growing impact of global variation in surgical safety

Complications of abdominal surgery are common and range in their therapeutic consequence from small deviations in pharmacological management, to reoperation, admission to critical care and death [22-26]. Risk of complications and death after surgery varies dependent on patient, disease, operation, perioperative care, and health system level factors [22-24]. Even accounting for all these factors, the risk of postoperative complications is two- to three-times higher after surgery in the lowest versus highest resource settings [22-24]. There is a global requirement to increase surgical volume to manage the growing burden of surgically treatable disease [7, 27-29]. This has become even more pressing following the SARS-CoV-2 pandemic, with millions of operations cancelled worldwide [20, 30, 31]. If surgical capacity increases without increasing the safety of surgery, the global burden of postoperative complications risks rising in parallel, and disproportionately affecting some of the world's most vulnerable populations [12, 32]. The implications of this are profound and represent a global healthcare crisis which urgently requires innovation, evidence, and investment in parallel.

1.2.2 Multifactorial reasons for variation in outcomes

The reasons for variability in surgical outcomes are multifactorial. First, lowest resource healthcare systems do not have capacity to rescue some patients with severe surgical complications, including lack of access to cross-sectional imaging, interventional radiology, 24h emergency theatres, critical care beds and organ support services [22, 33]. Second, there is variability in access to healthcare technologies to support safer perioperative and surgical care (for example, capacity to provide minimally invasive surgery) [34]. Third, there is an insufficient number of expert surgery, anaesthesia, and obstetric providers to manage

the volume of patients requiring surgery and perioperative care [10, 27]. Fourth, pathways for preoperative optimisation are often underdeveloped and underfunded for planned surgical patients, leading to suboptimal patient physiology at the time of the surgical insult [24, 35]. For emergency surgery, patients often face delayed access to care with increased physiological severity and more advanced disease stage at the time of surgery. Fifth, pathways for subspecialised surgical care in LMICs, whilst being urgently developed by national and regional organisations, remain under-resourced and access to specialised care is scarce [10, 27, 36-39]. Together this equates to higher risk patients, operated in more fragile surgical systems, that are less able to rescue patients when they have surgical complications.

1.2.3 What is needed

Access to safe surgery is an issue of high global importance [7]. If surgery were to be considered a cause of death, it would be the third leading cause worldwide, with greater impact than tuberculosis, malaria, and Human Immunodeficiency Virus (HIV) combined [32]. Co-prioritised research programmes are urgently needed to identify effective, contextually relevant, and frugal interventions to reduce the global burden of surgical complications [40-42]. This thesis aims to strengthen the quality and efficiency of these research systems.

1.2.4 NIHR Global Health Research Unit on Global Surgery

To support research that seeks to address these global health priorities, the United Kingdom Government allocated Official Development Assistance (ODA) funding to the National Institute for Health Research (NIHR). The NIHR Global Health Research Unit on Global Surgery is a partnership between the Universities of Birmingham and Edinburgh and was successful in securing funding in 2017 and 2022 (£15-million total). The NIHR Unit primarily aims to enhance international multidisciplinary partnership of surgeons, anaesthetists, and

research methodologists across seven LMICs, building capacity and research leadership for the future. The Unit network delivers a range of research programmes including randomised controlled trials [43-45], prospective cohort studies [22-24, 46-49], guidelines [33, 50, 51], and implementation studies. The Unit network provided the research platform and collaborator network for this thesis.

1.3 The global impact of surgical site infection

1.3.1 Pathoaetiology and incidence

An important example of a postoperative complication after abdominal surgery is surgical site (wound) infection (SSI). It is defined as an infection of the soft tissue around a surgical incision and can be superficial (affecting the skin and subcutaneous tissue), deep (also affecting the fascia and skeletal muscle layers) or organ space (intraabdominal) [52]. The causative pathogens depend on the type of surgery and organ system. Common organisms isolated from infected abdominal wound swabs include Staphylococcus aureus, coagulase-negative staphylococci, Enterococcus spp. and Escherichia coli [53]. It is the most common complication of surgery worldwide affecting between 1 in 20 patients after clean surgery and 1 in 2 patients where there is significant pus or faecal contamination of the abdomen [23].

1.3.2 The impact of SSI on patients

For patients that have SSI the consequences can range from (i) targeted antimicrobial therapy, sometimes with wound opening and delayed healing, to (ii) major reoperation, (iii) abdominal dehiscence (complete breakdown of the abdominal wound) and, (iv) long critical care admissions [23, 43, 54, 55]. This morbid condition impairs recovery from surgery and can have severe effects on wellbeing and quality of life for our patients [56-63].

1.3.3 The impact of SSI on global healthcare systems and economies

Approximately 9,800,000 surgical wound infections occur around the world each year, with a huge cost to patients, families, societies, and health systems [60, 64, 65]. Costs of treatment of SSI vary significantly from country to country but are extremely costly, particularly those for more severe manifestations [66]. With many patients in LMICs at high risk of financial catastrophe after their index surgery, the consequences of this 'second hit' where Universal Health Coverage is not available is often life altering [9, 65, 67, 68].

1.3.4 Risk factors for SSI

Common reported patient level risk factors for surgical wound infection include older age, smoking, comorbidities such as diabetes, obesity, vascular disease, autoimmune and malignant disease, and concurrent immunosuppression [23, 25, 69-78]. Larger operations, through bigger incisions typically also have greater risk [43, 56]. However, the most important consideration in risk stratification for SSI is the degree of intraabdominal contamination during surgery (*Table 1.1*) [23]. The greater degree of abdominal contamination, the higher the SSI risk the patient faces. This has been previously defined by Mangram et al, 1999 [52]. For this reason, in this thesis, intraabdominal contamination (clean-contaminated versus contaminated/dirty) will be explored as key strata throughout.

Table 1.1. Wound contamination classification [79]

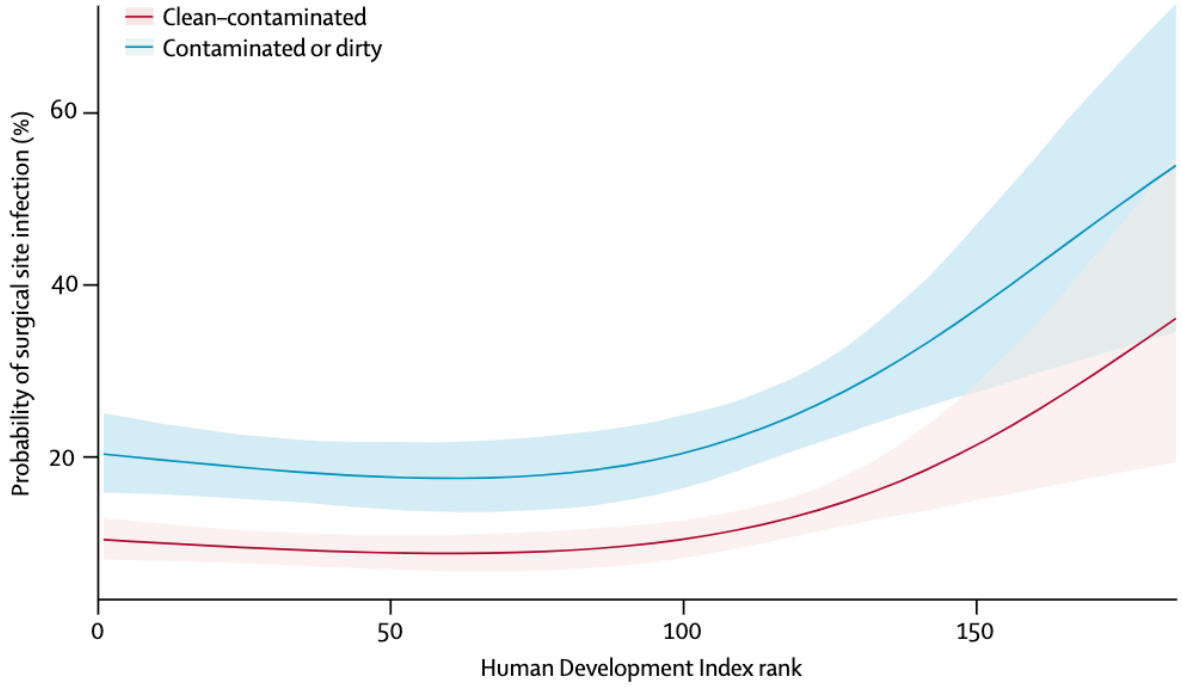
Classification	Description
Clean	An uninfected operative wound in which no inflammation is encountered and the respiratory, alimentary, genital, or uninfected urinary tract is not entered. In addition, clean wounds are primarily closed and, if necessary, drained with closed drainage. Operative incisional wounds that follow nonpenetrating (blunt) trauma should be included in this category if they meet the criteria.
Clean-Contaminated	An operative wound in which the respiratory, alimentary, genital, or urinary tracts are entered under controlled conditions and without unusual contamination. Specifically, operations involving the biliary tract, appendix, vagina, and oropharynx are included in this category, provided no evidence of infection or major break in technique is encountered.
Contaminated	Open, fresh, accidental wounds. In addition, operations with major breaks in sterile technique (e.g., open cardiac massage) or gross spillage from the gastrointestinal tract, and incisions in which acute, non-purulent inflammation is encountered are included in this category.
Dirty	Old traumatic wounds with retained devitalized tissue and those that involve existing clinical infection or perforated viscera. This definition suggests that the organisms causing postoperative infection were present in the operative field before the operation.

1.3.5 Global variation in rates of SSI

When examining global variation in SSI rates, researchers collect data on these key parameters which are believed to be both causally and biologically linked to SSI to allow for risk adjustment of estimates. In global outcomes studies from the GlobalSurg collaborative the adjusted odds of SSI was observed to be 1.6x higher (1·60, 95% credible interval 1·05–2·37) in low versus high income environments (*Figure 1.1*). SSI therefore disproportionately affects patients in low resource settings in health systems that have fewer resources to support patients and their recovery. Reflecting this, SSI was identified as the highest priority research area in multiple global co-prioritisation exercises [40], and is the target of several ongoing global randomised controlled trials (RCTs) [43-45] and quality improvement efforts [80].

Figure 1.1. Probability of SSI by human development index (HDI) country rank.

Reproduced from GlobalSurg Collaborative, The Lancet Infectious Diseases (2018) [23].



The Human Development Index (HDI) was developed by the United Nations and is a composite of life expectancy, education (mean years of schooling completed and expected years of schooling upon entering the education system), and per capita income. It is used to rate and rank countries based on their development and group them into four tiers (high, upper-middle, lower-middle, low). A higher HDI rank indicates a higher population lifespan, education level, and income. The shared area represents the 95% credible interval derived from Bayesian mixed-effects modelling.

1.4 The role of research in global surgery

1.4.1 The importance of research in global surgery

High quality research has a fundamental role in strengthening global surgery and anaesthesia systems. This includes pragmatic randomised controlled trials, prospective cohort studies, health service delivery research, mixed-methods, and implementation science projects. Evidence can inform health policy, intervention selection, investment cases, system reform, training models and behaviour change. Randomised controlled trials have a special place in this practice change, as they control for selection and confounding bias and so allow for causal interpretation. This encourages adoption into practice by clinicians for real-world patient benefit.

1.4.2 Problems with research in global surgery

Inequity in authorship and leadership and imbalance in power continue to be problematic in research in the area known as ‘global surgery’ [81-85]. Global research, however defined, must be co-designed and co-led with partners and patients from the health systems and communities that it hopes to represent [85, 86]. Partnerships between global North and South actors can be delivered ethically where the balance of benefits is equal, learning occurs bilaterally, and agreements are made on an equal footing, as equal partners [82]. ‘Parachute’ research models are hugely damaging in global health partnerships and serve to increase inequity rather than to improve it [87]. Multinational research efforts must therefore strive to balance power between partners, work collaboratively and share leadership at every level, and seek routes to dissemination that are locally sourced and locally relevant. Few international practice guidelines to date have included data from LMICs, largely because of a lack of trials conducted in these settings [42]. Building research capacity and infrastructure for LMIC-led research will grow equity in global health and more rapidly build towards a healthier future for the global population, with benefits for all [2]. The research in this thesis

has been designed, delivered, and interpreted as a deep and equal collaboration with partners from several countries around the world. All outputs from primary data will be published under a single corporate authorship structure, crediting all collaborating authors and with a joint corresponding author model between LMIC and HIC partners [24, 88, 89].

1.5 The problem with detection of surgical site infection

1.5.1. Challenges of measurement of SSI

Surgical site infection is not an easy complication to measure. There remains no single microbiological, biochemical, imaging, or invasive test for SSI diagnosis [52]. As such several diagnostic criteria for SSI have evolved including from the US Centres for Disease Control (CDC), ASEPSIS criteria and a definition from Public Health England [52, 90, 91]. The accepted 'gold standard' reference test for diagnosis is in-person evaluation according to the US CDC criteria [52]. These are summarised in *Box 1.1* below. For the purposes of this thesis, I have combined the definitions of superficial and deep SSI after primary closure of the abdominal surgical wound into a single concept, where deep SSI is a more severe manifestation of the same postoperative complication. Deep organ space SSI is not included within this definition, and has a differential biological rationale (e.g., anastomotic leak, fistula, residual fluid collection). Adoption of these criteria in high-quality research studies brings reproducibility to SSI assessment, but several parts of the criteria remain subjective. For example, the criterion "diagnosis of an SSI by a clinician or on imaging" includes diagnosis at any time after surgery by any clinician; this individual may not have been trained in the CDC criteria, or have applied them incorrectly. The subjective nature of the symptoms of SSI are also a common source of variation. This leads to non-random intra- and inter-observer variability in the diagnosis of SSI, even where this 'gold standard' assessment is adopted [56, 92].

Box 1.1. Centres for Disease Control definition of surgical site infection

The 'gold standard' reference diagnostic test for surgical site infection (SSI) during the 30-days after surgery is in-person review according to US Centres for Disease Control Criteria [93]. The following definition was used to identify deep incisional or superficial incisional SSIs:

- The infection must occur within 30-days of the index operation

AND

- The infection must involve the skin, subcutaneous, muscular, or fascial layers of the incision

AND

The patient must have at least one of the following:

- Purulent drainage from the wound
- Organisms are detected from a wound swab
- Wound opened spontaneously or by a clinician

AND

At the surgical wound, the patient has at least one of:

- pain or tenderness
- localised swelling
- redness
- heat
- systemic fever (>38°C)
- diagnosis of SSI by a clinician or on imaging

1.5.2 Importance of variation in SSI measurement

In clinical practice, variation in detection of SSI could lead to direct harm through overtreatment of borderline cases (e.g., with antibiotics) with higher rates of side effects and antimicrobial resistance, which is another global health crisis [94, 95]. It could also lead to undertreatment and delayed diagnosis of SSI, leading potentially to sepsis and more severe consequences of wound infection. In RCTs, differential misclassification of SSI between arms can lead to measurement bias, and incorrect effect estimates [96]. This leads to indirect harm for patients that may miss out on the benefit of novel therapies (a false negative), or inappropriately receive a harmful or ineffective treatment (a false positive).

1.5.3 The size of the problem

The risk of misdiagnosis (predominantly under-detection) of SSI is magnified significantly where the CDC criteria are not used. In population surveillance programmes which use routine electronic health records or ad hoc clinical reporting, reported rates of SSI after major abdominal surgery are as low as 2% to 5% which is likely to be a 5 to 6-fold underestimation of the true SSI rate [58, 60, 64, 97, 98]. In population level cohort studies of abdominal surgery, rates of 12% to 15% are reported, roughly half that of equivalent populations in randomised trials [25, 56, 69, 99, 100]. Even in high quality randomised controlled trials where SSI is included as a secondary outcome measure, the adjusted SSI rate is 50% of that which would be expected in a trial with SSI as a primary outcome [92]. Put simply, in the diagnosis of SSI, the harder you look, the more you see. ‘Missed diagnoses’ are likely to be early or milder SSI events, however, they may also include patients with very severe symptoms that are admitted to a different hospital for care (so their outcome is challenging to record) or for patients who have poor access to care, so may in fact be at highest risk. Better methods for the diagnosis of this globally important surgical complication are urgently needed.

1.6 The importance of post-discharge wound surveillance

1.6.1 Timing of presentation of SSI

SSI can occur any time after abdominal surgery, but typically has a bimodal peak with high incidence at 5-7 days postoperatively and a second peak around 10-14 days postoperatively [101, 102]. The biological rationale for this is unknown, but it is likely to be related to the balance between immunosuppression related to the acute surgical insult, colonisation of the surgical wound before primary healing is achieved, and manifestation of superficial SSI as a late herald of deep SSI in patients with suggestive signs and symptoms or deep collection on imaging [23, 102, 103]. A new diagnosis of SSI over 30-days after surgery in the absence of an intraabdominal complication or enterocutaneous fistula is extremely rare, and this is typically adopted for the time of cessation of follow-up in RCTs of SSI prevention strategies [43, 56, 104].

1.6.2 Variation in length of hospital stay across settings

Time to discharge after surgery varies from patient to patient and hospital to hospital [22-24]. In high resource settings this has generally followed a trend of decreasing length of hospital stay, even after major surgery, in the wake of Enhanced Recovery After Surgery (ERAS) programmes [105-107]. Where reliable and resilient pathways for readmission to hospital exist, these programmes theorise that early discharge will encourage drinking, eating, mobilising, and sleeping in the home environment, and reduce rates of nosocomial infection, with a strong and evolving evidence base to date [108-112]. Length of stay in lower resource settings may vary from this for a number of reasons; patients may have travelled a very long distance to access tertiary care, may not have access to specialised care in the community or easy pathways for readmission in the event of deterioration, may not be able to afford continuing ward based care, may face significant pressures for early return to work or care

duties, or where capacity for ERAS programmes are insufficient [22, 113, 114]. Together this means that around two thirds of SSI occurs after discharge but with significant variation between hospitals (from 13.5% to 91% in a 2016 systematic review)[101]. This highlights the importance of post-discharge surveillance of SSI for delivery of high-quality research and clinical care. This was recognised as a key quality criteria in a Lancet Infectious Disease meta-analysis which adapted the Cochrane Risk of Bias 2 tool for application to SSI trials [56].

1.7 Inadequacy of in-hospital SSI as a surrogate for 30-day post-discharge SSI

Despite variability in in-hospital SSI detection due to heterogeneity in practice across settings, we might consider whether in-hospital SSI measurement could be used as an early surrogate for 30-day SSI assessment. To act as an appropriate surrogate, in-hospital SSI measurement would have to meet four Prentice criteria [115, 116]:

1. Treatment must have a significant effect upon the surrogate endpoint (in-hospital SSI)
2. Treatment must have a significant effect upon the true endpoint (30-day SSI)
3. The surrogate endpoint (in-hospital SSI) must have a significant effect upon the true endpoint (30-day SSI)
4. The full effect of treatment upon the true endpoint (30-day SSI) must be mediated by the surrogate (in-hospital)

This was updated by Marc Buyse in 2000 to provide a statistical approach to assessment of surrogacy in binary-binary endpoints [115].

To explore the validity of in-hospital SSI assessment as a surrogate for 30-day SSI using these methods, we can explore the data from the FALCON trial published in *The Lancet* in 2021 [43]. This was a pragmatic 2×2 factorial RCT evaluating skin preparation and fascial sutures in 5788 patients from seven LMICs. Patients underwent both in-hospital and 30-day

SSI assessment. In this study 5.4% (311/5788) of participants died before 30-days and (310/5778) 0.7% (43/5788) of patients remained in-hospital at 30-days after surgery, with a very high SSI rate (62.8%, 27/43) in this high-risk group. Of those who were discharged and alive at 30-days after surgery (N=5470) the median day of discharge was 5 days (IQR: 3 to 8 days). The SSI rate at hospital discharge was 12.3% (639/5186, 284 missing) and 21.7% (1137/5248, 222 missing) at 30-days. Overall, 43.4% (497/1134) of SSI occurred after discharge.

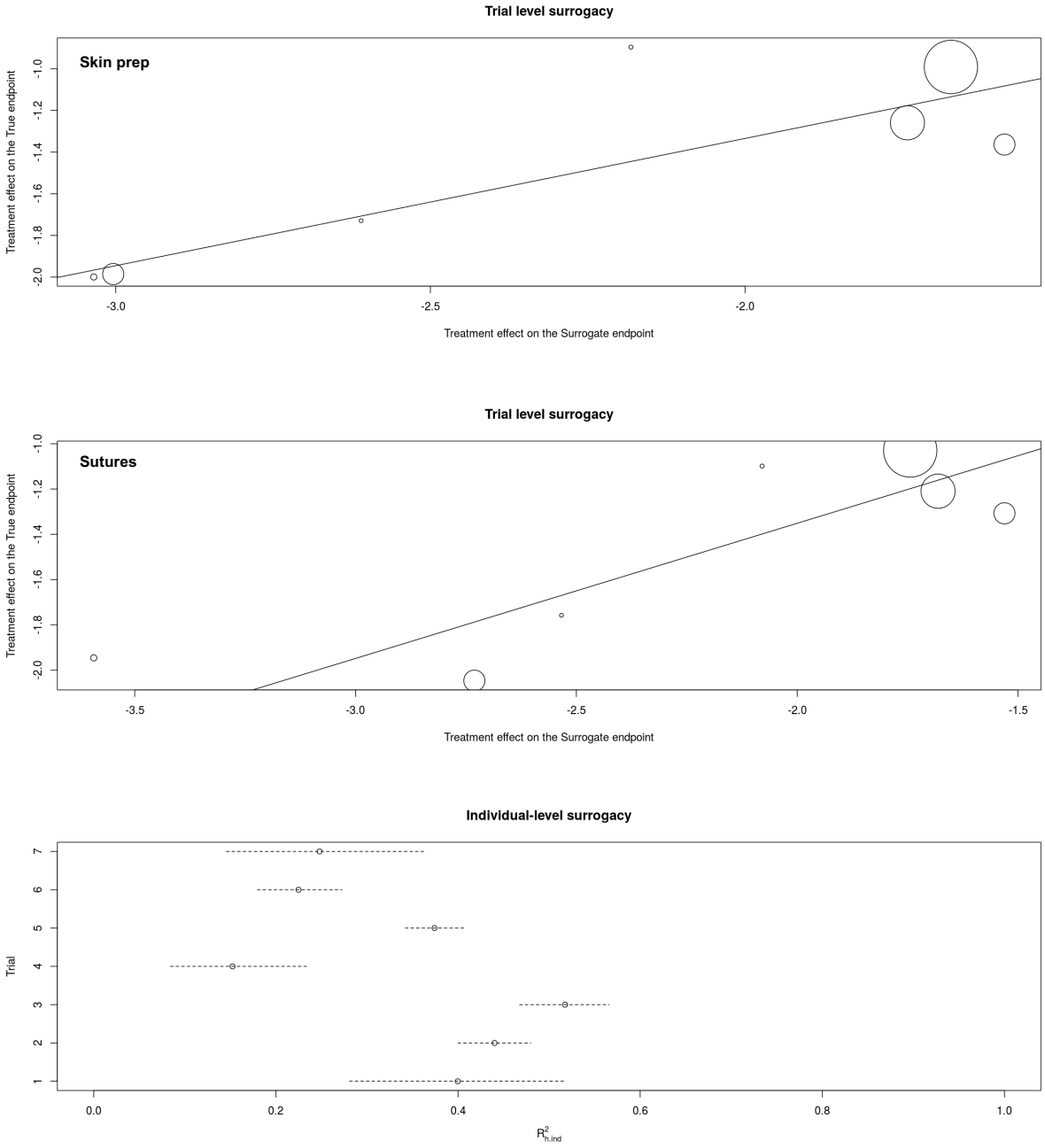
I used an information-theoretic approach to estimate trial- and individual-level surrogacy based on full fixed-effect models. I used subgroups by country as a proxy for data from different trials upon meta-analysis [117] (R Project for Statistical Computing v4.2.1, package: *surrogate*). Within-trial correlation (the extent to which the surrogate (in-hospital SSI) estimates the magnitude and variability of the effect estimate between trial arms in comparison to the true endpoint (30-day SSI)) can be explored using slope of the linear regression between the trial-level effects of treatment upon both endpoints (R^2_{trial}). A surrogate would be 'trial valid' if the R^2_{trial} was close to 1 (e.g., $\geq$0.8 [115]). In this example, one might expect within-trial correlation to be a low if there was a difference in discharge practices between patient groups with different baseline risks of SSI.

Within-patient association (the extent to which occurrence of the surrogate (in-hospital SSI) is predictive of the true endpoint (30-day SSI)) can be explored with the individual-level association between both endpoints ($R^2_{individual}$). Again, the surrogate would be 'individual valid' if the R^2_{trial} was close to 1. Both criteria must be met for a surrogate to be valid overall. In this example, one might expect within-patient correlation to be low if a high proportion of SSI was detected after discharge.

The R^2_{trial} and $R^2_{individual}$ for in-hospital SSI assessment in the FALCON trial are summarised in *Figure 1*. The R^2_{trial} was 0.69 (95 CI: 0.20 to 0.95) for skin prep arms, the R^2_{trial} was 0.44 (95 CI: 0.03 to 0.85) for the fascial suture arms and $R^2_{individual}$ was 0.38 (95% CI: 0.36 to 0.40). None of the three R^2 were $\geq$0.8 and I was unable to accept surrogate validity. The wide variation in trial-level surrogacy from country to country may have reflected differences in discharge practices between countries.

Together this analysis suggest that in-hospital SSI assessment is not a valid surrogate for 30-day SSI in multinational RCTs. Robust post-discharge surveillance pathways are required. This has been recognised as a key quality measure in an adapted Cochrane Risk of Bias-2 (ROB-2) tool for SSI research [56].

Figure 1.2 Trial and individual level surrogacy effects in the FALCON trial data



Each estimate plotted for the R^2 coefficient represents the slope of the linear regression plotted between the surrogate and true endpoint for a different country participating in the FALCON trial. The top two graphs represent the trial level surrogacy estimates defined as the extent to which the surrogate (in-hospital SSI) estimates the magnitude and variability of the effect estimate between trial arms in comparison to the true endpoint (30-day SSI). The surrogate is valid where the plotted estimates sit along the diagonal line. Visualising the graphs, trial level surrogacy appears superior for skin prep than sutures. The bottom graph represents individual level surrogacy defined as the extent to which occurrence of the surrogate (in-hospital SSI) is predictive of the true endpoint (30-day SSI). The surrogate is valid where the estimates are closer to 1.0 (i.e., farther to the right). Overall, there was poor individual level validity of in-hospital SSI as a surrogate with some variability between countries. This might be due to differences for example in discharge practices between countries; where hospitals keep patients in-hospital for a longer time, their SSI is more likely to manifest in-hospital, thus the surrogate endpoint (in-hospital SSI) is more likely to be same as the true endpoint (30-day SSI).

1.8 Remote detection of post-discharge SSI in low resource settings

1.8.1 Challenges with in-person assessment after discharge in LMICs

Post-discharge surveillance for SSI according to the accepted gold standard would require a patient to travel back to hospital for an in-person evaluation (e.g., at 30-days postoperatively for a research assessment) [52, 104, 118]. In research studies that means an extra in-hospital assessment for each participant recruited. This is problematic for several reasons in LMICs. First, patients often travel long distances to a tertiary centre for surgery before returning to their home location [113]. Second, patients commonly face the risk of catastrophic expenditure because of their index surgery and a return journey to hospital risks further expenditure and time away from work or care [9, 67, 68]. Third, there is a substantial opportunity cost for trained surgeons or allied clinicians to perform routine wound assessment for every patient without appropriate triage. This is particularly important considering the scarcity of trained surgeons, obstetricians, and anaesthetist providers in many LMICs [10, 27]. Where 70% of patients will not have an SSI at 30-days after surgery, this additional visit for reassurance only is wasteful and inefficient [23].

1.8.2 Loss to follow-up and attrition bias

1.8.2.1 What is attrition bias?

Loss to follow-up occurs in a randomised trial where a patient does not maintain their trial involvement up to the point of the end of scheduled follow-up. This can occur where the patient formally withdraws consent, or simply because of non-attendance or non-contact for a follow-up assessment in the context of surgical studies. Loss to follow-up is a major source of bias in randomised trials. It causes “attrition bias” when patients are differentially lost to follow-up between arms (effectively a problem of missing data not at random or ‘MNAR’). This could mean more patients lost to follow-up in one randomised arm than another, or different characteristics of patients lost to follow-up in one arm (i.e., a systematic error). This

occurs as loss to follow-up in trials is rarely random and the patients with missing outcome data are often different to those with complete outcome data. In one example of follow-up of a group of infants born preterm there was an 8-fold higher rate of excess disability in a small group of babies (47/795) who were more difficult to trace due to social mobility. If only patients who were easy to track were included in the study, the overall primary outcome rate would have halved [119]. In the example of SSI, patients with severe SSI may have too much pain or disability to reattend an in-person assessment or be admitted to another hospital without knowledge of the research team. Equally, patients with no SSI that recovered quickly may have returned to work and may be unable to attend follow-up due to limitations imposed by their employer. Both risk differential misclassification in a clinical effectiveness evaluation in SSI prevention.

1.8.2.2 The impact of attrition bias

Where attrition bias occurs, it can have a significant impact on the overall treatment effect in a randomised trial. In a systematic review of high-impact medical journals between 2005-2007, as many as 1 in 5 to 1 in 3 trials would have had a change in overall direction of effect (e.g., from a significant to non-significant estimate) in a 'worse-case scenario' when missing data were re-coded as either all negative or all positive events [120]. Attrition bias therefore can lead to patient harm, either when future patients miss out on an effective treatment that seemed ineffective in the trial population (i.e., a type II error), or when they are inappropriately exposed to an ineffective treatment (and its potential side effects, i.e., type I error). Where loss to follow-up is $\geq$5% this can lead to concerns with internal validity, and when $\geq$20% this can lead to serious risk of bias [121]. Trial retention has therefore been prioritised as a key research area in methodology research and was the focus of a James Lind Alliance priority setting partnership (PRIORITY-II) [122]. When patients are required to make a journey back to hospital after surgery this creates several potential logistical,

financial, physical, and psychological barriers to completion of outcome assessment. It is possible that by removing some of these barriers, remote data capture (e.g., using telephone, video, application, or text-message based follow-up) could reduce loss to follow-up in future research and subsequent risk of bias. Remote detection of SSI using telemedicine methods (i.e., digital solutions for remote patient communication, assessment and/or management) will be another key consideration in this thesis and explored in data and discussion.

1.8.2 Potential solutions for remote wound assessment

To date, no contextually-relevant, high-quality pathway for remote SSI detection exists for use in global randomised trials and clinical practice [123]. Several potential options for remote assessment exist, including questionnaire, telephone call, text message, photograph, video, virtual reality, online and app-based assessment and triage tools [124-132]. Broadly these can be considered to be components of 'telemedicine' as they all use some form of digital technology for remote outcome assessment. The evidence base for the safety and feasibility of these telemedicine methods remains immature, and no tools have been developed or tested in LMICs where perioperative pathways, cultural and contextual barriers and patient-provider interactions vary considerably [35, 129, 133, 134].

1.8.3 Opportunities for telemedicine in low-resource care pathways

Significant global opportunities now exist for remote methods for post-discharge surveillance. The World Bank estimates that over 80% of the global population now has access to a mobile phone, and this proportion is growing rapidly [129]. This level of connectivity facilitates accessible adoption of both established and novel digital technologies in remote assessment pathways. If valid tools can be identified and tested, they could be administered by non-expert doctors (or even non-clinicians) without compromising on the quality of assessment.

This minimises risk of research bias or harm to patients through false reassurance or overtreatment, whilst reducing the huge financial and opportunity costs of routine in-person follow-up after surgery [135, 136].

1.8.4 Potential applications of telemedicine in SSI detection

There are two principal ways in which telemedicine tools could be applied to SSI detection. First, they could be used in direct diagnosis of an SSI. This allows the patient to seek care through established local pathways and provides a definitive binary assessment of the SSI status for the purposes of research studies. Second, they could be used to triage patients based on their likelihood of SSI to seek in-person review for diagnosis (high or moderate likelihood), re-review at a future time (low likelihood), or provide reassurance (very low likelihood). This still brings efficiency to the SSI assessment pathway versus routine review of all patients in-person postoperatively but requires additional systems to be in-place for in-person review of a proportion of at-risk patients. In this thesis, I will consider both applications in detail, as either could be preferred in specific settings. I will discuss the relevant strengths and weakness of each throughout and provide data to support implementation of telemedicine tools using either method.

1.9. Relevance of telemedicine in post-pandemic recovery

1.9.1 Risks of in-person follow-up for surgical patients during COVID-19

In the early SARS-CoV-2 pandemic period, patients undergoing surgery were identified as being at high risk of severe COVID-19 related postoperative pulmonary complications and death. In large international datasets, 1 in 2 patients with SARS-CoV-2 within 30-days of surgery suffered a pulmonary complication and as a many as in 1 in 4 of these died [48, 49]. Pulmonary complications were responsible for 2 in every 3 deaths in surgical patients in 2021 [137]. No difference in risk was observed between patients that became infected

preoperatively and any time postoperatively up to 30-days after surgery, and risk was high even after minor operations [49]. It was hypothesised that this additional risk in comparison to the general population was due to the 'double-hit' of the stress of surgery and invasive mechanical ventilation in addition to SARS-CoV-2 infection [138, 139]. Healthcare systems developed several strategies to protect patients from infection including COVID-19 free surgical pathways, routine preoperative testing, and vaccination [140, 141]. Outpatient attendance to clinic for in-person follow-up therefore, for many, became an unnecessary additional risk with health providers initiating telemedicine methods without the evidence base to support their implementation.

1.9.2 Changes in patterns over time

Whilst the phenotype of perioperative SARS-CoV-2 is likely to have decreased in severity over time, related to improvements in critical care, widespread vaccination, and emergence of the Omicron variant, surgical patients remain at increased risk [46, 142]. Many hospitals have invested time and energy in behavioural change for clinicians to provide telemedicine services and for patients to accept and understand the limitations of telemedicine consultations [143, 144]. It is therefore very unlikely that the rapid adoption of telemedicine in the surgical setting will be reversed. Providing high-quality models for implementation of telemedicine is now an urgent priority for the global surgical community. Pathways must be developed that are applicable to all hospitals and patients worldwide to counter the growing inequalities in health and welfare that have been compounded by the pandemic [145-148].

1.9.3 Emerging examples of telemedicine in postoperative care

Several examples of postoperative use of telemedicine have gained traction over the past 2-years. However, heavy investment in software and infrastructure has not been matched by

investment in high-quality research to evaluate its use. A few prominent good-quality examples exist:

- **PVC-RAM-1 trial**: 8 hospitals in Canada, demonstrated increased identification and correction of prescribing errors and reduced pain in patients undergoing non-elective surgery randomised to a tablet computer for daily wound photographs and remote monitoring versus standard care [149]
- **TWIST trial**: a single hospital in Scotland, demonstrated earlier SSI detection and fewer readmissions in adult emergency surgery patients randomised to a smartphone-delivered wound assessment versus standard care [131]
- **STOMPA trial**: demonstrated a reduced readmission rate and burden of travel in new ostomates randomised to receive regular teleconsultations with a stoma nurse in comparison to standard care [150]
- **JAMA Surgery, 2021**: two hospitals in USA, demonstrated a reduction in time taken for follow-up overall with a virtual visit versus in-person visit after minimally invasive appendicectomy or cholecystectomy, with no reduction in provider contact time for patients [124]

No high-quality diagnostic accuracy studies have evaluated telemedicine in postoperative care in LMICs.

1.9.4 Telemedicine as a core component of resilient surgical systems

A final example of how telemedicine can support COVID-19 pandemic recovery is in building resilient surgical systems. Resilient surgical systems are able to continue surgery and anaesthesia provision during periods of high external stress, such as pandemics, natural disasters, war and other social and political disruption [11, 47, 151, 152]. Reducing resource usage and building more flexibility into perioperative care teams will allow agile working across multidisciplinary team members, reduce demands on physical space and clinical staff,

and significantly reduce administrative burden in outpatient services [136, 153, 154]. Remote pathways for postoperative surveillance were prioritised as one of 23 key Surgical Preparedness Indicators (SPIs) in a recent global consensus and international hospital assessment, highlighting its importance in preparedness for external shocks [47, 155].

1.10 Introduction summary

SSI is a problem of huge global importance. Robust pathways to detect SSI require postoperative surveillance up to 30-days after surgery, which currently requires patients to travel back to the hospital after surgery. This is particularly inefficient in low resource environments where time, space, staff, and resources are scarce. As the global community becomes increasing connected by mobile devices, there is a huge opportunity to leverage this to build high-quality, efficient postoperative surveillance pathways. This has been particularly notable during COVID-19, where hospitals around the world have adopted telemedicine despite an immature evidence base. This thesis aims to explore this area in depth, in the context of SSI, and develop a novel pathway for implementation around the world, including in LMICs. This will both strengthen future research studies and build resilience into surgical systems.

2. Aims and objectives

The overall aim of this thesis was to develop a high-quality pathway for remote surgical wound assessment using telemedicine that can be delivered flexibly across diverse healthcare settings. To achieve this, I had three aims, each with related objectives.

Aim 1. First, I aimed to evaluate the feasibility and accuracy of existing telemedicine methods in the detection of post-discharge SSI after abdominal surgery using published data.

Objectives

- To explore whether wound assessment using telemedicine was feasible across different country and income settings in data from an international cohort study
- To compare the rates of SSI reported using telemedicine and in-person follow-up in risk adjusted patient-level data
- To explore the rates of telemedicine adoption for wound assessment in the published literature
- To compare the rates of SSI detected using telemedicine and in-person follow-up in published data using meta-analysis

Aim 2. Second, I aimed to explore the cross-cultural equivalence of a Wound Healing Questionnaire (WHQ, developed and validated in the UK) across several LMICs and make recommendations for its adaptation for use in global surgery research and practice.

Objectives

- To explore cross-cultural equivalence, acceptability, and content validity of the WHQ across several LMICs using qualitative methods

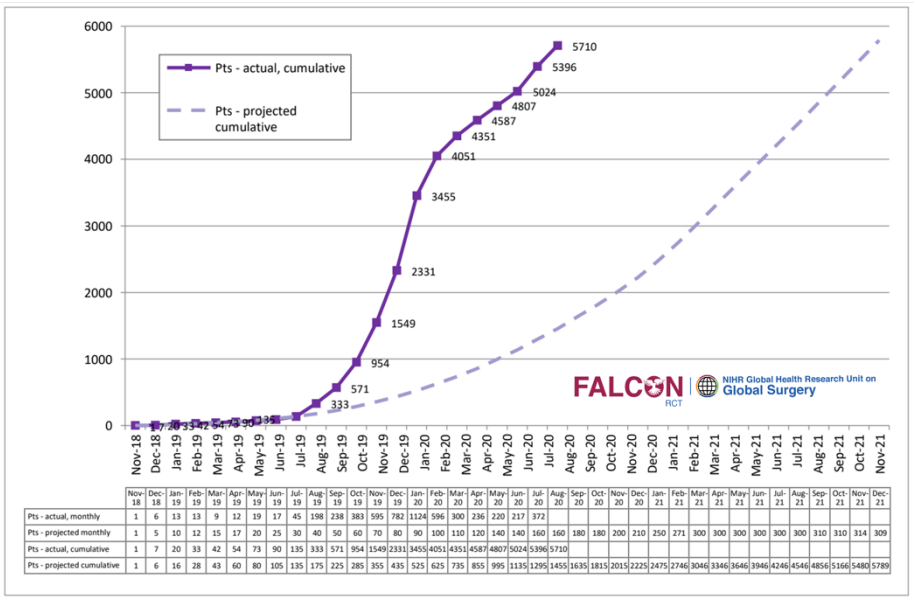
- To assess the translatability of the WHQ for use across relevant languages in the FALCON trial delivery network
- To perform cross-language translation of the WHQ according to best practice recommendations
- To assess the scaling and psychometric properties of the WHQ in quantitative data using Rasch analysis
- To triangulate these data to make recommendations for the adaptation of the WHQ for use in global surgical research and practice

Aim 3. Finally, I aimed to test the feasibility and accuracy of the adapted WHQ in remote diagnosis of post-discharge SSI.

Objectives

- To evaluate the feasibility of a telephone WHQ pathway for remote assessment of the surgical wound across seven LMICs
- To assess the accuracy of the telephone WHQ in diagnosis of SSI in adult patients undergoing major abdominal surgery
- To explore the diagnostic accuracy of the WHQ across key patient subgroups
- To work with patients and community members to co-produce an optimised pathway for telephone outcome measurement, and make recommendations for future researchers

Figure 2.1. Relationship between FALCON trial and TALON-1 and TALON-2 studies within a trial



	Nov-18	Dec-18	Jan-19	Feb-19	Mar-19	Apr-19	May-19	Jun-19	Jul-19	Aug-19	Sep-19	Oct-19	Nov-19	Dec-19	Jan-20	Feb-20	Mar-20	Apr-20	May-20	Jun-20	Jul-20	Aug-20	Sep-20	Oct-20	Nov-20	Dec-20	Jan-21	Feb-21	Mar-21	Apr-21	May-21	Jun-21	Jul-21	Aug-21	Sep-21	Oct-21	Nov-21	Dec-21
Pts - actual, monthly	1	6	13	13	9	12	19	17	45	198	238	383	595	782	1124	596	300	236	220	217	372																	
Pts - projected monthly	1	5	10	12	15	17	20	25	30	40	50	60	70	80	90	100	110	120	140	140	160	160	180	180	200	210	250	271	300	300	300	300	300	310	310	314	309	
Pts - actual, cumulative	1	7	20	33	42	54	73	90	135	333	571	954	1549	2331	3455	4051	4351	4587	4807	5024	5396	5710																
Pts - projected cumulative	1	6	16	28	43	60	80	105	135	175	225	285	355	435	525	625	735	855	995	1135	1295	1455	1635	1815	2015	2225	2475	2746	3046	3346	3646	3946	4246	4546	4856	5166	5480	5789

TALON-1: Mixed methods WHQ adaptation and translation

Planning | Delivery

TALON-2: Prospective cohort study of feasibility and validity

Planning | Delivery

3. Current use of telemedicine for post-discharge assessment of the surgical wound: an international cohort study, and systematic review with meta-analysis

3.1 Abstract

Background

Surgical site infection is the most common complication of surgery worldwide, and frequently occurs after hospital discharge. Evidence to support implementation of telemedicine during postoperative recovery will be an essential component of pandemic recovery. This chapter aimed to determine whether remote wound reviews using telemedicine can be safely upscaled, and if standardised assessment tools are needed.

Methods

The primary outcome of this study was surgical site infection reported up to 30-days after surgery. I compared rates of SSI reported on remote assessment using telemedicine (telephone and/or video assessment) versus those with in-person review. The first part of this study analysed primary data from an international cohort study of adult patients undergoing abdominal surgery who were discharged from hospital before 30-days after surgery. The second part combined this data with the results of a systematic review to perform a meta-analysis of SSI detection rates using telemedicine and in-person using all available data. This was reported in accordance with PRIMSA recommendations (PROSPERO:192596).

Results

The cohort study included 15,358 patients from 66 countries (8069 high, 4448 middle, 1744 low income). Of these, 6907 (45.0%) were followed up using telemedicine. The SSI rate reported using telemedicine was slightly lower than with in-person follow-up (11.1% versus 13.4%, $p<0.001$), which persisted after risk adjustment in a mixed-effects model (adjusted

odds ratio: 0.73, 95% confidence interval 0.63-0.84, p<0.001). This association was consistent across sensitivity and subgroup analyses, including a propensity-score matched model. In the systematic review, nine eligible non-randomised studies were identified. A pooled mean of 64% of patients underwent telemedicine follow-up. Upon meta-analysis, the SSI rate reported was lower with telemedicine (odds ratio: 0.67, 0.47-0.94) than in-person (reference) follow-up (I^2=0.45, p=0.12), although there a high risk of bias in the included studies.

Conclusions

Use of telemedicine to assess the surgical wound post-discharge is feasible, but risks underreporting of SSI. Standardised tools for remote assessment of SSI must be evaluated and adopted as telemedicine is upscaled globally.

3.2. Introduction

3.2.1 Background

3.2.1.1 Telemedicine to detect surgical wound complications

Surgical site infection (SSI) is the most common complication of surgery, with a high burden of morbidity, detriment to quality of life and economic consequences for both patients and providers [57, 61, 66, 93]. It has global impact with variation in risk across settings [23]. SSI often presents after patients have left hospital after surgery [102]. The current accepted standard in surgical site infection assessment requires an in-person review by an appropriately trained clinician, according to US Centre for Disease Control Criteria (CDC) [156]. In accordance with this framework, patients must travel back to hospital as an outpatient, or for a clinician to visit them in the community. Whilst telemedicine is an attractive target for assessment of the surgical wound, the evidence for its adoption remains limited. Quality of wound assessment is proportionate to the reported rate of SSI [92, 104]. Even in randomised trials, where SSI is a secondary rather than primary outcome the reported rate of SSI is twice as low [92]. Unstandardised telemedicine assessment therefore risks delay to timely intervention and introduction of research bias [22, 157].

3.2.1.2 The emergence of telemedicine during SARS-CoV-2

Telemedicine has now become a core component of health service delivery. During COVID-19 outbreaks, patients have been encouraged not to return to hospital for in-person assessment after surgery due to fear of SARS-CoV-2 transmission [48, 51, 139, 158]. Use of telemedicine in surgical follow-up has rapidly increased, but without opportunity for detailed evaluation [144, 159, 160]. If telemedicine assessment is not standardised it risks underreporting or misidentification of complications, and harm for patients. Better understanding the capacity to deliver telemedicine in the surgical setting and the accuracy of remote assessment for common complications will be fundamental to the pandemic recovery

effort [161, 162]. This may be particularly important in low-resource settings where, even pre-pandemic, patients had to travel longer distances to hospital and risk catastrophic expenditure as a result of a surgical episode [68].

3.2.1.3 Challenges to assessment of wound infection across settings

Hospitals in LMICs treat a high burden of surgical disease [14] and have high number of eligible patients for recruitment to pragmatic clinical trials. However, in-person assessment is labour and time intensive, and requires patients to take additional time-off work and incur costs of travel. This poses particular difficulty in LMICs where patients may live further from a specialist hospital and may already be at risk of financial catastrophe as a result of their index procedure [9, 29]. Remote methods for assessing SSI are therefore particularly important in low resource settings. Over 80% of the global population has access to a mobile telephone, opening an opportunity for remote and digital wound assessment pathways [129, 163]. Non-standardised telephone follow-up may risk reducing the validity of outcome assessment. Quality assured methods for remote wound evaluation are urgently required, both to deliver high quality research and for surveillance after hospital discharge.

3.2.2 Aims

The objectives of this cohort study, and systematic review with meta-analysis were to:

1. Explore whether wound assessment using telemedicine was feasible across different country and income settings
2. Compare the rates of SSI reported using telemedicine and in-person follow-up in risk adjusted data

3.3 Methods

3.3.1 Cohort study

This was a pre-planned, secondary analysis of a prospective, international, multi-centre cohort study conducted across the GlobalSurg Collaborative network (GlobalSurg-2) [164]. Detailed methodology for the study has been previously published [23]. The primary study aimed to describe the rates of SSI around the world and variability across United Nations Human Development Index (HDI) groups. Each contributing institution sought and obtained ethical and institutional approval according to local regulations. The cohort study was pre-registered on ClinicalTrials.gov (NCT02662231). I was a steering and writing group member for the study and co-founder of the GlobalSurg collaborative, so had direct access the cleaned source data. All data were fully anonymised, with individual patients given a unique study identifier.

3.3.1.1 Inclusion and exclusion criteria

Any centre performing elective and/or emergency abdominal surgery was invited to participate. Local investigators used consecutive sampling to include all patients undergoing elective (planned) or emergency (unplanned) gastrointestinal resection within discrete 2-week periods. Both open and minimally invasive approaches were eligible. Both adults and children (of any age) were eligible for inclusion. Patients were excluded where the primary identification for surgery was vascular, gynaecological, obstetric, urological, or for transplantation.

3.3.1.2 Data variables and data collection

Data were collected using a secure, password-encrypted, web-hosted Research Electronic Data Capture (REDCap) system. Participating centres were grouped into tertiles according to the United Nation's Human Development Index (HDI). A full description of the data variables

collected is available in the primary report of this study [23]. Data variables were chosen pragmatically to be objective, easily standardised and internationally relevant to minimise missing data and maximise data quality. Independent data validation was performed for case ascertainment and data accuracy.

3.3.1.3 Classification of follow-up method

Investigators were asked to actively monitor patients up to 30-day after surgery, and performed an assessment for SSI at 30-days after surgery by one of three methods: (1) Telemedicine review (telephone and/or video assessment), which was not standardised in the study, but was performed according to local practice and informed by CDC criteria; (2) In-person clinical review, either during an outpatient clinic appointment or a community visit in accordance with CDC criteria; (3) Inpatient only, with in-hospital assessment and review of patient notes and electronic records up to 30-days after surgery (i.e., no contact made after discharge). Patients that remained an inpatient at 30-days postoperatively were excluded from analysis (including those that were readmitted and were in hospital at 30-days after surgery). Patients that were readmitted and discharged before 30-days after surgery had an independent 30-day assessment of their SSI status.

3.3.1.4 Outcome measures

The primary outcome measure was surgical site infection reported up to 30-days after surgery defined according to the US Centre for Disease Control criteria [93]. I included both superficial and deep infections but excluded organ space infection, which has a different biological mechanism (e.g., anastomotic leak, gross contamination). Training in the CDC criteria for SSI diagnosis was provided to all investigators using an online training module. The secondary outcome measure was 30-day postoperative mortality rate (POMR) with day 0 as the day of surgery.

3.3.1.5 Statistical analysis

Differences in characteristics and the reported rates of SSI between telemedicine, in-person and inpatient only follow-up were tested with the Pearson $\chi 2$ test for categorical variables and with the Kruskal-Wallis test for continuous variables. There is likely to be variation in the methods of adoption of telemedicine across different resource settings. Global variation was explored by stratifying comparisons between high-HDI, middle-HDI and low-HDI countries to explore whether patterns were consistent across health systems.

Multilevel logistic regression models were constructed to explore associations between the method of follow-up and the SSI rate reported. Characteristics and outcomes of patients with no post-discharge assessment (inpatient only) were described for transparency, but were excluded from multivariable modelling. Adjustment for case mix was performed using patient, disease, and operation-specific factors, informed by a causal model constructed to inform covariable selection and presented using a Directed Acyclic Graph (DAG). Country was incorporated as random effects with a constrained gradient. Discrimination of the model was determined using the C-statistic (area under the receiver operating curve characteristic). Model coefficients were presented as adjusted odds ratios (OR) with 95% confidence intervals (CI).

To account for death after surgery as a competing risk, patients who died before 30-day follow-up were excluded in a sensitivity analysis of the primary analysis. A second sensitivity analysis was conducted including only patients with a postoperative length of stay of 14 days or less to explore associations in group who were unlikely to have experienced significant complications of surgery. A third sensitivity analysis included elective cancer surgery only, to explore associations in a more heterogenous, high-risk group.

The final sensitivity analysis was performed in propensity score matched (PSM) groups to address a risk of selection bias and counterfactuals [165, 166]. Propensity score matching (PSM) involved building a binary logistic regression model to explore predictors of 30-day SSI using a four-step approach.

- First, propensity scores were developed by including HDI tertile, age, sex, ASA grade, urgency, and approach in a binary logistic regression model.

- Second, this logistic regression model was then used to calculate propensity scores using the *MatchIt* package. The calculated weights represent the estimated probability of 30-day SSI based on each participant's matched characteristics. Participants with in-person review were matched to participants with telemedicine review using coarse exact matching (packages: cem, randomForests). This is a form of matching (without replacement) where covariates are coarsened into bins, and the coarsened covariates are used to created subclasses based on combinations of the coarsened covariate levels. It is deemed to be most appropriate: (1) where there are many covariates so exact matching may not be possible; (2) when evaluating extreme counterfactuals [165].

- Third, bias-corrected accelerated (BCa) bootstrap 95% confidence intervals were derived for model coefficients (package: boot).

- Finally, to ensure that the matching was effective, we checked the balance of covariates between the patients with in-person and telemedicine review in the matched data.

Subgroup analyses were performed for the primary analysis across high versus middle or low HDI countries to explore the effect of country resourcing on the association between method of follow-up and SSI detection rate.

Finally, to explore risk of reverse causation (i.e., patients with serious postoperative complications seeking in-person review) I looked for associations between follow-up method and 30-day POMR. All analyses were done using the R Foundation Statistical Program version 4.1.1 (packages: finalfit, tidyverse, boot, MatchIt, cem, randomForests).

3.3.2 Systematic review and meta-analysis

A systematic database search was performed according to a pre-published protocol (PROSPERO:192596) and followed Preferred Reporting Items for Systematic Reviews and Meta-analysis (PRISMA) guidance. Studies reporting surgical site infection rates reported using telemedicine and in-person assessment after non-cardiac surgery were included. Data extracted from published studies was combined with cohort data and a meta-analysis performed with all available data.

3.3.2.1 Database search and report characteristics

A search strategy was constructed using Medline, EMBASE and PubMed to identify two key concepts within published literature: (1) surgical site infection and (2) telemedicine. The full search strategy for the review is presented in *Box 3.1.*

All included studies assessed a proportion of patients both by telemedicine and in-person follow-up. Observational studies (retrospective or prospective) or prospective randomised trials in non-cardiac surgery (including caesarean section) were included where they reported surgical site infection rates in adult patients (greater than 16 years). Both planned elective

and unplanned emergency surgery were eligible. Studies were only included if they reported SSI rates up to a maximum 90-days after surgery, and assessed a proportion of patients using both telemedicine and in-person follow-up within the study. No limitations were imposed to the questionnaire schedule used or methodology for in-person follow-up.

Studies reporting either within-subject SSI rates (i.e., same patient received both telephone and in-person follow-up) or between-subject SSI rates (i.e. different patients received either telephone or in-person follow-up) at the same time point were both eligible. Studies were excluded where SSI rates were reported at different time points (e.g., in-hospital versus 30-days, where SSI rates could not be disaggregated between remote and in-person follow-up, between abdominal surgery and other surgery types, or from other postoperative complications. Studies were also excluded which reported ongoing follow-up of patients that had already suffered a surgical site infection. To provide contemporaneous estimates, studies published before 2010 (i.e., in the 10-years before study initiation) were excluded.

3.3.2.2 Outcome measures

The primary outcome measure was the rate of surgical site infection reported up to 30-days after surgery in the study. In the meta-analysis, this was defined pragmatically according to any classification system adopted (US CDC, ASEPSIS or Public Health England), or diagnosis by a clinician. The secondary outcome measure was the proportion of patients undergoing telemedicine versus in-person follow-up.

3.3.2.3 Data extraction and analysis

Abstracts were screened by two independent reviewers using Rayyan QCRI and full texts were retrieved for all studies that potentially met the inclusion criteria [167]. Any disagreement on eligibility of abstracts and/or full papers was resolved through consensus

discussion with a third reviewer. Data on the proportion of patients with an SSI reported by telemedicine and by in-person follow-up, and the proportion of patients that underwent telemedicine follow-up were extracted from eligible study and combined with data from the cohort study in *Part 1.* Data extraction was performed according to a pre-specified framework using Google Sheets® (Google LLC, Mountain View, USA) to support collaborative working. Data were cross-checked for accuracy by two independent researchers.

Data analysis was performed using R Foundation Statistical Program version 3.1 (packages: meta, metabin). Outcome measures were quantitatively summarised where data were available. Firstly, meta-analysis performed to estimate the pooled mean proportions of patients followed-up using telemedicine. Secondly, meta-analysis was performed to compare the reported SSI rates with telemedicine and in-person follow-up. Heterogeneity among study estimates was quantified using the I^2 and an associated test for heterogeneity. As heterogeneity was likely to be high, the DerSimonian and Laird random effects (RE) method was used to pool estimates, with inverse-variance weights. A subgroup analysis was performed of data from high versus low and middle-income countries. (packages: metaprop, meta).

3.3.2.4 Risk of bias assessment

Risk of bias was assessed for non-randomised studies using the ROBINS-I tool. As this was not a clinical effectiveness study, a GRADE level of evidence assessment was not deemed to be required.

Box 3.1. Full search terms included in systematic review

Concept 1: Surgical site infection

"Surgical site infect*" or "Surgical site complicat*" or "Surgical wound infect*" or "Surgical wound complicat*"or SSI or SSIs or SWI or SWIs or "post-operative infect*" or "postoperative infect*" or "post-op wound infect*" or "postop wound infect*" or "post-op infection" or "postop infection" or "post-operative wound infect*" or "postoperative wound infect*" or "Surgical wound dehiscence" OR "postoperative complication*"

Concept 2: Telemedicine

"phone" or "telephone" or "smartphone" or "cellphone" or "telemedicine" or "teleconsult" or "phone consult" or "tele* consult*" or "tele-nursing" or "mobile consult" or "remote consultation" or "phone interview" or "telephone interview*"or "phone call" or "tele* call" or "telehealth" or "tele-health" or "ehealth" or "e-health" or "mhealth" or "m-health" or "mobile health" or "telecommunication*" or "hotline" or "helpline" or "videoconference" or "mobile technolog*" or "telephone triage" or "e-referral" or "text messag*" or "text" or "sms" or "txt" or "interviews as topic"

3.4 Results

3.4.1 Cohort study

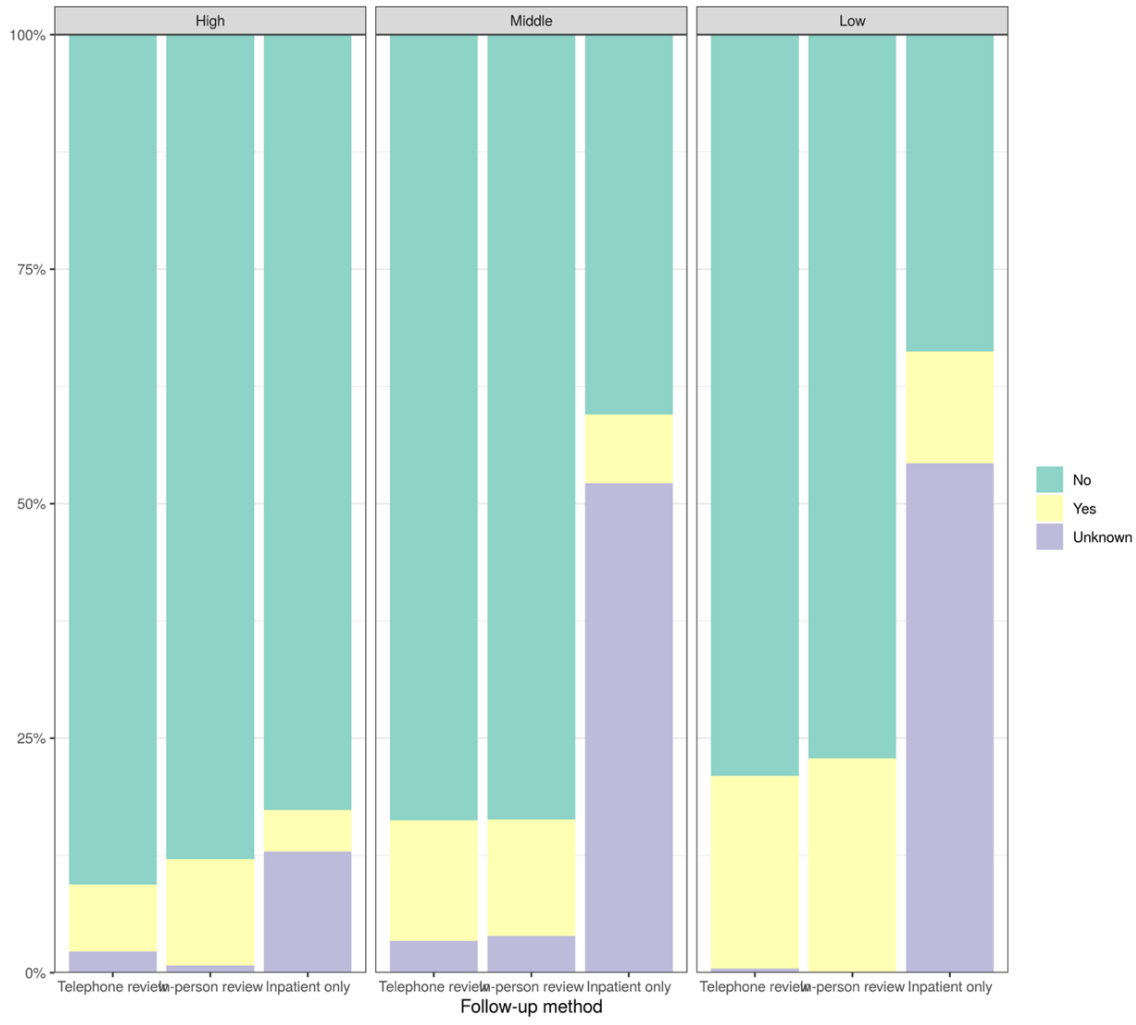
3.4.1.1 Methods of post-discharge follow-up

Overall, 15358 of 16015 patients (95.9%) were discharged before 30-days postoperatively and were included in this analysis. Of these patients, 6907 underwent telemedicine review (45.0%), 6171 in-person review (40.2%), and 2280 inpatient only assessment (14.8%).

3.4.1.2 Use of telemedicine

Telemedicine was used across 51 of 66 contributing countries spanning high (n=23), middle (n=16) and low-HDI (n=12) settings. In high-HDI settings 36.7% (3113/8492) of included patients were followed-up using telemedicine. The telemedicine follow-up rates were higher in both middle-HDI (61.4%, 3075/5006), and low-HDI settings (38.7%, 719/1860). Telemedicine was used for patients of both sexes (41.8%, of male patients, and 47.1% of female patients) and all age ranges, including both the youngest (2 to 20 years; 46.6%, 1003/2151) and oldest age groups (80 to 100 years; 31.8%, 154/485). Telemedicine was used to follow-up patients with a range of ASA grades, underlying pathologies and presenting for both elective and emergency care (*Table 3.1*).

Figure 3.1. Surgical site infection rates by method of follow-up across high-, middle- and low-income settings.



3.4.1.3 Characteristics of patients by follow-up group

There were significant differences in the baseline risk characteristics of the groups that underwent telemedicine, in-person, and inpatient only follow-up. Notably, patients that underwent surgery for malignancy were less likely to have telephone review than in-person clinical review (34.6% versus 55.6%, p<0.001). Patients undergoing emergency surgery were more likely to have telephone review than clinical review (43.8% versus 38.3%, p<0.001). Patients from high income countries (p<0.001), with gallstone disease or appendicitis as their indication for surgery (p<0.001), or that underwent emergency surgery (p<0.001) were most likely to have inpatient only assessment.

3.4.1.4 Reporting of surgical site infection

In this study, 11.2% (1721/15358) of patients had an SSI reported, and 5.5% (843/15358) had an unknown SSI status. The rate of SSI reported was slightly lower with telemedicine (11.1%, 766/6907) and lower with inpatient only follow-up (5.7%, 129/2280) than with in-person follow-up (13.4%, 826/6171, p<0.001). Of patients that had SSI reported, 44.5% (766/1721) of diagnoses were made using telemedicine, 48.0% (826/1721) in-person and 7.5% (129/1721) with inpatient only assessment.

Figure 3.1 shows the unadjusted SSI rates by method of follow-up, stratified by HDI tertile. 'Unknown' SSI status was higher in groups undergoing inpatient only assessment than telephone review or in-person clinical review groups; this difference was largest across middle- and low-HDI settings (p<0.001). Small differences were observed in reported SSI rates (unadjusted) following telemedicine and in-person review across high- (7.3% (222/3043) versus 11.4% (432/3793)), middle- (13.3% (396/2971) versus 12.9% (169/1305)) and low-HDI (20.7% (148/716) versus 22.8% (225/989)) countries. Unadjusted SSI rates with

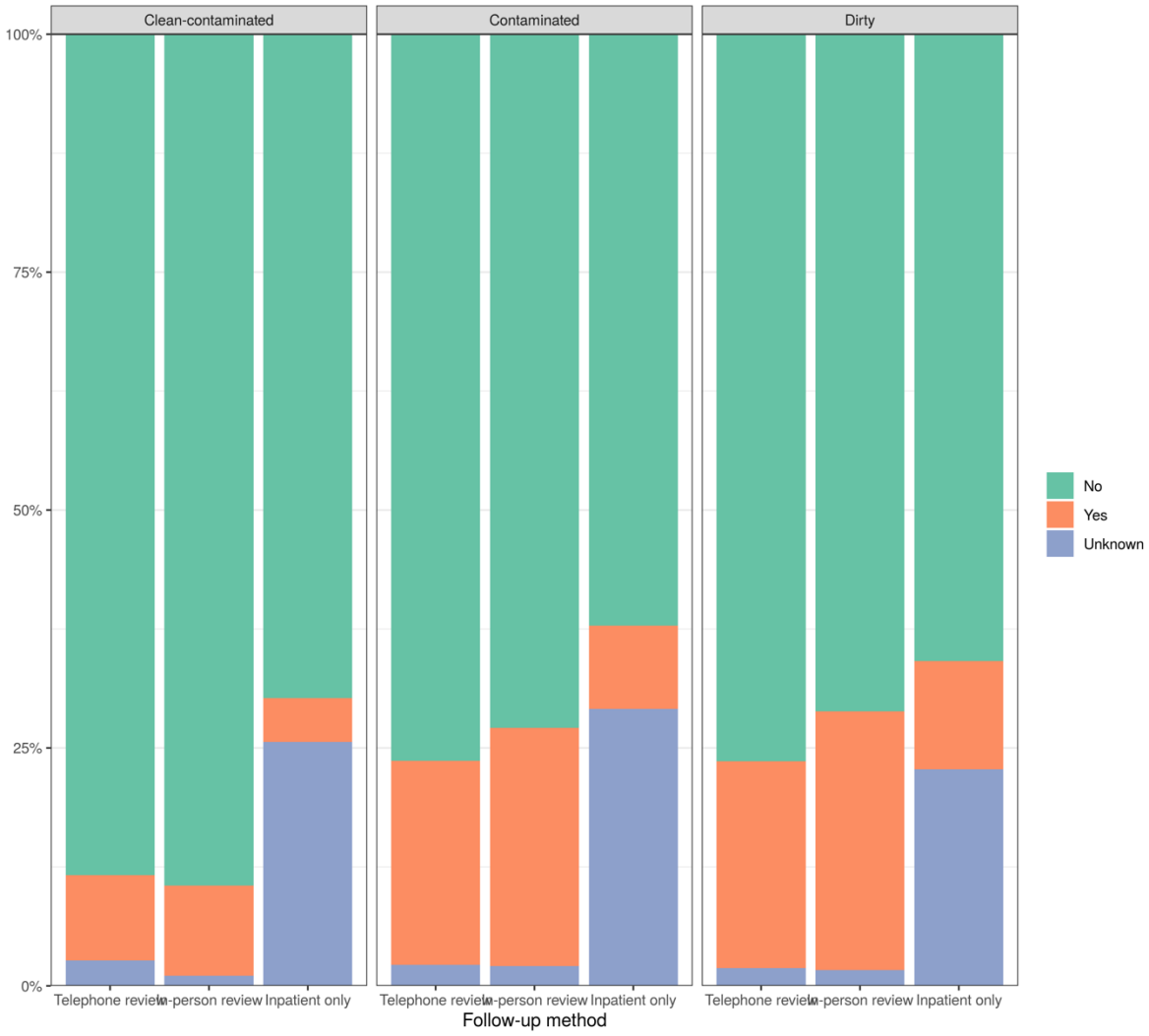
telephone follow-up and in-person clinical follow-up were comparable across strata of intra-abdominal contamination (*Figure 3.2*).

Inpatient only assessment had a lower recorded rate of SSI in high-income settings (5.1% (69/1355)), but a higher rate of SSI reported in middle- (15.3% (42/274)), and low-income settings (26.1% (18/69)) respectively.

3.4.1.5 Postoperative mortality rates

The overall 30-day postoperative mortality rate was 2.4% (352/14871, missing: 487). This was lower in high (1.9%, 157/8368) than in middle (2.3%, 105/4570) and low (4.7%, 90/1933) income countries, and lower after clean contaminated (1.4%, 161/11572) than after contaminated (4.3%, 79/1841) or dirty (8.4%, 109/1300) surgery.

Figure 3.2. Surgical site infection by method of follow-up in patients with different levels of intraabdominal contamination.



3.4.1.5 Multivariable models for reporting of surgical site infection

A Directed Acyclic Graph displaying a proposed casual model between method of follow-up and SSI test positive ('observed SSI') is displayed in *Figure 3.3*. Upon univariable analysis, the odds of reporting an SSI following telemedicine assessment (OR 0.81, 0.73-0.90, p<0.001) was lower than in-person (reference). After multivariable adjustment telemedicine assessment was associated with lower odds of reporting SSI than in-person review (OR 0.73, 0.64-0.84, p<0.001). *Figure 3.4* below displays a forest plot of the model. A lower odds ratio conveys a lower adjusted odds of reporting a surgical site infection (i.e., assumed to be under-detection of the true SSI rate). The full model summary is also presented in *Table 3.1*.

Figure 3.3. Proposed casual model

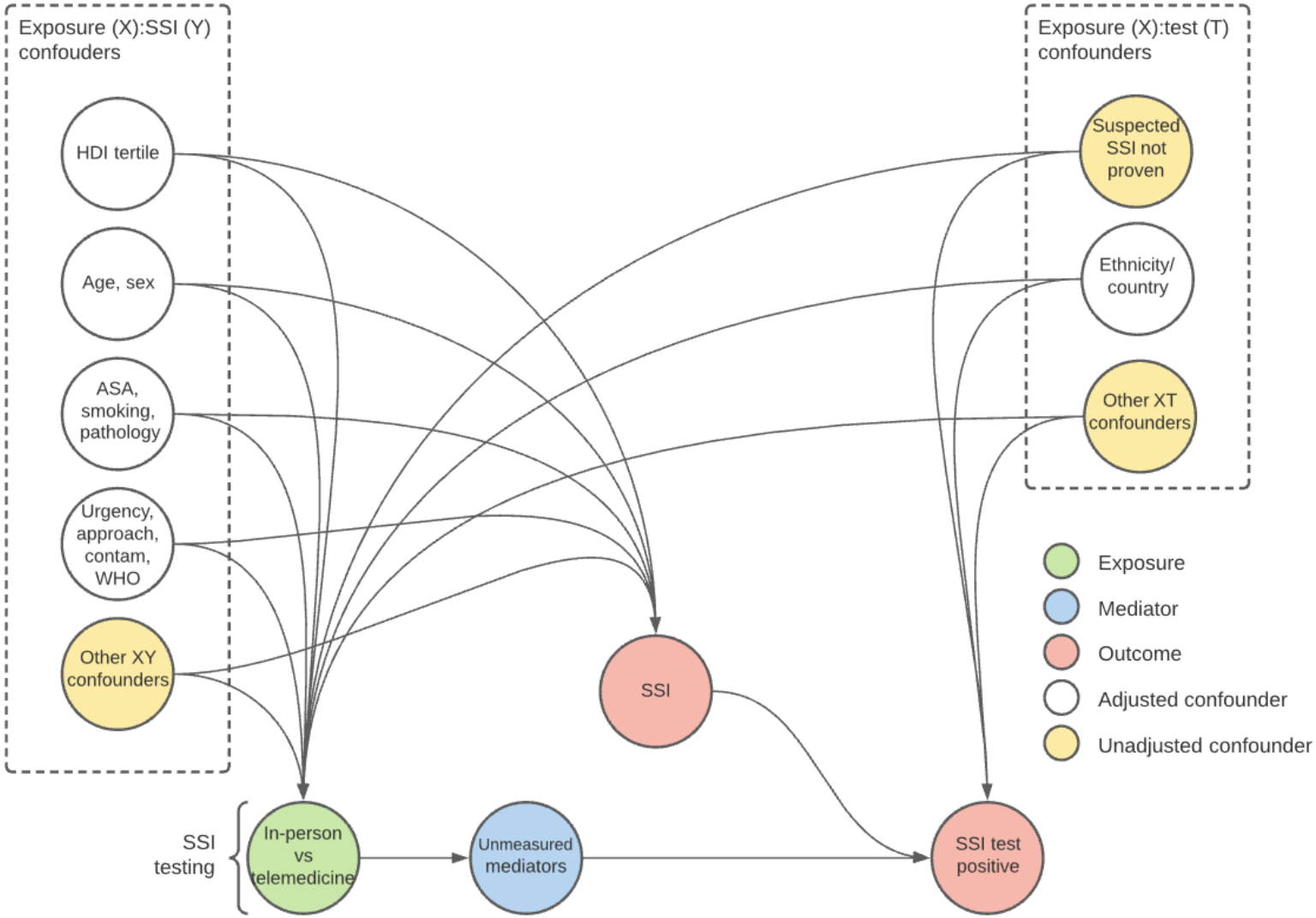


Figure 3.4. Forest plot of factors associated with reporting of post-discharge surgical site infection after abdominal surgery.

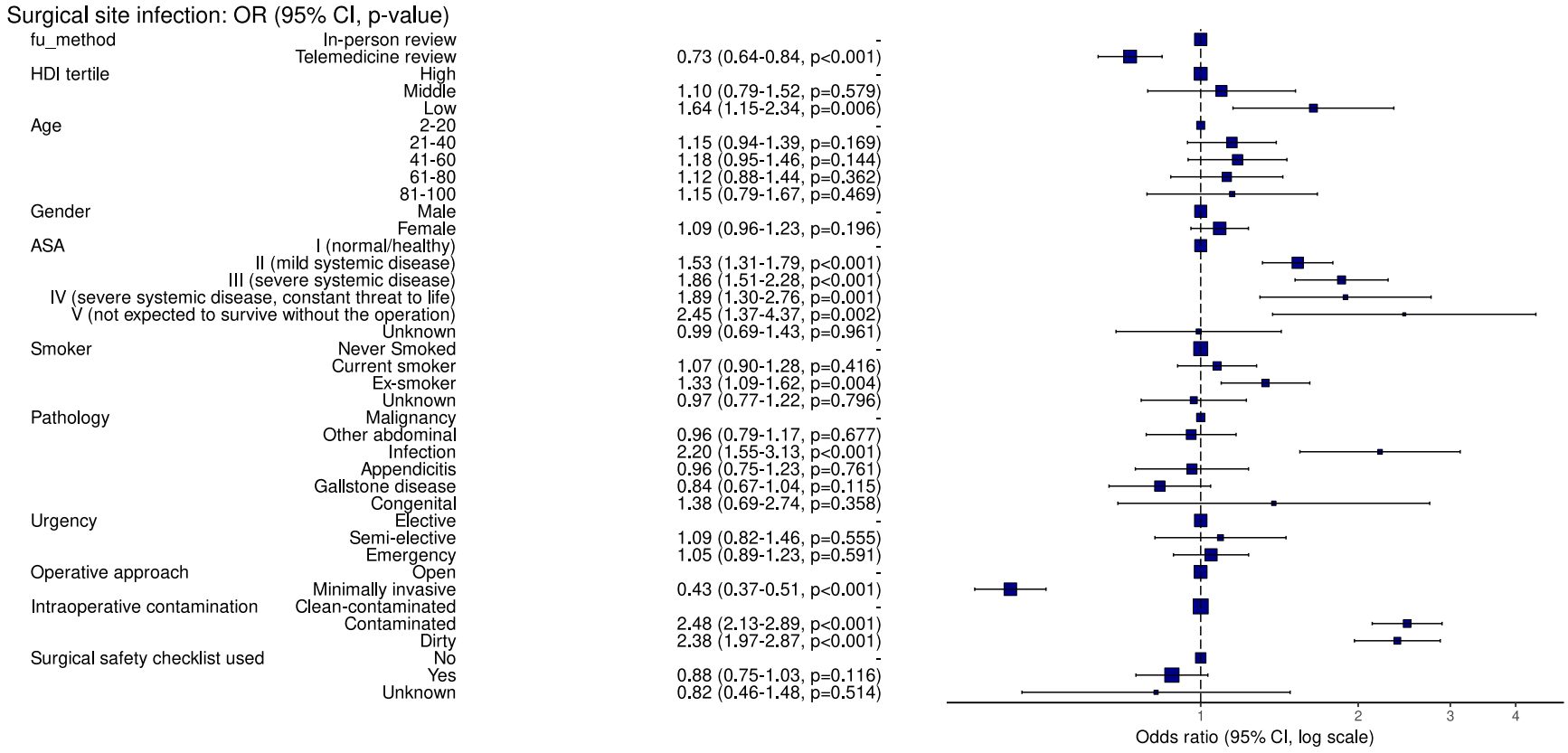


Table 3.1. Mixed effects model demonstrating the association between reported surgical site infection rates and method of follow-up.

Factor	Level	No SSI	SSI	OR (univariable)	OR (multilevel)
Follow-up	In-person review	5232 (86.4)	824 (13.6)	-	-
	Telemedicine review	5958 (88.6)	763 (11.4)	0.81 (0.73-0.90, p<0.001)	0.73 (0.64-0.84, p<0.001)
HDI tertile	High	6159 (90.4)	653 (9.6)	-	-
	Middle	3702 (86.8)	562 (13.2)	1.43 (1.27-1.61, p<0.001)	1.10 (0.79-1.52, p=0.579)
	Low	1329 (78.1)	372 (21.9)	2.64 (2.29-3.04, p<0.001)	1.64 (1.15-2.34, p=0.006)
Age group	<20	1550 (87.3)	226 (12.7)	-	-
	20-39	3501 (88.1)	471 (11.9)	0.92 (0.78-1.09, p=0.352)	1.15 (0.94-1.39, p=0.169)
	40-59	3153 (87.8)	439 (12.2)	0.95 (0.81-1.14, p=0.598)	1.18 (0.95-1.46, p=0.144)
	60-79	2393 (87.0)	356 (13.0)	1.02 (0.85-1.22, p=0.825)	1.12 (0.88-1.44, p=0.362)
	80-100	343 (85.1)	60 (14.9)	1.20 (0.88-1.62, p=0.246)	1.15 (0.79-1.67, p=0.469)
Gender	Male	4922 (86.3)	779 (13.7)	-	-
	Female	5615 (88.7)	716 (11.3)	0.81 (0.72-0.90, p<0.001)	1.09 (0.96-1.23, p=0.196)
ASA	I	5005 (89.3)	602 (10.7)	-	-
	II	4158 (88.0)	569 (12.0)	1.14 (1.01-1.28, p=0.038)	1.53 (1.31-1.79, p<0.001)
	III	1499 (83.6)	293 (16.4)	1.63 (1.40-1.89, p<0.001)	1.86 (1.51-2.28, p<0.001)
	IV	185 (78.1)	52 (21.9)	2.34 (1.68-3.19, p<0.001)	1.89 (1.30-2.76, p=0.001)
	V	47 (64.4)	26 (35.6)	4.60 (2.79-7.42, p<0.001)	2.45 (1.37-4.37, p=0.002)
	Unknown	296 (87.1)	44 (12.9)	1.24 (0.88-1.70, p=0.205)	0.99 (0.69-1.43, p=0.961)
Smoker	Never Smoked	7358 (88.0)	1006 (12.0)	-	-
	Current smoker	1657 (86.7)	255 (13.3)	1.13 (0.97-1.30, p=0.116)	1.07 (0.90-1.28, p=0.416)
	Ex-smoker	1164 (85.0)	205 (15.0)	1.29 (1.09-1.51, p=0.002)	1.33 (1.09-1.62, p=0.004)
	Unknown	1011 (89.3)	121 (10.7)	0.88 (0.71-1.06, p=0.191)	0.97 (0.77-1.22, p=0.796)
Pathology	Malignancy	1643 (85.8)	272 (14.2)	-	-
	Other abdominal	2700 (84.7)	488 (15.3)	1.09 (0.93-1.28, p=0.284)	0.96 (0.79-1.17, p=0.677)
	Infection	148 (61.7)	92 (38.3)	3.75 (2.80-5.01, p<0.001)	2.20 (1.55-3.13, p<0.001)
	Appendicitis	3052 (87.7)	429 (12.3)	0.85 (0.72-1.00, p=0.050)	0.96 (0.75-1.23, p=0.761)
	Gallstone disease	3475 (92.6)	277 (7.4)	0.48 (0.40-0.57, p<0.001)	0.84 (0.67-1.04, p=0.115)
	Congenital	170 (86.3)	27 (13.7)	0.96 (0.61-1.44, p=0.849)	1.38 (0.69-2.74, p=0.358)
Urgency	Elective	5620 (89.9)	634 (10.1)	-	-
	Semi-elective	564 (88.3)	75 (11.7)	1.18 (0.91-1.51, p=0.205)	1.09 (0.82-1.46, p=0.555)
	Emergency	5006 (85.1)	878 (14.9)	1.55 (1.39-1.73, p<0.001)	1.05 (0.89-1.23, p=0.591)
Approach	Open	5255 (81.7)	1176 (18.3)	-	-
	Minimally invasive	5935 (93.5)	411 (6.5)	0.31 (0.27-0.35, p<0.001)	0.43 (0.37-0.51, p<0.001)
Contamination	Clean-contaminated	9107 (90.6)	944 (9.4)	-	-
	Contaminated	1178 (75.9)	374 (24.1)	3.06 (2.68-3.50, p<0.001)	2.48 (2.13-2.89, p<0.001)
	Dirty	788 (74.8)	265 (25.2)	3.24 (2.78-3.78, p<0.001)	2.38 (1.97-2.87, p<0.001)
WHO Checklist	No	3258 (84.9)	579 (15.1)	-	-
	Yes	7783 (88.7)	994 (11.3)	0.72 (0.64-0.80, p<0.001)	0.88 (0.75-1.03, p=0.116)
	Unknown	148 (91.4)	14 (8.6)	0.53 (0.29-0.89, p=0.026)	0.82 (0.46-1.48, p=0.514)

3.4.1.6 Sensitivity analyses

This association was consistent across several sensitivity analyses. Firstly, in patients that were alive at 30-days after surgery only (*Table 3.2*). This was performed to address competing risk of death. Secondly, in a sensitivity analysis in patients that had a postoperative length of stay of 14 days or less (*Table 3.3*). This was performed to identify a lower risk group, and address a potential risk of selection bias between those that were undergoing in-person and telemedicine follow-up. Thirdly, in a sensitivity analysis in patients undergoing elective cancer surgery only (*Table 3.4*). This was chosen to select a heterogeneous, higher risk group and explore whether the association remained robust.

Table 3.2. Sensitivity analysis for the primary model of the association between reported surgical site infection rates and method of follow-up in patients that were alive at 30-days after surgery.

Factor	Level	No SSI	SSI	OR (univariable)	OR (multilevel)
Follow-up	In-person review	5176 (86.6)	799 (13.4)	-	-
	Telemedicine review	5903 (88.8)	748 (11.2)	0.82 (0.74-0.91, p<0.001)	0.74 (0.64-0.85, p<0.001)
HDI tertile	High	6112 (90.6)	635 (9.4)	-	-
	Middle	3655 (87.0)	547 (13.0)	1.44 (1.28-1.63, p<0.001)	1.08 (0.77-1.50, p=0.664)
	Low	1312 (78.2)	365 (21.8)	2.68 (2.32-3.08, p<0.001)	1.65 (1.15-2.37, p=0.006)
Age group	Feb-20	1545 (87.3)	225 (12.7)	-	-
	21-40	3488 (88.3)	464 (11.7)	0.91 (0.77-1.08, p=0.297)	1.14 (0.94-1.39, p=0.179)
	41-60	3130 (87.9)	432 (12.1)	0.95 (0.80-1.13, p=0.541)	1.19 (0.96-1.49, p=0.111)
	61-80	2349 (87.4)	338 (12.6)	0.99 (0.83-1.18, p=0.896)	1.14 (0.89-1.46, p=0.303)
	81-100	333 (86.0)	54 (14.0)	1.11 (0.80-1.52, p=0.510)	1.13 (0.77-1.66, p=0.538)
Gender	Male	4861 (86.6)	753 (13.4)	-	-
	Female	5571 (88.8)	705 (11.2)	0.82 (0.73-0.91, p<0.001)	1.10 (0.97-1.25, p=0.150)
ASA	I	4980 (89.3)	598 (10.7)	-	-
	II	4145 (88.1)	560 (11.9)	1.13 (1.00-1.27, p=0.059)	1.52 (1.30-1.77, p<0.001)
	III	1469 (84.2)	276 (15.8)	1.56 (1.34-1.82, p<0.001)	1.80 (1.46-2.22, p<0.001)
	IV	164 (78.5)	45 (21.5)	2.29 (1.61-3.18, p<0.001)	1.88 (1.27-2.80, p=0.002)
	V	27 (52.9)	24 (47.1)	7.40 (4.22-12.92, p<0.001)	3.81 (1.97-7.40, p<0.001)
	Unknown	294 (87.2)	43 (12.8)	1.22 (0.86-1.68, p=0.243)	0.97 (0.67-1.41, p=0.886)
Smoker	Never Smoked	7279 (88.0)	988 (12.0)	-	-
	Current smoker	1647 (86.7)	252 (13.3)	1.13 (0.97-1.31, p=0.113)	1.08 (0.91-1.29, p=0.389)
	Ex-smoker	1151 (85.6)	194 (14.4)	1.24 (1.05-1.46, p=0.011)	1.32 (1.08-1.61, p=0.007)
	Unknown	1002 (89.9)	113 (10.1)	0.83 (0.67-1.02, p=0.077)	0.96 (0.76-1.21, p=0.718)
Pathology	Malignancy	1608 (86.0)	261 (14.0)	-	-
	Other abdominal	2642 (84.9)	470 (15.1)	1.10 (0.93-1.29, p=0.272)	0.95 (0.78-1.16, p=0.627)
	Infection	145 (62.5)	87 (37.5)	3.70 (2.74-4.96, p<0.001)	2.10 (1.46-3.02, p<0.001)
	Appendicitis	3050 (87.7)	429 (12.3)	0.87 (0.74-1.02, p=0.089)	0.95 (0.74-1.22, p=0.692)
	Gallstone disease	3469 (92.7)	272 (7.3)	0.48 (0.40-0.58, p<0.001)	0.82 (0.65-1.03, p=0.088)
	Congenital	163 (86.2)	26 (13.8)	0.98 (0.62-1.49, p=0.937)	1.38 (0.70-2.76, p=0.355)
Urgency	Elective	5589 (89.9)	625 (10.1)	-	-
	Semi-elective	557 (88.6)	72 (11.4)	1.16 (0.89-1.49, p=0.273)	1.09 (0.81-1.46, p=0.568)
	Emergency	4933 (85.3)	850 (14.7)	1.54 (1.38-1.72, p<0.001)	1.05 (0.89-1.24, p=0.573)
Approach	Open	5160 (81.9)	1142 (18.1)	-	-
	Minimally invasive	5919 (93.6)	405 (6.4)	0.31 (0.27-0.35, p<0.001)	0.43 (0.37-0.51, p<0.001)
Contamination	Clean-contaminated	9042 (90.7)	929 (9.3)	-	-
	Contaminated	1157 (76.3)	360 (23.7)	3.03 (2.64-3.47, p<0.001)	2.49 (2.13-2.90, p<0.001)
	Dirty	763 (75.0)	254 (25.0)	3.24 (2.77-3.79, p<0.001)	2.38 (1.96-2.89, p<0.001)
WHO Checklist	No	3218 (85.1)	563 (14.9)	-	-
	Yes	7712 (88.8)	970 (11.2)	0.72 (0.64-0.80, p<0.001)	0.90 (0.77-1.06, p=0.192)
	Unknown	148 (91.4)	14 (8.6)	0.54 (0.30-0.91, p=0.030)	0.83 (0.46-1.50, p=0.534)

Table 3.3. Sensitivity analysis for the primary model of the association between reported surgical site infection rates and method of follow-up in patients that had a postoperative length of stay of 14 days or less.

Factor	Level	No SSI	SSI	OR (univariable)	OR (multilevel)
Follow-up	In-person review	4972 (89.3)	597 (10.7)	-	-
	Telemedicine review	5795 (89.6)	675 (10.4)	0.97 (0.86-1.09, p=0.609)	0.82 (0.70-0.96, p=0.011)
HDI tertile	High	5883 (92.2)	500 (7.8)	-	-
	Middle	3622 (88.1)	488 (11.9)	1.59 (1.39-1.81, p<0.001)	1.12 (0.78-1.60, p=0.535)
	Low	1262 (81.6)	284 (18.4)	2.65 (2.26-3.10, p<0.001)	1.59 (1.07-2.35, p=0.022)
Age group	<20	1526 (88.9)	191 (11.1)	-	-
	20-39	3455 (89.3)	415 (10.7)	0.96 (0.80-1.15, p=0.657)	1.21 (0.99-1.49, p=0.066)
	40-59	3036 (89.6)	351 (10.4)	0.92 (0.77-1.11, p=0.405)	1.22 (0.97-1.54, p=0.093)
	60-79	2237 (89.6)	261 (10.4)	0.93 (0.77-1.14, p=0.486)	1.18 (0.90-1.54, p=0.230)
	80-100	300 (91.5)	28 (8.5)	0.75 (0.48-1.11, p=0.166)	0.91 (0.56-1.46, p=0.689)
Gender	Male	4693 (88.6)	603 (11.4)	-	-
	Female	5452 (90.0)	603 (10.0)	0.86 (0.76-0.97, p=0.014)	1.12 (0.98-1.29, p=0.098)
ASA	I	4924 (90.2)	533 (9.8)	-	-
	II	3997 (90.0)	445 (10.0)	1.03 (0.90-1.17, p=0.678)	1.52 (1.29-1.80, p<0.001)
	III	1377 (87.2)	202 (12.8)	1.36 (1.14-1.61, p=0.001)	1.82 (1.45-2.29, p<0.001)
	IV	147 (81.7)	33 (18.3)	2.07 (1.39-3.02, p<0.001)	2.00 (1.28-3.15, p=0.003)
	V	38 (66.7)	19 (33.3)	4.62 (2.59-7.97, p<0.001)	3.07 (1.59-5.92, p=0.001)
	Unknown	284 (87.9)	39 (12.1)	1.27 (0.88-1.77, p=0.178)	0.99 (0.68-1.45, p=0.963)
Smoker	Never Smoked	7108 (89.4)	844 (10.6)	-	-
	Current smoker	1596 (88.4)	210 (11.6)	1.11 (0.94-1.30, p=0.210)	1.09 (0.91-1.32, p=0.355)
	Ex-smoker	1092 (88.9)	137 (11.1)	1.06 (0.87-1.28, p=0.573)	1.20 (0.96-1.50, p=0.118)
	Unknown	971 (92.3)	81 (7.7)	0.70 (0.55-0.89, p=0.004)	0.87 (0.67-1.15, p=0.330)
Pathology	Malignancy	1478 (88.8)	186 (11.2)	-	-
	Other abdominal	2525 (88.1)	341 (11.9)	1.07 (0.89-1.30, p=0.466)	0.92 (0.73-1.16, p=0.471)
	Infection	134 (68.4)	62 (31.6)	3.68 (2.61-5.14, p<0.001)	2.11 (1.41-3.15, p<0.001)
	Appendicitis	3031 (88.3)	402 (11.7)	1.05 (0.88-1.27, p=0.577)	1.09 (0.83-1.44, p=0.546)
	Gallstone disease	3441 (93.1)	257 (6.9)	0.59 (0.49-0.72, p<0.001)	0.89 (0.70-1.15, p=0.379)
	Congenital	157 (87.2)	23 (12.8)	1.16 (0.72-1.82, p=0.520)	1.61 (0.77-3.37, p=0.205)
Urgency	Elective	5426 (91.4)	508 (8.6)	-	-
	Semi-elective	533 (89.7)	61 (10.3)	1.22 (0.92-1.61, p=0.160)	1.23 (0.91-1.68, p=0.181)
	Emergency	4808 (87.2)	703 (12.8)	1.56 (1.38-1.76, p<0.001)	1.07 (0.89-1.28, p=0.482)
Approach	Open	4939 (84.6)	897 (15.4)	-	-
	Minimally invasive	5828 (94.0)	375 (6.0)	0.35 (0.31-0.40, p<0.001)	0.48 (0.40-0.57, p<0.001)
Contamination	Clean-contaminated	8825 (91.6)	811 (8.4)	-	-
	Contaminated	1110 (80.1)	276 (19.9)	2.71 (2.33-3.14, p<0.001)	2.32 (1.95-2.74, p<0.001)
	Dirty	719 (79.7)	183 (20.3)	2.77 (2.31-3.30, p<0.001)	2.21 (1.79-2.74, p<0.001)
WHO Checklist	No	3162 (86.9)	478 (13.1)	-	-
	Yes	7463 (90.5)	784 (9.5)	0.69 (0.62-0.78, p<0.001)	0.88 (0.74-1.05, p=0.147)
	Unknown	141 (93.4)	10 (6.6)	0.47 (0.23-0.85, p=0.022)	0.77 (0.39-1.51, p=0.441)

Table 3.4. Sensitivity analysis for the primary model of the association between reported surgical site infection rates and method of follow-up in elective cancer surgery patients.

Factor	Level	No SSI	SSI	OR (univariable)	OR (multilevel)
Follow-up	In-person review	2354 (88.9)	293 (11.1)	-	-
	Telemedicine review	2728 (90.9)	273 (9.1)	0.80 (0.68-0.96, p=0.014)	0.75 (0.60-0.94, p=0.012)
HDI tertile	High	3073 (91.4)	288 (8.6)	-	-
	Middle	1602 (89.4)	190 (10.6)	1.27 (1.04-1.53, p=0.017)	0.92 (0.61-1.40, p=0.696)
	Low	407 (82.2)	88 (17.8)	2.31 (1.77-2.98, p<0.001)	1.99 (1.25-3.18, p=0.004)
Age group	<20	0 (0.0)	0 (0.0)	-	-
	20-39	1501 (91.0)	149 (9.0)	-	-
	40-59	1859 (89.9)	209 (10.1)	1.13 (0.91-1.41, p=0.269)	0.97 (0.75-1.24, p=0.787)
	60-79	1554 (89.4)	184 (10.6)	1.19 (0.95-1.50, p=0.129)	0.86 (0.64-1.15, p=0.308)
	80-100	168 (87.5)	24 (12.5)	1.44 (0.89-2.24, p=0.121)	0.95 (0.56-1.62, p=0.853)
Gender	Male	1832 (88.3)	243 (11.7)	-	-
	Female	2888 (91.0)	284 (9.0)	0.74 (0.62-0.89, p=0.001)	0.93 (0.76-1.15, p=0.508)
ASA	I	1689 (91.5)	157 (8.5)	-	-
	II	2455 (90.3)	265 (9.7)	1.16 (0.95-1.43, p=0.157)	1.42 (1.11-1.83, p=0.006)
	III	869 (87.2)	127 (12.8)	1.57 (1.23-2.01, p<0.001)	1.84 (1.33-2.53, p<0.001)
	IV	69 (80.2)	17 (19.8)	2.65 (1.48-4.52, p=0.001)	2.11 (1.12-3.96, p=0.021)
	V	0 (0.0)	0 (0.0)	-	-
Smoker	Never Smoked	3220 (90.5)	339 (9.5)	-	-
	Current smoker	721 (89.5)	85 (10.5)	1.12 (0.87-1.43, p=0.377)	1.10 (0.82-1.46, p=0.525)
	Ex-smoker	718 (87.6)	102 (12.4)	1.35 (1.06-1.70, p=0.013)	1.26 (0.96-1.67, p=0.097)
	Unknown	423 (91.4)	40 (8.6)	0.90 (0.63-1.25, p=0.540)	1.00 (0.68-1.47, p=0.981)
Approach	Open	1699 (83.6)	334 (16.4)	-	-
	Minimally invasive	3383 (93.6)	232 (6.4)	0.35 (0.29-0.42, p<0.001)	0.42 (0.34-0.52, p<0.001)
Contamination	Clean-contaminated	4663 (91.3)	443 (8.7)	-	-
	Contaminated	299 (75.5)	97 (24.5)	3.41 (2.65-4.37, p<0.001)	2.55 (1.93-3.38, p<0.001)
	Dirty	48 (66.7)	24 (33.3)	5.26 (3.14-8.58, p<0.001)	3.81 (2.22-6.54, p<0.001)
WHO Checklist	No	1475 (87.8)	205 (12.2)	-	-
	Yes	3607 (90.9)	361 (9.1)	0.72 (0.60-0.86, p<0.001)	0.76 (0.59-0.97, p=0.027)
	Unknown	0 (0.0)	0 (0.0)	-	-

3.4.1.7 Propensity score matched analysis

After propensity score matching using coarsened exact matching there remained some significant different in risk profiles between the group that underwent in-person and telemedicine review, although there was better balance that in the original sample (*Table 3.5*). In the propensity score matched model there remained a significant association between telemedicine review and a reduced odds of reporting of SSI (OR 0.87, 0.78-0.98, 0.019; *Table 3.6*).

Table 3.5. Balance of propensity score matched groups.

		Matched groups		
		In-person review	Telemedicine review	P-value
SSI	No	4669 (86.5)	5323 (88.7)	0.001
	Yes	727 (13.5)	680 (11.3)	
HDI tertile	High	3459 (64.1)	2760 (46.0)	<0.001
	Middle	1137 (21.1)	2604 (43.4)	
	Low	800 (14.8)	639 (10.6)	
Age	2-20	714 (13.2)	905 (15.1)	<0.001
	21-40	1571 (29.1)	2078 (34.6)	
	41-60	1545 (28.6)	1750 (29.2)	
	61-80	1349 (25.0)	1149 (19.1)	
	81-100	217 (4.0)	121 (2.0)	
Gender	Male	2730 (50.6)	2589 (43.1)	<0.001
	Female	2666 (49.4)	3414 (56.9)	
ASA	I	2104 (39.0)	2957 (49.3)	<0.001
	II	2213 (41.0)	2128 (35.4)	
	III	889 (16.5)	668 (11.1)	
	IV	90 (1.7)	65 (1.1)	
	V	14 (0.3)	15 (0.2)	
	Unknown	86 (1.6)	170 (2.8)	
Urgency	Elective	2635 (48.8)	2969 (49.5)	0.135
	Semi-elective	243 (4.5)	226 (3.8)	
	Emergency	2518 (46.7)	2808 (46.8)	
Operative approach	Open	2743 (50.8)	2862 (47.7)	0.001
	Minimally invasive	2653 (49.2)	3141 (52.3)	

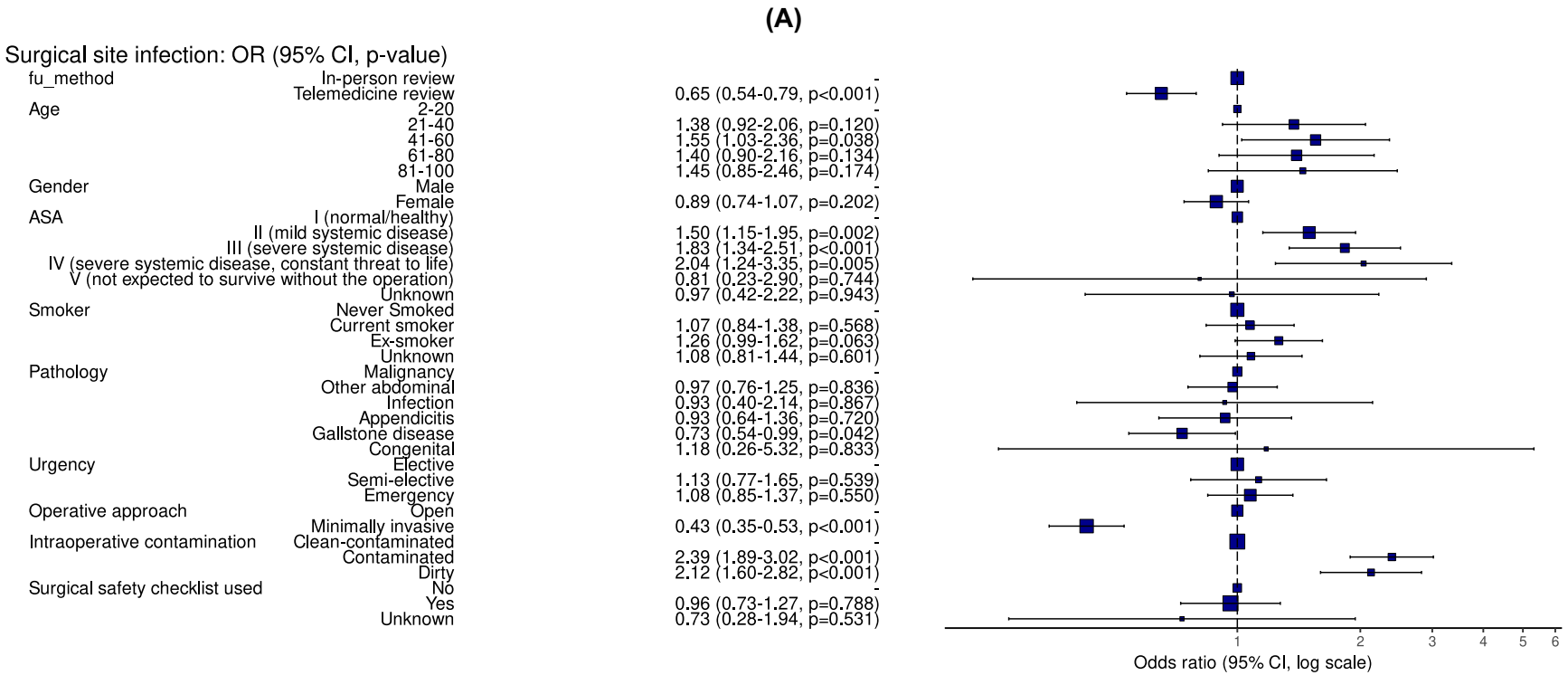
Table 3.6. Sensitivity analysis for the primary model in propensity score matched groups using coarsened exact matching.

			95% confidence intervals		
		Odds ratio	**Lower**	**Upper**	**P-value**
Follow-up method	In-person review	-	-	-	-
	Telemedicine review	0.871	0.776	0.977	***0.019***
HDI tertile	High	-	-	-	-
	Middle	1.469	1.272	1.696	***0.000***
	Low	2.104	1.750	2.531	***0.000***
Age	2-20	-	-	-	-
	21-40	1.082	0.904	1.296	0.391
	41-60	1.203	0.988	1.466	0.066
	61-80	1.085	0.860	1.368	0.492
	81-100	1.051	0.667	1.657	0.831
Gender	Male	-	-	-	-
	Female	1.004	0.891	1.132	0.945
ASA	I	-	-	-	-
	II	1.454	1.257	1.682	***0.000***
	III	1.801	1.457	2.225	***0.000***
	IV	1.929	1.194	3.118	***0.007***
	V	4.241	1.967	9.146	***0.000***
	Unknown	1.066	0.746	1.524	0.725
Urgency	Elective	-	-	-	-
	Semi-elective	1.133	0.827	1.553	0.435
	Emergency	1.236	1.079	1.414	***0.002***
Operative approach	Open	-	-	-	-
	Minimally invasive	0.372	0.323	0.428	***0.000***

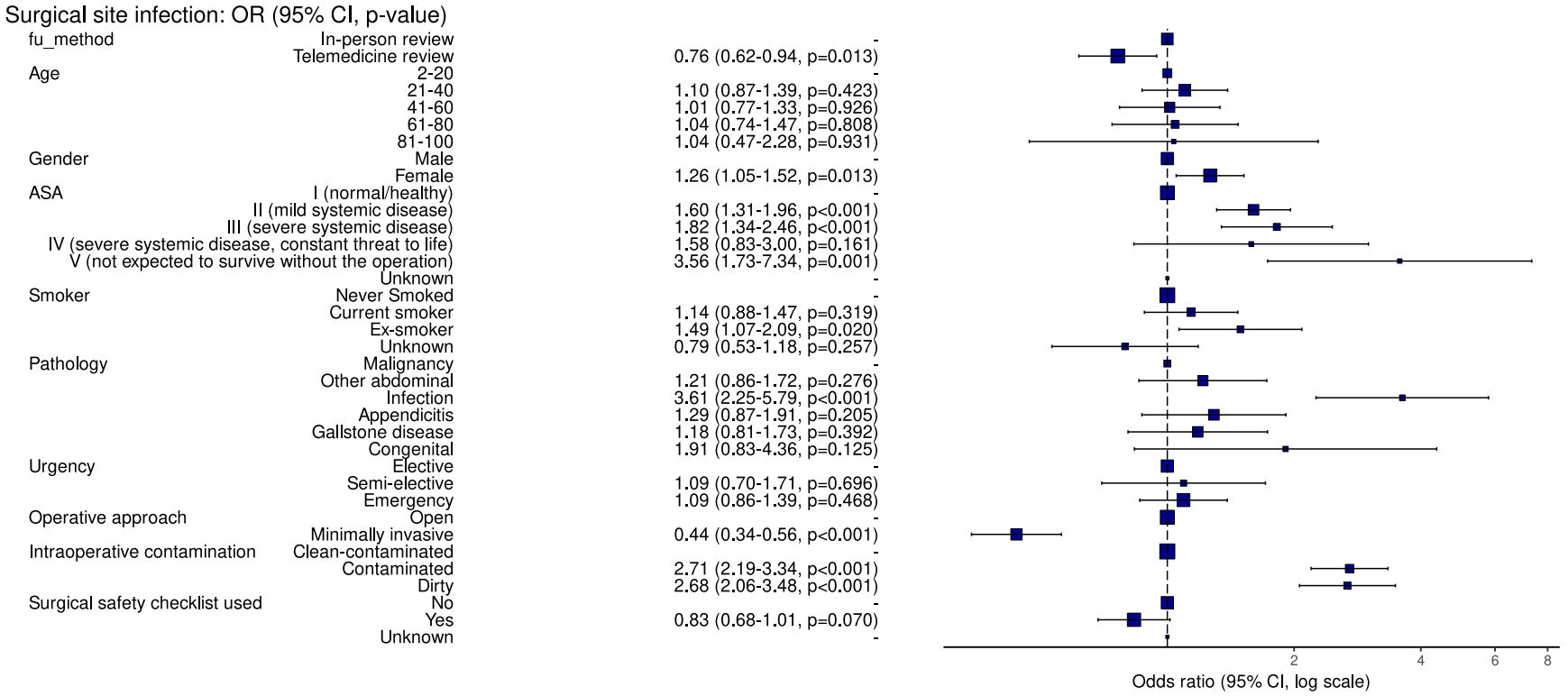
3.4.1.8 Subgroup analyses

The association between telemedicine review and a reduced odds of reporting SSI than in-person review was consistent across both high (OR 0.65, 0.54-0.79, p<0.001) and low-middle (OR 0.76, 0.62-0.94, p=0.013) income countries (*Figure 3.5*).

Figure 3.5. Subgroup analysis for primary model demonstrating the association between reported surgical site infection rates and method of follow-up across (A) high HDI versus (B) middle or low HDI countries



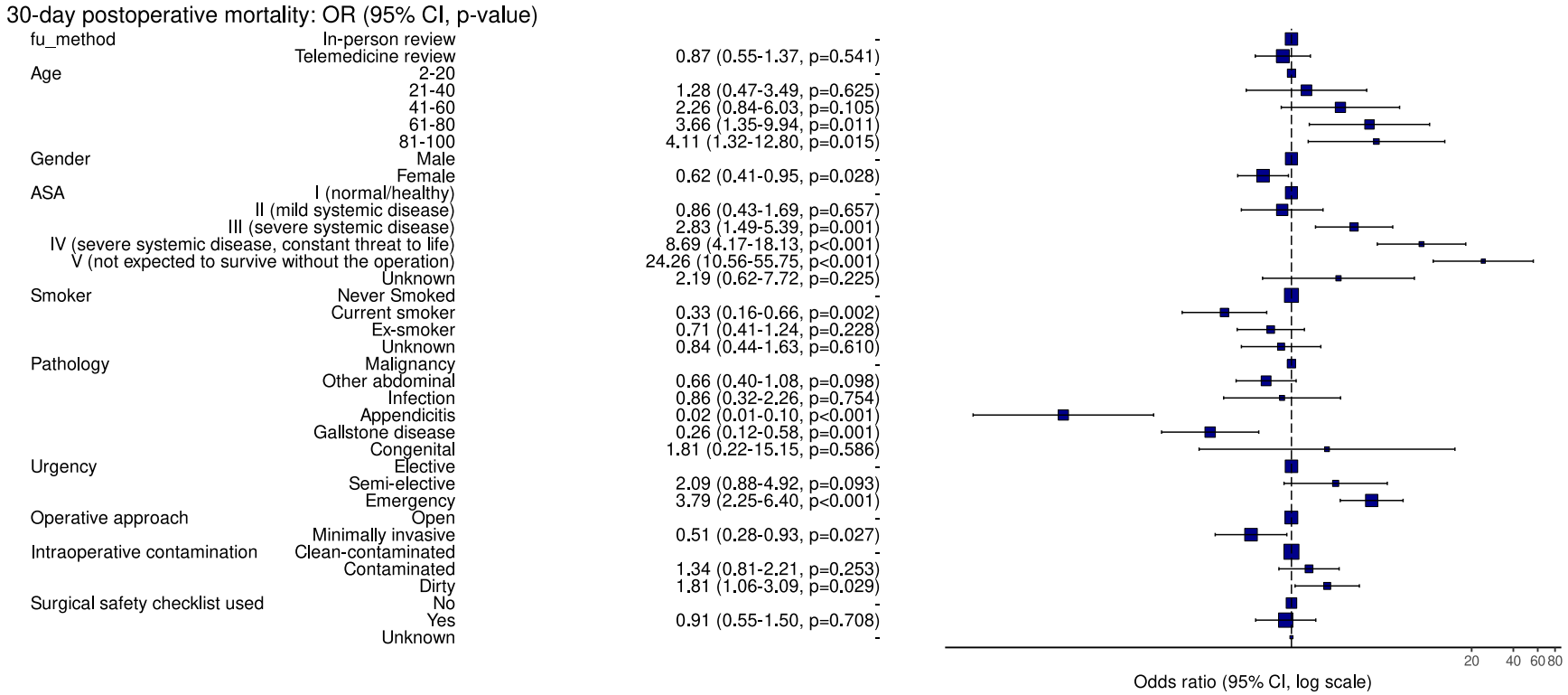
(B)


Surgical site infection: OR (95% CI, p-value)
fu_method
In-person review
Telemedicine review
0.76 (0.62-0.94, p=0.013)
Age
2-20
21-40
1.10 (0.87-1.39, p=0.423)
41-60
1.01 (0.77-1.33, p=0.926)
61-80
1.04 (0.74-1.47, p=0.808)
81-100
1.04 (0.47-2.28, p=0.931)
Gender
Male
Female
1.26 (1.05-1.52, p=0.013)
ASA
I (normal/healthy)
II (mild systemic disease)
1.60 (1.31-1.96, p<0.001)
III (severe systemic disease)
1.82 (1.34-2.46, p<0.001)
IV (severe systemic disease, constant threat to life)
1.58 (0.83-3.00, p=0.161)
V (not expected to survive without the operation)
3.56 (1.73-7.34, p=0.001)
Unknown
Smoker
Never Smoked
Current smoker
1.14 (0.88-1.47, p=0.319)
Ex-smoker
1.49 (1.07-2.09, p=0.020)
Unknown
0.79 (0.53-1.18, p=0.257)
Pathology
Malignancy
Other abdominal
1.21 (0.86-1.72, p=0.276)
Infection
3.61 (2.25-5.79, p<0.001)
Appendicitis
1.29 (0.87-1.91, p=0.205)
Gallstone disease
1.18 (0.81-1.73, p=0.392)
Congenital
1.91 (0.83-4.36, p=0.125)
Urgency
Elective
Semi-elective
1.09 (0.70-1.71, p=0.696)
Emergency
1.09 (0.86-1.39, p=0.468)
Operative approach
Open
Minimally invasive
0.44 (0.34-0.56, p<0.001)
Intraoperative contamination
Clean-contaminated
Contaminated
2.71 (2.19-3.34, p<0.001)
Dirty
2.68 (2.06-3.48, p<0.001)
Surgical safety checklist used
No
Yes
0.83 (0.68-1.01, p=0.070)
Unknown
2
4
6
8
Odds ratio (95% CI, log scale)

3.4.1.9 Multivariable model for postoperative mortality

There was no association between the method of follow-up and the 30-day POMR (OR 0.87, 0.55-1.37, p=0.54; *Figure 3.6*). This suggested that in-person follow-up was not reserved solely for those at very high risk of surgical complications and death (i.e., low signal of significant selection bias).

Figure 3.6. The association between method of follow-up and 30-day postoperative mortality rate (POMR).



3.4.2 Systematic review and combined meta-analysis

3.4.2.1 Search results

From 1299 de-duplicated search results, 25 full papers reported an SSI rate reported using telemedicine. 28.0% had no comparator group (7/25), and 36.0% (9/25) compared telemedicine to assessment at a different time point (e.g., in-hospital versus 30-day telemedicine assessment). Nine eligible studies were therefore included [168-175]. Summary data from the cohort study in *Part 1* was combined with these nine studies for qualitative synthesis. A PRIMSA flowchart for the review is displayed in *Figure 3.7*.

3.4.2.2 Study characteristics

Of the included studies, 66.7% (6/9) were published within the last five years (2015-2020) [168, 172-174, 176, 177]. Eight were prospective cohort studies, with one retrospective study [177]. Most reported data from high-income countries (55.6%, 5/9) [170, 171, 174, 175, 177]; 1 was from an upper-middle income country [173] and 3 from lower-middle income countries [168, 172, 178]. No data from low-income countries or multi-country studies were reported. Of included articles, 44.4% (4/9) reported outcome assessment in patients undergoing general surgery [172, 173, 175, 178], 22.2% (2/9) in trauma and orthopaedics [170, 171], 11.1% (1/9) in obstetric surgery [177], and 22.2% (2/9) in all non-cardiac surgery [168, 174]. There was a moderate or severe risk of bias in all included studies (*Figure 3.8*).

Figure 3.7. PRISMA flowchart of studies included in meta-analysis.

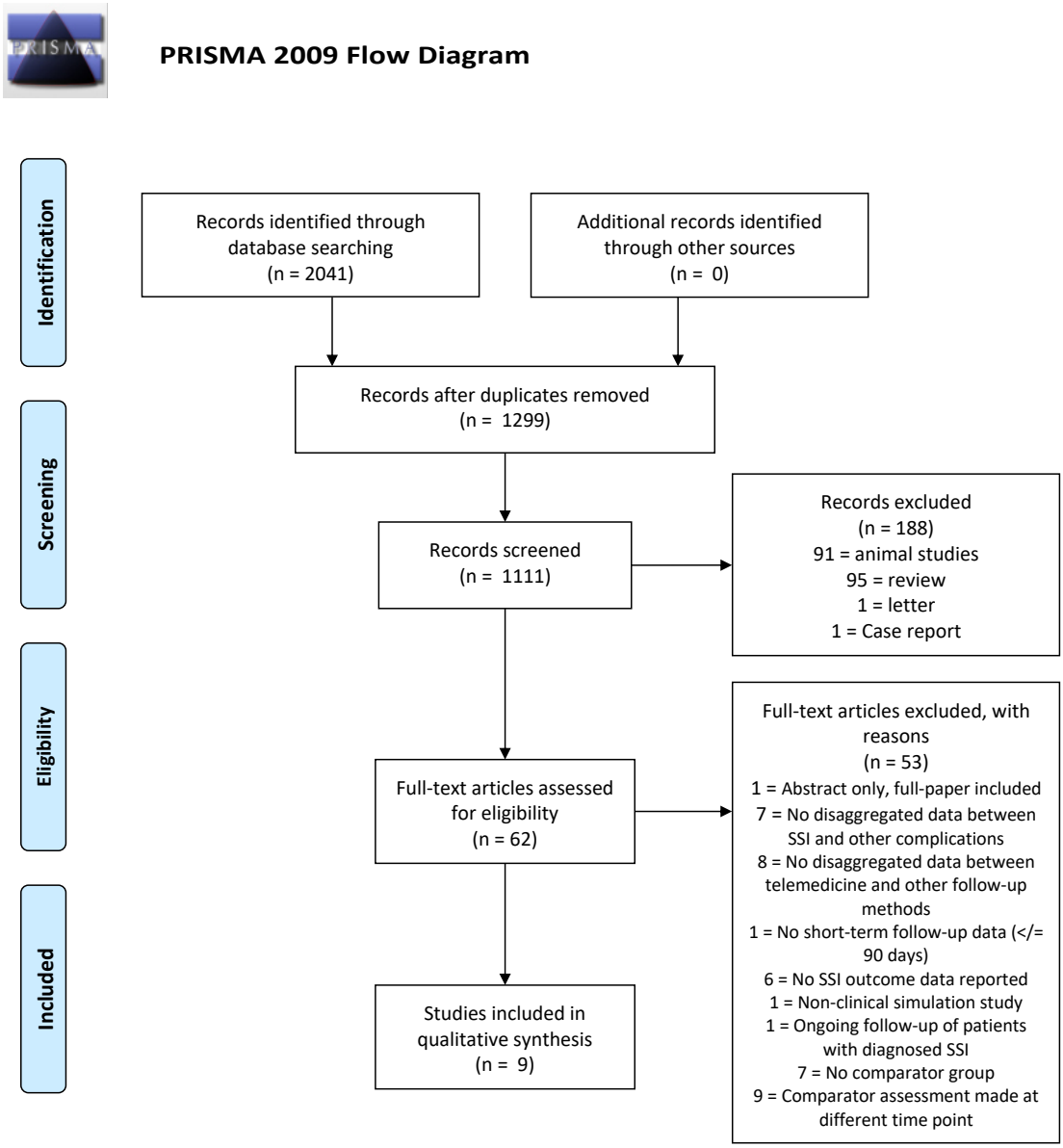


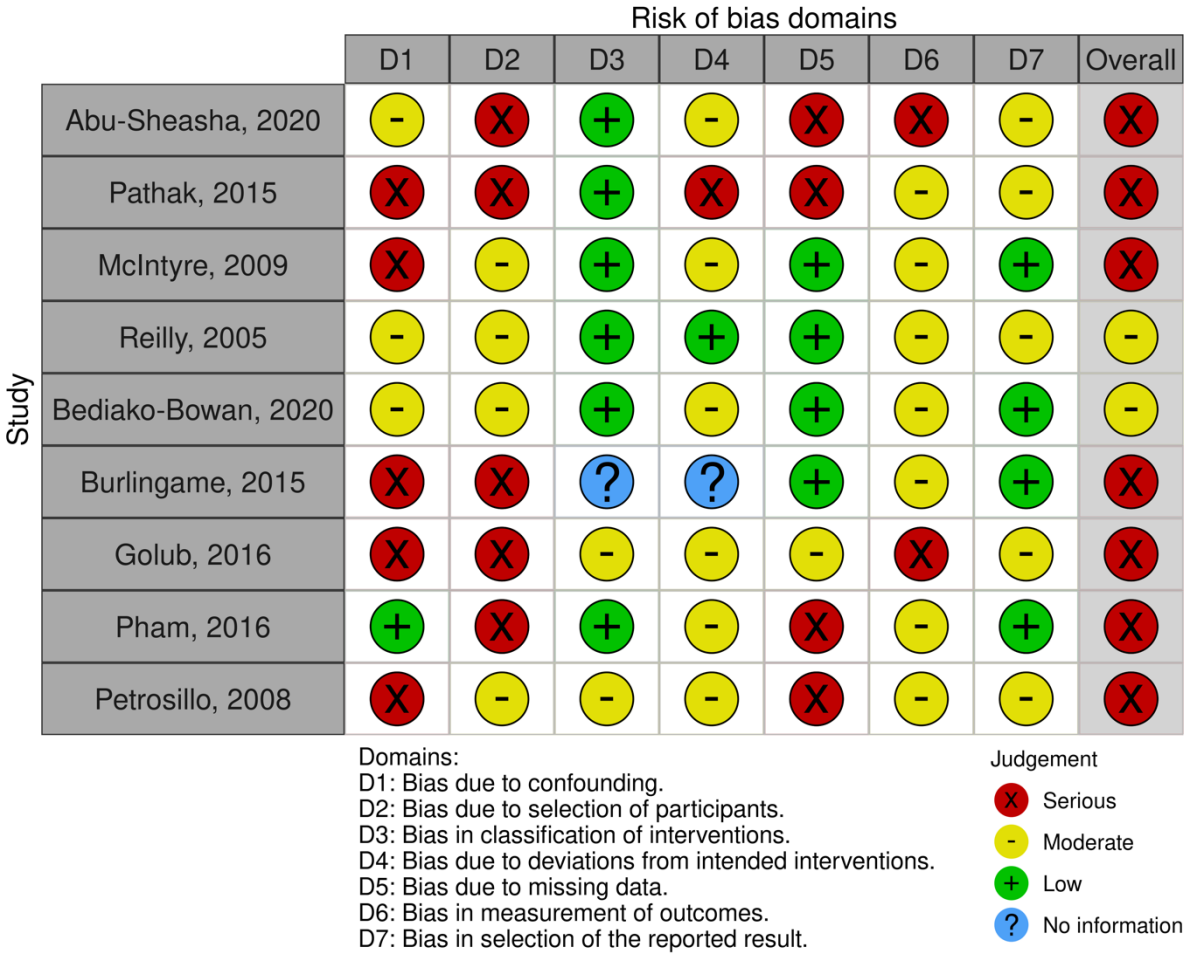
Table 3.7. Study and patient characteristics included in meta-analysis

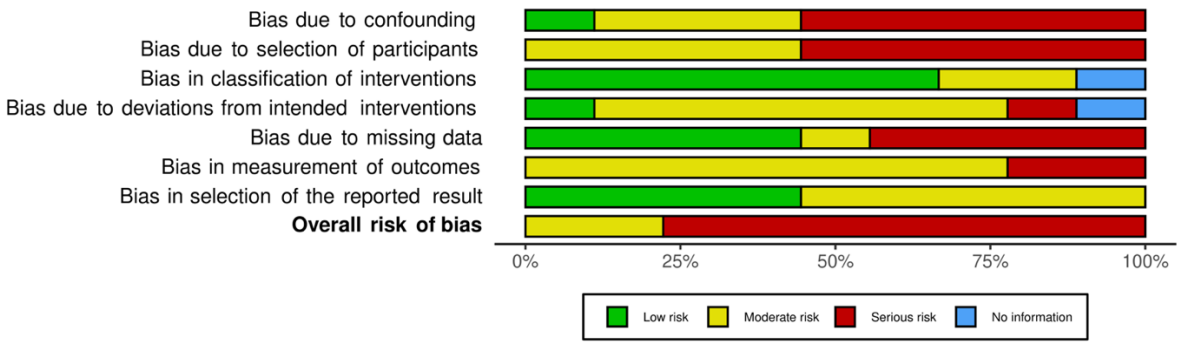
Author(s)	Year	Paper type	Study type	Country	Language	World bank quartile	Specialty	Urgency of surgery
Abu-Sheasha et al.	2020	Full Paper	Prospective cohort	Egypt	English	Lower-Middle Income	All	Not stated
Bediako-Bowan et al.	2020	Full Paper	Prospective cohort	Ghana	English	Lower-Middle Income	General surgery	Elective and Emergency
Burlingame et al.	2015	Abstract	Retrospective cohort	USA	English	High-Income	Obstetrics	Elective and Emergency
Golub et al	2016	Full Paper	Prospective cohort	Russia	Russian	Upper-Middle Income	General surgery	Not stated
McIntyre et al.	2009	Full Paper	Randomised trial	USA	English	High-Income	Trauma	Emergency
Pathak et al.	2015	Full Paper	Prospective cohort	India	English	Lower-Middle Income	General surgery	Elective and emergency
Petrosillo et al.	2008	Full Paper	Prospective cohort	Italy	English	High-Income	General surgery & gynaecology	Elective and Emergency
Pham et al	2016	Full Paper	Retrospective cohort	USA	English	High Income	Non-cardiac	Not stated
Reilly et al.	2005	Full Paper	Prospective cohort	UK	English	High-Income	Orthopaedics	Elective and emergency

Table 3.8. Follow-up methods included in meta-analysis. PHE=Public Health England questionnaire for surveillance of SSI post discharge. CDC= US Centres for Disease Control.

Follow-up modality	Diagnostic criteria	Follow-up	Frequency of assessment	Time points
Telephone	PHE	30 days	One-off	30d
Telephone	PHE	30 days	Sequential (intervals)	3d, 15d, 30d
Telephone	Not stated	30 days	One-off	30d
Telephone	Not stated	30 days	One-off	30d
Telephone	CDC	28 days	One-off	28d
Telephone	Not stated	30 days	One-off	30d
Telephone	CDC	30 days	One-off	30d
Telephone	Not stated	30 day	One-off	30d
Telephone	CDC (modified)	30 days	Sequential (intervals)	10d, 20d, 30d

Figure 3.8. Summary of risk of bias assessment (ROBINS-I) for included studies





3.4.2.3 Use of telemedicine

The proportion of patients with follow-up using telemedicine ranged from 45% to 96%. Study sizes ranged from 141 to 11225 patients. The pooled proportion of patients with telemedicine follow up on meta-analysis was 64% (95% C.I. 55% to 73%). There was very high heterogeneity (I^2=100%, p<0.001).

3.4.2.4 Delivery of telemedicine

Four included studies did not state a standardised schedule for outcome assessment. 22.2% (2/9) used the Public Health England Post-discharge Surveillance Questionnaire [168, 172]and 33.3% (3/9) used questions based on CDC criteria [170, 171, 175]. 77.8% (7/9) were used as a one-off assessment at 30 postoperative days [168, 170, 173-175, 177, 178], with two using serial postoperative assessments [171, 172].

3.4.2.5 Comparison of telemedicine to in-person follow-up

Four studies involved a comparator of telemedicine to in-person follow-up method [168, 170, 171, 178] and were included in meta-analysis of SSI rates reported, combined with the cohort study data (5 studies in total). Two studies had paired within-subject measurements at the same time point [168, 171], and two had measurements at the same time point but in different patient groups [170, 178]. Only two (50%) compared telemedicine to an in-person assessment according to US Centre for Disease Control criteria (*Table 3*). In the random effects meta-analysis, the rate of SSI reported using telemedicine was significantly lower in the telemedicine group than the in-person group (0.67, 95% C.I. 0.47 to 0.94; *Figure 3.9*). The odds ratio describes a comparison of the odds of patients having an SSI reported with telemedicine versus in-person follow-up (i.e., a reduced odds ratio conveys a lower rate of SSI reported with one method in comparison to the other, and vice versa). There was some

evidence of between-study heterogeneity, but this did not have a significant effect on the random effects meta-analysis (I^2=0.45, 0.00-0.78, p=0.12; Tau=0.27, 0.00-0.93). There was no significant evidence of funnel plot asymmetry (p=0.326; *Figure 3.10*).

Figure 3.9. Forest plot of rates of SSI reported by telemedicine and in-person follow-up on meta-analysis.

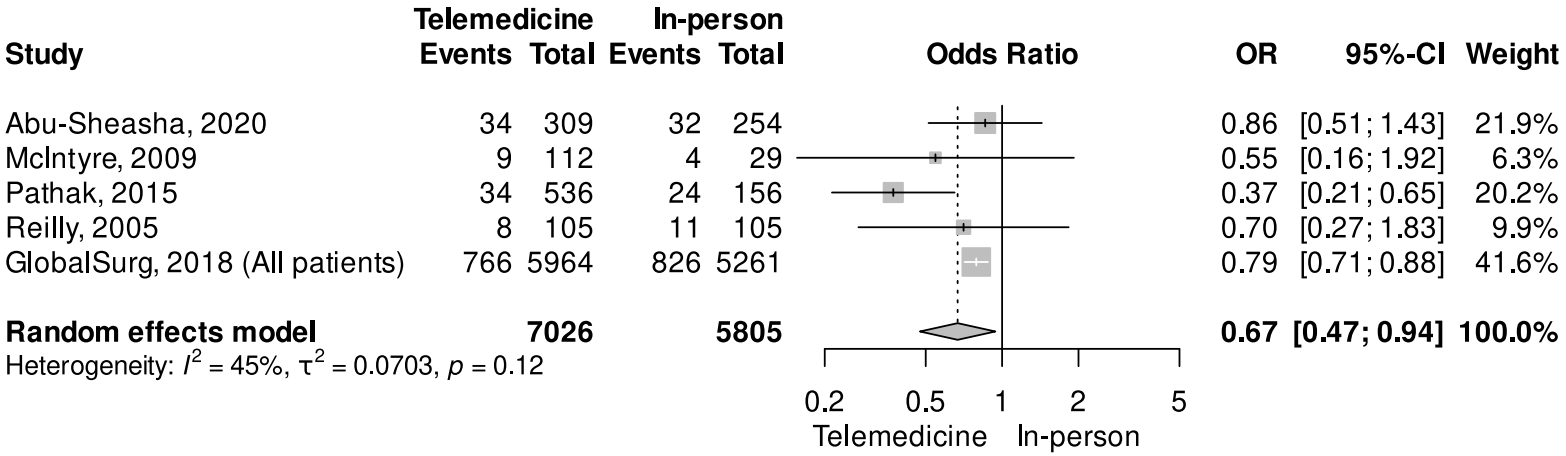
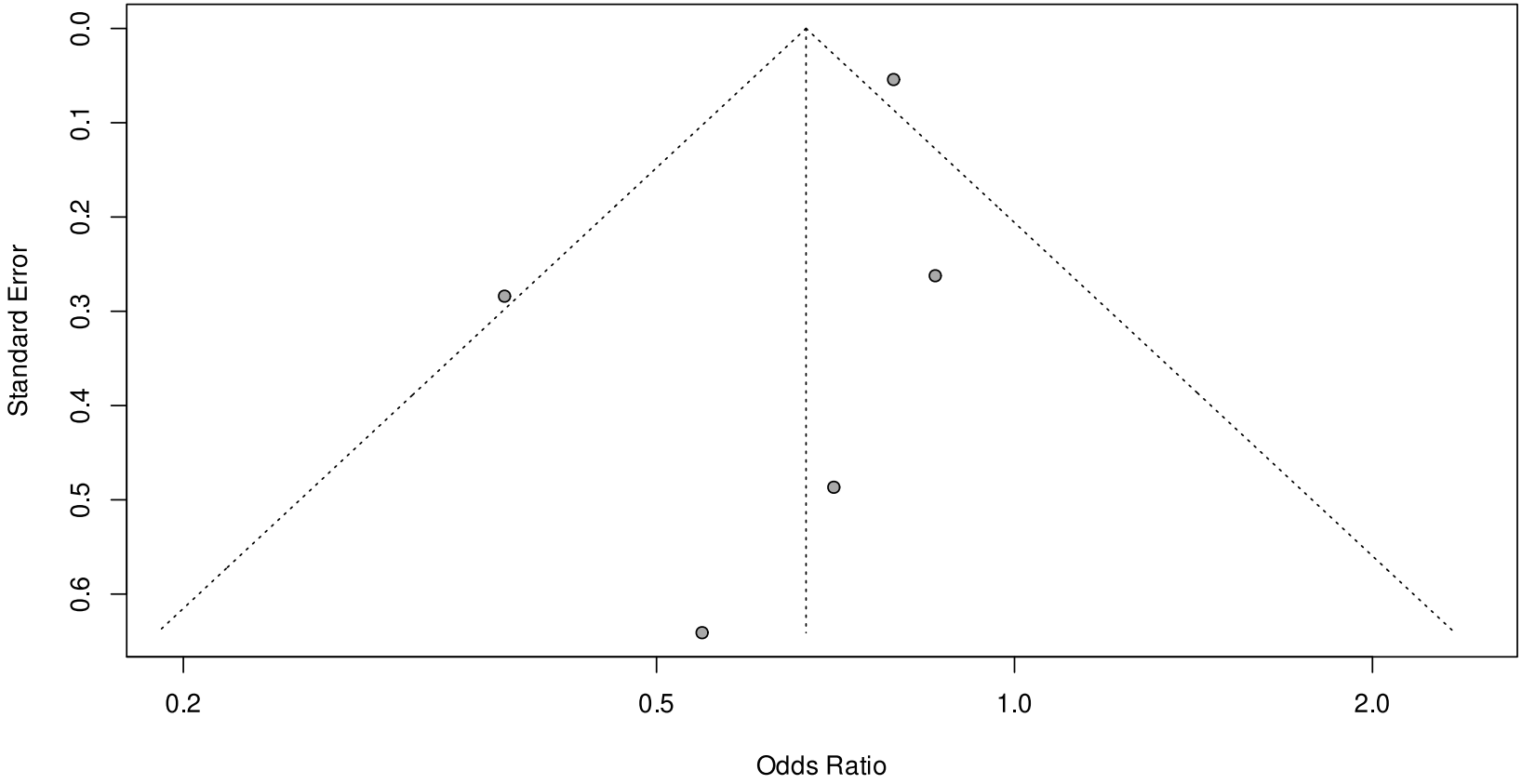


Figure 3.10. Funnel plot of estimates included in meta-analysis of SSI rates reported with telemedicine and in-person follow-up



3.4.2.6 Comparison of telemedicine to other follow-up methods

Five studies compared telemedicine to a follow-up method that did not require in-person review (e.g., Electronic Health Records or postal questionnaire) [172-175, 177]. These are excluded in the meta-analysis of SSI rates. Four of five studies had a higher rate of reporting of SSI in the telemedicine group than the electronic health records or postal questionnaire group. One study had a much lower SSI rate reported by telemedicine than electronic health records (1.1% versus 11.2%) , but the two methods were applied in clearly different patient populations (responders to a postal questionnaire versus non-responders) [172].

3.5 Discussion

3.5.1 Summary of key findings

This cohort study and meta-analysis identified that use of telemedicine for wound assessment post-discharge is feasible across settings. The adjusted rate of SSI reported using telemedicine in patients that underwent post-discharge assessment was lower than with in-person follow-up in the international cohort study, raising concerns of underreporting of SSI. This was robust to several sensitivity analyses, a propensity score-matched model and across HDI settings. This analysis of real-world, global data suggests that telemedicine methods used in the pre-pandemic setting may risk patient safety or introduce bias to research studies. This was corroborated in the combined meta-analysis. The studies included were of low quality, and rarely used standardised tools. High-quality frameworks for remote assessment of SSI must be evaluated and adopted as telemedicine is upscaled globally.

3.5.2 Findings in context

3.5.2.1 Relevance to SARS-CoV-2 pandemic recovery

Telemedicine for follow-up of surgical patients holds significant promise during the SARS-CoV-2 pandemic recovery effort. The high connectivity of global telecommunication networks opens opportunities for telemedicine in both well-resourced and resource constrained settings [129]. Efficient methods for surgical follow-up may be most relevant in LMICs where patients may already travel long distances or take time out of work to return to hospital after discharge, and health systems face severe resource limitations [7, 9, 14, 113, 179]. During future SARS-CoV-2 outbreaks, use of telemedicine may reduce the risk of exposure in hospital outpatient settings [48]. During the post-pandemic recovery, it may help alleviate the growing backlog of outpatient appointments and investigations that health systems face around the world [180-182]. However, as the use of these methods increases, it is important

that the quality of assessment does not decrease. Delayed or missed identification of postoperative complications can lead to failure to rescue and death, more severe sequalae, and increased costs [22, 183]; these events, whilst rare, have the potential to undermine the benefits of telemedicine, particularly for higher-risk patients and operations [124].

3.5.2.2 Comparison of current tools

Two different standardised tools for identification of SSI using telemedicine (Centre for Disease Control Criteria and Public Health England Post-Discharge Questionnaire) were identified in the systematic review, but neither have been formally adapted or validated for use in telemedicine. A universal outcome reporter 'Bluebelle' Wound Healing Questionnaire has demonstrated promise as tool for remote detection of SSI, demonstrating excellent discrimination and reliability [184, 185]; however, this has only undergone evaluation in a single language in one country, and cultural and linguistic adaptation and validation to support international application [186]. No included studies used videography to help identify SSI; this may prove a useful adjunct to future development in this area.

3.5.2.3 Relevance to other postoperative complications

SSI has been identified as a key priority to improve the health of patients undergoing surgery worldwide, particularly in low resource settings [40, 50]. Lessons from use of telemedicine for wound assessment may be generalisable to other common complications of surgery, but bespoke tools may be required for each to ensure accurate identification of different events. Quality-assured digital methods for remote assessment will also have high value for use in pragmatic international trials, where delivery can be made more efficient, and more benefit for more patients can be realised at a lower time and resource cost [187, 188].

3.5.3 Limitations

This study has several limitations.

3.5.3.1 Design limitations

I inferred that the 'gold standard' in-person assessment represents the true SSI rate. I am unable to assert from my data whether SSI is over-reported using in-person follow-up or under-reported using telemedicine where a difference is observed. Second, I assume that the differences in reported SSI rates are unrelated to differences in patient characteristics after risk-adjustment. Whilst I used multi-level models to adjust for several confounders, there is a risk of residual selection bias.

3.5.3.2 Interpretation limitations

First, the quality of studies included in meta-analysis was low. I excluded studies that reported SSI when telemedicine was used for a clearly different patient populations (e.g., different subgroups of patients, responders versus non-responders, different geographical areas), with no comparator group, or a comparator group at a different time point (e.g., in-hospital versus 30-day remote assessment). However, remaining studies demonstrated some 'selection' of patients for telemedicine follow-up, no studies were randomised, and all were at moderate or severe risk of bias. Second, I do not have paired within-patient measures of SSI in-hospital and post-discharge at 30-days, and I am therefore unable to fully account for changes in patient selection to a particular follow-up modality as a result of inpatient infection. This may have exaggerated the difference between telephone and in-person follow-up, however my analysis of postoperative mortality did not indicate a serious risk of reverse causation. Third, as the patients with 'inpatient only' follow-up had no post-discharge wound assessment they were effectively 'lost to follow-up' for the purposes of the primary 30-day analysis. As I do not know the intended follow-up method (i.e., whether

telemedicine or in-person follow-up was planned, but only inpatient data were collected) I am unable to fully explore the impact of attrition bias on the primary comparison. Fourth, there is a further risk of reverse causation in linking patients with inpatient only assessment and a lower observed SSI rate (i.e., those without features of SSI postoperatively may be less likely to re-interact with clinical services). As such, this group were excluded from multivariable analyses, and I recommend caution in interpretation. Fifth, there was very high between study heterogeneity in the pooled estimate of proportions of patients receiving telemedicine follow-up; this was expected *a priori* due to differences in local pathways, resourcing, and preferences. Whilst this limits the interpretation and precision of the central estimate value, we did adopt mixed-effects meta-analysis, and present this analysis as an exploratory estimate of the opportunity to expand telemedicine services rather than informing, for example, health technology assessment which could be more problematic [189].

3.5.3.3 Scope limitations

I was unable to differentiate here between different methods of remote wound assessment (i.e., telephone versus video), although from ongoing work across this network it is likely that a majority of assessment would have been telephone-based [186]. Eight, the cohort study used a pragmatic, observational methodology so did not standardise the training or delivery of telemedicine methods. This should therefore be interpreted as the real-world effectiveness of telemedicine, rather than the potential efficacy of telemedicine in an optimised system [190]. Finally, I have only included one, common postoperative complication in my synthesis. These data set the scene for a broad research agenda to identify and validate tools for remote digital assessment across diverse patient groups and operation types. The rapid upscaling of telemedicine during the SARS-CoV-2 pandemic highlights this as an urgent research priority for the global surgical community.

3.6 Conclusions

This chapter has demonstrated that current methods for remote assessment of the healing surgical wound miss between 1 in 3 and 1 in 5 patients with infection. This can lead to direct and indirect harm to patients. Direct harm may be caused through missed diagnosis, delay to seeking care and increased rates of sepsis and serious complications of SSI. Indirect harm may be even greater, for example by introducing measurement bias in randomised clinical trials that leads to systematic error, and a misleading evidence base for clinical guidelines and service commissioning.

Patients in low- and middle-income countries were at greater risk of SSI, and the use of telemedicine was widespread across these settings. Remote follow-up methods may be particularly important in resource poor environments, reducing burden on outpatient services, and reducing costs, travel, and time out of work for patients. High quality tools for remote wound assessment using telemedicine methods that are relevant across cultures, contexts, and languages in LMICs are urgently needed to strengthen postoperative care pathways. One potential solution is the Bluebelle Wound Healing questionnaire.

The Bluebelle Wound Healing Questionnaire (WHQ) has been developed and validated in the UK (English language) to assess post-discharge infections following abdominal surgery (HTA: 12/200/04) and is attractive for use in randomised trials [184]. The WHQ was designed to be completed either by healthcare professionals, or self-reported by patients [185], and as such has been described as a ‘universal-reporter’ outcome measure (UROM) [191]. In a UK validation study, the WHQ demonstrated good reliability and excellent discrimination [184, 192, 193]. The WHQ was completed both in-person and over the telephone by a healthcare professional trained in wound assessment (e.g., nurse, junior doctor), demonstrating feasibility of telephone delivery. However, no external validation has been performed in

LMICs where health literacy, language and cultural contexts, and digital infrastructure differ substantially. If the WHQ can be administered remotely (e.g., over the telephone) with satisfactory diagnostic accuracy this would reduce resource usage, making surgical research more effective and more sustainable. Other digital adjuncts to surgical site evaluation such as video assessment may further enhance accuracy [128].

3.6.1 Link to next chapter

In the next Chapter, I will describe the adaptation of the WHQ for use in global surgery research and practice. I will adopt a mixed method approach by triangulating qualitative data from structured interviews and focus groups and performing Rasch analysis on data from a prospective cohort study implementing the WHQ across several LMICs.

4. Adaptation of the Wound Healing Questionnaire universal-reporter outcome measure for use in global surgery trials (TALON-1 study): A mixed-methods study and Rasch analysis

4.1 Abstract

Introduction

The Bluebelle Wound Healing Questionnaire is a universal reporter outcome measure developed in the UK for remote detection of surgical site infection (SSI) after abdominal surgery. This study aimed to explore cross-cultural equivalence, acceptability, and content validity of the WHQ for use across low- and middle-income countries (LMICs) and make recommendations for its adaptation.

Methods

This was a mixed-methods study within a trial (SWAT) embedded in an international randomised trial, conducted according to best practice guidelines, and co-produced with community and patient partners (TALON-1). First, structured interviews and focus groups were used to gather data regarding cross-cultural, cross-contextual equivalence of the individual items and scale and conduct a translatability assessment. Second, translation was completed in accordance with *Mapi* recommendations into five languages. Next, data from a prospective cohort SWAT were interpreted using Rasch analysis to explore scaling and measurement properties of the WHQ. Finally, qualitative and quantitative data were triangulated using a modified, exploratory, instrumental design model.

Results

In the qualitative phase, 10 structured interviews and 6 focus groups were performed with a total of 47 investigators across six countries. Themes related to comprehension, response mapping, retrieval and judgment were identified with rich cross-cultural insights. In the

quantitative phase, an exploratory Rasch model was fitted to data from 537 patients (369 excluding extremes) across 5 class intervals. Due to the number of extreme (floor) values the overall level of power was low. The single WHQ scale satisfied tests of unidimensionality indicating validity of the ordinal total WHQ score. There was significant overall model misfit, misfit of five items (5, 9, 14, 15, 16) and local dependency in 11 item pairs. The person separation index (PSI) was estimated as 0.48 suggesting weak discrimination between classes, whereas Cronbach's alpha was high at 0.81. Triangulation of qualitative data with the Rasch analysis supported recommendations for cross-cultural adaptation of the WHQ of items 1 (Redness), 3 (Clear fluid), 7 (Deep wound opening), 10 (Pain), 11 (Fever), 15 (Antibiotics), 16 (Debridement), 18 (Drainage), 19 (Reoperation). Changes to three item response categories (1 = Not at all, 2 = A little, 3 = A lot) were adopted for symptom items 1 to 10 and two categories (0 = No, 1 = Yes) for item 11 (Fever).

Conclusion

This study made recommendations for the cross-cultural adaptation of the WHQ for use in global surgical research and practice, using co-produced mixed methods data from three continents. Translations are now available for implementation into remote wound assessment pathways.

4.2 Introduction

4.2.1 Importance of surgical site infection research

Surgical site infection (SSI) is the most common complication of abdominal surgery, and has a cross-societal, global impact on patients and their families [23, 25, 55, 66, 194]. Delayed return to work, readmission or reoperation leads to substantial effects on quality of life during recovery, and has spill-over effects on mental, economic, and social wellbeing for our patients [2]. This is particularly relevant in low-resource settings, where patients are more likely to suffer catastrophic expenditure around the time of surgery [68]. Consequently, research in SSI prevention has been prioritised by patients, researchers, and clinicians in low- and middle-income countries (LMICs) [40].

4.2.2 Importance of post-discharge wound surveillance

Timely identification of SSI is essential in maintaining patient safety after hospital discharge. Missed SSI diagnoses or misclassification of SSI can directly and indirectly impact patient safety [195]; directly, through delayed intervention for patients with an active infection, or indirectly, by introducing bias to randomised studies that feed into best practice guidelines [23, 56]. Post-discharge surveillance is therefore considered to be a key quality marker in SSI research and is an important component of postoperative care pathways [56]. Chapter 3 demonstrated that current methods for remote SSI detection are not fit for purpose.

4.2.3 Candidate tools for wound surveillance

The Bluebelle Wound Healing Questionnaire (WHQ) was developed and validated in the UK in English Language to support post-discharge surveillance for SSI following abdominal surgery [184, 192]. However, this instrument has not yet been adapted for cross-cultural and cross-language implementation in low- and middle-income countries (LMICs). High-quality, contextually relevant tools for remote wound evaluation are urgently needed. Firstly, to build

resilient and sustainable surgical systems and support safe upscaling of capacity during pandemic recovery [11, 47]. Secondly, to reduce loss to follow-up and risk of attrition bias in randomised trials by developing contextually relevant pathways for remote assessment [195].

4.2.4 Objectives

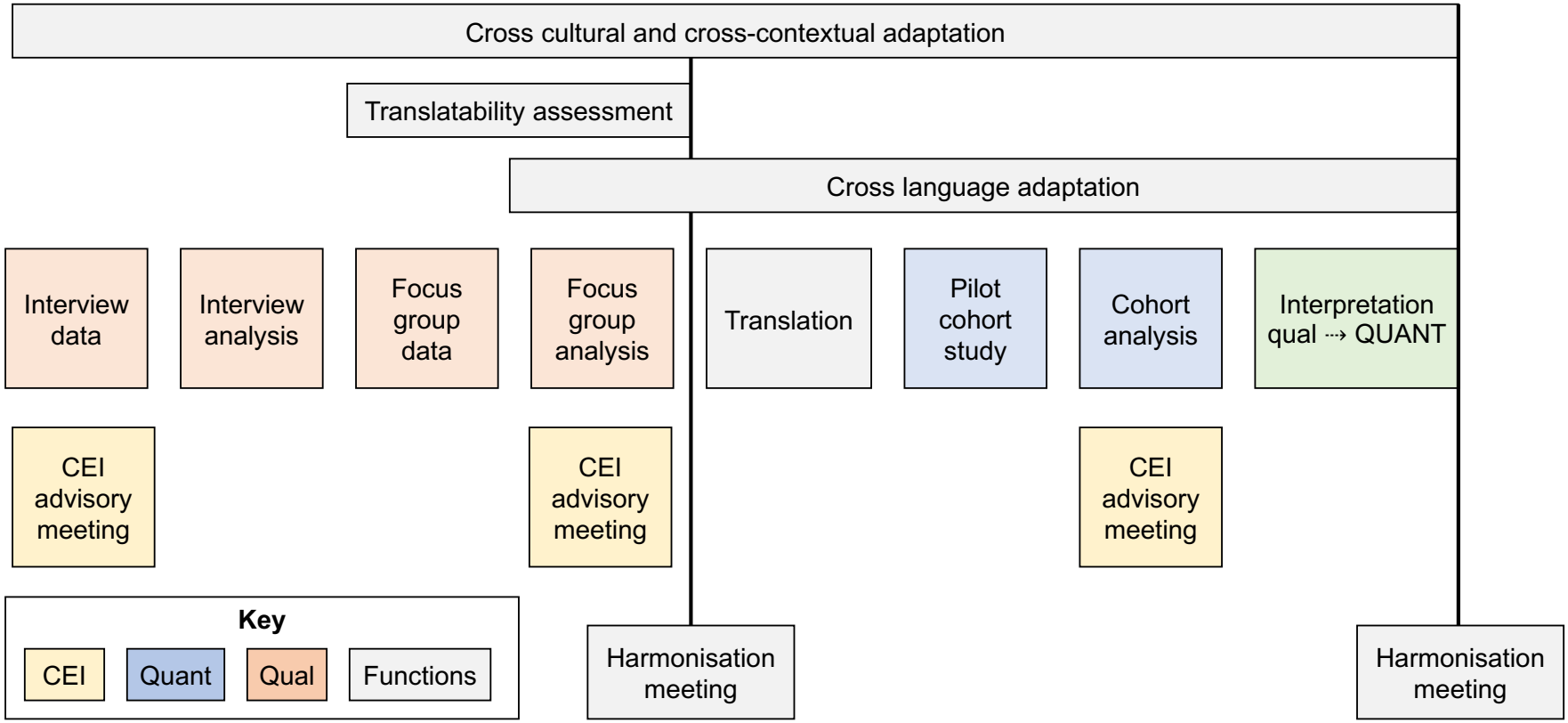
This mixed-methods study (TALON-1) aimed first to explore cross-cultural and cross-language equivalence, acceptability, and content validity of the WHQ across several LMICs. Then, to assess the scaling and psychometric properties of the WHQ when used across different patient populations and subgroups using Rasch analysis. Finally, to consolidate recommendations for the adaptation of the WHQ for use in global surgical research by triangulating qualitative and quantitative data.

4.3 Methods

4.3.1 Overview

TALON-1 was a mixed-methods study embedded in an international randomised trial, conducted according to best practice guidelines, and co-produced with community and patient partners [86, 196, 197]. The study used qualitative and quantitative data to explore the extent to which the WHQ measured SSI as a concept, and the parameters of the latent trait (i.e., an underlying outcome of interest) in the target (i.e., the low resource context) and source (i.e., United Kingdom, a high resource Universal Healthcare System) cultures. It then aimed to assess how accurately items could transfer meaning across languages [198]. Adaptation of the standard methodology was required to progress the qualitative phase during the SARS-CoV-2 pandemic. First, expert review of the WHQ was conducted using structured interviews and focus groups with surgeons and site researchers involved in wound evaluation. These were used to gather rich data regarding cross-cultural, cross-contextual equivalence of the individual items and scale and conduct a baseline translatability assessment. Second, data from a prospective cohort study were interpreted using Rasch unidimensional measurement modelling approach to explore scaling and measurement properties of the questionnaire, including cross-cultural differential item functioning. Next, qualitative and quantitative data were triangulated using a modified, exploratory, instrumental design model to recommend adaptations for use of the WHQ in global surgery research and practice [199]. Finally, translation was completed in accordance with *Mapi* recommendations into five languages. An overview of the study methodology is shown in *Figure 4.1* and detailed in *Table 4.1*.

Figure 4.1. Study design overview



TALON-1 adopted a modified exploratory instrument design methodology. The qualitative phase was conducted first to explore cross-cultural and cross-contextual equivalence in interview and focus groups. This phase also included a baseline translatability assessment. Data were reviewed at a harmonisation meeting. The WHQ was then translated according to *Mapi* recommendations. The quantitative phase followed next with a prospective cohort study and Rasch analysis. The qualitative and quantitative data were then triangulated to make final recommendations for WHQ adaptation, agreed at a final harmonisation meeting.

Table 4.1. Summary of Wound Healing Questionnaire adaptation methodology

Methodology	Yes	No	Details
Concept definition (protocol)	**Y**		Protocol agreed between international Study Management Group, developers, community, and patient partners. Pre-registered and published in *Trials [186]* and in SWAT store registry (ID126)
Qualitative: Cross-cultural and cross-contextual adaptation (in Source language, English)			
Consultant identified	**Y**		An in-country consultant was identified in each target country who was fluent in both the source and target language(s). This was typically the national Principal Investigator (i.e., a surgeon involved directly in wound assessment) for the study, or else a clinical nominee.
Structured interviews (expert review)	**Y**		Structured interviews were designed to review the instrument validity, items, and scaling. The topic guide was directed item-by-item, learning from cognitive theory. In each country, 2 to 3 interviews were conducted with site investigators directly involved in wound assessment.
Focus groups (reconciliation and translatability assessment)	**Y**		A focus group was held with each country to review coding and analysis from the expert review phase ('member checking'). This included several investigators fluent in both the source and target language. An item-by-item translatability assessment was made in parallel. Any further iterative modifications were made before moving into the harmonisation meeting.
Community and patient partner review	**Y**		Patient advisory group meeting with representation from 4 of the target countries (Nigeria, South Africa, India, Ghana) was convened to review the recommendations for adaptation of the instrument item-by-item, and to co-design the cohort study including co-production of the telephone follow-up pathway and supporting documentation.
Harmonisation meeting	**Y**		Virtual meeting on Zoom platform with national principal investigators to 'sign-off' final adaptation of the adapted English language WHQ to move into cross-language translation.
Cross-language translation (performed for each Target language)			
Dual forward translations	**Y**		Performed by translators fluent in both the source and target language, and native to the target country.
Forward translation reconciliation	**Y**		Comparison of translations with any discrepancies resolved with discussion between translators and in-country consultant.
Back translation x1	**Y**		Performed by translator fluent in both the source and target language, and native to the target country.
Back translation reconciliation	**Y**		Comparison of back-translated source language questionnaire with original. Discussion within in-country consultant to review and resolve any consistencies.
Developer's review	**Y**		Developers collaborated as members of the Study Management Group and co-authors on this manuscript.
Cognitive interviewing *(replaced with data review)*		**N**	Cognitive interviewing with patients was not possible during SARS-CoV-2. Qualitative data from the expert review and transability assessment was used instead to inform translation, led by the consultant.
Clinician review and proofreading	**Y**		Clinicians involved in wound care embedded in the adaptation and translation process. Two native speaking clinicians provided the final review and proofreading
Pilot testing	**Y**		Target language delivery tested during follow-up with 5 to 10 patients to test comprehension, phrasing, and delivery. A monitoring call was held with the investigators to review feedback before progression to the quantitative study.
Harmonisation meeting	**Y**		Virtual meeting on Zoom platform with national principal investigators to act as a final quality check and share review lessons learnt during translation.
Published	**Y**		Final version published in Appendix E and F
Quantitative: Cohort study of adapted and translated WHQ (Source and target languages)			
Cohort study	**Y**		Study within a trial within the FALCON RCT to test feasibility, acceptability, and measurement properties of the Wound Healing Questionnaire. Minimum sample size target of 100 patients per country.
Rasch analysis of cohort study data	**Y**		Rasch unidimensional measurement modelling in data used to evaluate scaling, measurement properties and differential item functioning across key subgroups.
Reconciliation and reporting			
Triangulation	**Y**		Triangulation of qualitative and quantitative data to inform final recommendations for WHQ adaptation.
Community and patient partner review	**Y**		Presented findings of cohort study to patient advisory group to co-interpret patterns in data and share insight on final recommendations for WHQ adaptation. Co-produced a lay abstract summary of the research findings for dissemination to the public.
Final harmonisation meeting	**Y**		Virtual meeting on Zoom platform with national principal investigators
Validation report	**Y**		A full prospective validation study for the adapted global Wound Healing Questionnaire in the target languages in seven low- and middle-income countries is reported elsewhere.

Adapted from Oxford University Innovation outcomes centre checklist, and *Mapi* process for cross-cultural and cross-language adaptation

4.3.2 Ethics and ethical approvals

This study within a trial was first approved within the FALCON trial protocols by a University of Birmingham Research Ethics Committee (ERN_18-0230_A and ERN_19-0719). Additional approvals were then obtained from national, regional and/or hospital-level ethics committees for selected centres in all participating countries, in accordance with local protocols. In the qualitative phase, an Information Sheet for was provided to all participants. Verbal consent was taken and recorded. In the quantitative phase, written (or fingerprint) informed consent to participate was obtained from all participants.

4.3.3 Data management

Participant data were pseudonymised for storage securely within a password protected NVivo V12 data management system. Quantitative data were stored in a secure REDCap server, hosted at the University of Birmingham, UK, and held in line with General Data Protection Regulation (GDPR) principles [200].

4.3.4 Host trial

FALCON was a stratified, pragmatic, multi-centre, 2x2 factorial trial testing two measures (skin preparation and antimicrobial sutures) to reduce superficial or deep skin infection following abdominal surgery in seven low- and middle-income countries (NCT03700749) [194]. FALCON provided a platform for this study to both identify eligible site investigators for interviews and focus groups, and co-recruitment of patients to the embedded prospective cohort study.

4.3.5 Study instrument

2.3.5.1 UK development and validation

The WHQ was developed with the aim of detecting post-discharge surgical site infection following abdominal surgery in a large feasibility study in a pilot RCT (Bluebelle) in the UK [118, 184, 193]. Development included three phases: (1) analysis of existing tool and semi-structured interviews; (2) item development; (3) pre-testing for acceptability and understanding. The WHQ includes 19 items (18 items and 1 sub-item) related to the construct of surgical wound healing (i.e., SSI), with 11 items (10 items and 1 conditional sub-item) related to 'symptoms' of SSI and 8 items related to interaction with the treatment 'pathway' for SSI. It was designed so that it could either be administered by a healthcare professional, or self-reported by patients [185] (i.e., 'universal-reporter' outcome measure (UROM) [191]). In an English language validation study of 792 patients, the WHQ demonstrated an acceptable scale structure, acceptable inter-rater reliability (Kappa for items ranged between 0·40 and 0·74) and excellent discrimination (area under receiver operating curve characteristic (AUROC) 0·91, 95% confidence interval (CI) 0·83 to 0·98) of patients with and without SSI [184, 192, 193].

4.3.5.2 Global applicability

No adaptation, translation, or validation of the WHQ has been performed for patients in the Global South where health literacy, language and cultural contexts, and digital infrastructure differ substantially. Two principal developers of the WHQ were collaborating members of the Study Management Group (RM, JB).

4.3.6 Reporting and registration

This study was reported with reference to recommendations from the Global Health Network for qualitative research in LMICs, consolidated criteria for reporting qualitative research

(COREQ) framework [196, 201], PCORI recommendations [197] for best practices in mixed-methods adaptation of outcome measures (see *Appendix A* for PCORI checklist). Primary data from FALCON was published in *The Lancet* in 2021 [43]. The protocol for TALON-1 was pre-registered on the MRC Hubs for Trial Methodology Research database [202] (Queen's University Belfast) (SWAT ID:126) and published in *Trials [186]*.

4.3.7 Qualitative phase

4.3.7.1 Cross-cultural and cross-contextual adaptation

Due to the number of target languages for the questionnaire in the host trial, cross-cultural adaptation was initially performed in English language. A study protocol was developed and approved by the cross-disciplinary SMG. Whilst cognitive interviews with patients remains the optimal methodology for cross-cultural and cross-language adaptation of an outcome measure [133, 203], modification was required to progress the study during the SARS-CoV-2 pandemic. Travel was prohibited, and face-to-face outpatient appointments were typically avoided due to concerns related to SARS-CoV-2 transmission [49, 204]. Cross-cultural and cross-contextual adaptation was therefore performed through expert review and structured interviews with site researchers. Structured interviews were conducted with two to three research staff in each country, according to a template from the Social Research Association based on Willis, 2005 [205].

4.3.7.2 Interview participant sampling

Purposive sampling is a non-probabilistic sampling method to produce a sample of individuals that are data-rich and are logically assumed to be representative of a target population [206, 207]. It was selected for sampling in this thesis to balance the potential diversity of the study population with pragmatism in international research delivery. Participants were purposively sampled from sites participating in the FALCON trial (research

nurses, or doctors directly involved in postoperative wound assessment) with view to including a mix of sex, country, patient population served (urban or rural) and experience in face-to-face and telephone follow-up assessments.

4.3.7.3 Interview structure and topic guides

Semi-structured interviews were conducted to explore universality of the construct of SSI, cross-cultural relevance of concepts and construct validity of the questionnaire [198]. The topic guide was structured around four pre-defined categories (*Appendix B*): item comprehension (patients' understanding of the idea and item), response mapping (relating a patients' internally generated answer to response categories provided), retrieval (patients' ability to remember and recall their response) and judgement (patients' overall ability to respond it the item and how they came to this answer) [205].

4.3.7.4 Interview data analysis

Interview notes and a reflexive diary were also maintained as an additional data source. Coding was performed using thematic content analysis with a generic pragmatic qualitative approach informed by cognitive theory. First, unrefined data (transcripts and interview notes) from each interview was reviewed during familiarisation. Second, structured WHQ item-by-item summaries were generated for each interview during a charting phase. Thirdly, themes across the four structured categories were coded and explored. To ensure the depth of experience of participants was reflected, we allowed flexibility to include novel, inductive codes related to wound healing, SSI and/or measurement procedures in global contexts, and I presented these in this thesis using thick description [208]. A sample of 5% of the data were double coded by an experienced qualitative researcher (Mathers) and a clinical researcher from South Africa (Hyman) to ensure themes generated were representative of an LMIC perspective. Finally, themes were aggregated for each WHQ item to compare and contrast

perspectives [205]. The reflexive diary supported interpretation of the interviewer's role as a questionnaire developer and the potential impact on data collection. To ensure credibility, member checking was performed with the final summary themes with representative participants and in-country consultants to ensure meaning was correctly interpreted and maintained [209].

4.3.7.5 Focus group sampling

Focus groups were then held with investigators from each country to review and discuss the thematic coding (1 to 2 per country). The focus groups were conducted in English language and led by the lead researcher (JG) with one or more in-country consultant co-leads. These sampled 8 to 12 participants, adopting a similar sampling frame to the structured interviews. A new sample of participants (separate to those participating in interviews) were approached for the focus group phase. Focus groups were conducted in addition to interviews to explore consensus and contrasting opinions between different stakeholders around themes emerging in the semi-structured interviews; the overall objective was to obtain a single cross-culturally adapted questionnaire to move into cross-language adaptation [210, 211]. Recruitment continued until the research team judged that both the data and sample had sufficient depth and breadth overall [212]. Within the pragmatic limitations of this study, we did not attempt to reach data saturation at a country-level. Where required, iterative adaptation of the WHQ was made until a point of saturation according to accepted best practice principles for adaptation of instruments [133, 197, 213].

4.3.7.6 Initial recommendations from qualitative phase

Recommendations from the qualitative phase were made overall, specific to an individual item, or related to questionnaire administration. The focus group also included several investigators fluent in both the source and target language to serve as a baseline

translatability assessment. This process produced an English language questionnaire which had been adapted to broadly ensure cross-cultural equivalence across the participating countries, was acceptable to all national principal investigators (PI) and highlighting potential translatability issues during cross-language adaptation. The procedures for remote, telephone administration of the WHQ were also explored using targeted questions based on investigators experience within the FALCON trial.

4.3.8 Cross-language adaptation

In some countries, English was a primary or prevalent secondary language amongst the host trial participants. In these countries, feasibility of single-language administration of the questionnaire was tested at sites during the cohort study. Where translation of the WHQ was required, this was performed according to the *Mapi* process for standard linguistic validation to verify conceptual equivalence across languages [133, 214, 215].

4.3.8.1 Translation methodology

This involved a seven-step process alongside clinicians directly involved in wound assessment. Firstly, an in-country consultant (typically the national PI, or other clinical nominee) was recruited and briefed to oversee the process of translation in the target country and language(s). Secondly, forward translation was performed by two independent translators fluent in both the target and source language (native to the target country). Thirdly, the two versions were compared in detail with any differences in wording highlighted. These were reconciled by discussion between the translators and the in-country consultant. This produced a new translation of the instrument which was reviewed and signed off by the in-country consultant. Fourthly, the translated instrument underwent backwards translation by a third independent translator fluent in both the target and source language (native to the target country). Fifth, the back-translated document was compared to the original instrument.

Again, any differences were explored with discussion between the backwards translator, the in-country consultant and a forwards translator (where this was possible). Any further changes to improve clarity in the target language or cross-contextual relevance when then reconciled in the final translated instrument. Sixth, the translated instrument was piloted with site investigators and 2 and 4 patients per country. Further clinician review was not deemed to be necessary, with deep clinical involvement throughout the translation process. Finally, a harmonisation meeting was held in an online focus group using the Zoom platform (Zoom Corporation, Tokyo, Japan) to review the in-country consultants' experience with the instrument and translation process. The instrument was reviewed item-by-time to ensure conceptual equivalence and share new learning in cross-cultural similarities and differences.

4.3.9 Quantitative phase

4.3.9.1 Cohort study design

Data for the quantitative phase were collected during a prospective, international cohort SWAT. Consecutive adult patients (greater than 18 years) recruited to the FALCON trial were eligible. This included a broad range of abdominal operations with a predicted clean-contaminated, or contaminated or dirty operating field and a planned skin incision of greater than 5cm. Operations could be performed for benign, malignant, trauma, or obstetric indications.

4.3.9.2 Informed consent and patient information

Consent for an additional telephone follow-up call to administer the WHQ was taken at the same time as trial consent, using a targeted Informed Consent Form and Patient Information Sheet. Patient and community partners supported co-production of these resources to ensure culturally attuned language and delivery.

4.3.9.3 Measurement procedures

Telephone administration of the translated WHQ was performed at 28-30 days after surgery (i.e., in the 72 hours prior to in-person follow-up) integrated into the host trial pathway. The telephone WHQ was administered by a researcher, doctor or research nurse (non-consultant or attending grade), that was independent of the assessment for the trial primary outcome at 30-days after surgery. In response to Community Engagement and Involvement (CEI [86]) partner feedback, patients were asked to provide two or three contact numbers, which could include a family member or community worker. The researcher was blinded to the outcome of the in-person wound assessment within the FALCON trial, and underwent training from the Study Management Group (SMG).

4.3.9.4 Quality assurance and monitoring

A monitoring call was performed using the Zoom platform (Zoom Corporation, Tokyo, Japan) after each local researcher had completed the WHQ for 5 to 10 patients to allow feedback, troubleshooting and quality assurance. Where recordings were available, these were reviewed by a member of the SMG fluent in the target language. A WhatsApp group was also created for all site investigators participating in each country to share early experience and best practice. The pathway for telephone WHQ follow-up was co-designed with patient and community partners to ensure culturally sensitive delivery.

4.3.9.5 Sample size considerations

No minimum sample size was set, but a target of 100 patients per country was discussed with each of the national principle investigators for use in Rasch unidimensional measurement modelling, based on published recommendations [216].

4.3.10 Psychometric testing using Rasch analysis

4.3.10.1 Justification for use of Rasch analysis

Unidimensionality, measurement properties and cross-cultural item functioning of the WHQ was tested using exploratory Rasch analysis [134]. In brief [217], Rasch modelling is a statistical method for exploring a characteristic that is not directly observable (the 'latent trait', in this case remote detection of surgical wound infection). A questionnaire measures this by looking at representative behaviours (measured separately in each question or 'item') that are combined into one of more scales. Questionnaires provide a raw score for a response to each item that are then summarised to create an overall score. If the questionnaire is to be interpreted correctly, then it should behave like a ruler, where each point score increase indicates the same increase (equal 'intervals') in the overall strength of the 'trait' (here, the severity of SSI).

4.3.10.2 Problems with use of 'raw' sum questionnaire scores

A 'raw' sum score can be a misleading if: (1) different items convey more information ('difficulty') about the trait than others; (2) if more than one trait is being measured ('single versus multiple domains'); (3) if patients have a very low ('floor') or very high ('ceiling') score; (4) if items are very highly correlated so are asking for the same information about the trait; (5) if patients with different characteristics respond in different ways ('differential item functioning'); (6) in the case of missing item response data. The Rasch model allows exploration (and adjustment) for these patterns and takes into account random variation in item responses, allowing deeper understanding of the 'person-metric' properties of a questionnaire and uncertainty around them. Adoption has been widely encouraged to increase quality during questionnaire development and adaptation [134, 218, 219].

4.3.10.3 Application of Rasch methods in this study

The Rasch unidimensional measurement model was fitted to examine the psychometric properties of the WHQ, identify anomalies in the data and evaluate the extent to which the WHQ items are measuring the latent trait of wound infection [219, 220]. Individual items were assessed for excessive misfit (i.e., not measuring the trait in question) and response dependency (i.e., where items are related by more than just the underlying trait). Additionally, appropriate use of item response categories was checked using category probability curves and threshold mapping. Where probability curves were disordered response categories were rescored and item fit was then re-examined. Where residual correlations between items were high, subtesting was performed with re-evaluation of item and model fit. Differential item functioning was examined for each item by country, language, and patient home location (urban / rural). Exploration of DIF was performed only where a subgroup included $\geq$50 complete WHQ responses.

4.3.11 Triangulation

Qualitative and quantitative data were triangulated using data (i.e., between countries) and methodological (i.e., between qualitative interviews and psychometric analysis of quantitative data) triangulation adopting a modified, exploratory, instrumental design model. Triangulation was performed item by item to inform a final version of the instrument in both source (English) and target languages to be finalised and consolidated [197, 221-224]. Finally, there was a phase of proofreading, before completion of a final report of the adapted WHQ, and adoption of this version for further prospective validation. Data were also triangulated regarding measurement procedures to optimise future implementation of remote follow-up pathways.

4.3.12 Community engagement and involvement (CEI)

Patients and community members from LMICs were engaged in all phases of the design and delivery of this study. The interview topic guide was co-designed with input from a representative global surgery patient forum. Practicable methods for conducting interviews, and patient compensation for time in participation were determined with the support of local community leaders. The Guidance for Reporting Involvement of Patients and the Public (GRIPP-2) short form was used to track and report the impact of CEI [225].

4.4 Results

4.4.1 Overview of qualitative data

In total, 10 structured interviews and 6 focus groups were performed with a total of 47 investigators across six countries. Sampling included surgeons (N=34), anaesthetists (N=5) and research staff (N=8), male (N=32) and female (N=15) investigators, caring for patients in both urban and rural populations, and across a range of abdominal surgery disciplines. Interview lengths ranged from 34 minutes to 112 minutes, and focus groups from 92 minutes to 126 minutes. Interview and focus group data from site investigators confirmed that the assumption of a universalist approach to SSI was acceptable, and that symptomology and treatment paradigms were shared across settings. We did not identify any divergence from this during thematic analysis. This was also explored with our CEI partners; together, allowing us to confirm content validity across settings. No new domains or concepts related to symptoms or treatment of SSI arose, suggesting content validity across contexts.

4.4.2 Translation outcomes

Translation was successfully completed in accordance with the schema presented in *Table 1* in five target languages after the qualitative phase: French (Benin), Hindi (India), Kinyarwanda (Rwanda), Punjabi (India), Tamil (India). For some potential languages of delivery, there was no written version of the dialect (e.g., Goun in Benin, Fante in Ghana), and on rare occasions patients would travel a very long distance for treatment and spoke an uncommon language to the local area (e.g., Malayam in Northern India). Here, the questionnaire was translated *ad hoc* from English (source language) by the assessor in the cohort study.

4.4.3 Overview of quantitative data

Of 655 patients attempted for contact in the cohort study across five countries, 5 had died by 30-days (15 missing status). Of those 635 confirmed alive, 537 were contactable for WHQ completion (84.5%). All recorded calls (N=14, 3 languages) demonstrated accurate and consistent delivery of the WHQ. Monitoring calls supported refinement of the delivery pathway and identification of contextually-attuned approaches. Features of included patients are summarised in *Table 4.2* and measurement procedures in *Table 4.3*.

Table 4.2 Patient characteristics (quantitative phase)

Feature	Level	Ghana	India	Benin	Mexico	Nigeria	Total
Timing of WHQ	Per protocol	224 (99.1)	3 (3.8)	100 (100.0)	12 (10.1)	13 (100.0)	352 (65.5)
	Outside of protocol	1 (0.4)	76 (96.2)	0 (0.0)	107 (89.9)	0 (0.0)	184 (34.3)
	(Missing)	1 (0.4)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	1 (0.2)
Age	<18 years	33 (14.6)	0 (0.0)	0 (0.0)	2 (1.7)	3 (23.1)	38 (7.1)
	18-39	115 (50.9)	65 (82.3)	82 (82.0)	79 (66.4)	4 (30.8)	345 (64.2)
	40-59	51 (22.6)	10 (12.7)	13 (13.0)	27 (22.7)	4 (30.8)	105 (19.6)
	60-79	24 (10.6)	4 (5.1)	5 (5.0)	9 (7.6)	2 (15.4)	44 (8.2)
	≥80	3 (1.3)	0 (0.0)	0 (0.0)	2 (1.7)	0 (0.0)	5 (0.9)
Sex	Male	142 (62.8)	7 (8.9)	55 (55.0)	16 (13.4)	7 (53.8)	227 (42.3)
	Female	84 (37.2)	72 (91.1)	45 (45.0)	103 (86.6)	6 (46.2)	310 (57.7)
Home location	Urban	137 (60.6)	53 (67.1)	92 (92.0)	89 (74.8)	11 (84.6)	382 (71.1)
	Rural	89 (39.4)	25 (31.6)	8 (8.0)	30 (25.2)	2 (15.4)	154 (28.7)
	(Missing)	0 (0.0)	1 (1.3)	0 (0.0)	0 (0.0)	0 (0.0)	1 (0.2)
Level of education	Below high school level	157 (69.5)	32 (41.6)	29 (29.0)	23 (19.3)	6 (46.2)	247 (46.2)
	High school or above	69 (30.5)	45 (58.4)	71 (71.0)	96 (80.7)	7 (53.8)	288 (53.8)
Known diabetes	Yes	4 (1.8)	3 (3.8)	1 (1.0)	9 (7.6)	0 (0.0)	17 (3.2)
	No	222 (98.2)	76 (96.2)	99 (99.0)	110 (92.4)	13 (100.0)	520 (96.8)
HIV status	Known negative	17 (7.5)	78 (98.7)	4 (4.0)	36 (30.3)	6 (46.2)	141 (26.3)
	Known positive	1 (0.4)	0 (0.0)	0 (0.0)	2 (1.7)	0 (0.0)	3 (0.6)
	Status not known	208 (92.0)	1 (1.3)	96 (96.0)	81 (68.1)	7 (53.8)	393 (73.2)
Smoking status	Never smoked	218 (96.5)	78 (98.7)	97 (97.0)	107 (89.9)	12 (92.3)	512 (95.3)
	Ex-smoker	5 (2.2)	1 (1.3)	0 (0.0)	10 (8.4)	0 (0.0)	16 (3.0)
	Current smoker	3 (1.3)	0 (0.0)	3 (3.0)	2 (1.7)	1 (7.7)	9 (1.7)
Urgency	Elective (planned)	20 (8.8)	24 (30.4)	0 (0.0)	94 (79.0)	2 (15.4)	140 (26.1)
	Emergency (unplanned)	206 (91.2)	55 (69.6)	100 (100.0)	25 (21.0)	11 (84.6)	397 (73.9)
Indication	Malignant disease	11 (4.9)	7 (8.9)	2 (2.0)	3 (2.5)	2 (15.4)	25 (4.7)
	Benign disease	201 (88.9)	9 (11.4)	97 (97.0)	64 (53.8)	10 (76.9)	381 (70.9)
	Trauma	9 (4.0)	0 (0.0)	1 (1.0)	0 (0.0)	0 (0.0)	10 (1.9)
	Obstetric	5 (2.2)	63 (79.7)	0 (0.0)	52 (43.7)	1 (7.7)	121 (22.5)
Operation location	Foregut	73 (32.3)	2 (2.5)	8 (8.0)	31 (26.1)	2 (15.4)	116 (21.6)
	Hindgut	25 (11.1)	8 (10.1)	2 (2.0)	6 (5.0)	2 (15.4)	43 (8.0)
	Appendix	75 (33.2)	0 (0.0)	85 (85.0)	9 (7.6)	5 (38.5)	174 (32.4)
	Urogenital	6 (2.7)	65 (82.3)	0 (0.0)	67 (56.3)	1 (7.7)	139 (25.9)
	Other	47 (20.8)	4 (5.1)	5 (5.0)	6 (5.0)	3 (23.1)	65 (12.1)
ASA grade	Grade I	144 (63.7)	23 (29.1)	77 (77.0)	28 (23.5)	2 (15.4)	274 (51.0)
	Grade II	60 (26.5)	51 (64.6)	21 (21.0)	85 (71.4)	4 (30.8)	221 (41.2)
	Grade III	21 (9.3)	3 (3.8)	2 (2.0)	6 (5.0)	6 (46.2)	38 (7.1)
	Grade IV/V	0 (0.0)	2 (2.5)	0 (0.0)	0 (0.0)	1 (7.7)	3 (0.6)
	(Missing)	1 (0.4)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	1 (0.2)
WHO Checklist	Yes	214 (94.7)	79 (100.0)	99 (99.0)	116 (97.5)	10 (76.9)	518 (96.5)
	No	12 (5.3)	0 (0.0)	1 (1.0)	3 (2.5)	3 (23.1)	19 (3.5)
Operation grade	Intermediate/Minor	79 (35.7)	0 (0.0)	85 (85.0)	9 (7.6)	5 (38.5)	178 (33.6)
	Major	142 (64.3)	78 (100.0)	15 (15.0)	109 (92.4)	8 (61.5)	352 (66.4)
Contamination	Clean/Clean-contaminated	44 (19.5)	73 (92.4)	14 (14.0)	110 (92.4)	2 (15.4)	243 (45.3)
	Contaminated	106 (46.9)	6 (7.6)	38 (38.0)	7 (5.9)	5 (38.5)	162 (30.2)
	Dirty	74 (32.7)	0 (0.0)	48 (48.0)	1 (0.8)	6 (46.2)	129 (24.0)
	(Missing)	2 (0.9)	0 (0.0)	0 (0.0)	1 (0.8)	0 (0.0)	3 (0.6)
Approach	Open midline	175 (77.4)	11 (13.9)	33 (33.0)	28 (23.5)	7 (53.8)	254 (47.3)
	Open non-midline	50 (22.1)	65 (82.3)	67 (67.0)	89 (74.8)	6 (46.2)	277 (51.6)
	Laparoscopic attempted	0 (0.0)	3 (3.8)	0 (0.0)	2 (1.7)	0 (0.0)	5 (0.9)
	(Missing)	1 (0.4)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	1 (0.2)
Stoma formation	Yes	9 (4.0)	5 (6.3)	0 (0.0)	2 (1.7)	0 (0.0)	16 (3.0)
	No	215 (95.1)	74 (93.7)	100 (100.0)	117 (98.3)	13 (100.0)	519 (96.6)
	(Missing)	2 (0.9)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	2 (0.4)

Table 4.3. Measurement processes (quantitative phase)

Feature	Levels	Ghana	India	Benin	Mexico	Nigeria	Total
Language of delivery *(translated WHQ)*	English	32 (14.2)	2 (2.5)	0 (0.0)	0 (0.0)	9 (69.2)	43 (8.0)
	French	0 (0.0)	0 (0.0)	88 (88.0)	0 (0.0)	0 (0.0)	88 (16.4)
	Hindi	0 (0.0)	52 (65.8)	0 (0.0)	0 (0.0)	0 (0.0)	52 (9.7)
	Punjabi	0 (0.0)	20 (25.3)	0 (0.0)	0 (0.0)	0 (0.0)	20 (3.7)
	Spanish	0 (0.0)	0 (0.0)	0 (0.0)	119 (100.0)	0 (0.0)	119 (22.2)
	Tamil	0 (0.0)	3 (3.8)	0 (0.0)	0 (0.0)	0 (0.0)	3 (0.6)
Language of delivery *(ad hoc translation)*	Dagbani	38 (16.8)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	38 (7.1)
	Fante	8 (3.5)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	8 (1.5)
	Fon	0 (0.0)	0 (0.0)	6 (6.0)	0 (0.0)	0 (0.0)	6 (1.1)
	Goun	0 (0.0)	0 (0.0)	6 (6.0)	0 (0.0)	0 (0.0)	6 (1.1)
	Malayalam	0 (0.0)	2 (2.5)	0 (0.0)	0 (0.0)	0 (0.0)	2 (0.4)
	Twi	148 (65.5)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	148 (27.6)
	Yoruba	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	4 (30.8)	4 (0.7)
Phone owner	Patient themselves	142 (62.8)	23 (29.1)	84 (84.0)	85 (71.4)	10 (76.9)	344 (64.1)
	Healthcare worker	0 (0.0)	1 (1.3)	0 (0.0)	0 (0.0)	0 (0.0)	1 (0.2)
	Friend or relative	83 (36.7)	54 (68.4)	16 (16.0)	34 (28.6)	2 (15.4)	189 (35.2)
	Other	1 (0.4)	1 (1.3)	0 (0.0)	0 (0.0)	1 (7.7)	3 (0.6)
Phone type	Landline phone	0 (0.0)	1 (1.3)	0 (0.0)	2 (1.7)	0 (0.0)	3 (0.6)
	Mobile phone (with a camera)	118 (52.2)	70 (88.6)	77 (77.0)	104 (87.4)	11 (84.6)	380 (70.8)
	Mobile phone (without a camera)	108 (47.8)	8 (10.1)	23 (23.0)	13 (10.9)	2 (15.4)	154 (28.7)
Questionnaire administrator	Consultant (doctor)	0 (0.0)	0 (0.0)	0 (0.0)	7 (5.9)	2 (15.4)	9 (1.7)
	Other doctor	132 (58.4)	0 (0.0)	100 (100.0)	75 (63.0)	0 (0.0)	307 (57.2)
	Research nurse	65 (28.8)	76 (96.2)	0 (0.0)	0 (0.0)	1 (7.7)	142 (26.4)
	Other	28 (12.4)	3 (3.8)	0 (0.0)	37 (31.1)	10 (76.9)	78 (14.5)
	(Missing)	1 (0.4)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	1 (0.2)

4.4.3.1 Unidimensionality of scale

The exploratory Rasch model was fitted using these data from 537 patients (369 excluding extremes) across 5 class intervals (*Table 4.4*). Both analysis of principal components between positively and negatively loading items (1.86%, N=10 independent t-tests <5%) and symptom and pathway items (0.56%, N=8) suggested unidimensionality of the WHQ instrument in detection of surgical site infection.

Table 4.4. Class interval structure in Rasch analysis

		Class interval				
Item	**Description**	**1**	**2**	**3**	**4**	**5**
I0001	Redness	110	96	75	74	14
I0002	Heat	110	96	75	74	14
I0003	Clear fluid	110	96	75	74	14
I0004	Blood-stained fluid	110	96	75	74	14
I0005	Purulent fluid	110	96	75	74	14
I0006	Wound opening	110	96	74	74	14
I0007	Deep wound opening	10	11	10	9	6
I0008	Local swelling	110	96	75	74	14
I0009	Smell	110	96	75	75	12
I0010	Tenderness	110	96	75	74	14
I0011	Fever	110	96	74	74	14
I0012	Advice	110	96	75	74	14
I0013	Dressing	110	96	75	74	14
I0014	Readmission	110	96	75	74	14
I0015	Antibiotics	110	96	75	74	14
I0016	Deliberate opening	110	96	75	74	14
I0017	Wound scraping	110	96	75	74	14
I0018	Wound drained	110	96	75	74	14
I0019	Reoperated	110	96	75	74	14

4.4.3.2 Model fit and targeting

Overall, the model did not fit well with a high probability of item-trait interaction (χ^2 209.2, DF: 76, $p<0.0000001$), and a poor person-separation index (0.48, low power of analysis). Conversely, Cronbach's alpha (with missing data excluded) demonstrated acceptable internal consistency (0.86). There was a strong positive skew of person location values with the mean person location of -2.91 (s.d. 1.05), demonstrating some mistargeting of the WHQ as may be expected in a diagnostic or screening tool (*Figure 4.2)*. The item-location map reflected clinical severity (*Figure 4.3*) with 31.3% (168/537) of participants at the floor of the scale (i.e., no signs or symptoms of SSI) and item locations reflecting degrees of infection at the ceiling.

Figure 4.2. Person-threshold distribution map of the WHQ

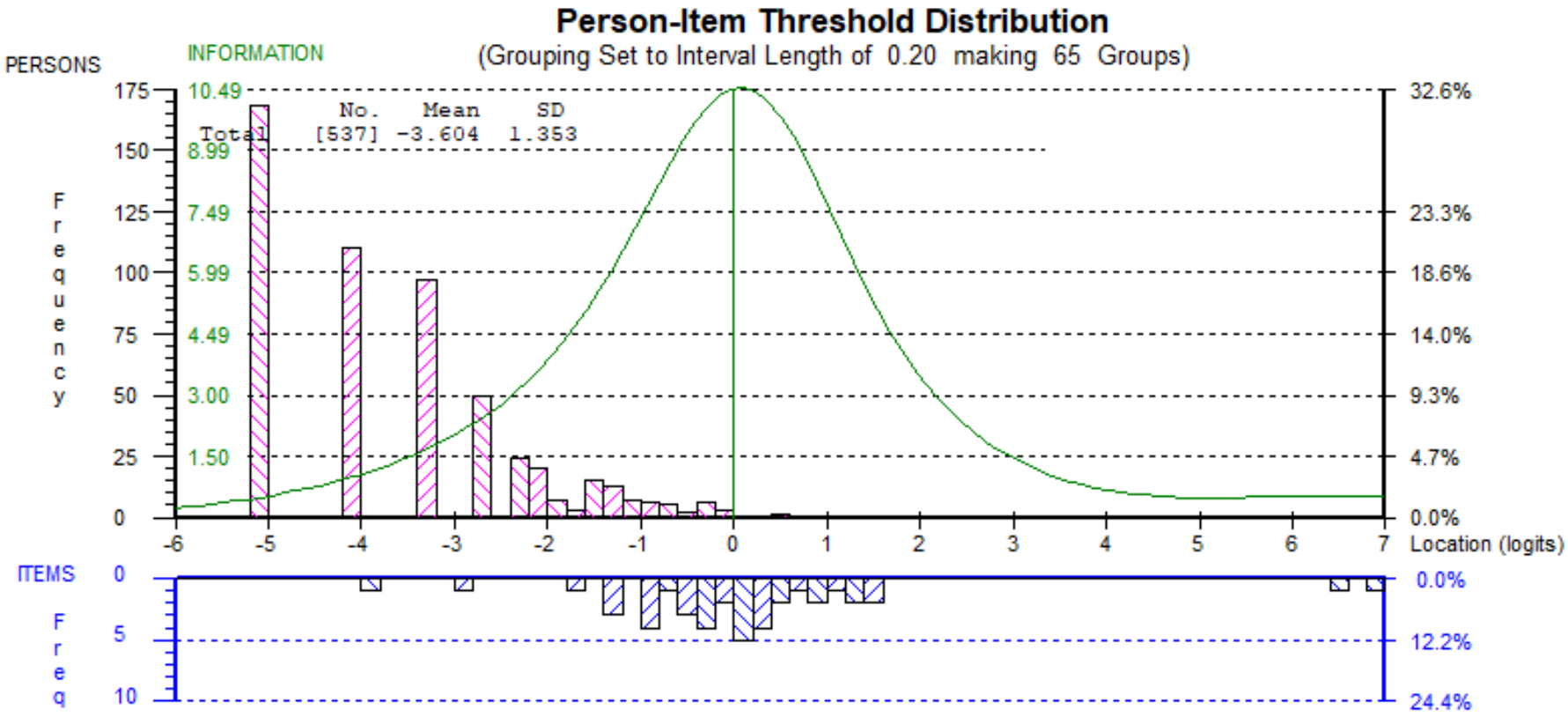
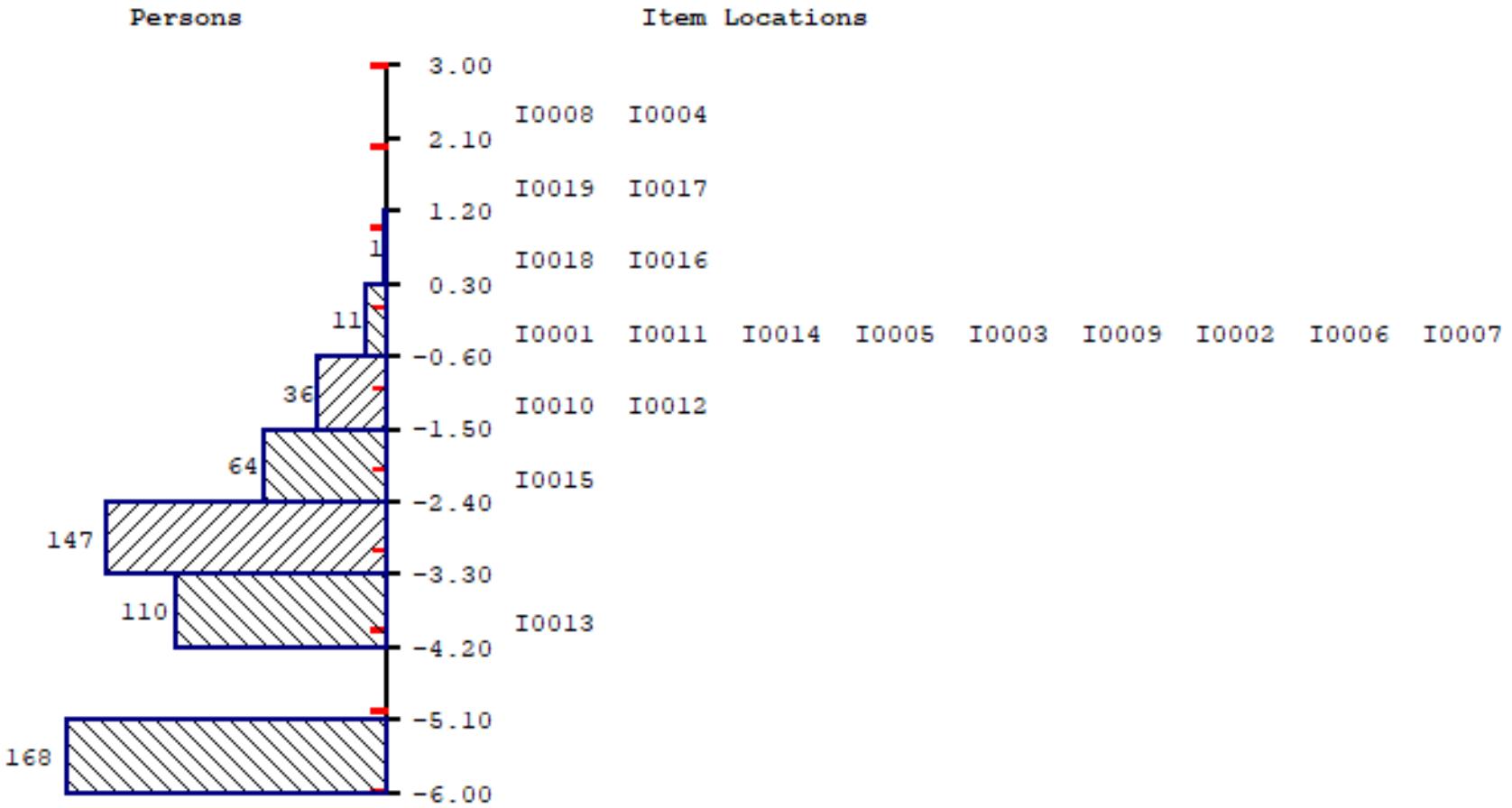


Figure 4.3. Item-location map for the adapted WHQ



Item 13 (dressing) was the lowest item, indicating many of the participants would have scored on this item. Items such as 8 (Local swelling), 4 (Blood stained fluid), 19 (Reoperation) and 17 (Wound debridement) were at the ceiling and participants were more likely to affirm these if items if they had a more severe infection.

4.4.3.3 Individual item fit and dependency

Several items displayed significant misfit to the model (mean item fit residual: -1.61, s.d. 1.75, *Table 4.5)*, but person-fit was acceptable (mean person fit residual -0.52, s.d. 0.69). Examination of individual person fit did not reveal any significant misfit (s.d. of fit residual greater than +2.5 or less than -2.5). Individual item fit (*Table 4.5*) and between-item residual correlation (*Table 4.8*) is triangulated with qualitative data below. There was a high degree of correlation and dependence between items with local dependency in 11 item pairs.

4.4.3.4 Differential item functioning

There was significant evidence of uniform DIF by country in items 1, 3, 5, 8, 10 and 13 and non-uniform DIF by country in items 4, 10, 13, 16, 17 and 19 (*Table 4.6*). There was no significant DIF observed by patient home location (*Table 4.7*).

Table 4.5. Individual item fit in Rasch analysis

Item	Description	Location	SE	Fit residual	DF	χ^2	DF	Prob	F-statistic	DF1	DF2	Prob
I0001	Redness	-0.54	0.106	-0.459	346.42	6.129	4	0.189686	1.28	4	364	0.277285
I0002	Heat	0.148	0.123	-0.574	346.42	5.142	4	0.27303	0.965	4	364	0.426828
I0003	Clear fluid	-0.012	0.125	-0.402	346.42	2.158	4	0.706659	0.387	4	364	0.817702
I0004	Blood-stained fluid	2.358	0.156	-1.863	346.42	5.396	4	0.249023	1.461	4	364	0.213624
I0005	***Purulent fluid***	-0.063	0.126	-2.837	346.42	11.977	4	0.017525	4.809	4	364	0.000864
I0006	Wound opening	0.166	0.137	-1.769	345.49	7.386	4	0.116833	1.513	4	363	0.197776
I0007	Deep wound opening	0.294	0.243	0.407	43.19	5.71	4	0.221911	1.232	4	41	0.31248
I0008	Local swelling	2.196	0.134	-1.251	346.42	7.844	4	0.097462	1.905	4	364	0.108954
I0009	***Smell***	0.115	0.137	-3.519	345.49	14.846	4	0.005032	6.62	4	363	0.000044
I0010	***Tenderness***	-1.332	0.083	2.535	346.42	31.471	4	0.000002	7.273	4	364	0.000015
I0011	***Fever***	-0.503	0.107	0.172	345.49	24.686	4	0.000059	6.607	4	363	0.000041
I0012	***Advice***	-0.979	0.153	-2.772	346.42	11.437	4	0.022073	3.403	4	364	0.009473
I0013	Dressing	-3.992	0.122	0.028	346.42	5.976	4	0.200922	1.691	4	364	0.151444
I0014	***Readmission***	-0.433	0.178	-4.5	346.42	18.184	4	0.001137	10.985	4	364	0
I0015	***Antibiotics***	-1.617	0.134	-3.891	346.42	25.537	4	0.000041	8.655	4	364	0
I0016	***Clinician opening***	0.847	0.277	-3.213	346.42	9.51	4	0.04954	8.658	4	364	0.000002
I0017	Wound scraping	1.489	0.361	-2.335	346.42	4.568	4	0.334617	3.285	4	364	0.011551
I0018	Wound drained	0.467	0.239	-1.883	346.42	4.61	4	0.329672	1.781	4	364	0.132017
I0019	***Reoperated***	1.392	0.347	-2.497	346.42	6.673	4	0.154187	6.222	4	364	0.000076

Items that appeared to misfit the Rasch model based on one or more assessment criteria highlighted in bold. P-values highlighted for Bonferroni-adjusted value of P<0.000877 (base alpha 0.05) to account for multiplicity in testing.

Table 4.6. Exploration of differential item functioning by country

		Class interval (ANOVA)				Country (uniform DIF)				Class interval by country (non-uniform DIF)			
Item	**Description**	**MS**	**F-statistic**	**DF**	**Prob**	**MS**	**F-statistic**	**DF**	**Prob**	**MS**	**F-statistic**	**DF**	**Prob**
I0001	Redness	1.10905	1.44985	4	0.217174	5.95565	7.78575	4	0.000001	1.8406	2.40619	15	0.002484
I0002	Heat	0.81154	0.97723	4	0.419959	1.32731	1.59832	4	0.17429	0.95976	1.15572	15	0.305279
I0003	Clear fluid	0.33813	0.40548	4	0.804693	5.13824	6.1617	4	0.000076	0.63418	0.7605	15	0.721338
I0004	Blood-stained fluid	0.82023	1.58117	4	0.178817	0.58955	1.13649	4	0.339085	1.53853	2.96583	15	0.000184
I0005	Purulent fluid	2.03576	5.2885	4	0.000383	2.55904	6.64789	4	0.00004	0.736	1.91197	15	0.021178
I0006	Wound opening	0.97441	1.57546	4	0.180351	2.25915	3.65269	4	0.006259	0.79936	1.29243	15	0.204236
I0007	Deep wound opening	1.42662	1.7364	4	0.169002	2.40595	2.9284	3	0.050296	1.82772	2.2246	9	0.049837
I0008	Local swelling	1.47699	2.13467	4	0.076134	5.33207	7.70634	4	0.000005	1.47582	2.13297	15	0.008346
I0009	Smell	2.47752	6.98484	4	0.000021	1.35706	3.82596	4	0.004665	0.56007	1.57901	15	0.077303
I0010	Tenderness	7.85005	8.81738	4	0	12.63569	14.19274	4	0.000006	2.34485	2.63379	15	0.000872
I0011	Fever	6.05665	6.72738	4	0.000032	2.17053	2.4109	4	0.04896	0.95995	1.06626	15	0.386969
I0012	Advice	1.95626	3.47161	4	0.008488	0.55603	0.98674	4	0.414748	0.83991	1.49052	15	0.106046
I0013	Dressing	1.58484	2.09556	4	0.080998	11.54038	15.2593	4	0	2.27201	3.00417	15	0.000153
I0014	Readmission	3.30776	10.88698	4	0	0.18136	0.59693	4	0.665092	0.27042	0.89004	15	0.575761
I0015	Antibiotics	4.74728	8.73443	4	0.000006	1.06867	1.96623	4	0.099221	0.52497	0.96588	15	0.491144
I0016	Clinician opening	1.7993	9.72982	4	0.000005	0.59012	3.19112	4	0.013583	0.63243	3.4199	15	0.000022
I0017	Wound scraping	0.7204	3.65057	4	0.006284	0.68523	3.47234	4	0.008476	0.6001	3.04095	15	0.000128
I0018	Wound drained	0.83979	1.85386	4	0.118133	0.73257	1.61718	4	0.16943	0.82874	1.82947	15	0.029577
I0019	Reoperated	1.28757	7.12483	4	0.000023	0.36665	2.02889	4	0.089948	0.76782	4.24876	15	0

Table 4.7. Exploration of differential item functioning by patient home location (urban versus rural)

		Class interval (ANOVA)				Home location (uniform DIF)				Class interval by home location (non-uniform DIF)			
Item	**Description**	**MS**	**F-statistic**	**DF**	**Prob**	**MS**	**F-statistic**	**DF**	**Prob**	**MS**	**F-statistic**	**DF**	**Prob**
I0001	Redness	1.10906	1.28119	4	0.276925	1.97513	2.28168	1	0.131792	0.64867	0.74935	4	0.558954
I0002	Heat	0.81154	0.96149	4	0.428636	0.05993	0.07101	1	0.790022	0.78441	0.92934	4	0.446887
I0003	Clear fluid	0.33813	0.38624	4	0.818478	0.00328	0.00375	1	0.951239	0.86892	0.99256	4	0.411526
I0004	Blood-stained fluid	0.82023	1.45809	4	0.214462	0.55646	0.9892	1	0.320601	0.47444	0.8434	4	0.498362
I0005	Purulent fluid	2.03576	4.8977	4	0.000742	0.30434	0.73218	1	0.392754	1.13895	2.74012	4	0.028591
I0006	Wound opening	0.97441	1.51754	4	0.196467	0.61448	0.95699	1	0.328607	0.82575	1.28603	4	0.275033
I0007	Deep wound opening	1.42662	1.30231	4	0.28761	2.07281	1.8922	1	0.177453	1.49612	1.36576	4	0.265085
I0008	Local swelling	1.47699	1.92532	4	0.105648	1.86336	2.42897	1	0.119991	1.22648	1.59877	4	0.174069
I0009	Smell	2.47751	6.56173	4	0.000034	0.04832	0.12797	1	0.720756	0.15683	0.41536	4	0.797571
I0010	Tenderness	7.85005	7.53315	4	0.000005	6.50787	6.24515	1	0.012894	3.064	2.94031	4	0.02055
I0011	Fever	6.05665	6.56487	4	0.000045	0.19971	0.21647	1	0.642034	0.5747	0.62292	4	0.646432
I0012	Advice	1.95626	3.38917	4	0.009717	0.05805	0.10057	1	0.751322	0.48865	0.84657	4	0.496397
I0013	Dressing	1.58484	1.70809	4	0.147585	0.66633	0.71815	1	0.397315	1.84948	1.99331	4	0.095016
I0014	Readmission	3.30776	11.28915	4	0.000008	0.00005	0.00016	1	0.989568	1.10343	3.76593	4	0.005144
I0015	Antibiotics	4.74728	9.06901	4	0.000008	1.85356	3.54097	1	0.060681	2.47129	4.72105	4	0.001009
I0016	Clinician opening	1.7993	8.64645	4	0.000006	0.21679	1.04179	1	0.308094	0.18073	0.86851	4	0.482919
I0017	Wound scraping	0.7204	3.30981	4	0.011091	0.35397	1.62629	1	0.20304	0.33296	1.52974	4	0.192945
I0018	Wound drained	0.83979	1.78495	4	0.131242	0.02946	0.06261	1	0.802543	0.678	1.44107	4	0.219887
I0019	Reoperated	1.28757	6.1947	4	0.000072	0.18986	0.91344	1	0.339846	0.13069	0.62876	4	0.642267

4.4.4 Cross-contextual and cross-language equivalence of items and response scales

A summary of qualitative data across four pre-defined categories is presented for 'symptoms' items in *Appendix C* and 'treatment' items in *Appendix D*. Triangulation of qualitative and quantitative data was performed item-by-item for the 11 'symptom' items (10 items and 1 sub-item) and 8 'pathway' items to recommend modifications, and explore individual item fit and correlations below.

4.4.4.1 Item response categories

Investigators from all six countries raised concerns about translatability, comprehension, and judgement between two item response levels for 'symptoms' items 1 to 10: (1) 'A little bit' and (2) 'Quite a lot':

> *"The difference between 'a little bit' and 'quite a lot' is for the very "English" English" (Surgeon, Focus group NG002F, Nigeria).*
>
> *"I think it is sometimes challenging trying to explain to patients to find the balance where it was... a little versus quite a bit" (Research nurse, Focus group GH001F, Ghana).*
>
> "Quite a bit, a little means the same (in Hindi) I think" (*Surgeon*).
>
> "Yes, differentiating between quite a bit and a lot will be a bit difficult for the patient". (*Research nurse, Focus group IN002F, India*)

Feedback from focus groups also highlighted that many patients might struggle to understand the scalar nature of four item response levels when delivered over the telephone:

> *"Questionnaires in Nigeria are more like a conversation than a very structured interaction – it is what is needed to keep the patient engaged. They won't understand the question as a scale, and the assessor would have to make their best guess" (Research nurse, Interview NG001I, Nigeria)*

In contrast, one included country (Mexico) felt translation of the item would be possible into Spanish language. However, investigators agreed that comprehension would remain a problem for less health literate patients:

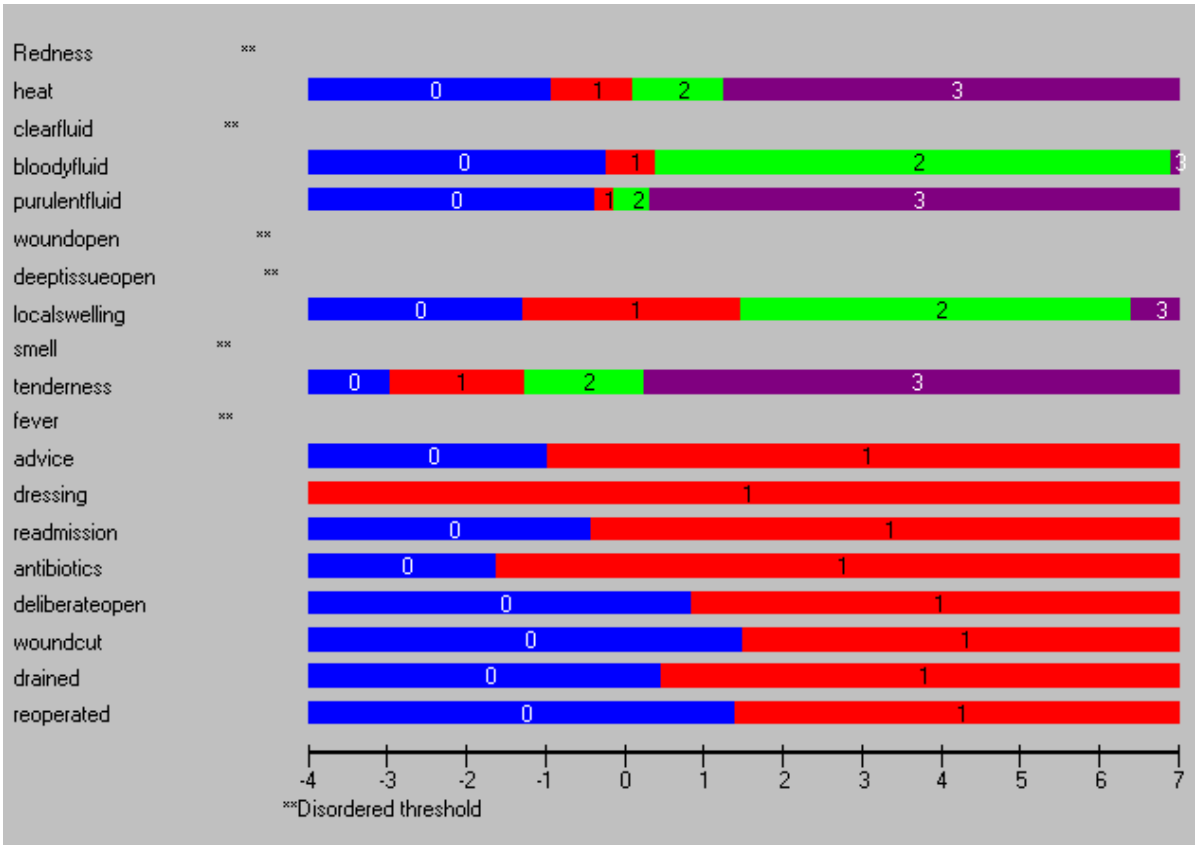
> *"You can translate both 'a little bit' and 'quite a lot' into Spanish, but its whether they (patients) will understand the difference in real life" (Research nurse, Interview MX001I, Mexico).*

Similarly, investigators from Nigeria, Mexico and Rwanda highlighted that local patients were typically unable to discern whether they had a true fever (Item 11), and many felt 'hot' or 'not right' after surgery without knowing why:

> *"Typically, patients either feel either yes or no, very hot or normal" (Research Nurse, Interview NG002I, Nigeria).*
>
> *"The one with the fever, it's the answers are a little weird, because you cannot say not at all, a little, quite a bit or a lot, because it's you had or you have it or you don't. (Surgical trainee, Focus group MX001F, Mexico).*

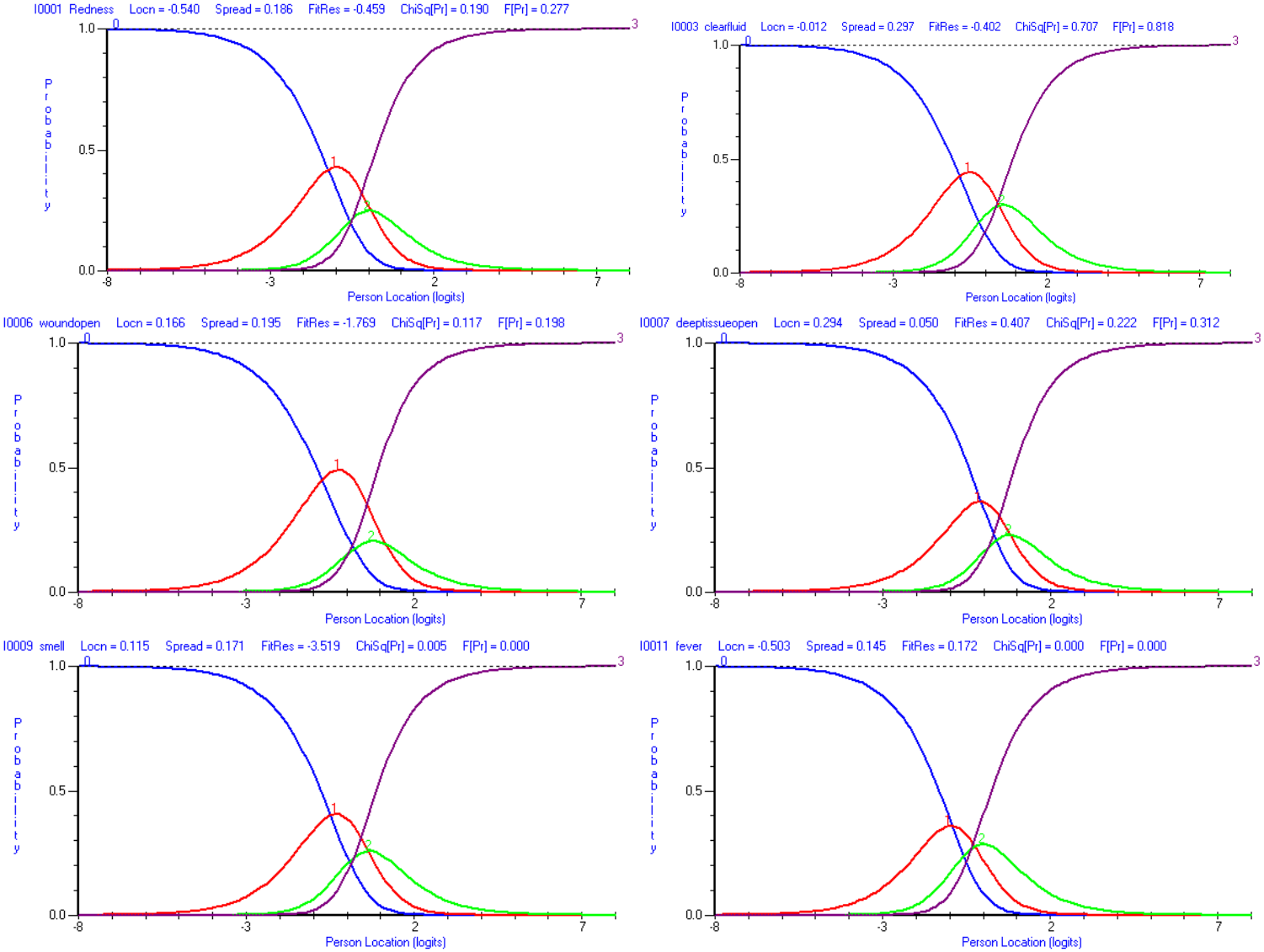
In the qualitative harmonisation meeting, investigators proposed a three-level scale ("None", "A little", "A lot") for 'symptoms' items in the adapted questionnaire, and two-level scale ("No", "Yes") for Item 11 (Fever). Data were collected in the cohort study using the original WHQ item response categories. Upon Rasch analysis, disordered thresholds were detected for a majority of 'symptoms' items: Item 1 (Redness), Item 3 (Clear fluid), Item 6 (Wound opening), Item 7 (Deep tissue opening), Item 9 (Smell) and Item 11 (Fever). A threshold map is shown in *Figure 4.4* and relevant probability curves in *Figure 4.5.* Triangulated with qualitative data, this supported reconfiguration of the item response categories in the adapted Wound Healing Questionnaire and was agreed in the final harmonisation meeting.

Figure 4.4. Threshold map for WHQ



Wound Healing Questionnaire (WHQ) tems 1 to 19 are listed along the y-axis. The logit threshold values for each item response are listed on the x-axis. Items with disordered thresholds are indicated with **.

Figure 4.5. Category probability curves with overlapping response thresholds



Items displayed (from top, reading left to right) are Item 1 (Redness), Item 3 (Clear fluid), Item 6 (Wound opening), Item 7 (Deep wound opening), Item 9 (Smell), Item 11 (Fever).

4.4.4.2 Item 1. Redness

Researchers from several countries reported poor cross-contextual relevance of ‘redness of the skin’ and perceived challenges in translating this in a way that would be comprehensible to patients in a local language, particularly for patients with dark skin tones (e.g., in Sub-Saharan Africa):

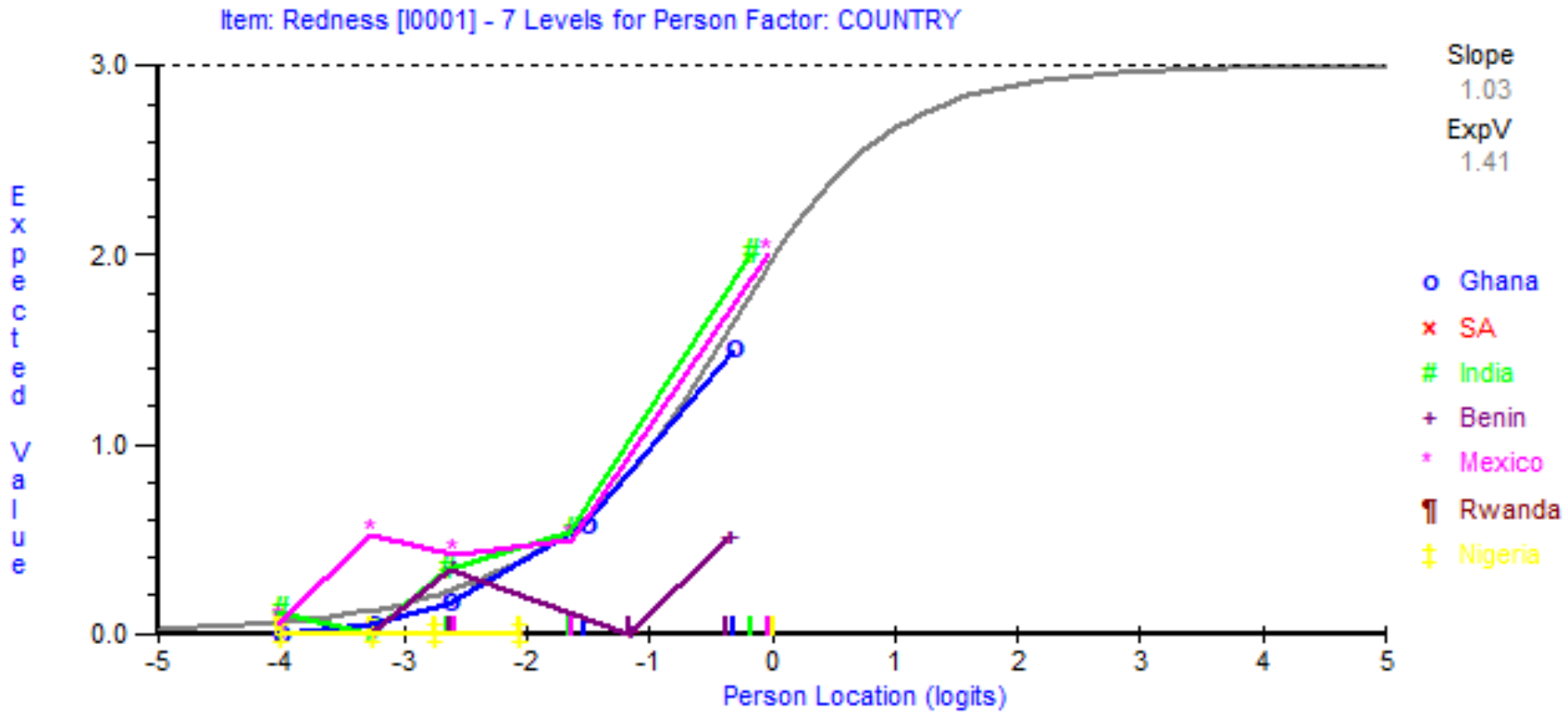
> *“Redness around the wound, from my experience it becomes a bit of a challenge for many of our patients to recognise that redness, bearing in mind the colour of the skin.” (Surgeon, Focus group NG002F, Nigeria).*
>
> *“The red colour is difficult to find in black skin” (Surgeon, Interview BN001I, Benin)*

This was also reflected in cohort study data from Benin, South Africa, Ghana, and Nigeria where redness of the skin demonstrated poor discrimination with significant non-uniform DIF (*Figure 4.6*). Investigators suggested that an equivalent concept would be ‘shining’ or ‘tightness’ of the skin and amended wording for this item reached agreement during the final harmonisation meeting.

> *“I suggest we substitute shining skin with redness, with dark skin, inflamed skin we tend to ask or we see it shining - that is an adequate replacement” (Surgeon, Focus group NG002F, Nigeria).*

Recommendation 1: Amend wording to capture ‘shining of the skin’ to ensure conceptual relevance to dark skin tones. Modify wording during translation into Ghanian dialects to ensure conceptual equivalence.

Figure 4.6. Differential item functioning by country for symptoms items 1 (Redness)



4.4.4.3 Item 2. Warmth

Some concerns were raised about patients' comprehension, retrieval and judgment for this item, with some investigators worried that they would only be aware of 'abnormal warmth' if a healthcare worker had informed them of this:

> *"It may be (that) people only notice heat if doctor (sic) has told them this (Research nurse, Ghana, Interview GH001I)".*

However, this item demonstrated acceptable fit to the Rasch model (fit residual -0.574, p=0.427), and there was no significant DIF by country or language.

Recommendation 2: No adaptation required.

4.4.4.4 Item 3. Clear fluid

Clear fluid was generally considered to be well understood by patients with acceptable retrieval and judgement. A cross-language translatability issue for Ghanian dialects was raised in an interview, with patients describing clear serous fluid as 'water':

> *Patients describe this as 'water' from the wound in Ghana (Research nurse, Ghana, Interview GH001I)".*

Recommendation 3: Clarification of consistency of 'thin' clear fluid.

4.4.4.5 Item 4. Blood-stained fluid

Item 4 was considered to be comprehensible, with no perceived issues in retrieval or judgement. Item fit was satisfactory (fit residual: -1.863, p=0.214) and with no significant DIF. Interesting contextual data were provided from an investigator in Ghana to support retrieval, whereby family members sometimes report blood-stained fluid on dressings:

"Family members may also notice bloody leakage from the skin and on the dressings so will be useful help (to the patient) in assessment (Surgeon, Ghana, Interview GH002I)"

Recommendation 4: No cross-cultural adaptation required.

4.4.4.6 Item 5. Purulent fluid

No themes were identified in the qualitative data related to this item. However, there was some item misfit on Rasch analysis (fit residual -2.837, p=0.0009). Despite perceived structural dependency by the WHQ developers [184] there was no significant local dependency seen in residual correlations of Items 3, 4 and 5 (*Table 4.8*). Where all three 'fluid' items were combined in a subtest (aligned to the structural dependency suggested by the WHQ developers) there was improvement in item fit (fit residual -2.079, p=0.050) and improved ordering of thresholds (*Figure 4.7*), but no improvement in overall model fit (χ^2=199.00, DF:68, p<0.0000001).

Recommendation 5: No cross-cultural adaptation required. In future analyses, consider accounting for structural dependency of Item 3, 4 and 5 using subtesting.

Table 4.8. Exploration of item correlations and local dependency between items in Rasch analysis

Item	I0001	I0002	I0003	I0004	I0005	I0006	I0007	I0008	I0009	I0010	I0011	I0012	I0013	I0014	I0015	I0016	I0017	I0018	I0019
I0001																			
I0002	0.032																		
I0003	-0.128	-0.082																	
I0004	-0.042	-0.061	0.041																
I0005	-0.131	-0.097	0.028	-0.067															
I0006	-0.014	-0.094	0.021	0.071	0.016														
I0007	-0.159	-0.217	-0.062	0.182	-0.305	0.333													
I0008	-0.021	0.049	-0.054	-0.079	-0.12	-0.103	0.098												
I0009	-0.074	-0.1	-0.121	0.107	0.085	-0.019	-0.258	-0.017											
I0010	-0.209	-0.065	-0.104	-0.08	-0.118	-0.163	-0.158	-0.036	-0.129										
I0011	-0.098	-0.065	-0.084	-0.128	-0.049	-0.075	-0.023	-0.174	-0.071	-0.057									
I0012	-0.139	-0.056	-0.019	-0.087	0.043	-0.051	-0.3	-0.104	-0.085	-0.145	-0.029								
I0013	0.054	-0.192	-0.174	-0.121	-0.036	-0.115	-0.073	-0.088	0.077	-0.442	-0.069	0.023							
I0014	-0.108	-0.126	-0.09	-0.028	0.037	0.004	0.174	-0.003	-0.041	-0.147	-0.171	0.178	0.088						
I0015	-0.174	-0.085	0.055	-0.071	0.101	-0.076	-0.036	-0.116	-0.012	-0.216	-0.092	0.106	-0.01	0.209					
I0016	-0.06	-0.102	-0.094	-0.005	-0.016	0.053	0.037	-0.041	0.162	-0.164	-0.18	0.039	0.034	0.167	0.072				
I0017	0.009	-0.087	-0.108	-0.007	-0.031	0.074	-0.009	-0.059	0.111	-0.101	-0.119	-0.02	0.024	0.155	-0.003	0.413			
I0018	-0.08	-0.04	0.099	0.086	-0.024	-0.037	-0.223	-0.073	-0.041	-0.163	-0.128	0.119	0.017	0.032	0.143	0.212	0.015		
I0019	-0.123	-0.098	-0.077	-0.016	-0.028	0.051	0.403	-0.018	0.01	-0.1	-0.149	-0.015	0.025	0.248	0.064	0.67	0.409	0.155	

Highlighted residual correlation coefficients are +0.2 above the mean average of all item residual correlations (-0.027) [226]

Figure 4.7. Threshold probability map after subtest analysis of item 3 (clear fluid), item 4 (blood-stained fluid) and item 5 (purulent fluid)

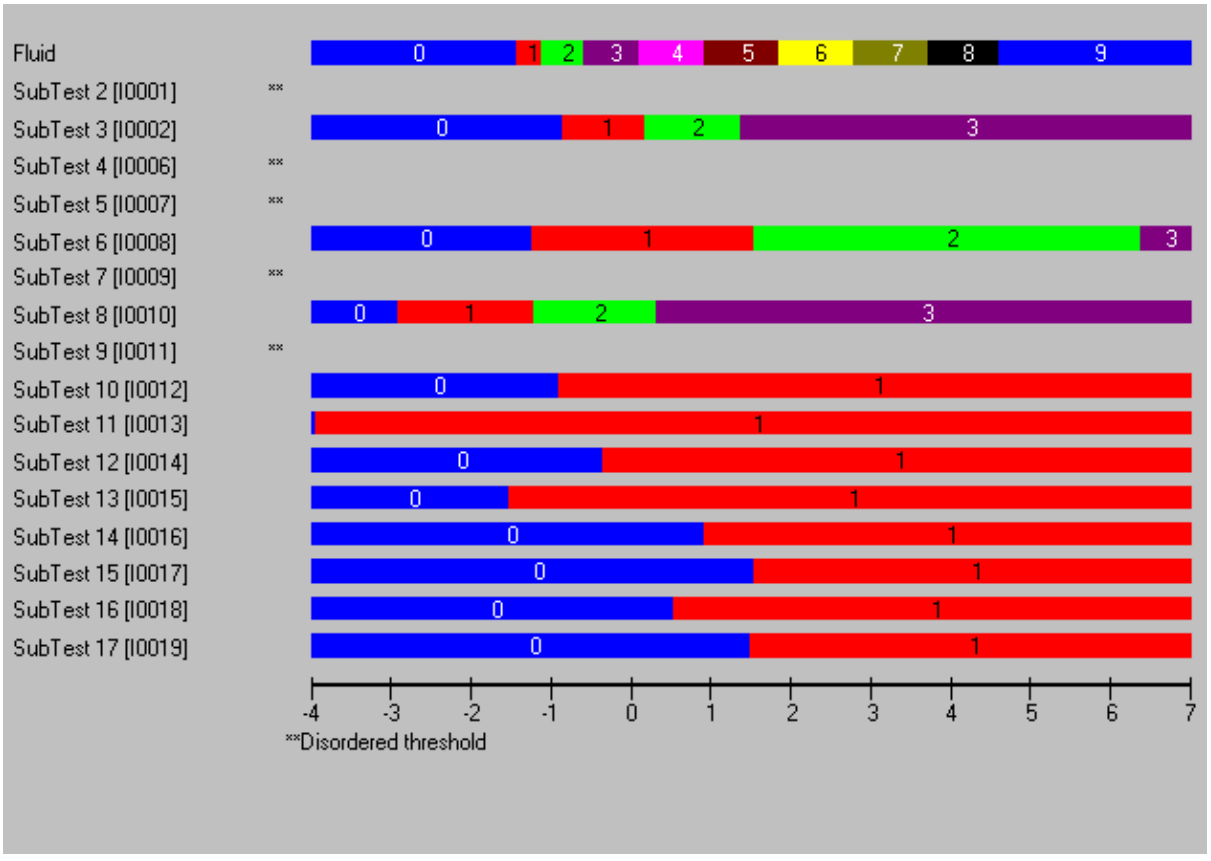
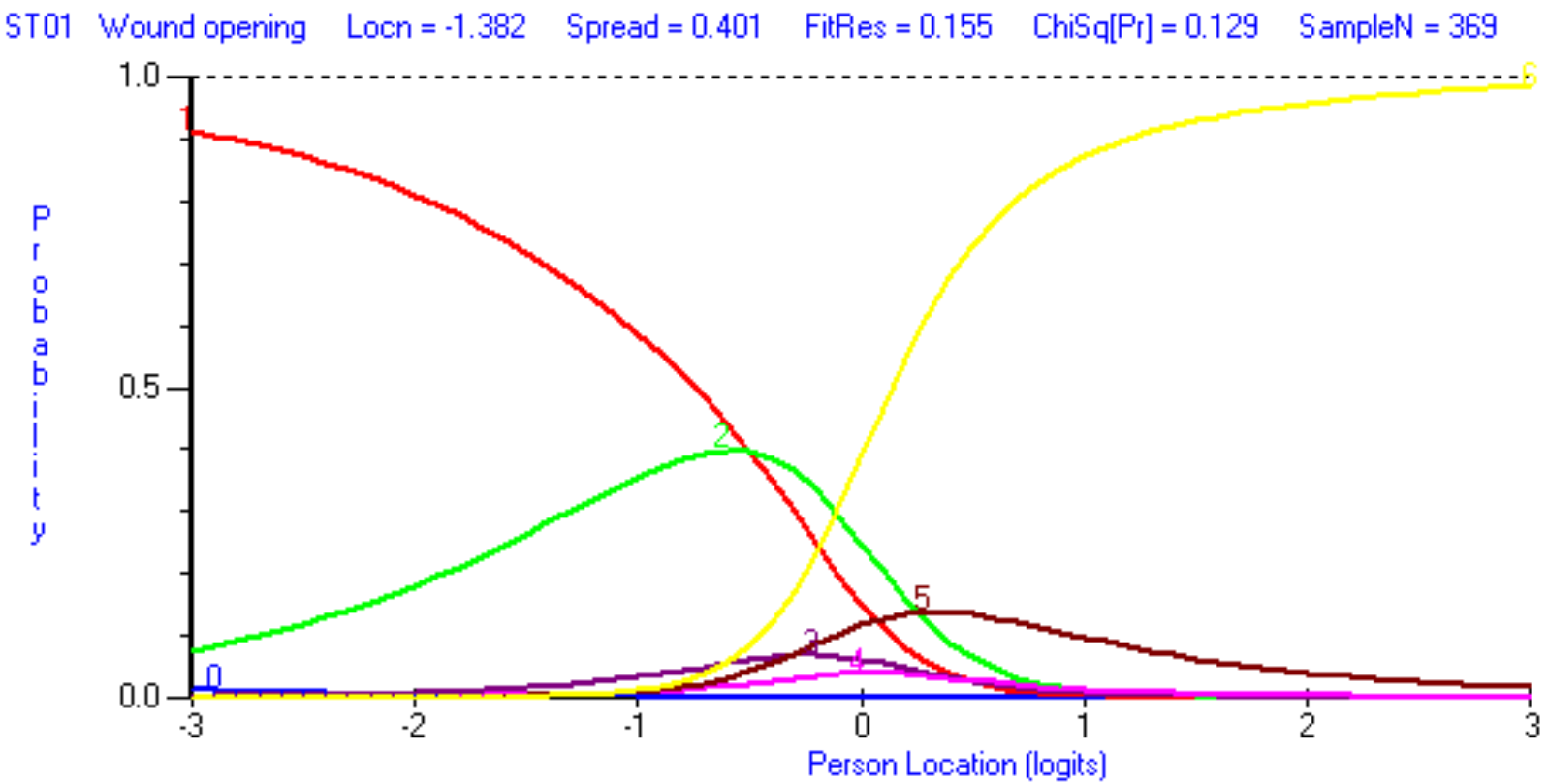


Figure 4.8. Category probability curve upon subtest analysis of item 6 (wound opening) and item 7 (deep wound opening)



4.4.4.7 Item 6. Wound opening

Item 6 largely raised issues with retrieval and judgement in the qualitative data. Investigators were particularly concerned about judgement of lower abdominal incisions which can be hidden from direct view:

> *"Patients might sometimes need someone to help with a caesarean section or low (abdominal) wound. They can't tell if a family member isn't present" (Research nurse, Nigeria, Interview NG001I)*.

Another concern was in retrieval where a wound had opened spontaneously but closed ahead of the 30-day postoperative assessment:

> *"(The wound) may have opened, then later closed, so you might need to ask others" (Research nurse, Nigeria, Interview NG001I).*

A translatability issue was raised in Kinyarwanda where the word for 'opening' can be interpreted as 'healing' (i.e., dressings removed and left to the open air) or as suture removal *(Surgeon, Rwanda, Interview RW001I).* Despite these concerns, item fit was acceptable (fit residual -1.769, p=0.198) and with no significant DIF.

Recommendation 6: No cross-cultural adaptation required. Assessors should ask patients to communicate with family members or carers when assessing Item 6 if the wound cannot be directly visualised. Modify wording during translation into Kinyarwanda to ensure conceptual equivalence.

4.4.4.8 Item 7. Deep wound opening

There were several issues reported with comprehension and judgment for this question. Investigators felt that both the term 'wound' and what constituted 'opening' would be challenging for local populations:

“Deeper tissue is too scientific” (Surgeon, Nigeria, Interview NG002I)

“Patients will not understand the word tissue or be able to tell the difference between superficial and deeper tissues” (Ghana, Focus group GH001F)

“What if they don’t understand or they think the edges are deeper tissue?” (Surgical trainee, Focus group MX001F, Mexico)

This was highlighted particularly for patients with low health literacy:

“People with low literacy will be unable to appreciate differences” (Surgeon, Mexico, Interview MX002I).

Others reflected that some patients with prior medical knowledge may have a greater comprehension of these concepts:

“(I) don’t think they will be able to say it is a tissue unless it is a medical person, unless they are medical or have been told” (Research nurse, India, Interview IN002I).

However, data completeness for this item was high (completed in 46 of 46 patients with ‘a little’ or more wound opening in Item 6) and item fit was acceptable (fit residual 0.407, p=0.312). A solution proposed by investigators in India and Rwanda was to highlight ‘the inside sutures’ in the item description, as something clearly visible in the deep wound space:

“In Rwanda, we commonly ask patients ‘did you also see the blue sutures’ (deep sutures) to help make this clear” (Surgeon, Rwanda, Interview RW002I).

Investigators also highlighted the need for safety netting in settings with low health literacy and limited access to care:

"Patients will want a 'solution'… 'next-steps'…so their needs to be clear pathway for safety netting" (Research nurse, Nigeria, Interview NG001I).

This item was structurally dependent with Item 6 (wound opening) and found to be locally dependent when exploring residual correlations (co-efficient: +0.33, *table 3*). It had a high degree of correlation with several others: item 4 (Blood-stained fluid), item 14 (readmission), item 19 (reoperation). Where item 6 and 7 were analysed together as a subtest, there were still overlapping probability categories (*figure 3*), but the individual item fit improved. Issues with translatability into Spanish where also raised, with the clarification of 'inside sutures' perceived to be important:

"It's hard to translate for 'deeper tissues' and I don't think they (patients) will understand it" (Researcher, Mexico, Interview MX001I).

Recommendation 7: Reword item to support comprehension and judgement by including 'the flesh beneath the skin or the 'inside sutures'. In future analyses, consider accounting for structural and local dependency of Item 6 and 7 using subtesting and collapsed item response categories, or exploring Rasch model fit with this item excluded. The assessor should ensure safety netting via a local clinical pathway if there are any concerns about deep wound opening, and the patient has not yet sought care.

4.4.4.9 Item 8. Local swelling

In general, comprehension and retrieval were thought to be acceptable for Item 8:

"Patients will easily be able to report areas of swelling and whether or not this is around the wound" (Research Nurse, India, Interview IN002I).

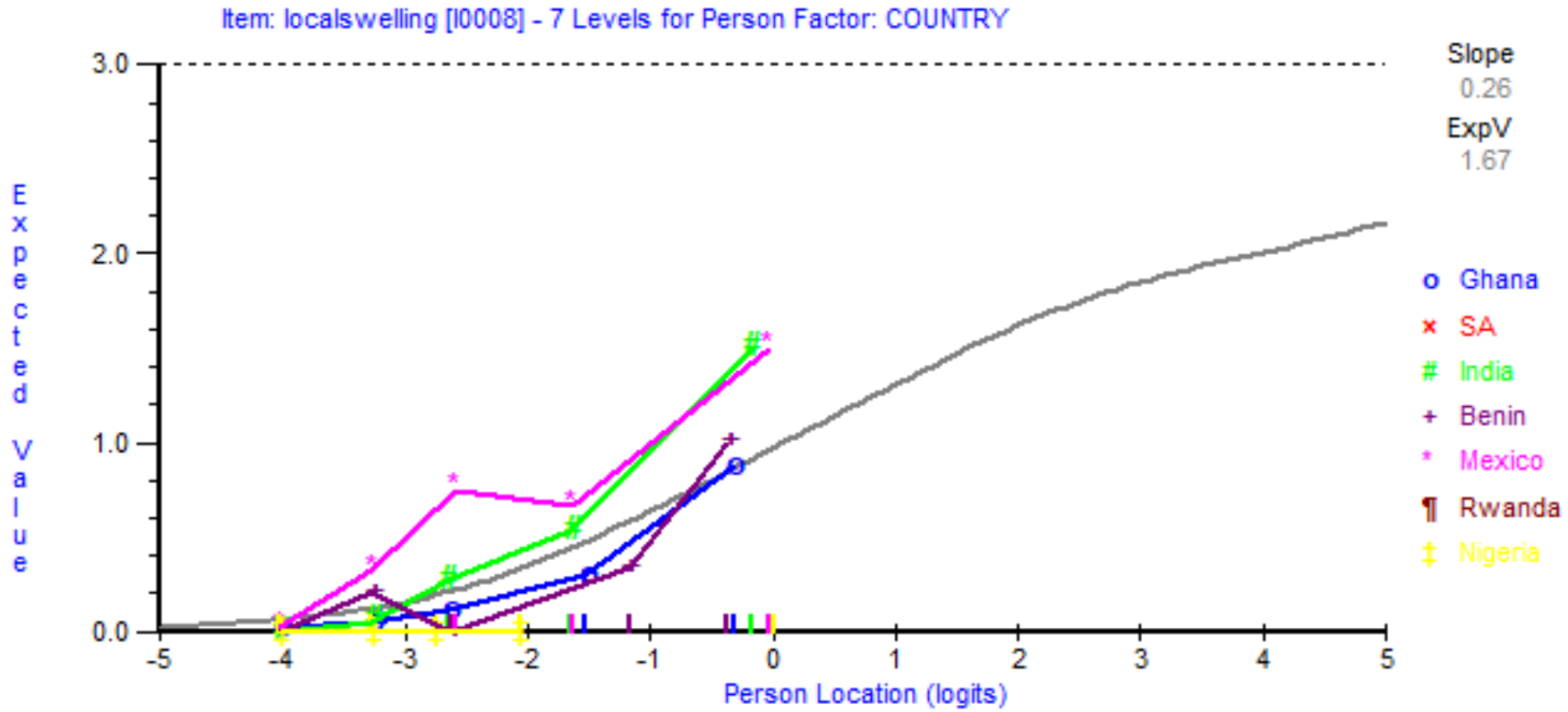
However, there were some issues raised with judgment in comparison to normal healing processes:

"Patients may confuse (it with) swelling or hardness of the wound related to regular healing" (Surgeon, Ghana, Interview GH002I).

Potential variation in judgement across contexts was corroborated in DIF analysis, displaying uniform DIF by country (F-statistic 7.71, P=0.000005, *Figure 4.9*). However, item fit was acceptable (fit residual -1.25, p=0.109).

Recommendation 8: No cross-cultural adaptation required. Consider splitting item for DIF by country in future analyses.

Figure 4.9. Differential item functioning by country for symptoms item 8 (Local swelling)



4.4.4.10 Item 9. Smell

The inclusion of this item was felt to be very important across cultures:

> *"(This is a)….useful symptom to collect as (it's) commonly reported in patients with SSI in Mexico but not part of current SSI assessment in FALCON" (Researcher, Mexico Interview MX002I)*.

Although some investigators, in particular research nurses, were worried about eliciting this information from patients:

> "*Patients may be very ashamed to share this information, and only provide the answer if trust is given (to the interviewer). Especially the female gender, some males too, and probably young adults." (Research nurse, Nigeria, Interview NG001I)*

Others felt that something this notable would be highly likely to be reported:

> *"Patients hate to feel abnormal and they will definitely report something so disturbing" (Surgeon, Ghana, Interview GH002I).*
>
> *"No, they will not be offended by this and be able to answer, because they know it's a part of the routine examination or post-operative follow-up." (Surgeon, Focus group IN002F, India).*

Investigators also acknowledged the important role of healthcare workers and family members in accurate assessment of this item:

> *"(Patients are)….most likely to notice (the wound being smelly) during wound dressing changes by a relative or health workers" (Research nurse, Nigeria, Interview NG001I).*

One potential issue with judgment was proposed in distinguishing wound smells from other sources:

> *""I think it depends on the smell, because the lotions used to dress the wound sometimes they may come with… for example povidone iodine… though it's not an offensive smell." (Research nurse, Nigeria, Interview NG001I).*

In keeping with this there was significant item misfit (fit residual -3.52, p=0.000044), although no significant DIF and no local dependency.

Recommendation 9: No cross-cultural adaptation required. Advise assessor to build rapport with patients early during measurement procedure, ask family members whether they have noticed the wound being 'smelly' during dressing changes, and reassure patient to share symptoms if possible.

4.4.4.11 Item 10. Tenderness

Investigators were concerned with patients' ability to judge normal postoperative pain from pain associated with SSI:

> "*Patients struggle to tell what is 'normal pain' and pain related to SSI" (Research nurse, Ghana, Interview GH001I).*

Others highlighted that patients are discouraged from touching their wound after surgery, but

> *"…their wound may also be painful during daily activities if patients had chosen not to touch the wound or area around it" (Research nurse, India, Interview IN001I).*

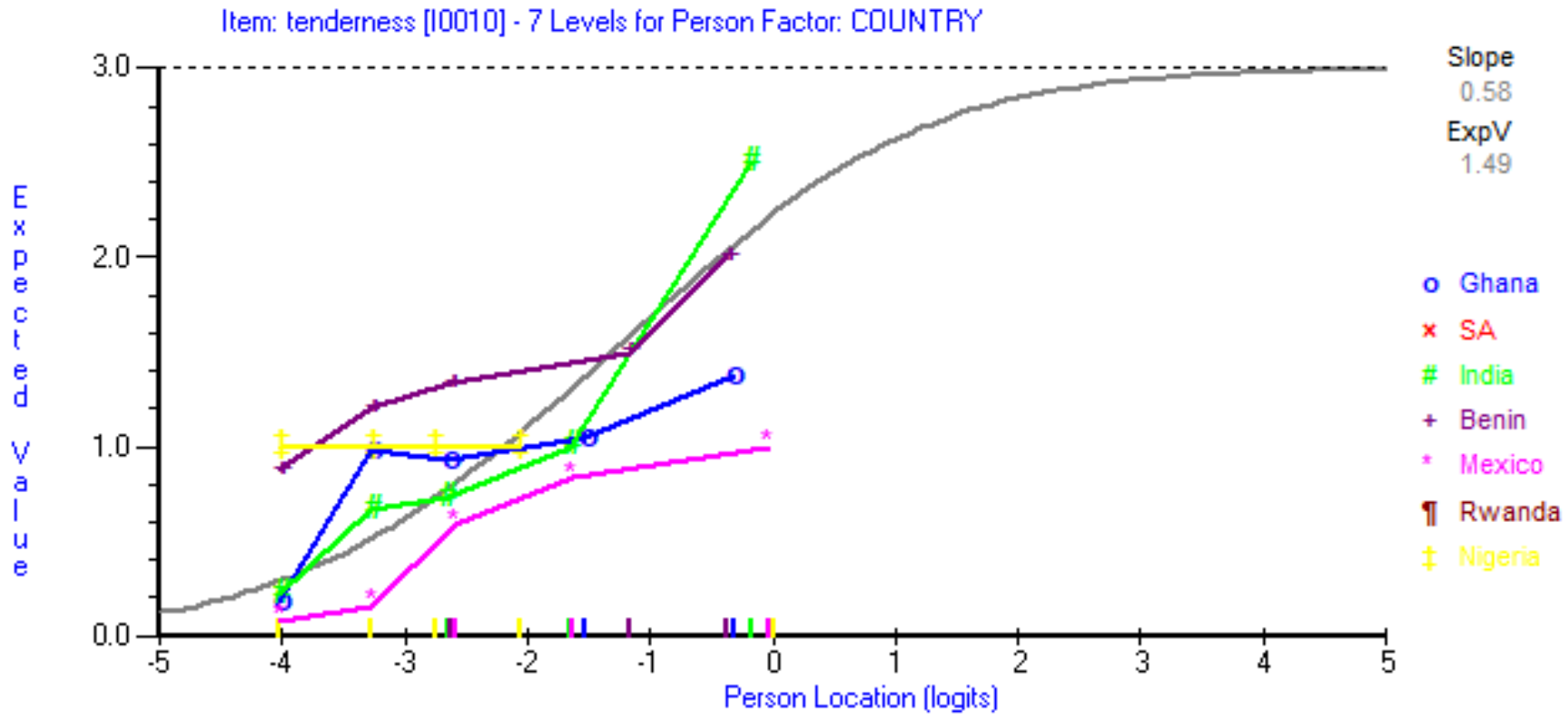
Other activities where patients may have their wound touched by others were also highlighted:

> *"In Ludhiana most of the ladies just answer for these questions while changing the dressings they feel pain. Yeah, otherwise they don't touch so much." (Research nurse, Focus group IN002F, India).*

The item fitted poorly with the Rasch model (fit residual 2.54, p=0.000015), and demonstrated significant uniform (F-statistic 14.19, p=0.000005) and non-uniform DIF (F-statistic 2.63, p=0.0009) by country. It demonstrated underdiscrimination in most countries (*Figure 4.10*). The overall fit for the model only slightly improved when Item 10 was removed (χ^2 176.55, DF 55, p<0.0000001).

Recommendation 10: No cross-cultural adaptation required. Consider dropping item if replicated in future validation studies.

Figure 4.10. Differential item functioning by country for symptoms items 10 (Tenderness)



4.4.4.12 Item 11. Fever

Conflicting themes arose related to comprehension versus judgement for Item 11. Comprehension was generally considered to be good:

> *"Patients have lots of experience reporting temperature as this is a malaria area…patients are usually able to report this well" (Surgeon, Ghana, Interview GH002I).*

However, many investigators reported concerns about judgement. This largely followed two themes: (1) patients very commonly reporting this symptom, particularly in warm environments; (2) lack of access to medical thermometers in the community:

> *"Patients all feel temperature rises, as the weather is very hot, and they are very anxious after surgery. Unless they have a thermometer it's very difficult for them to know" (Surgeon, Rwanda, Interview RW002I).*

Another investigator in Rwanda reported:

> *"(Fever)… may not be useful to wound healing in Rwanda. It doesn't adapt well to our patients" (Focus group, Rwanda, RW001F).*

The inclusion of a degree Celsius cut-off in the item was universally considered to have low contextual relevance in low resource settings:

> *"Patients just do not have access to thermometers, so temperatures are never accurate" (Focus group, India, IN001F).*
>
> *"In Mexico, it's patients say that they have a temperature or fever but they don't really measure the temperature. They just feel a little hot and they say oh I have temperature, I have fever." (Surgeon, Focus group MX001F, Mexico)*

Corroborating this, item 11 demonstrated significant misfit (fit residual 0.172, p=0.00004), but no significant DIF or dependency. The model fit did not improve when Item 11 was deleted (χ^2 191.30, DF 72, p<0.0000001).

Recommendation 11. Remove reference to >38^{o}C as patients do not have the ability to record this in low resource settings.

4.4.4.13 Item 12. Advice

No themes were identified to improve cross-contextual relevance of Item 12. Comprehension and judgement were considered to be satisfactory:

> *"This is easy for patients. They will be given a routine clinic date before discharge, so will be told to come back early if needed" (Research nurse, Ghana, Interview GH001I).*

Some interesting contextual themes arose related to where patients sought help, ranging from traditional healers, community health workers, primary care, district hospitals, and the surgical centre:

> *"…some of them they are in the very remotest areas, they can see these herbalist, or herbal occupation - they may worsen a disease" (Anaesthetist, Focus group GH002F, Ghana)*
>
> *"The majority of patients now go to the nearby clinic or they invite the health worker to their house, or traditional healers. But the majority, say 80 to 90% visit the nearby clinic, or ask the health worker to come down to their house to assess their wound" (Surgeon, Nigeria, Interview NG002I).*

Item 12 displayed acceptable fit with the Rasch model with no significant DIF.

Recommendation 12: No cross-cultural adaptation required.

4.4.4.14 Item 13. Dressing

Variation in standard of wound care between included countries led to concern about judgement and retrieval. In India:

> *"patients often go home with wound care packages, or visit the hospital or healthcare centre for wound care" (Research nurse, India, Interview IN001I),*

whereas in Nigeria:

> *"the practice here is after the surgery most of the patients when they are discharged on the ward the wound is healed and most of them don't even go home with dressing - like without the wound covered" (Surgeon, Nigeria, Interview NG002I).*

This was supported by quantitative data where there was significant uniform (F-statistic 15.26, p<0.0000001) and non-uniform (F-statistic 2.27, p=0.00001) DIF by country (*Figure 4.11*). However, the item demonstrated good model fit with no local dependency. There was also concern about comprehension for patients in Ghana, who use the term 'washing the wound' to describe wound care and dressing.

Recommendation 13: No cross-cultural adaptation required. Consider splitting for DIF in future analyses. Modify wording during translation into Ghanian dialects to ensure conceptual equivalence.

4.4.4.15 Item 14. Readmission

Investigators generally felt comprehension and retrieval for Item 14 would be satisfactory as readmission was a substantial 'event' during their postoperative journey. However, judgement related to the cause of readmission was considered to be challenging for some:

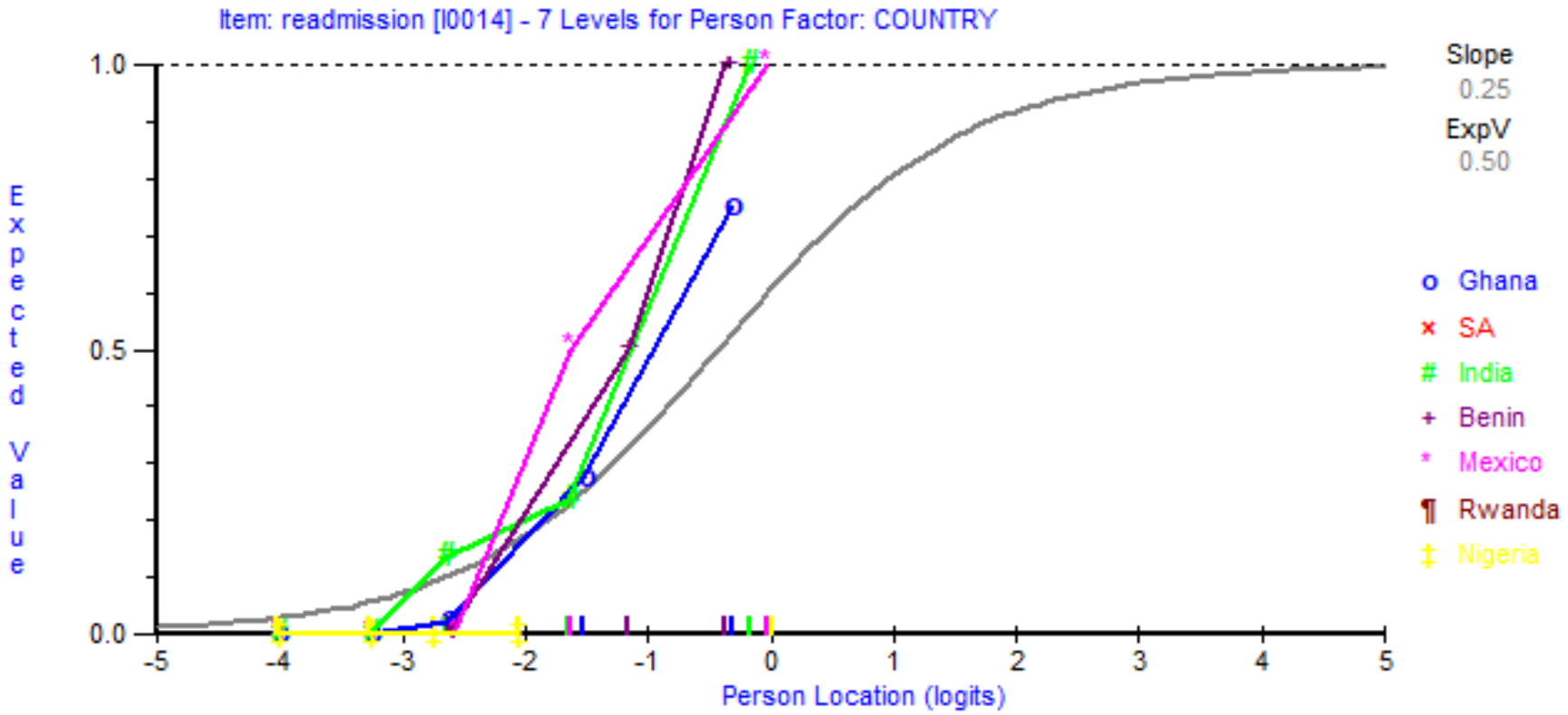
> *"Telling the difference between seeking advice related to a wound problem and going back to hospital for another problem might be difficult" (Research nurse, Ghana, Interview GH001I)*.

In keeping with this, item 14 fitted poorly with the Rasch model (fit residual -4.5, $p<0.0000001$) with evidence of overdiscrimation (*Figure 4.12*). It also demonstrated a high degree of correlation with Item 7 (deep wound opening) and 15 (antibiotics). Some contextual themes related to access to care demonstrated variability in patients' behaviours across settings:

> *"patients… always return to the hospital with any postoperative problems, as traditional healers wont tamper with surgical wounds" (Focus group, Ghana, GH001F)*.

Recommendation 14: No cross-cultural adaptation required. Consider subtesting to account for local dependency with Item 7 and 14 in future analyses.

Figure 4.12. Differential item functioning by country for pathway items 14 (Readmission)



4.4.4.16 Item 15. Antibiotics

Several issues were raised with the cross-cultural and cross-contextual application of Item 15. Researchers described that:

> *"I think most of the patients the medicines they take home they can't tell which one is antibiotic" (Focus group, Ghana, GH001F)*

in particular that there might be:

> *"…confusion between pain (medicines) and antibiotics" (Surgeon, Rwanda, Interview RW002I).*

There was particular concern in judgement for patients with low health literacy or socioeconomic status:

> *"…accuracy may be variable depending on their level of education and the area they live" (Research nurse, India, Interview IN001I).*
>
> *"Patients are unable to understand the word antibiotic care, especially daily wages workers." (Research nurse, Focus group IN002F, India).*

Supporting this, item 15 misfit the Rasch model (fit residual 3.89, p=0.000004) with evidence of overdiscrimination (*Figure 4.13)*. In contrast, researchers in Mexico reported that:

> *"I think there will be no problem with the patients understanding what antibiotics are, because here in Mexico apparently all patients love antibiotics, that's all what they want, and they are happy if the doctor gave, it's like… (laughter)" (Researcher, Mexico, Interview* MX002I)

thus:

> *"…patients will be very aware of antibiotics" (Researcher, Mexico, Interview MX001I).*

Retrieval was also considered to be challenging:

> *"Patients might struggle to remember which medications they were on early after surgery" (Focus group, Ghana, GH001F).*

Potential solutions to support judgement were proposed:

“It may require further explanation or the names of the antibiotics if literate” (Focus group NG002F, Nigeria).

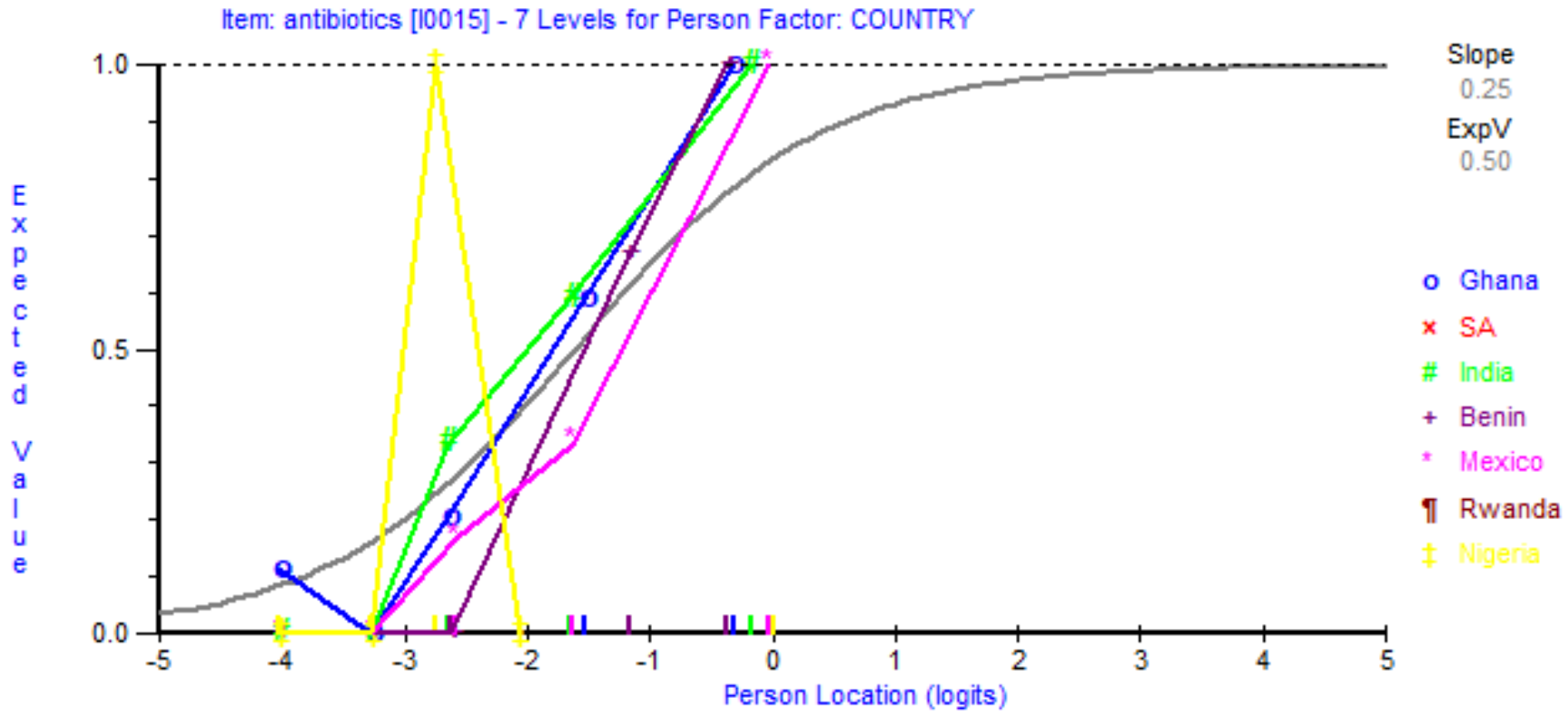
“We can ask them what medication was prescribed to them and we can just that yes this is antibiotic and we can put yes here.” (Surgeon, Focus group IN002F, India).

A surgeon from Rwanda also recommended

“The form is also useful. In Rwanda, antibiotics usually take the form of capsules” (Focus group, Rwanda, RW001F).

Recommendation 15: Change wording to ‘medicines (antibiotics)’, mirroring other universal reporter outcome measure style items (13 and 18). Assessors should ask patients to read out the name of their medications or describe the colour and form, if possible, to check that they are correctly identified as antibiotics.

Figure 4.13. Overdiscrimination in pathway item 15 (Antibiotics)



4.4.4.17 Item 16. Clinician wound opening

In general, comprehension and retrieval were both considered acceptable for Item 16. One participant from Ghana reported some ambiguity in patients differentiating suture removal and deliberate wound opening in her practice *(Research nurse, Ghana, Interview GH001I)*. Others highlighted the need to differentiate item 16 from item 19 (Reoperation), which captures a similar concept of clinician intervention for a wound complication:

> *"The problem is, it can happen both on wards and in theatre. We need to make sure patients have the difference with the general anaesthetic question" (Focus group, Nigeria, NG002F).*

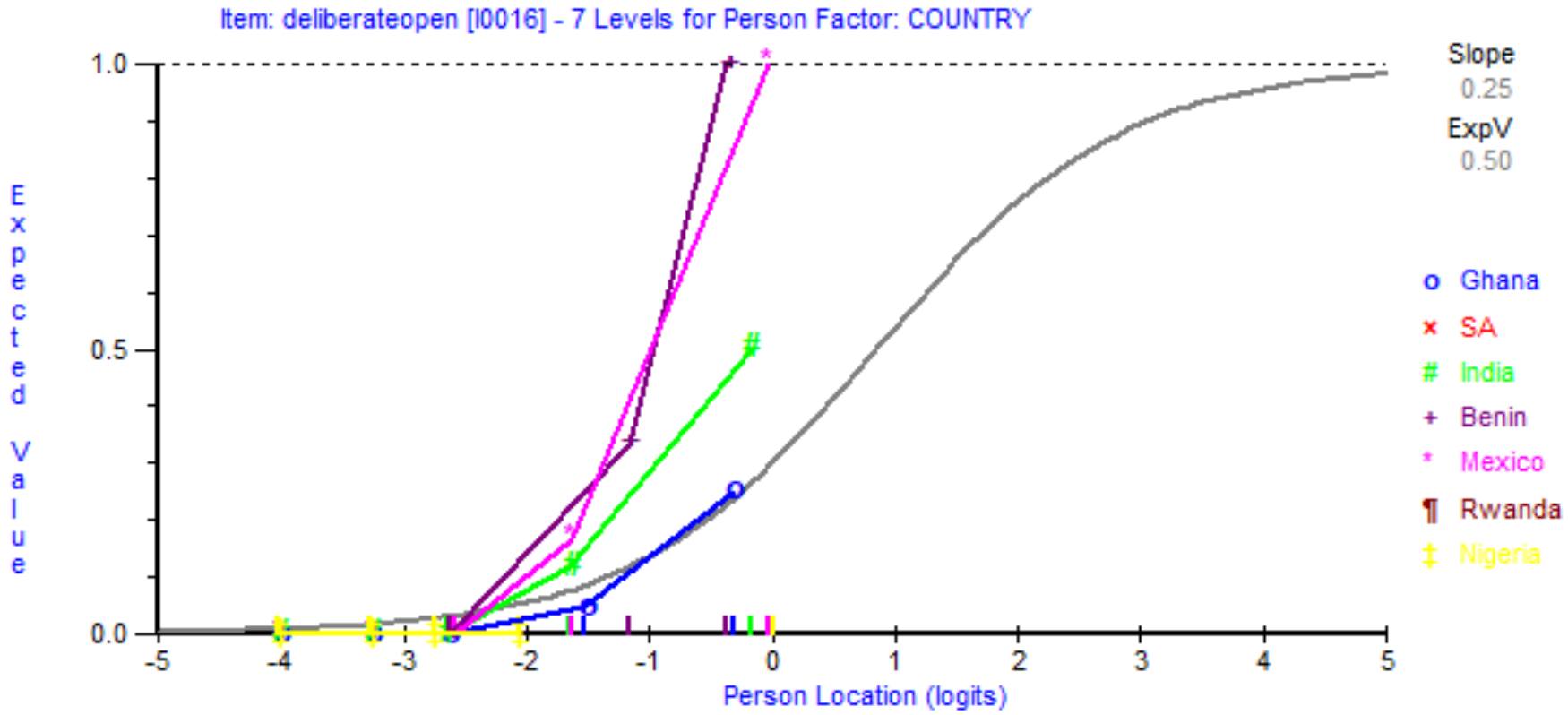
The word 'deliberate' was considered to have specific negative connotations in Nigeria, where both an interview and focus group discussion perceived it to imply medical harm *(Focus group, Nigeria, NG001F):*

> *"…this sounds like a malicious act" (Surgeon, Nigeria, NG003).*

On Rasch analysis, there was significant item misfit (fit residual -3.213, p=0.000002) with evidence of overdiscrimination and local dependency with items 17 (Wound scraping), 18 (drained) and 19 (reoperation). A subtest of Item 16, 17, 18 and 19 slightly improved model fit overall (χ^2 185.08, DF 64, p<0.0000001). Non-uniform DIF was detected by country (F-statistic 3.04, p=0.000128; *Figure 4.14)*.

Recommendation 16: Remove the term 'deliberate' from the item description to avoid negative connotations. Where necessary assessors should clarify that this item refers to clinician wound opening without general anaesthesia (i.e., outside of the operating room). Consider subtesting to account for local dependency with Item 17, 18 and 19 or splitting this item for DIF in future analyses.

Figure 4.14. Differential item functioning by country for pathway item 16 (Clinician opening)



4.4.4.20 Item 17. Wound scraping

No issues with comprehension or retrieval were raised. Judgment was considered challenging by an investigator from Nigeria who highlighted:

> *"Patients will only be able to answer if the doctor explains it to them at the time of debridement" (Surgeon, Nigeria, Interview NG003I).*

There were also issues raised again with the term 'tissue', particularly for patients using Ghanian dialects:

> *"The word tissue is not used in Ghana. I think unwanted flesh would be okay, they will understand it much better rather than tissue" (Research nurse, Ghana, Interview GH001I).*

Item 17 demonstrated acceptable fit with the Rasch model with no significant DIF.

Recommendation 17. Word 'tissue' changed to 'flesh' to improve translatability and comprehension for patients with lower health literacy.

4.4.4.19 Item 18. Wound drained

Comprehension was highlighted as a major issue for Item 18, particularly in understanding of the concept of an abscess:

> *"Abscess is not a word that is commonly used or understood by Ghanian patients" (Ghana, Focus group GH001F).*

However, pus was generally considered to be a concept that was well understood:

> *"Patients in general will know the word pus but probably not the term 'abscess'" (Research Nurse, India, Interview IN002I).*

and acceptable to patients:

> *"maybe for the word abscess they don't understand what it is, but knowing that it's pus it might be easier" (Researcher, Mexico, Interview MX002I)*

with good judgment:

> *"…pus is well understood, and patients know abnormal fluid" (Surgeon, Rwanda, Interview RW001I)*

and that translated well across settings:

> *"Pus translates well into Kinyarwanda" (Rwanda, Focus group RW001F).*

A suggestion to improve the item was made to increase standardisation with Item 5 (Thick, green fluid):

> *"I suggest that 'yellow or green' pus would improve the description (Nigeria, Focus group NG002F)."*

Concerns were also raised about crossover with Item 5, which captures a similar concept (drainage of pus) but in a passive (i.e., spontaneous, without intervention) rather than an active (i.e., performed by a clinician) way:

> *"It might be hard to notice between wound being actively drained and passively draining 'water'" (Surgeon, Ghana, Interview GH002I).*

Specific translatability issues were highlighted to improve across language adaptation in Hindi:

> *"When translating into Hindi, this translates as 'bad blood'" (India, Focus group IN001F).*

And Ghanian dialects:

> *"Patients may stay 'water coming from wound', but would be able to say whether this is bloody, or yellow and green colour water" (Research nurse, Ghana, Interview GH001I)*

Despite these perceived challenges, item fit was acceptable and with no significant DIF.

Recommendation 18: Change wording to yellow or green fluid (pus), mirroring Item 5. Addition of "drained from your wound by a doctor or nurse" to emphasis active event of item 18. Assessor should not state the term 'abscess' to improve comprehension but can apply this concept during measurement.

4.4.4.20 Item 19. Reoperation

Comprehension was identified as a major cross-cultural issue with the term 'general anaesthetic' for item 19:

> *"Patients are unlikely to understand the term general anaesthesia" (Mexico, Focus group MX001F).*
>
> *"A patient told me anaesthesia is a medical term, he does not understand it" (Anaesthetist, Focus group GH001F, Ghana).*

An improvement to the item wording was proposed:

> *"Have you been put to sleep for an operation on the wound or for treatment on the wound – that would encompass it. Knowing that process where you are asleep, that's what I can say, not anaesthetic" (Surgeon, Nigeria, Interview NG002I).*

However, an important cross-contextual clinical point was also made by several investigators about the higher proportion of patients undergoing spinal rather than general anaesthesia in LMICs, due to training, safety and capacity issues. This would not be captured by the current item.

Although retrieval and judgment were generally perceived to be good:

> *"Patients likely to be able to answer, as it's a serious event to return to surgery" (Research nurse, Ghana, Interviewer GH001I).*

Ghanian and Nigerian investigators raised the challenge for patients of judging between reoperation for a wound complication and another problem, as both would include wound revision and/or relaparotomy:

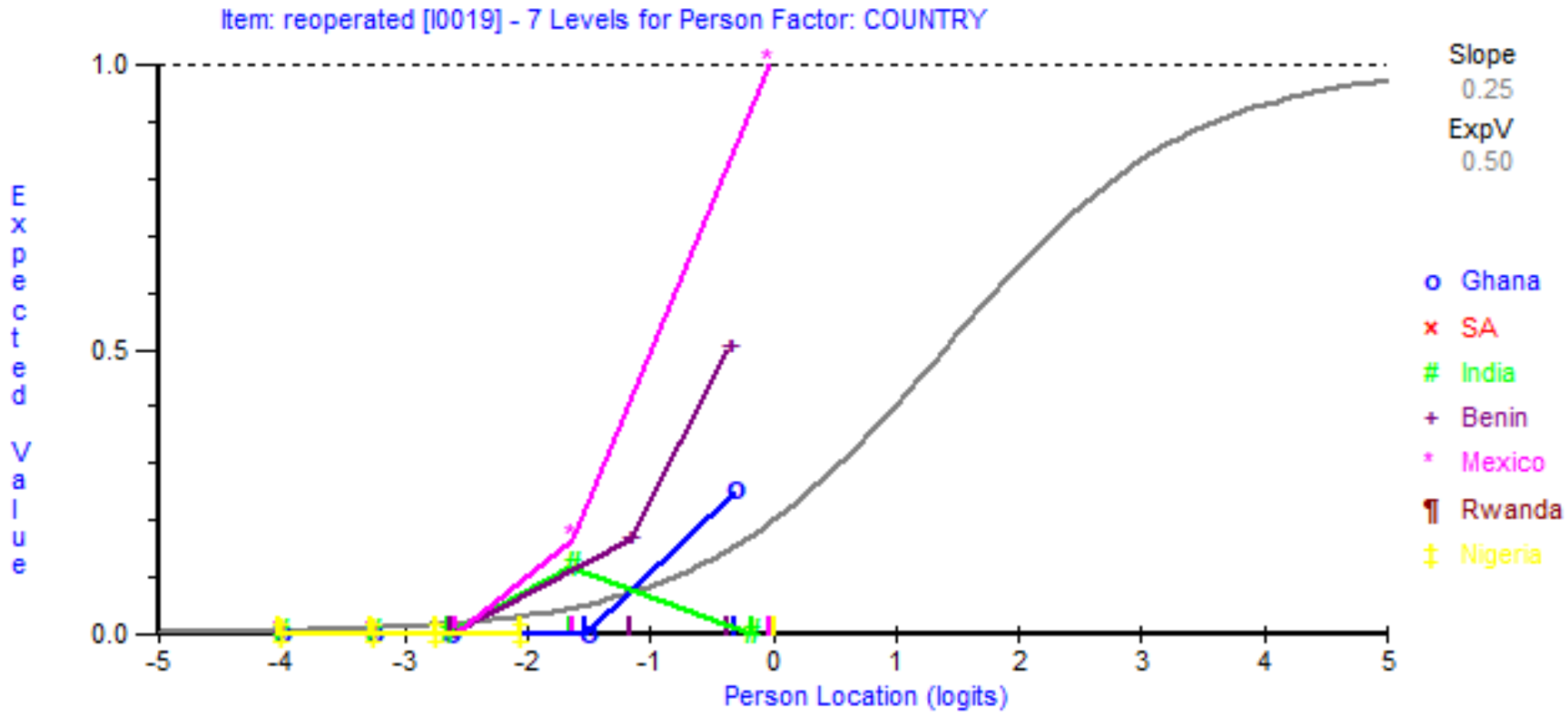
> *"Patients will definitely remember this as it's such a big event. But it will be difficult to tell 'for a wound problem' versus another reason" (Interviewer GH002I)*

A Rwandan surgeon proposed assessors considering clarification of a 'second operation' to support patient's comprehension (Surgeon, Rwanda, Interview RW002I).

Corroborating some potential issues with comprehension or judgement there was significant item misfit (fit residual -2.497, p=0.000076) and non-uniform DIF (F-statistic 4.25, p<0.0000001, *Figure 4.15*). Item 19 had the highest degree of local dependency of all WHQ items, with high correlation with item 7 (Deep wound opening), 14 (Readmission), 16 (Clinician opening), 17 (Wound scraping). A subtest with all these items together improved overall model fit (χ^2 167.25, DF 64, p<0.0000001).

Recommendation 19: Consensus that general anaesthesia would not be understood across contexts. As spinal anaesthesia is common in some LMICs, this concept would not currently be captured. Wording adapted to collect information about any procedure carried out in the operating room. Assessors may wish to use the term ‘second operation’ to aid comprehension. Consider subtesting to account for local dependency with item 7, 14, 16 and 17 in future analyses.

Figure 4.15. Differential item functioning by country for pathway item 19 (Reoperated)



4.4.5. Summary of recommendations

A summary of cross-cultural and cross-contextual adaptations to the English language WHQ is provided in *Table 4*, and the final adapted questionnaire in *Appendix E*. Translated versions of the final adapted WHQ are provided in the study *Appendix F*.

Table 4.9. Summary of recommendations for adaptation of WHQ (English language)

Item number	Original item	Original response categories	Adapted item	Adapted response categories
1	Was there redness spreading away from the wound?	1 = Not at all; 2 = A little; 3 = Quite a bit; 4 = A lot.	Was there redness (or shining of the skin) spreading away from the wound?	1 = Not at all; 2 = A little; 3 = A lot.
2	Was the area around the wound warmer than the surrounding skin?	1 = Not at all; 2 = A little; 3 = Quite a bit; 4 = A lot.	-	1 = Not at all; 2 = A little; 3 = A lot.
3	Has any part of the wound leaked clear fluid?	1 = Not at all; 2 = A little; 3 = Quite a bit; 4 = A lot.	Has any part of the wound leaked thin clear fluid?	1 = Not at all; 2 = A little; 3 = A lot.
4	Has any part of the wound leaked blood-stained fluid?	1 = Not at all; 2 = A little; 3 = Quite a bit; 4 = A lot.	-	1 = Not at all; 2 = A little; 3 = A lot.
5	Has any part of the wound leaked thick and yellow or green fluid?	1 = Not at all; 2 = A little; 3 = Quite a bit; 4 = A lot.	-	1 = Not at all; 2 = A little; 3 = A lot.
6	Have the edges of any part of the wound separated or gaped open of their accord?	1 = Not at all; 2 = A little; 3 = Quite a bit; 4 = A lot.	-	1 = Not at all; 2 = A little; 3 = A lot.
7	If the wound edges opened, did the deeper tissue also separate?	1 = Not at all; 2 = A little; 3 = Quite a bit; 4 = A lot.	If the wound edges opened, did the flesh beneath the skin or the inside sutures also separate?	1 = Not at all; 2 = A little; 3 = A lot.
8	Has the area around the wound become swollen?	1 = Not at all; 2 = A little; 3 = Quite a bit; 4 = A lot.	-	1 = Not at all; 2 = A little; 3 = A lot.
9	Has the wound been smelly?	1 = Not at all; 2 = A little; 3 = Quite a bit; 4 = A lot.	-	1 = Not at all; 2 = A little; 3 = A lot.
10	Has the wound been painful to touch?	1 = Not at all; 2 = A little; 3 = Quite a bit; 4 = A lot.	Has the wound been painful to touch?	1 = Not at all; 2 = A little; 3 = A lot.
11	Have you had, or felt like you have had, a raised temperature or fever (>38oC)?	1 = Not at all; 2 = A little; 3 = Quite a bit; 4 = A lot.	Have you had, or felt like you have had, a raised temperature or fever?	1 = No; 2 = Yes
12	Have you sought advice because of a problem with your wound, other than at a planned follow-up appointment?	1 = No; 2 = Yes	-	-
13	Has anything been put on the skin to cover the wound? (dressing)	1 = No; 2 = Yes	-	-
14	Have you been back into hospital for a problem with your wound?	1 = No; 2 = Yes	-	-
15	Have you been given antibiotics for a problem with your wound?	1 = No; 2 = Yes	Have you been given medicines (antibiotics) for a problem with your wound?	-
16	Have the edges of your wound been deliberately separated by a doctor or nurse?	1 = No; 2 = Yes	Have the edges of your wound been separated by a doctor or nurse?	-
17	Has your wound been scraped or cut to remove any unwanted tissue?	1 = No; 2 = Yes	-	-
18	Has your wound been drained? (drainage of pus or an abscess)	1 = No; 2 = Yes	Has thick, yellow, or green fluid (pus) been drained from your wound by a doctor or nurse (abscess)?	-
19	Have you had an operation under general anaesthetic for treatment of a problem with your wound?	1 = No; 2 = Yes	Have you had to go back to the operating room for treatment of a problem with your wound?	-

4.4.6 Measurement procedures

Despite concerns with mobile phone connectivity in qualitative data, telephone WHQ completion was feasible (84.5%, 537/635) with high data completeness (99.0% instruments complete overall, range by item: 99.1% to 100.0%). 99.2% (533/537) of patients reported the telephone WHQ pathway to be ‘very satisfactory’ or ‘satisfactory’:

> *“Early feedback that the questionnaire is highly acceptable to patients. Patients say they are receiving a 'VIP' treatment.” (Junior doctor, Focus group GH001F, Ghana)*
>
> *“People were very impressed that I was calling them and still following up on the surgeries and were willing to talk very happily.” (Research nurse, Focus group IN002F, India)*

Often the phone owner was a friend or relative rather than the patient themselves (35.2%, 189/537), and commonly this was a mobile phone (99.5%, 534/537). 28.7% had a mobile phone with video capability (154/537). Feedback from CEI partners alongside interview data supported optimisation of the telephone follow-up pathway for future implementation. Specific recommendations were combined into an infographic and toolkit presented in *Chapter 5 (Figure 5.8)*.

4.5 Discussion

4.5.1 Key findings

Pathways for remote assessment of common complications after surgery in low resource settings are essential in improving safety and resilience of surgical care systems. This mixed methods study made recommendations for cross-cultural and cross-language adaptation of the WHQ for use in LMICs. This improved its relevance across cultures and for patients with lower levels of health literacy. Conceptual equivalence, content and construct validity was confirmed across languages using qualitative and translation methods. Unidimensionality, measurement properties and use of the total WHQ score were seen to be valid within the Rasch framework, although the overall power was low. The telephone pathway was demonstrated to be feasible with high data completeness, and highly acceptable. Working with CEI partners we made recommendations for optimisation of telephone follow-up in research and postoperative surveillance programmes. This study provides a large, international, high-quality proof of concept for rapid adaptation and implementation of patient reported measures in emerging global health arenas such as surgery.

4.5.2 Mixed methodology in cross-cultural adaptation

The use of mixed methods here added strength and depth. The qualitative data were used primarily to inform cross-cultural adaptation ahead of translation [133, 134, 215]. Whilst this was based on cognitive theory, data were collected indirectly about patient experience from frontline clinicians involved in wound assessment [203, 205, 221]. The Rasch analysis supplemented this and allowed patient-level data to enrich and inform final recommendations for adaptation [134, 219, 220]. In most instances, the qualitative and quantitative data were supportive of one another, demonstrating coherence during triangulation. Where conflict arose, qualitative findings were softened and/or caveated (i.e., changes were recommended

where there was coherence on triangulation, and further exploration recommended where there was conflict between the qualitative and quantitative data).

4.5.3 Critical analysis of application of Rasch methods

Rasch analysis is an established method for instrument development and cross-cultural refinement [134, 218, 220]. Here its principal utility was to confirm the validity of the use of the total WHQ score as an ordinal scale and in enhancing understanding of the response structure and local dependency. However, properties of the WHQ make it a rather unusual application of the Rasch model. Firstly, it is principally a diagnostic tool for SSI rather than an interval-level tool measuring a spectrum of severity of a latent trial. This was best seen in 'mistargeting' of the WHQ to the study population with many patients at the 'floor' that added low information value to the model; however, this would largely be expected in a screening tool where many patients are asymptomatic for the condition in question. This reduced the overall power of fit as many participants contributed little information about item locations. Secondly, as expected in a diagnostic test many items had high levels of local dependency which may have contributed to the overall model misfit. Thirdly, several items misfit the Rasch model and person separation index was poor, with a conversely high Cronbach's alpha. Again, this is highly likely to be due to the extreme 'floor' of respondents in the setting of a diagnostic tool. It was not the overall aim to fit this diagnostic tool closely to the Rasch model and it would not be required to be valid for use if it demonstrates a satisfactory psychometric structure, unidimensionality and sufficient sensitivity and specificity upon clinical application. This highlights the importance of further work to externally validate the tool in a diagnostic test accuracy study.

4.5.4 Areas for exploration in future development of the WHQ

Exploring complex relationships between items and optimising the measurement properties using subtesting and adjusting for DIF was not the aim here, but warrants further investigation. It is feasible that the instrument could be simplified, or its diagnostic accuracy could be improved using Rasch by better accounting for differences in the symptomology and health seeking behaviours for patients with SSI across countries. DIF by country observed for several items here supports methods to ensure balance in randomised trials such as stratification or minimisation of randomisation by country.

4.5.5 Results in context

The use of patient (PROMs) and universal reporter (UROMs) outcome measures in low-income settings is complex; many instruments have not yet undergone cross-cultural and cross-language adaptation and there is uncertainty about the feasibility of remote, digital methods. Whilst examples exist from established global health fields such cardiovascular disease, few studies in global surgery have adopted PROMs to date [227-229]. Health technology assessments thus neglect important insights into quality of recovery and health utility that could affect policy decisions [42]. This study provides a proof of concept for rapid, pragmatic adaptation of instruments in the surgical setting that can be used across other measures and emerging contexts. Developing culturally-attuned, remote follow-up pathways is particularly important during pandemic recovery in building resilience in resource poor health systems [155, 230]. The co-produced pathway for telephone follow-up in LMICs described is ready for wider adoption. Recommendations from this mixed-methods study can now to be used for further exploration of the diagnostic accuracy of the adapted WHQ in low resource contexts.

4.5.5 Limitations

There are several limitations of this study, explored by domain below.

4.5.5.1 Design limitations

First related to design. Due to safety and ethical concerns during the SARS-CoV-2 pandemic, we were unable to perform cognitive interviewing directly with patients. Instead, we explored aggregate perspectives of frontline clinicians involved in the care of surgical patients. This meant data represented clinicians' impressions of patients' responses and challenges in retrieval and judgement rather than direct exploration with patients in typical cognitive interviewing [205]. Sampling of researchers directly involved in the same portfolio of trials was a pragmatic decision but may have reduced the transferability of themes across other hospital types (e.g., remote rural hospitals), resource settings (e.g., hospitals with less research infrastructure) or differing populations (e.g., less literate populations, with poorer healthcare access). We aimed for thematic saturation overall when ending recruitment to the qualitative phase, but this is unlikely to have been reached at an individual country level [212]. It is therefore possible that important insights were missed during adaptation, although recommendations were strengthened by triangulation with quantitative data to reduce over-reliance on qualitative data alone [221].

4.5.5.2 Analysis limitations

Second related to analysis, as the WHQ did not meet all the Rasch assumptions for model fit we did not develop a logit-adjusted scale. Further development could improve the measurement properties of the questionnaire to allow direct patient-to-patient comparisons in future research. We have not accounted for complex patterns of DIF in measurement that could lead to differences in point score equivalence across different patients with differing characteristics when applied clinically. We also handled Item 6 (Wound opening) and Item 7

(Deep wound opening) as independent items for the purposes of Rasch analysis, despite structural dependency (i.e., patients ineligible to respond to Item 7 were scored as 0). This could have been managed used a partial credit item or as a subtest to avoid falsely inflating the total score.

4.5.5.3 Interpretation limitations

Finally related to interpretation, the most important metric of clinical utility in a screening tool such as this would be criterion validity. A formal diagnostic test accuracy study comparing the WHQ to a gold standard reference test for SSI is now required [52, 186]. A choice of cut-point score for the adapted WHQ is likely to favour sensitivity to triage all patients with a likelihood of SSI to seek medical care.

4.6 Conclusion

This chapter adopted a mixed-methods approach to explore the cross-cultural and cross-language equivalence of the WHQ. Recommendations have been made to adapt the wording of 9 items and the response structure for 11 items for use in global surgery research and practice.

4.6.1 Link to next chapter

The two most important features of a pathway for remote detection of SSI, are that it is feasible to implement across a variety of settings and patient groups, and that it can accurately discriminate between patients with and without SSI. In Chapter 5, I will evaluate the feasibility and diagnostic accuracy of the adapted WHQ in a study within a randomised trial across seven LMICs.

5. Feasibility and diagnostic accuracy of a telephone Wound Healing Questionnaire in detection of surgical site infection following abdominal surgery (TALON-2): A study within a trial in seven low and middle-income countries

5.1 Abstract

5.1.1 Scientific abstract

Introduction

Telemedicine is being rapidly adopted in postoperative surveillance with little formal evaluation. This international study within a trial aimed to evaluate the feasibility and diagnostic accuracy of telephone administration of an adapted Wound Healing Questionnaire (WHQ) in the detection of surgical site infection after abdominal surgery in low- and middle-income countries.

Methods

A multi-centre, international, non-randomised prospective validation study was embedded in a factorial, randomised controlled trial testing measures to reduce surgical site infection in low- and middle-income countries (FALCON, NCT03700749). The study was conducted according to a published protocol and statistical analysis plan (SWAT126) and reported according to STARD guidelines. The reference test was in-person review by a trained clinician at 30-days after surgery according to the US Centres for Disease Control criteria. The index test was telephone administration of the adapted WHQ by an independent researcher at 27 to 30-days after surgery, with item responses summed to create an overall point score between 0 and 29. The primary outcome measure was diagnostic accuracy of the WHQ, defined as the proportion of surgical site infections correctly identified by the telephone WHQ, and summarised using the area under the receiving operator characteristic

curve (AUROC) and diagnostic test accuracy statistics at an optimal cut-off derived using Youden's index.

Results

Patients were included from three upper-middle income (396 patients, 13 hospitals), three lower-middle income (746 patients, 19 hospitals), and one low-income country (54 patients, 4 hospitals). Successful telephone contact was achieved in 90.3% (1088/1196) of patients. Those with non-midline incisions (adjusted odds ratio: 0.36, 95% CI 0.17-0.73, p=0.005) or with a confirmed reference test diagnosis of SSI (OR: 0.42, 95% CI 0.20-0.92, p=0.006) were less likely to be contactable. The WHQ discriminated patients with and without SSI (AUROC 0.869, 95% CI 0.824-0.914). An adapted WHQ cut-off point score of $\geq$4 demonstrated sensitivity of 0.701 (0.610-0.792), specificity of 0.911 (0.878-0.9430), positive predictive value of 0.723 (0.633-0.814) and negative predictive value of 0.901 (0.867-0.935). Some differences were seen in discrimination in rural (AUROC 0.818, 0.721-0.914) versus urban populations (AUROC 0.886, 0.836-0.937) and after emergency (AUROC 0.871, 0.826-0.916) versus elective surgery (AUROC 0.966, 0.895-1.000).

Conclusion

This study demonstrated feasibility and validity of telephone assessment for post-discharge SSI diagnosis in low-resource environments. Postoperative telemedicine pathways must focus on strategies for retention to avoid attrition bias in patients uncontactable for follow-up.

5.1.2 Co-produced lay abstract

The TALON study: Keeping track of your healing from surgery

What was this research about? A wound infection happens when germs enter the cut made in your body by the doctor when you were operated on. Germs are small organisms that cannot be seen by your eyes but can cause problems healing of the cut. Infection is the most common problem after surgery and can delay you getting out of hospital, and back to normal life.

The current way to check whether you have an infection is for a doctor or nurse to look at the cut made on your tummy, and see how it is healing. For example, the doctor may check if the cut has a green, liquid oozing from it or if the area of the wound is red or swollen. A month after you leave hospital, the doctor may ask you to come back for a follow-up visit. However, this will require you to travel to the hospital, take a day off work or away from your families, and can be expensive and time-consuming if you travel far.

What did we want to do? We wanted to find out if talking to a doctor over the phone would work as well as you travelling to the hospital to show the wound to a doctor or nurse in person.

To do this, we asked over 1000 patients that had recently undergone surgery to be checked using both methods – to take a phone call from one doctor and be checked in-person by a different doctor. We were able to compare the phone follow-up and in-person check to see if the doctor came to a different conclusion. We also looked whether patients were able to receive a phone call at home, and their experience of the process.

What did we find out? For most patients, the phone call from the doctor was just as good at seeing if a patient had an infection as a face-to-face check-up by the doctor. However, the phone call was not perfect all the time, particularly for patients with very mild infections. Most patients were able to receive the call after a few tries, and all patients were very happy with the process. As an international research team, we are now trying new ways to improve the phone call, including looking at the wound over video if possible.

What does this mean for you as a patient? A phone call to check how your wound is healing can now be used as a substitute for face-to-face check-up by a doctor. If you have any worries about your wound after the phone call you should still seek help from a doctor or nurse. We hope that the telephone call will be more convenient for patients like you to avoid travelling back to the hospital and taking time away from your work and family. We wish you all the best for your ongoing recovery from surgery.

5.2. Introduction

5.2.1. Importance of surgical site infection research

Surgical site infection (SSI) is a global issue. It is the most common healthcare-associated infection in low- and middle-income countries (LMICs) [64, 93], and carries a huge burden to patients, doctors, and health systems around the world [63, 231, 232]. Reported rates vary, but SSI is particularly prevalent in abdominal and colorectal surgery; as many as one in three patients get an SSI when the operation involves the large bowel [23]. It was highlighted as the key research priority to improve surgical care worldwide in an international prioritisation process [40] and is the focus of several ongoing global randomised trials [45, 80, 194, 233].

5.2.2 Importance of post-discharge surveillance

Whilst some SSI occurs while patients are in-hospital, the majority occurs after discharge [234]. In this thesis introduction, I demonstrated that in-hospital only measurement was not a sufficient surrogate for 30-day SSI. Post-discharge surveillance of SSI is therefore an important quality marker in wound infection research [56]. The accepted reference standard of assessment for SSI during the 30-days after surgery is an in-person review according to US Centre for Disease Control (CDC) criteria [93]. However, in-person assessment is labour and time intensive, and requires patients to take additional time-off work and incur costs of travel. This is particularly challenging in resource-limited environments where there are shortages in the surgical workforce, and patients are already at risk of catastrophic expenditure as a direct and indirect result of their surgical care [9]. In Chapter 4, I made recommendations for adaptations to the Bluebelle Wound Healing Questionnaire for use in global surgery research and practice. This provides an attractive candidate for remote SSI detection, but evaluation of the feasibility and accuracy of a telephone WHQ pathway is required.

5.2.3 Relevance to SARS-CoV-2 pandemic

Remote follow-up methods have been rapidly adopted during the SARS-CoV-2 pandemic to reduce risk of in-hospital transmission, and conserve resources for surges in COVID-19 admissions and to address elective surgical backlogs [124, 235]. Whilst telephone follow-up may offer greater efficiency and cost-savings, missed SSI events may lead directly to patient harm through care delays or indirectly through inefficiencies in SSI prevention research [195].

5.2.4 Justification of SWAT study design

Studies within a trial (SWATs) have gained significant attention from trial methodologists and funders over the past 3-years, and are now the focus of a Trial Methodology Research Partnership working group (Trial Forge) [202] and National Institute for Health Research (NIHR) funding stream. SWATs exploit the delivery network and infrastructure of major randomised trials to efficiently answer methodological research questions. Ongoing large international trials in global surgery provide a unique opportunity to improve the quality and efficiency of global wound infection research [44]. However, as SWATs only examine patients recruited to trials, one must consider their generalisability to the broader population and implications of low screening : recruitment ratios in the host trial.

5.2.5 Objectives

The overall aim of this study was to evaluate the feasibility and diagnostic accuracy of telephone administration of a wound healing questionnaire for remote detection of SSI after abdominal surgery in low- and middle-income countries (LMICs). The results of this study will inform efficient design and conduct of future randomised trials and postoperative surveillance programmes.

5.3 Methods

5.3.1 Overview

This was a prospective, multi-centre, international, non-randomised cohort Study Within a Trial (SWAT) exploring the feasibility and accuracy of remote follow-up pathways for surgical site infection assessment (TALON-2). It was embedded within a pragmatic multicentre factorial randomised controlled trial testing measures to reduce surgical site infection in low- and middle-income countries (FALCON). FALCON was a stratified, pragmatic, multi-centre, 2x2 factorial trial testing two measures (skin preparation and antimicrobial sutures) to reduce superficial or deep skin infection following abdominal surgery of 5788 patients in 54 hospitals in 7 low- and middle-income countries (NCT03700749) [194]. In this trial superiority of the intervention groups over the control group, either alone or in combination, was not demonstrated overall, either alone or in combination, or in any pre-planned subgroup [236].

5.3.2 Protocol and registration

The study protocol was pre-registered online on the MRC Hubs for Trial Methodology Research SWAT store database [202] (Queen's University Belfast) (SWAT ID:126) and published in *Trials [186]*. This report was prepared with reference to SAMPL (Statistical Analyses and Methods in the Published Literature) guidelines [237], Patient-Centred Outcomes Research Institute's (PCORI) methodology standards [197], STARD guidelines for diagnostic test accuracy studies [238], and COSMIN guidelines for patient reported outcomes research [239].

5.3.3 Ethical approvals and consent

A protocol amendment to embed TALON-2 in the host trial (FALCON) was obtained from the University of Birmingham International Ethics Committee. All individual participating countries

obtained local, or national ethical approval in accordance with local protocols. Written (or fingerprint) informed consent to participate was obtained from all participants.

5.3.4 Inclusion and exclusion criteria

Consecutive adult patients (greater than 16 years old) recruited to the FALCON trial between 10 Dec 2018 and 6 Sep 2020 were eligible for recruitment to TALON-2. Any centre participating in FALCON was eligible to participate. Centres were given flexibility to include patients over different date ranges depending on their local capacity and infrastructure, so long as sampling was consecutive. This included a broad range of abdominal operations with a predicted clean-contaminated, or contaminated or dirty operating field and a planned skin incision of greater than 5cm, for benign, malignant, trauma, and obstetric indications. This aimed to be representative of patients undergoing emergency or elective surgery in LMICs. Patients that were unlikely to be contactable for 30-day follow-up were excluded from the FALCON trial. Patients with a missing FALCON 30-day follow-up assessment (either in-person or by telephone), or that died before 30-days after surgery were excluded from analysis in this study.

5.3.5 Reference diagnostic test

The reference diagnostic test for surgical site infection (SSI) during the 30-days after surgery was in-person review according to US Centre for Disease Control Criteria (CDC) [93]. This is widely accepted as a quality standard in SSI research, and has been used by most, major international RCTs [56]. A full description of the definition used in the FALCON trial is presented in *Box 1.1*.

5.3.6 Index diagnostic test

The index diagnostic test under evaluation was a telephone-administered Bluebelle Wound Healing Questionnaire (WHQ) [186], adapted for use in LMICs. The WHQ was originally

developed and validated in the UK (English language) to assess post-discharge infections following abdominal surgery [184, 192]. The WHQ was designed to be completed either by healthcare professional, or self-reported by patients [185], and as such has been described as a 'universal-reporter' outcome measure (UROM) [191]. In a UK validation study, the WHQ demonstrated good reliability and high sensitivity and specificity when discriminating between SSI and no SSI in comparison to an in-person CDC assessment [184, 192].

5.3.6.1 Adaptation of WHQ for use in global surgery

The original WHQ was adapted for use in global surgery trials for use across language and resource settings using recognised practices for translating outcome measures, reported in Chapter 4. A summary of adaptations and the full adapted WHQ instrument is provided in *Table 5.1*.

Briefly, this involved two phases. First, an adaptation phase with structured interviewing and translatability assessment with local researchers, triangulated with analysis of the scaling and measurement properties of the WHQ in cohort study data, and informed by Rasch unidimensional measurement modelling [133, 215]. Second, a nine-phase translation phase for each language of delivery following *Mapi* recommendations [133]. In the adapted version of the WHQ, the response options and subsequent scoring were also modified.

Here, 'WHQ' cites this adapted questionnaire. In the adapted WHQ scale, items assessing SSI signs and symptoms were scored between 0 and 2 (Not at all, A little, A lot), and items assessing wound care interventions were scored between 0 and 1 (No, Yes). These were added together create an overall score between 0 and 29.

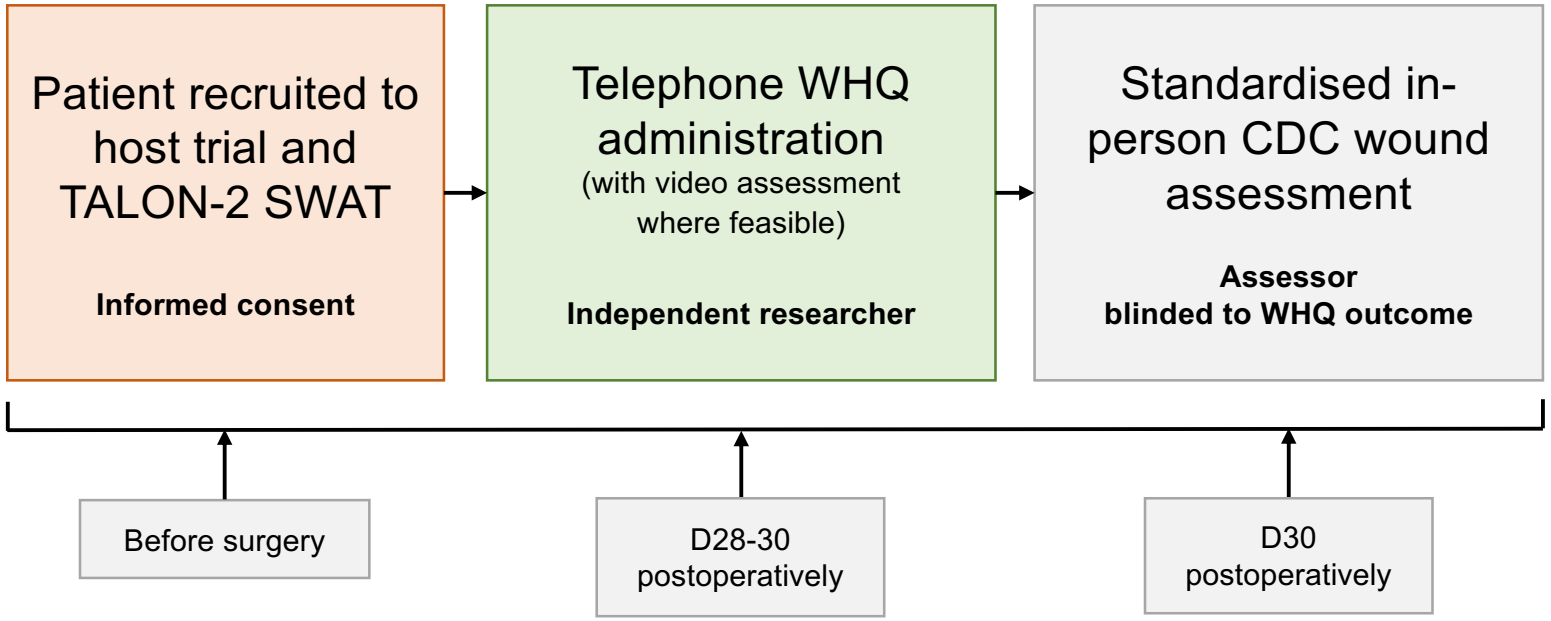
Table 5.1 Summary of adaptation of Wound Healing Questionnaire

Item number	Original item	Original response categories	Adapted item	Adapted response categories
1	Was there redness spreading away from the wound?	1 = Not at all; 2 = A little; 3 = Quite a bit; 4 = A lot.	Was there redness (or shining of the skin) spreading away from the wound?	1 = Not at all; 2 = A little; 3 = A lot.
2	Was the area around the wound warmer than the surrounding skin?	1 = Not at all; 2 = A little; 3 = Quite a bit; 4 = A lot.	-	1 = Not at all; 2 = A little; 3 = A lot.
3	Has any part of the wound leaked clear fluid?	1 = Not at all; 2 = A little; 3 = Quite a bit; 4 = A lot.	Has any part of the wound leaked thin clear fluid?	1 = Not at all; 2 = A little; 3 = A lot.
4	Has any part of the wound leaked blood-stained fluid?	1 = Not at all; 2 = A little; 3 = Quite a bit; 4 = A lot.	-	1 = Not at all; 2 = A little; 3 = A lot.
5	Has any part of the wound leaked thick and yellow or green fluid?	1 = Not at all; 2 = A little; 3 = Quite a bit; 4 = A lot.	-	1 = Not at all; 2 = A little; 3 = A lot.
6	Have the edges of any part of the wound separated or gaped open of their accord?	1 = Not at all; 2 = A little; 3 = Quite a bit; 4 = A lot.	-	1 = Not at all; 2 = A little; 3 = A lot.
7	If the wound edges opened, did the deeper tissue also separate?	1 = Not at all; 2 = A little; 3 = Quite a bit; 4 = A lot.	If the wound edges opened, did the flesh beneath the skin or the inside sutures also separate?	1 = Not at all; 2 = A little; 3 = A lot.
8	Has the area around the wound become swollen?	1 = Not at all; 2 = A little; 3 = Quite a bit; 4 = A lot.	-	1 = Not at all; 2 = A little; 3 = A lot.
9	Has the wound been smelly?	1 = Not at all; 2 = A little; 3 = Quite a bit; 4 = A lot.	-	1 = Not at all; 2 = A little; 3 = A lot.
10	Has the wound been painful to touch?	1 = Not at all; 2 = A little; 3 = Quite a bit; 4 = A lot.	Has the wound been painful to touch?	1 = Not at all; 2 = A little; 3 = A lot.
11	Have you had, or felt like you have had, a raised temperature or fever (>38oC)?	1 = Not at all; 2 = A little; 3 = Quite a bit; 4 = A lot.	Have you had, or felt like you have had, a raised temperature or fever?	1 = No; 2 = Yes
12	Have you sought advice because of a problem with your wound, other than at a planned follow-up appointment?	1 = No; 2 = Yes	-	-
13	Has anything been put on the skin to cover the wound? (dressing)	1 = No; 2 = Yes	-	-
14	Have you been back into hospital for a problem with your wound?	1 = No; 2 = Yes	-	-
15	Have you been given antibiotics for a problem with your wound?	1 = No; 2 = Yes	Have you been given medicines (antibiotics) for a problem with your wound?	-
16	Have the edges of your wound been deliberately separated by a doctor or nurse?	1 = No; 2 = Yes	Have the edges of your wound been separated by a doctor or nurse?	-
17	Has your wound been scraped or cut to remove any unwanted tissue?	1 = No; 2 = Yes	-	-
18	Has your wound been drained? (drainage of pus or an abscess)	1 = No; 2 = Yes	Has thick, yellow, or green fluid (pus) been drained from your wound by a doctor or nurse?	-
19	Have you had an operation under general anaesthetic for treatment of a problem with your wound?	1 = No; 2 = Yes	Have you had to go back to the operating room for treatment of a problem with your wound?	-

5.3.6.2 Administration of the adapted WHQ

According to the TALON-2 protocol, the WHQ was to be administered over the telephone by a non-surgeon (consultant, attending or equivalent) researcher (i.e., a junior doctor, research nurse, or other non-clinical personnel) between 27 and 30-days after surgery (i.e., before the reference diagnostic test) as the index diagnostic test in this study (*Figure 5.1*). The clinician/researcher administering the questionnaire was independent of the 30-day wound assessment in the FALCON trial (i.e., each was blinded to the reference and index test result respectively) and underwent standardised training from the Study Management Group (SMG).

Figure 5.1 Timing of administration of the telephone WHQ



5.3.7 Monitoring and quality assurance of the WHQ administration

A monitoring call was performed after the local researcher had completed the WHQ for 5 to 10 patients for quality assurance . Where recordings were available during the pilot, these were reviewed by a member of the SMG fluent in the target language. A WhatsApp group was also created for all site investigators participating in each country to share early experience and best practice during the pilot phase. Pathways for questionnaire administration were co-designed between patient partners, site investigators and research managers.

5.3.8 Adaptation for WHQ delivery during the SARS-CoV-2 pandemic

This study ran over the emergence of the SARS-CoV-2 pandemic. This had several operational consequences. Firstly, sites were asked to extend their recruitment where this was possible, in case fewer patients were able to return for in-person FALCON trial follow-up due to concerns about nosocomial SARS-CoV-2 infection. Secondly, some centres chose to administer the WHQ to consecutive patients that were farther away from their date of surgery (i.e., with a longer interval between FALCON trial follow-up and WHQ administration), handled in sensitivity analyses for the primary outcome measure. Thirdly, sites were allowed to pause and restart recruitment to TALON-2 if needed, so long as when the site was ‘live’ consecutive patients had attempted WHQ follow-up.

5.3.9. FALCON trial follow-up

Due to personal (mobility, deterioration, psychological reasons) and environmental (cost, transport links, SARS-CoV-2 transmission risk) reasons, not all patients were able to return to hospital for the reference test assessment in the FALCON trial (in-person 30-day follow-up according to CDC criteria).

Eligible patients were therefore categorised according to their corresponding FALCON-trial follow-up as:

(1) in-person FALCON trial follow-up

(2) telephone FALCON trial follow-up only.

5.3.10 Primary outcome measure

The primary outcome measure was diagnostic accuracy of the telephone WHQ in identification of SSI up to 30-days after surgery. I summarised the performance of the test using discrimination (area under the receiving operator curve characteristic or AUROC) and diagnostic test accuracy statistics (sensitivity, specificity, accuracy, positive predictive value, and negative predictive value).

5.3.11 Secondary outcome measures

The secondary outcome measure was the feasibility of telephone WHQ follow-up and was characterised using: (1) Telephone contact: successful contact of a patient on the telephone by the research team; (2) Return rate: successful completion of the WHQ where telephone contact was made; (3) Patient satisfaction: patient's self-reported satisfaction with the telephone WHQ follow-up; (4) Data completion rate: complete item response data. The estimated 'retention benefit' of using a telephone pathway versus in-person follow-up was estimated as the difference between the proportion of patients for whom the telephone WHQ was successfully completed and/or a telephone FALCON trial follow-up was completed, and the proportion for whom an in-person FALCON trial follow-up was completed [240].

5.3.12 Sample size

A range of sample sizes and their impact on the precision of estimates of sensitivity and specificity, from 95% confidence intervals (CI) were investigated. Calculations assumed a 30-

day SSI prevalence of 21.0% using the binomial exact formula and were pre-specified. Sample sizes were adjusted to allow for 15.0% predicted loss to follow-up from FALCON trial follow-up, and 15.0% of patients predicted not to undergo in-person FALCON trial follow-up. In patients with successful telephone contact and in-person FALCON trial follow-up, 87 events and 325 non-events would estimate sensitivity of 0.92 with a 95% CI of 0.84-0.97 and specificity of 0.95 with a 95% CI of 0.92-0.97. A target of 100 or more patients per country were recommended to be recruited, however no minimum or maximum sample size limitations per site or per country were imposed (*Table 5.2*).

5.3.13 Exploration of telephone WHQ administration pathway

Data were collected about the pathway for telephone WHQ follow-up to describe variability in administration across contexts. This included: Questionnaire translation (pre-translated questionnaire / ad hoc, translated by questionnaire administrator / ad hoc, translated by formal translator); Language of delivery; Phone owner (patient themselves / healthcare worker / friend or relative / other); Phone type (Landline / Mobile phone with a camera / Mobile phone without a camera); Questionnaire administrator (Consultant (doctor) / Junior doctor / Research nurse / Other non-clinical); Duration (minutes).

Table 5.2 Protocol sample size considerations [186]

					Precision around test accuracy measures	
Patients recruited to FALCON trial	**Patients retained in FALCON trial follow-up**	**In-person FALCON trial follow-up**	**Patients with SSI[∅]**	**Patients without SSI[♦]**	**Sensitivity (95% C.I.)**	**Specificity (95% C.I.)**
714	607	516	108	408	0.92 (0.85-0.96)	0.95 (0.93-0.97)
571	485	412	87	325	0.92 (0.84-0.97)	0.95 (0.92-0.97)
428	364	309	65	244	0.92 (0.83-0.97)	0.95 (0.92-0.97)
285	242	206	43	163	0.92 (0.81-0.99)	0.95 (0.91-0.98)

[∅]0.21 * number in-person FALCON trial follow-up. [♦]0.79 * number in-person FALCON trial follow-up. Estimates around sensitivity and specificity were derived from a UK validation study of the WHQ [192]

5.3.14 Statistical analysis

A full statistical analysis plan (SAP) was published online on 8th March 2021 [241]. All analyses were performed using R Studio V4.1.1 (R Foundation for Statistical Computing, Vienna, Austria), packages: tidyverse, finalfit, reportROC, predictr, bcROCcurve. Country-income level was defined according to the World Bank's 2018 definitions, and classified into upper-middle (UMIC), lower-middle (LMIC) or low-income (LIC) based on annual Gross Domestic Product per capita ($).

5.3.14.1 Missing data

The overall rate of missing data was anticipated to be low**.** A sensitivity analysis for the primary validation model was pre-planned to be performed with missing item response data imputed using Multiple Imputation by Chained Equations if the level of missingness was above 5% overall (i.e., per questionnaire) or for any individual item.

5.3.14.2 Patient inclusion in analyses

Data from patients with both (1) in-person FALCON trial follow-up or (2) telephone FALCON trial follow-up only were included in evaluation of feasibility outcome measures. Data from patient with (1) in-person FALCON trial follow-up (i.e., both the reference and index test available) were included in the evaluation of diagnostic accuracy. A potential risk of partial verification bias by including only patients with in-person FALCON trial follow-up in the diagnostic accuracy analysis was identified *a priori*, and addressed in a sensitivity analysis.

5.3.14.3 Presentation of data

Baseline demographics and feasibility outcomes were presented overall, by country, by patient home location and by FALCON trial follow-up group. Distributions of continuous variables were visually inspected for normality. Differences between these groups were

explored using Student's T-test for normal data and the Mann-Whitney U test for non-normal data. The χ2 test was used for categorical data with Fisher's exact modification where required. The proportion of patients included by FALCON trial follow-up group over the study period was summarised graphically.

5.3.14.4 Risk adjustment in feasibility outcome analysis

An exploratory mixed-effects binary regression model was used to explore factors associated with successful telephone contact, with patients nested within countries. The casual pathway for telephone contact was mapped, and patient, disease, operation, and location specific factors were selected *a priori* for inclusion in risk adjustment.

5.3.14.5 Primary outcome analysis

Cross-tabulations of the reference test diagnosis ('no SSI' or 'SSI') against a binary outcome variable derived from the total score of the index test (created by a cut-off score; a WHQ total score of less than or equal to specified values between 1 and 10) were presented. Criterion validity was examined against the reference test to evaluate the performance of the WHQ in discriminating between individuals with and those without SSI. I plotted a receiver operating characteristic (ROC) curve showing test performance across all thresholds, with overall discrimination presented as area under the ROC curve (AUROC) with 95% CIs overall and across several subgroups. The "optimal" cut-points of WHQ scores are obtained by using Youden index method, which maximising the sum of sensitivity and specificity. Diagnostic test accuracy statistics (Accuracy, Sensitivity, Specificity, Positive Predictive Value, Negative Predictive Value) were presented at these cut-points. Calibration of the WHQ was presented as the proportion of patients with SSI diagnosis in the reference test at each WHQ point score interval.

5.3.14.6 Sensitivity analyses

Several sensitivity analyses were performed for the primary model:

1. To allow flexibility during the SARS-CoV-2 pandemic, administration of the WHQ was permitted after FALCON follow-up. The effect of a longer duration after surgery between the WHQ and telephone assessment was explored in a sensitivity analysis including both per-protocol and out-of-protocol patients

2. To evaluate the diagnostic accuracy of the WHQ in post-discharge SSI diagnosis only, a second analysis excluded patients with an in-hospital, pre-discharge SSI diagnosis.

3. To address a risk of partial verification bias, I conducted an inverse probability weighted (IPW) sensitivity analysis for the primary model. In brief, this bias represents a missing data problem, where the reference test is missing for a subset of the sample [242]. Under an assumption of missing data at random (MAR), the IPW method weights each observation in the verified sample by the inverse of the probability of verification to provide a corrected estimate of sensitivity and specificity. The estimated probability of verification is then obtained using a logistic regression model [243, 244].

5.3.14.7 Subgroup analyses

Subgroups included urban versus rural home location, UMIC versus LMIC versus LIC, patient age >60 versus $\leq$60, elective versus emergency surgery which were pre-specified, and pre-translated questionnaire versus ad hoc translation, and no reoperation (mild SSI only) which were added post-hoc for exploratory analysis.

5.3.14.8 Changes from the published Statistical Analysis Plan

Some small, iterative changes were required to the published statistical analysis plan (available at: https://globalsurg.org/resources/phd-research-projects/talon/) related to the SARS-CoV-2 pandemic, the sample size of proposed comparator groups, and patterns observed in the data. There were no major changes to the primary comparison groups, definition of outcomes, primary analysis methods or handling of missing data.

These included:

1. The number of events and non-events was too small per country to justify per-country analyses. As such:
    a. Cut-off WHQ point scores were presented overall and not by country
    b. Subgroup analyses were presented by country income group rather than by individual country

2. Multiple comparisons of patient characteristics and outcomes between urban versus rural patient home location were too extensive for a single peer-reviewed publication. I presented a subgroup analysis of the model discrimination by urban versus rural settings, and will explore patient home location further in future work.

3. I did not anticipate a significant proportion of the cohort to be outside of the protocol timing window for administration of the WHQ. Due to the SARS-CoV-2 pandemic, I relaxed the pre-specified time window to allow flexibility for overburdened site investigators and a time of system strain, and presented the primary analysis per-protocol, with a sensitivity analysis including all patients including those out-of-protocol.

4. I did not anticipate a large proportion of the cohort to not receive the index test (in-person FALCON trial follow-up), but face-to-face follow-up was deemed not to be safe or feasible in many settings during the SARS-CoV-2 pandemic. I therefore introduced an inverse probability weighted sensitivity analysis to account for a risk of verification bias.
5. In response to investigator and CEI partner feedback, I added two exploratory subgroup analyses: (1) formal versus adhoc translation of the WHQ and (2) mild SSI only (i.e., no reoperation)

5.3.15 Community Engagement and Involvement (CEI)

The aim of CEI in this study was to optimise the pathway for telephone WHQ administration to ensure cultural and contextual acceptability and maximise both the telephone contact and questionnaire completion rate. Patient and community partners were involved in study prioritisation, design, steering, and reporting using three methods. Firstly, through direct involvement in the Study Management Group. Secondly, through a UK-based advisory group with expatriate partners from collaborating countries. Thirdly, an extended network of patient and community partners were consulted through the NIHR Unit on Global Surgery network. CEI in this study is reported according to the GRIPP-2 short form [225].

5.4 Results

5.4.1 Overview of study inclusion

Overall, 1240 patients were included with telephone WHQ follow-up attempted, of whom 29 had died by 30-days after surgery (1 missing status) and 14 more had no FALCON trial follow-up. 1196 patients were therefore eligible for inclusion in analyses (*Figure 5.4*). Patients were from three UMICs (396 patients, 13 hospitals), three LMICs (746 patients, 19 hospitals), and one LIC (54 patients, 4 hospitals). The largest contributing countries were Ghana (532/1196, 44.5%), Mexico (216/1196, 18.1%) and India (120/1196, 10.0%). Baseline demographics are shown in *Table 5.3*. 17.5% (209/1196) had an SSI diagnosis within the 30-days after surgery in the FALCON trial. A comparison of patients included in the TALON-2 study versus the FALCON trial overall is presented in *Table 5.4*. Of note, there were fewer patients undergoing elective surgery, fewer female patients, more intermediate/minor operations, and more contaminated/dirty surgery in TALON-2 than in FALCON overall. The completed STARD checklist is in *Appendix G*.

Table 5.3 Baseline demographics by country (N=1196)

		Country							
Factor	**Levels**	**Ghana N=532**	**South Africa N=60**	**India N=120**	**Benin N=103**	**Mexico N=216**	**Rwanda N=54**	**Nigeria N=111**	**Total N=1196**
FALCON trial follow-up	In-person	199 (37.4)	27 (45.0)	76 (63.3)	82 (79.6)	129 (59.7)	35 (64.8)	40 (36.0)	588 (49.2)
	Telephone only	333 (62.6)	33 (55.0)	44 (36.7)	21 (20.4)	87 (40.3)	19 (35.2)	71 (64.0)	608 (50.8)
Timing of telephone WHQ	Per protocol	517 (97.2)	53 (88.3)	29 (24.2)	100 (97.1)	17 (7.9)	54 (100.0)	108 (97.3)	878 (73.4)
	Outside of protocol	2 (0.4)	0 (0.0)	91 (75.8)	0 (0.0)	112 (51.9)	0 (0.0)	0 (0.0)	205 (17.1)
	(Missing)	13 (2.4)	7 (11.7)	0 (0.0)	3 (2.9)	87 (40.3)	0 (0.0)	3 (2.7)	113 (9.4)
Home location	Urban	299 (56.2)	52 (86.7)	70 (58.3)	92 (89.3)	98 (45.4)	8 (14.8)	88 (79.3)	707 (59.1)
	Rural	221 (41.5)	2 (3.3)	49 (40.8)	8 (7.8)	31 (14.4)	46 (85.2)	23 (20.7)	380 (31.8)
	(Missing)	12 (2.3)	6 (10.0)	1 (0.8)	3 (2.9)	87 (40.3)	0 (0.0)	0 (0.0)	109 (9.1)
Age (years)	<18	68 (12.8)	0 (0.0)	2 (1.7)	0 (0.0)	2 (0.9)	1 (1.9)	21 (18.9)	94 (7.9)
	18-39	261 (49.1)	32 (53.3)	80 (66.7)	85 (82.5)	137 (63.4)	32 (59.3)	43 (38.7)	670 (56.0)
	40-59	140 (26.3)	20 (33.3)	26 (21.7)	13 (12.6)	56 (25.9)	10 (18.5)	29 (26.1)	294 (24.6)
	60-79	56 (10.5)	8 (13.3)	12 (10.0)	5 (4.9)	17 (7.9)	9 (16.7)	15 (13.5)	122 (10.2)
	≥80	7 (1.3)	0 (0.0)	0 (0.0)	0 (0.0)	4 (1.9)	2 (3.7)	3 (2.7)	16 (1.3)
Sex	Male	350 (65.8)	37 (61.7)	32 (26.7)	57 (55.3)	39 (18.1)	36 (66.7)	64 (57.7)	615 (51.4)
	Female	182 (34.2)	23 (38.3)	88 (73.3)	46 (44.7)	177 (81.9)	18 (33.3)	47 (42.3)	581 (48.6)
Level of education	Below high school level	337 (64.9)	15 (27.8)	64 (54.2)	29 (29.0)	24 (18.6)	42 (77.8)	46 (41.4)	557 (51.3)
	High school or above	182 (35.1)	39 (72.2)	54 (45.8)	71 (71.0)	105 (81.4)	12 (22.2)	65 (58.6)	528 (48.7)
Known diabetes	Yes	11 (2.1)	2 (3.3)	7 (5.8)	1 (1.0)	16 (7.4)	1 (1.9)	2 (1.8)	40 (3.3)
	No	521 (97.9)	58 (96.7)	113 (94.2)	102 (99.0)	200 (92.6)	53 (98.1)	109 (98.2)	1156 (96.7)
HIV status	Known negative	42 (7.9)	11 (18.3)	114 (95.0)	4 (3.9)	63 (29.2)	40 (74.1)	70 (63.1)	344 (28.8)
	Known positive	4 (0.8)	14 (23.3)	1 (0.8)	0 (0.0)	3 (1.4)	1 (1.9)	0 (0.0)	23 (1.9)
	Status not known	486 (91.4)	35 (58.3)	5 (4.2)	99 (96.1)	150 (69.4)	13 (24.1)	41 (36.9)	829 (69.3)
Smoking status	Never smoked	504 (94.7)	40 (66.7)	113 (94.2)	100 (97.1)	187 (86.6)	46 (85.2)	99 (89.2)	1089 (91.1)
	Ex-smoker	16 (3.0)	11 (18.3)	6 (5.0)	0 (0.0)	24 (11.1)	5 (9.3)	4 (3.6)	66 (5.5)
	Current smoker	12 (2.3)	9 (15.0)	1 (0.8)	3 (2.9)	5 (2.3)	3 (5.6)	8 (7.2)	41 (3.4)
Urgency	Elective (planned)	27 (5.1)	8 (13.3)	38 (31.7)	0 (0.0)	171 (79.2)	3 (5.6)	21 (18.9)	268 (22.4)
	Emergency (unplanned)	505 (94.9)	52 (86.7)	82 (68.3)	103 (100.0)	45 (20.8)	51 (94.4)	90 (81.1)	928 (77.6)
Indication	Malignant disease	23 (4.3)	5 (8.3)	20 (16.7)	2 (1.9)	13 (6.0)	2 (3.7)	20 (18.0)	85 (7.1)
	Benign disease	485 (91.2)	39 (65.0)	31 (25.8)	100 (97.1)	123 (56.9)	46 (85.2)	83 (74.8)	907 (75.8)
	Trauma	18 (3.4)	13 (21.7)	0 (0.0)	1 (1.0)	1 (0.5)	5 (9.3)	6 (5.4)	44 (3.7)
	Obstetric	5 (0.9)	3 (5.0)	69 (57.5)	0 (0.0)	79 (36.6)	1 (1.9)	2 (1.8)	159 (13.3)
	(Missing)	1 (0.2)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	1 (0.1)
Operation location	Foregut	183 (34.4)	11 (18.3)	12 (10.0)	8 (7.8)	70 (32.4)	9 (16.7)	22 (19.8)	315 (26.3)
	Hindgut	45 (8.5)	9 (15.0)	21 (17.5)	2 (1.9)	11 (5.1)	3 (5.6)	21 (18.9)	112 (9.4)
	Appendix	166 (31.2)	19 (31.7)	6 (5.0)	88 (85.4)	14 (6.5)	11 (20.4)	28 (25.2)	332 (27.8)
	Urogenital	9 (1.7)	1 (1.7)	71 (59.2)	0 (0.0)	110 (50.9)	4 (7.4)	9 (8.1)	204 (17.1)
	Other	128 (24.1)	19 (31.7)	10 (8.3)	5 (4.9)	11 (5.1)	27 (50.0)	28 (25.2)	228 (19.1)
	(Missing)	1 (0.2)	1 (1.7)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	3 (2.7)	5 (0.4)
ASA grade	Grade I	320 (60.2)	30 (50.0)	43 (35.8)	79 (76.7)	58 (26.9)	17 (31.5)	13 (11.7)	560 (46.8)
	Grade II	158 (29.7)	20 (33.3)	69 (57.5)	22 (21.4)	145 (67.1)	20 (37.0)	36 (32.4)	470 (39.3)
	Grade III	50 (9.4)	8 (13.3)	6 (5.0)	2 (1.9)	11 (5.1)	17 (31.5)	56 (50.5)	150 (12.5)
	Grade IV/V	3 (0.6)	2 (3.3)	2 (1.7)	0 (0.0)	2 (0.9)	0 (0.0)	6 (5.4)	15 (1.3)
	(Missing)	1 (0.2)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	1 (0.1)
WHO checklist	Yes	500 (94.0)	57 (95.0)	120 (100.0)	102 (99.0)	210 (97.2)	53 (98.1)	69 (62.2)	1111 (92.9)
	No	32 (6.0)	3 (5.0)	0 (0.0)	1 (1.0)	6 (2.8)	1 (1.9)	42 (37.8)	85 (7.1)
Operation grade	Intermediate/Minor	185 (35.8)	20 (34.5)	8 (6.7)	88 (85.4)	18 (8.4)	12 (24.5)	29 (26.9)	360 (30.8)
	Major	332 (64.2)	38 (65.5)	111 (93.3)	15 (14.6)	197 (91.6)	37 (75.5)	79 (73.1)	809 (69.2)
Contamination	Clean/Clean-contaminated	75 (14.1)	14 (23.3)	89 (74.2)	14 (13.6)	200 (92.6)	1 (1.9)	14 (12.6)	407 (34.0)
	Contaminated	225 (42.3)	30 (50.0)	26 (21.7)	40 (38.8)	14 (6.5)	26 (48.1)	52 (46.8)	413 (34.5)
	Dirty	230 (43.2)	16 (26.7)	5 (4.2)	49 (47.6)	2 (0.9)	27 (50.0)	45 (40.5)	374 (31.3)
	(Missing)	2 (0.4)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	2 (0.2)
Approach	Open midline	435 (81.8)	48 (80.0)	41 (34.2)	34 (33.0)	43 (19.9)	49 (90.7)	83 (74.8)	733 (61.3)
	Open non-midline	96 (18.0)	11 (18.3)	74 (61.7)	69 (67.0)	169 (78.2)	5 (9.3)	28 (25.2)	452 (37.8)
	Laparoscopic attempted	0 (0.0)	1 (1.7)	5 (4.2)	0 (0.0)	4 (1.9)	0 (0.0)	0 (0.0)	10 (0.8)
	(Missing)	1 (0.2)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	1 (0.1)
Stoma formation	Yes	15 (2.8)	4 (6.7)	17 (14.2)	0 (0.0)	10 (4.6)	6 (11.1)	7 (6.3)	59 (4.9)
	No	514 (96.6)	55 (91.7)	102 (85.0)	103 (100.0)	205 (94.9)	47 (87.0)	102 (91.9)	1128 (94.3)
	(Missing)	3 (0.6)	1 (1.7)	1 (0.8)	0 (0.0)	1 (0.5)	1 (1.9)	2 (1.8)	9 (0.8)

*FALCON, a stratified, pragmatic, multi-centre, 2x2 factorial trial testing two measures (skin preparation and antimicrobial sutures) to reduce superficial or deep skin infection following abdominal surgery in seven low- and middle-income countries (NCT03700749) (2). WHQ = Wound Healing Questionnaire. HIV = Human immunodeficiency virus.

Table 5.4 Comparison of patients included in TALON-2 study within a trial and FALCON trial overall

Factor	Levels	Included in TALON-2 study within a trial N=1088	Included in FALCON study overall N=5788
Country	Ghana	520 (47.8)	1424 (24.6)
	South Africa	54 (5.0)	177 (3.1)
	India	120 (11.0)	800 (13.8)
	Benin	100 (9.2)	145 (2.5)
	Mexico	129 (11.9)	238 (4.1)
	Rwanda	54 (5.0)	839 (14.5)
	Nigeria	111 (10.2)	2165 (37.4)
Age (years)	<18	90 (8.3)	811 (14.0)
	18-39	606 (55.7)	3183 (55.0)
	40-59	262 (24.1)	1186 (20.5)
	60-79	115 (10.6)	534 (9.2)
	≥80	15 (1.4)	74 (1.3)
Sex	Male	582 (53.5)	2200 (38.4)
	Female	506 (46.5)	3534 (61.6)
Known diabetes	Yes	33 (3.0)	217 (3.8)
	No	1055 (97.0)	5513 (96.2)
HIV status	Known negative	319 (29.3)	3207 (55.9)
	Known positive	21 (1.9)	128 (2.2)
	Status not known	748 (68.8)	2399 (41.8)
Smoking status	Never smoked	996 (91.5)	5344 (93.2)
	Ex-smoker	53 (4.9)	220 (3.8)
	Current smoker	39 (3.6)	168 (2.9)
Urgency	Elective (planned)	197 (18.1)	1915 (33.1)
	Emergency (unplanned)	891 (81.9)	3873 (66.9)
Indication	Malignant disease	80 (7.4)	442 (7.8)
	Benign disease	834 (76.7)	3326 (59.0)
	Trauma	40 (3.7)	442 (7.8)
	Obstetric	133 (12.2)	1702 (30.2)
	(Missing)	1 (0.1)	1 (0.0)
Operation location	Foregut	272 (25.0)	1084 (18.7)
	Hindgut	106 (9.7)	529 (9.1)
	Appendix	324 (29.8)	933 (16.1)
	Urogenital	162 (14.9)	2255 (39.0)
	Other	219 (20.1)	834 (14.4)
	(Missing)	5 (0.5)	153 (2.6)
ASA grade	Grade I	519 (47.7)	2540 (45.0)
	Grade II	415 (38.1)	2121 (37.6)
	Grade III	142 (13.1)	788 (14.0)
	Grade IV/V	11 (1.0)	196 (3.4)
	(Missing)	1 (0.1)	3 (0.0)
WHO Checklist	Yes	1006 (92.5)	680 (12.0)
	No	82 (7.5)	4965 (88.0)
Operation grade	Intermediate/Minor	350 (33.0)	1023 (18.4)
	Major	711 (67.0)	4534 (81.6)
Contamination	Clean/Clean-contaminated	322 (29.6)	3123 (55.4)
	Contaminated	399 (36.7)	1235 (21.9)
	Dirty	365 (33.5)	1282 (22.7)
	(Missing)	2 (0.2)	3 (0.0)
Approach	Open midline	707 (65.0)	2764 (49.0)
	Open non-midline	371 (34.1)	2827 (50.1)
	Laparoscopic attempted	9 (0.8)	51 (0.8)
	(Missing)	1 (0.1)	1 (0.1)
Stoma formation	Yes	55 (5.1)	290 (5.1)
	No	1025 (94.2)	5318 (94.8)
	(Missing)	8 (0.7)	12 (0.1)

5.4.2. Feasibility outcomes

5.4.2.1 Telephone contact rate

Baseline demographics grouped by whether telephone contact was made or not is presented in *Table 5.5*. Overall, the telephone contact rate was high at 90.3% (1088/1196) with 9.7% (108/1196) lost to follow-up with some variability by country (*Table 5.6*).

The WHQ was completed for all but one patient where successful contact was made (99.9%, 1087/1088). The rate of telephone contact reduced as time from date of surgery increased (*Figure 5.2*).

The most significant factor associated with lower odds of telephone contact-up in the multivariable model was time from surgery (*Figure 5.3*). Importantly, patients with non-midline incisions (adjusted odds ratio: 0.36, 95% CI 0.17-0.73, p=0.005) or with a confirmed reference test diagnosis of SSI (OR: 0.42, 95% CI 0.20-0.92, p=0.006) were less likely to be contactable. Where data were available, most patients were followed-up with 1 (47.7%, 267/560) or 2 to 3 (33.0%, 185/560) attempts at telephone follow-up (missing: 636).

5.4.2.2 Patient satisfaction

Patients overall felt very satisfied (71.5%, 393/550) or satisfied (27.6%, 152/550) with undergoing telephone WHQ follow-up (missing: 646).

Figure 5.2 Proportion of patients with successful telephone contact with increasing time from surgery (days)

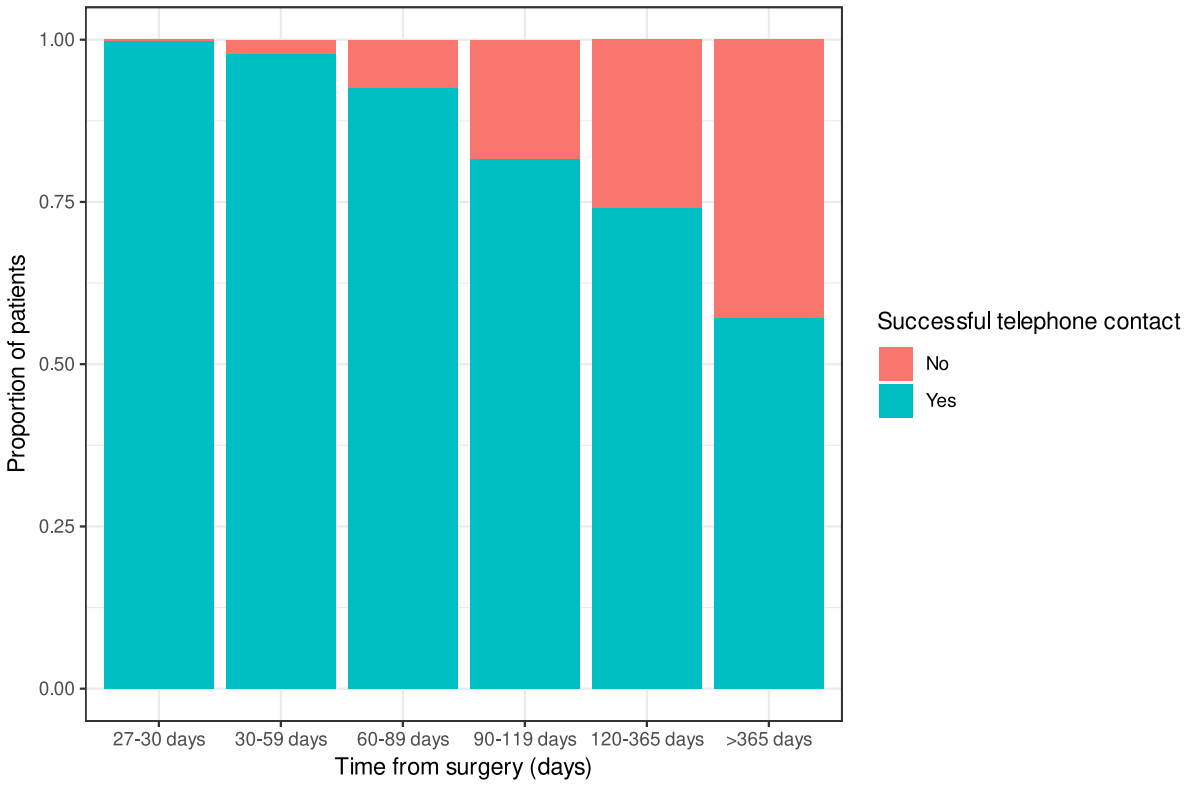


Table 5.5 Baseline demographics of patients with and without successful telephone contact (N=1196)

		Successful telephone contact			
Factor	**Levels**	**No N=108**	**Yes N=1088**	**Total**	**P-value for difference**
Age	<18 years	4 (3.7)	90 (8.3)	94 (7.9)	0.213
	18-39	64 (59.3)	606 (55.7)	670 (56.0)	
	40-59	32 (29.6)	262 (24.1)	294 (24.6)	
	60-79	7 (6.5)	115 (10.6)	122 (10.2)	
	≥80	1 (0.9)	15 (1.4)	16 (1.3)	
Sex	Male	33 (30.6)	582 (53.5)	615 (51.4)	<0.001
	Female	75 (69.4)	506 (46.5)	581 (48.6)	
Known diabetes	Yes	7 (6.5)	33 (3.0)	40 (3.3)	0.105
	No	101 (93.5)	1055 (97.0)	1156 (96.7)	
HIV status	Known negative	25 (23.1)	319 (29.3)	344 (28.8)	0.395
	Known positive	2 (1.9)	21 (1.9)	23 (1.9)	
	Status not known	81 (75.0)	748 (68.8)	829 (69.3)	
Smoking status	Never smoked	93 (86.1)	996 (91.5)	1089 (91.1)	0.006
	Ex-smoker	13 (12.0)	53 (4.9)	66 (5.5)	
	Current smoker	2 (1.9)	39 (3.6)	41 (3.4)	
Urgency	Elective (planned)	71 (65.7)	197 (18.1)	268 (22.4)	<0.001
	Emergency (unplanned)	37 (34.3)	891 (81.9)	928 (77.6)	
Indication	Malignant disease	5 (4.6)	80 (7.4)	85 (7.1)	0.006
	Benign disease	73 (67.6)	834 (76.7)	907 (75.8)	
	Trauma	4 (3.7)	40 (3.7)	44 (3.7)	
	Obstetric	26 (24.1)	133 (12.2)	159 (13.3)	
	(Missing)	0 (0.0)	1 (0.1)	1 (0.1)	
Operation location	Foregut	43 (39.8)	272 (25.0)	315 (26.3)	<0.001
	Hindgut	6 (5.6)	106 (9.7)	112 (9.4)	
	Appendix	8 (7.4)	324 (29.8)	332 (27.8)	
	Urogenital	42 (38.9)	162 (14.9)	204 (17.1)	
	Other	9 (8.3)	219 (20.1)	228 (19.1)	
	(Missing)	0 (0.0)	5 (0.5)	5 (0.4)	
ASA grade	Grade I	41 (38.0)	519 (47.7)	560 (46.8)	0.003
	Grade II	55 (50.9)	415 (38.1)	470 (39.3)	
	Grade III	8 (7.4)	142 (13.1)	150 (12.5)	
	Grade IV/V	4 (3.7)	11 (1.0)	15 (1.3)	
	(Missing)	0 (0.0)	1 (0.1)	1 (0.1)	
WHO Checklist	Yes	105 (97.2)	1006 (92.5)	1111 (92.9)	0.101
	No	3 (2.8)	82 (7.5)	85 (7.1)	
Operation grade	Intermediate/Minor	10 (9.3)	350 (33.0)	360 (30.8)	<0.001
	Major	98 (90.7)	711 (67.0)	809 (69.2)	
ontamination	Clean/Clean-contaminated	85 (78.7)	322 (29.6)	407 (34.0)	<0.001
	Contaminated	14 (13.0)	399 (36.7)	413 (34.5)	
	Dirty	9 (8.3)	365 (33.5)	374 (31.3)	
	(Missing)	0 (0.0)	2 (0.2)	2 (0.2)	
Approach	Open midline	26 (24.1)	707 (65.0)	733 (61.3)	<0.001
	Open non-midline	81 (75.0)	371 (34.1)	452 (37.8)	
	Laparoscopic attempted	1 (0.9)	9 (0.8)	10 (0.8)	
	(Missing)	0 (0.0)	1 (0.1)	1 (0.1)	
Stoma formation	Yes	4 (3.7)	55 (5.1)	59 (4.9)	0.703
	No	103 (95.4)	1025 (94.2)	1128 (94.3)	
	(Missing)	1 (0.9)	8 (0.7)	9 (0.8)	

Table 5.6 Feasibility outcomes by country (N=1196)

		Country								
Factor	**Levels**	**Ghana N=532**	**South Africa N=60**	**India N=120**	**Benin N=103**	**Mexico N=216**	**Rwanda N=54**	**Nigeria N=111**	**Total N=1196**	**P-value**
Successful telephone contact	No	12 (2.3)	6 (10.0)	0 (0.0)	3 (2.9)	87 (40.3)	0 (0.0)	0 (0.0)	108 (9.0)	<0.001
	Yes	520 (97.7)	54 (90.0)	120 (100.0)	100 (97.1)	129 (59.7)	54 (100.0)	111 (100.0)	1088 (91.0)	
Attempts at contact	1 attempt	145 (49.8)	25 (41.7)	28 (68.3)	0 (0.0)	10 (62.5)	12 (22.2)	47 (48.0)	267 (47.7)	<0.001
	2-3 attempts	93 (32.0)	22 (36.7)	8 (19.5)	0 (0.0)	1 (6.2)	29 (53.7)	32 (32.7)	185 (33.0)	
	4-5 attempts	26 (8.9)	6 (10.0)	5 (12.2)	0 (0.0)	0 (0.0)	11 (20.4)	14 (14.3)	62 (11.1)	
	>5 attempts	27 (9.3)	7 (11.7)	0 (0.0)	0 (0.0)	5 (31.2)	2 (3.7)	5 (5.1)	46 (8.2)	
	(Missing)*	241	0	79	103	200	0	13	636	
Patient satisfaction[$]	Very Satisfied	243 (84.1)	42 (77.8)	13 (28.9)	0 (0.0)	1 (9.1)	39 (72.2)	55 (56.7)	393 (71.5)	<0.001
	Satisfied	45 (15.6)	9 (16.7)	32 (71.1)	0 (0.0)	10 (90.9)	15 (27.8)	41 (42.3)	152 (27.6)	
	Neutral	1 (0.3)	3 (5.6)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	1 (1.0)	5 (0.9)	
	Unsatisfied	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	
	Very unsatisfied	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	
	(Missing)*	243	6	75	103	205	0	14	646	

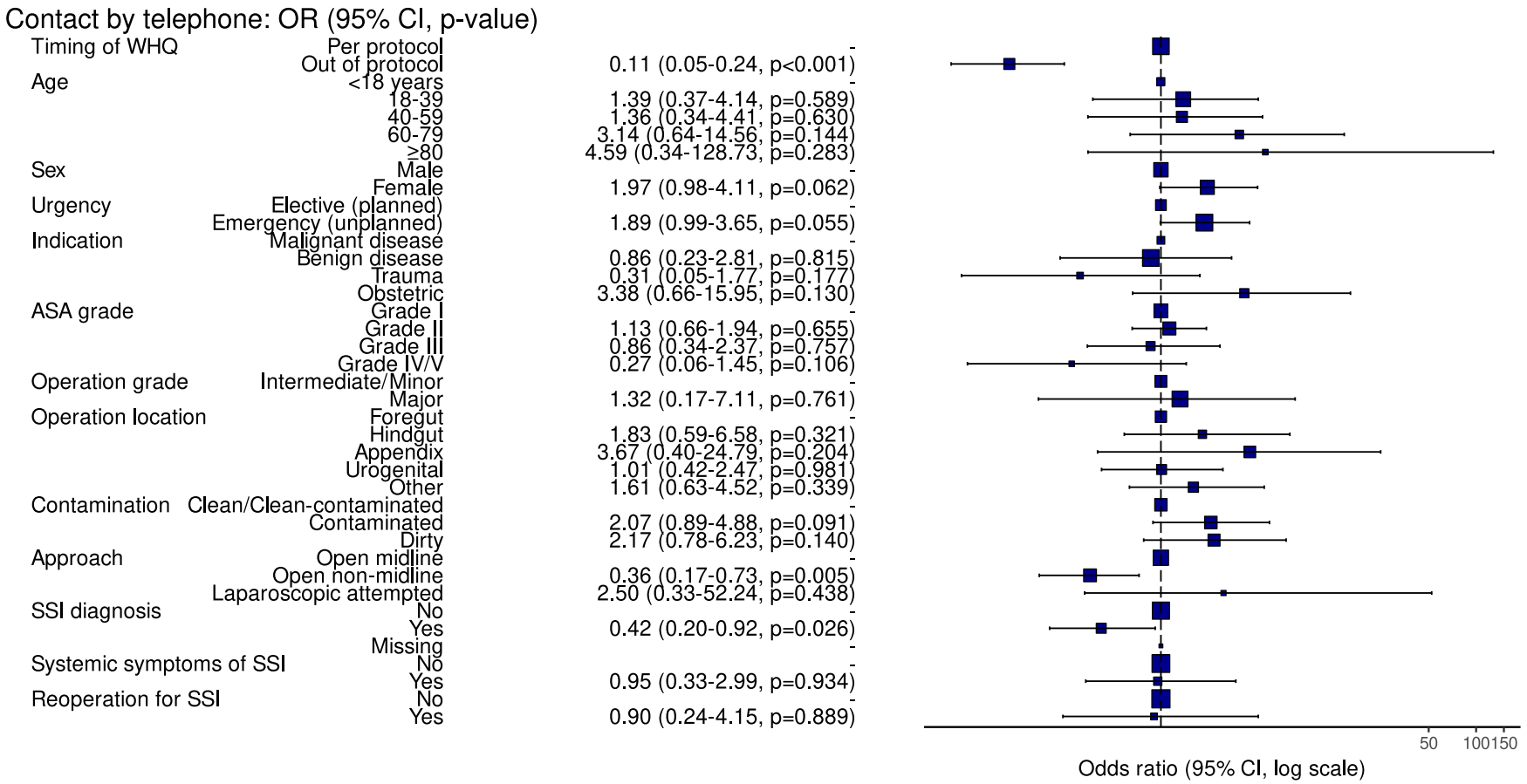
*Question added after pilot phase in response to Community Engagement and Involvement group feedback, so not available for patients recruited in pilot phase. [$] Represents patient's self-reported satisfaction with telephone administration of the WHQ.

Table 5.7. Factors associated with successful telephone contact in a multivariable model

		Successful telephone contact			
Factor	**Levels**	**No**	**Yes**	**OR (univariable)**	**OR (multivariable)**
Timing of WHQ	Per protocol	25 (2.8)	881 (97.2)	-	-
	Out of protocol	83 (28.8)	205 (71.2)	0.07 (0.04-0.11, p<0.001)	***0.11 (0.05-0.24, p<0.001)***
Patient factors					
Age	<18 years	4 (4.3)	90 (95.7)	-	-
	18-39	64 (9.6)	606 (90.4)	0.42 (0.13-1.05, p=0.101)	1.39 (0.37-4.14, p=0.589)
	40-59	32 (10.9)	262 (89.1)	0.36 (0.11-0.95, p=0.063)	1.36 (0.34-4.41, p=0.630)
	60-79	7 (5.7)	115 (94.3)	0.73 (0.19-2.49, p=0.624)	3.14 (0.64-14.56, p=0.144)
	≥80	1 (6.2)	15 (93.8)	0.67 (0.09-13.51, p=0.725)	4.59 (0.34-128.73, p=0.283)
Sex	Male	33 (5.4)	582 (94.6)	-	-
	Female	75 (12.9)	506 (87.1)	0.38 (0.25-0.58, p<0.001)	1.97 (0.98-4.11, p=0.062)
Urgency	Elective (planned)	71 (26.5)	197 (73.5)	-	-
	Emergency (unplanned)	37 (4.0)	891 (96.0)	8.68 (5.70-13.41, p<0.001)	1.89 (0.99-3.65, p=0.055)
Indication	Malignant disease	5 (5.9)	80 (94.1)	-	-
	Benign disease	73 (8.0)	834 (92.0)	0.71 (0.25-1.65, p=0.480)	0.86 (0.23-2.81, p=0.815)
	Trauma	4 (9.1)	40 (90.9)	0.63 (0.16-2.64, p=0.501)	0.31 (0.05-1.77, p=0.177)
	Obstetric	26 (16.4)	133 (83.6)	0.32 (0.10-0.80, p=0.025)	3.38 (0.66-15.95, p=0.130)
ASA grade	Grade I	41 (7.3)	519 (92.7)	-	-
	Grade II	55 (11.7)	415 (88.3)	0.60 (0.39-0.91, p=0.017)	1.13 (0.66-1.94, p=0.655)
	Grade III	8 (5.3)	142 (94.7)	1.40 (0.68-3.29, p=0.396)	0.86 (0.34-2.37, p=0.757)
	Grade IV/V	4 (26.7)	11 (73.3)	0.22 (0.07-0.81, p=0.012)	0.27 (0.06-1.45, p=0.106)
Operation factors					
Operation grade	Intermediate/Minor	10 (2.8)	350 (97.2)	-	-
	Major	98 (12.1)	711 (87.9)	0.21 (0.10-0.38, p<0.001)	1.32 (0.17-7.11, p=0.761)
Operation location	Foregut	43 (13.7)	272 (86.3)	-	-
	Hindgut	6 (5.4)	106 (94.6)	2.79 (1.24-7.49, p=0.023)	1.83 (0.59-6.58, p=0.321)
	Appendix	8 (2.4)	324 (97.6)	6.40 (3.12-14.91, p<0.001)	3.67 (0.40-24.79, p=0.204)
	Urogenital	42 (20.6)	162 (79.4)	0.61 (0.38-0.97, p=0.038)	1.01 (0.42-2.47, p=0.981)
	Other	9 (3.9)	219 (96.1)	3.85 (1.92-8.58, p<0.001)	1.61 (0.63-4.52, p=0.339)
Contamination	Clean/Clean-contaminated	85 (20.9)	322 (79.1)	-	-
	Contaminated	14 (3.4)	399 (96.6)	7.52 (4.33-14.05, p<0.001)	2.07 (0.89-4.88, p=0.091)
	Dirty	9 (2.4)	365 (97.6)	10.71 (5.59-23.19, p<0.001)	2.17 (0.78-6.23, p=0.140)
Approach	Open midline	26 (3.5)	707 (96.5)	-	-
	Open non-midline	81 (17.9)	371 (82.1)	0.17 (0.10-0.26, p<0.001)	***0.36 (0.17-0.73, p=0.005)***
	Laparoscopic attempted	1 (10.0)	9 (90.0)	0.33 (0.06-6.22, p=0.303)	2.50 (0.33-52.24, p=0.438)
Surgical site infection status					
SSI diagnosis (reference test)	No	90 (9.1)	897 (90.9)	-	-
	Yes	18 (8.6)	191 (91.4)	1.06 (0.64-1.86, p=0.817)	***0.42 (0.20-0.92, p=0.026)***
Systemic symptoms of SSI	No	99 (9.1)	986 (90.9)	-	-
	Yes	9 (8.1)	102 (91.9)	1.14 (0.59-2.48, p=0.722)	0.95 (0.33-2.99, p=0.934)
Reoperation for SSI	No	104 (9.1)	1041 (90.9)	-	-
	Yes	4 (8.2)	45 (91.8)	1.12 (0.45-3.78, p=0.826)	0.90 (0.24-4.15, p=0.889)

Number in dataframe = 1196, Number in model = 1162, Missing = 34, AIC = 500, C-statistic = 0.917

Figure 5.3 Forest plot for factors associated with successful telephone contact in a multivariable model



A lower odds ratio conveyed a lower likelihood of telephone contact successfully being made by telephone to complete the TALON-2 questionnaire.

5.4.2.3 Implementation of telephone WHQ follow-up

Telephone WHQ administration was performed across diverse settings and patient groups, in 22 languages and 36 hospitals. Overall, 65.5% (707/1087) of contactable patients lived in urban and 35.0% (380/1087, missing: 1) in rural settings. 64.4% (701/1087) patients received the call using their own phone, whereas 33.7% (367/1087) used a family member's. 669 patients (64.2%) used a smartphone with video capability. Importantly the WHQ was mainly delivered by non-consultant (attending) grade researchers (other doctor: N=367, 33.7%, research nurse: N=327, 30.1%, other non-clinical: N=385, 35.4%) and largely took less than 20 minutes to complete for 96.0% (528/550, missing: 538). There were several differences in the implementation of telephone WHQ follow-up across participating countries (*Table 5.8*), demonstrating the diversity of the delivery network.

Table 5.8 Follow-up for patients contactable with the telephone WHQ (N=1088)

		Country							
Factor	**Levels**	**Ghana N=525**	**South Africa N=54**	**India N=120**	**Benin N=100**	**Mexico N=130**	**Rwanda N=54**	**Nigeria N=111**	**Total N=1094**
Questionnaire translation	Pre-translated questionnaire	150 (28.8)	9 (16.7)	14 (11.7)	87 (87.0)	120 (93.0)	43 (79.6)	65 (58.6)	424 (39.0)
	Adhoc, translated from English by questionnaire administrator	368 (70.8)	30 (55.6)	101 (84.2)	13 (13.0)	9 (7.0)	11 (20.4)	46 (41.4)	579 (53.2)
	Adhoc, translated from English by formal translator	0 (0.0)	1 (1.9)	5 (4.2)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	6 (0.6)
	(Missing)	2 (0.4)	14 (25.9)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	79 (7.3)
Language of delivery	English	79 (15.2)	24 (44.4)	5 (4.2)	0 (0.0)	0 (0.0)	0 (0.0)	65 (58.6)	173 (15.9)
	Fante	20 (3.8)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	20 (1.8)
	Fon	0 (0.0)	0 (0.0)	0 (0.0)	6 (6.0)	0 (0.0)	0 (0.0)	0 (0.0)	6 (0.6)
	Goun	0 (0.0)	0 (0.0)	0 (0.0)	6 (6.0)	0 (0.0)	0 (0.0)	0 (0.0)	6 (0.6)
	Hausa	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	2 (1.8)	2 (0.2)
	Igbo	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	16 (14.4)	16 (1.5)
	Malayalam	0 (0.0)	0 (0.0)	2 (1.7)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	2 (0.2)
	Sotho	0 (0.0)	10 (18.5)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	10 (0.9)
	Swati	0 (0.0)	1 (1.9)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	1 (0.1)
	Telegu	0 (0.0)	0 (0.0)	1 (0.8)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	1 (0.1)
	Tswana	0 (0.0)	3 (5.6)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	3 (0.3)
	Xhosa	0 (0.0)	3 (5.6)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	3 (0.3)
	Yoruba	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	28 (25.2)	28 (2.6)
	Zulu	0 (0.0)	13 (24.1)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	13 (1.2)
	Dagbani	56 (10.8)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	56 (5.1)
	French	0 (0.0)	0 (0.0)	0 (0.0)	88 (88.0)	0 (0.0)	0 (0.0)	0 (0.0)	88 (8.1)
	Hindi	0 (0.0)	0 (0.0)	69 (57.5)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	69 (6.3)
	Kinyarwanda	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	54 (100.0)	0 (0.0)	54 (5.0)
	Punjabi	0 (0.0)	0 (0.0)	20 (16.7)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	20 (1.8)
	Spanish	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	129 (100.0)	0 (0.0)	0 (0.0)	129 (11.9)
	Tamil	0 (0.0)	0 (0.0)	23 (19.2)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	23 (2.1)
	Twi	364 (70.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	364 (33.5)
	(Missing)	1 (0.2)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	2 (1.8)	1 (0.1)
Phone owner	Patient themselves	335 (64.4)	43 (79.6)	44 (36.7)	84 (84.0)	91 (70.5)	23 (42.6)	81 (73.0)	701 (64.4)
	Healthcare worker	0 (0.0)	1 (1.9)	1 (0.8)	0 (0.0)	0 (0.0)	3 (5.6)	0 (0.0)	5 (0.5)
	Friend or relative	183 (35.2)	9 (16.7)	74 (61.7)	16 (16.0)	38 (29.5)	20 (37.0)	27 (24.3)	367 (33.7)
	Other	1 (0.2)	1 (1.9)	1 (0.8)	0 (0.0)	0 (0.0)	8 (14.8)	3 (2.7)	14 (1.3)
	(Missing)	1 (0.2)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	1 (0.1)
Phone type	Landline phone	0 (0.0)	0 (0.0)	1 (0.8)	0 (0.0)	12 (9.3)	0 (0.0)	0 (0.0)	13 (1.2)
	Mobile phone (with a camera)	295 (56.7)	41 (75.9)	94 (78.3)	77 (77.0)	104 (80.6)	6 (11.1)	82 (73.9)	699 (64.2)
	Mobile phone (without a camera)	225 (43.3)	13 (24.1)	25 (20.8)	23 (23.0)	13 (10.1)	48 (88.9)	29 (26.1)	376 (34.6)
Questionnaire administrator	Consultant (doctor)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	6 (4.7)	0 (0.0)	2 (1.8)	8 (0.7)
	Other doctor	175 (33.7)	0 (0.0)	2 (1.7)	100 (100.0)	86 (66.7)	0 (0.0)	4 (3.6)	367 (33.7)
	Research nurse	104 (20.0)	48 (88.9)	102 (85.0)	0 (0.0)	0 (0.0)	46 (85.2)	27 (24.3)	327 (30.1)
	Other	240 (46.2)	6 (11.1)	16 (13.3)	0 (0.0)	37 (28.7)	8 (14.8)	78 (70.3)	385 (35.4)
	(Missing)	1 (0.2)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	1 (0.1)
Duration of telephone assessment	< 10 minutes	222 (42.7)	43 (79.6)	37 (30.8)	0 (0.0)	9 (7.0)	8 (14.8)	56 (50.5)	375 (34.5)
	11-20 minutes	66 (12.7)	11 (20.4)	7 (5.8)	0 (0.0)	2 (1.6)	29 (53.7)	38 (34.2)	153 (14.1)
	21-30 minutes	1 (0.2)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	15 (27.8)	1 (0.9)	17 (1.6)
	>30 minutes	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	2 (3.7)	3 (2.7)	5 (0.5)
	(Missing)	231 (44.4)	0 (0.0)	76 (63.3)	100 (100.0)	118 (91.5)	0 (0.0)	13 (11.7)	538 (49.4)

*Question added after pilot phase in response to CEI group feedback, so not available for patients recruited in pilot phase.

5.4.3 Patterns of FALCON trial follow-up

An overview of the grouping of patients included in this study is shown in *Figure 5.4*. Of the 1209 patients that were contactable for telephone WHQ follow-up, 531 (47.5%) had a FALCON trial in-person follow-up.

The proportion of patients with in-person follow-up over the study period is shown in *Figure 5.5*. Having a telephone follow-up pathway (telephone FALCON trial follow-up and/or telephone WHQ) led to 52.5% (557/1088) additional patients with complete outcome assessment (estimated 'retention benefit') than in-person FALCON trial follow-up alone.

No adverse events were reported related either to completion of in-person FALCON trial follow-up, nor the telephone WHQ.

Figure 5.4 Study flowchart

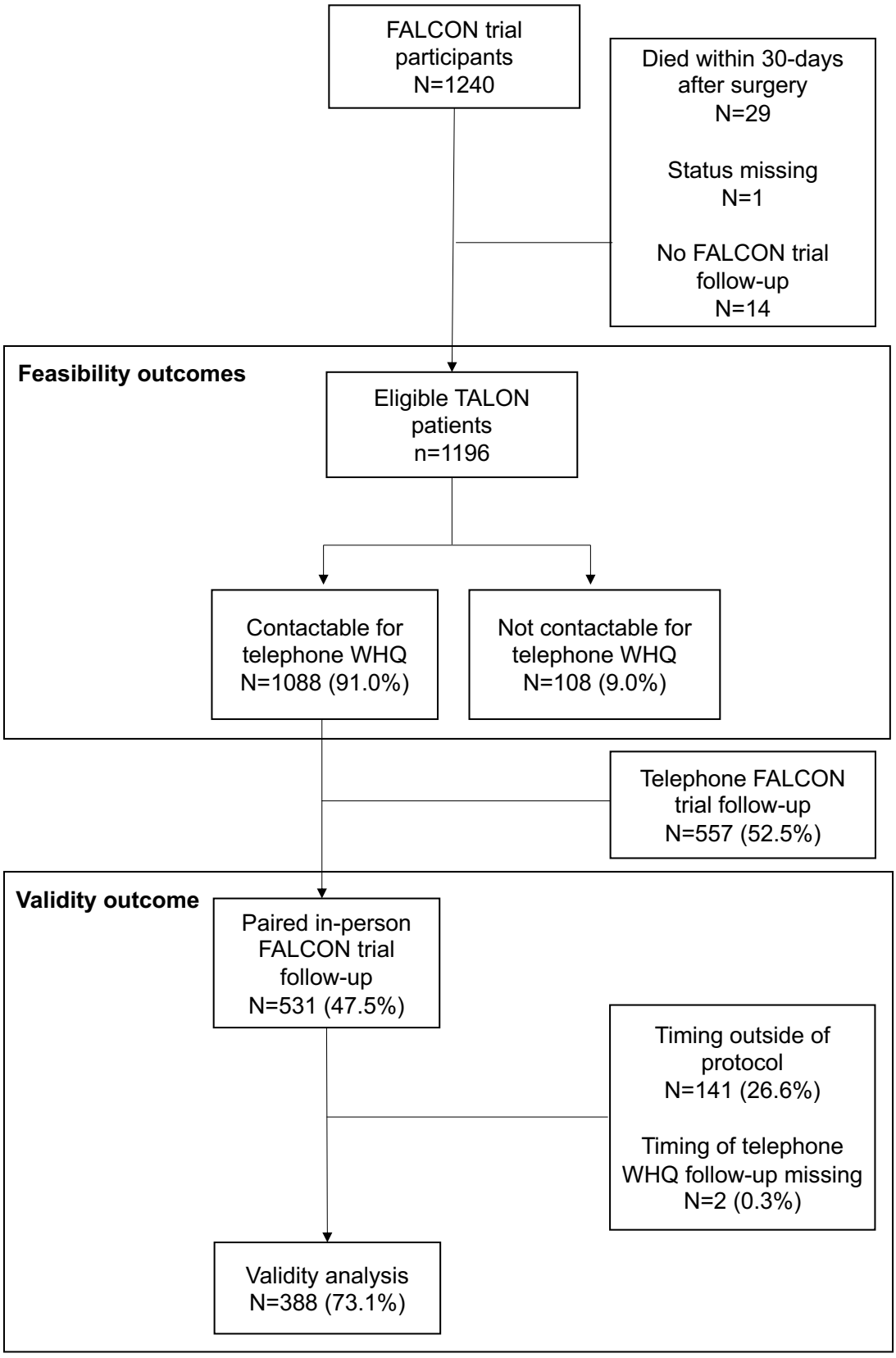
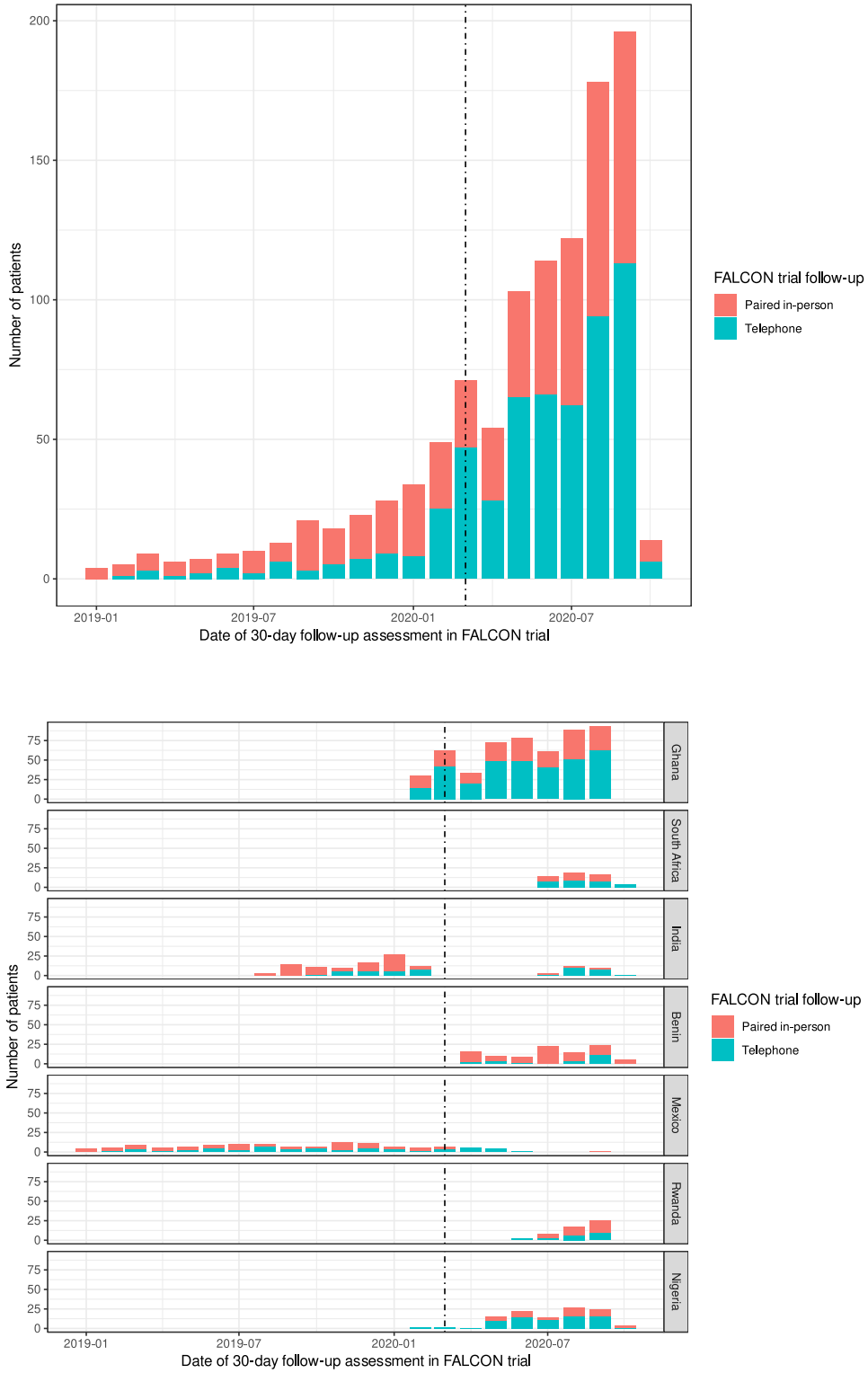


Figure 5.5 Proportion of patients with in-person FALCON trial versus telephone only follow-up (N=1088)



Grey line estimates the date of emergence of the first cases of SARS-CoV-2 outside of mainland china

5.4.3.1 Comparison of patients with in-person and telephone FALCON trial follow-up

There were some differences in the patients that had in-person FALCON trial follow-up, and telephone FALCON trial follow-up only or no trial follow-up (*Table 5.9*). Of note, there were fewer patients in rural settings (30.7% versus 39.0%, p<0.001) and fewer male patients (48.0% versus 58.7%, p<0.001), and more patients with obstetric indication (18.8% versus 5.9%, p<0.001), urogenital (22.0% versus 8.1%, p<0.001), clean-contaminated (36.7% versus 22.8%, p<0.001) and open non-midline (44.6% versus 24.1%, p<0.0001) surgery that returned for in-person versus telephone FALCON follow-up. However, patients from all participating countries and of a mix of baseline risk and operation type were included in both groups.

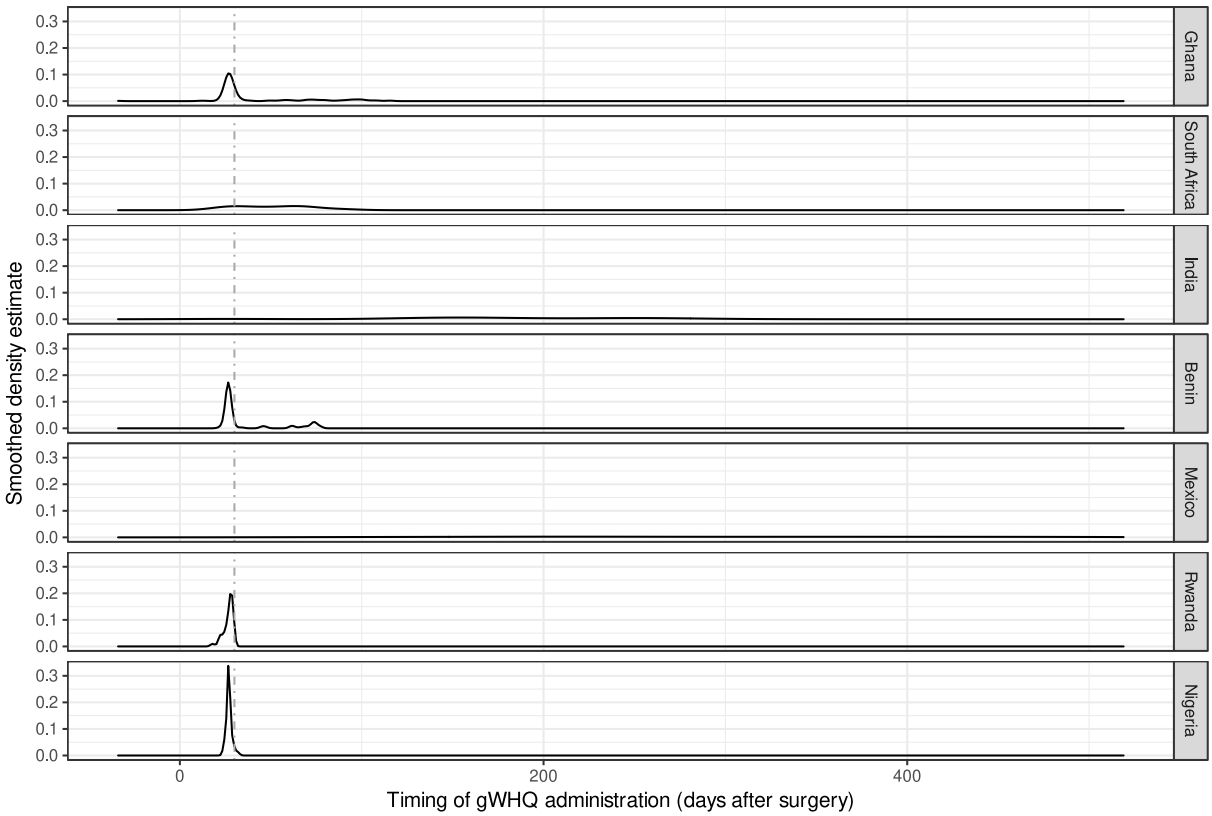
5.4.3.2 Timing of WHQ administration

Of the patients that had a FALCON trial in-person follow-up (N=531), 388 (73.1%) were conducted in a per-protocol time window, and 141 (26.6%) outside of protocol timing (*Figure 5.6*).

Table 5.9 Differences in baseline demographics between patients contactable by telephone that had in-person FALCON trial follow-up and telephone FALCON trial follow-up (N=1088).

Factor	Levels	In-person FALCON trial follow-up N=531	Telephone FALCON trial follow-up only N=557	Total N=1088	P-value
Country	Ghana	195 (36.7)	325 (58.3)	520 (47.8)	<0.001
	South Africa	25 (4.7)	29 (5.2)	54 (5.0)	
	India	76 (14.3)	44 (7.9)	120 (11.0)	
	Benin	80 (15.1)	20 (3.6)	100 (9.2)	
	Mexico	80 (15.1)	49 (8.8)	129 (11.9)	
	Rwanda	35 (6.6)	19 (3.4)	54 (5.0)	
	Nigeria	40 (7.5)	71 (12.7)	111 (10.2)	
Timing of telephone WHQ	Per protocol	388 (73.1)	490 (88.0)	878 (80.7)	<0.001
	Outside of protocol	141 (26.6)	64 (11.5)	205 (18.8)	
	(Missing)	2 (0.4)	3 (0.5)	5 (0.5)	
Home location	Urban	367 (69.1)	340 (61.0)	707 (65.0)	0.006
	Rural	163 (30.7)	217 (39.0)	380 (34.9)	
	(Missing)	1 (0.2)	0 (0.0)	1 (0.1)	
Age (years)	<18	42 (7.9)	48 (8.6)	90 (8.3)	0.27
	18-39	313 (58.9)	293 (52.6)	606 (55.7)	
	40-59	121 (22.8)	141 (25.3)	262 (24.1)	
	60-79	48 (9.0)	67 (12.0)	115 (10.6)	
	≥80	7 (1.3)	8 (1.4)	15 (1.4)	
Sex	Male	255 (48.0)	327 (58.7)	582 (53.5)	<0.001
	Female	276 (52.0)	230 (41.3)	506 (46.5)	
Level of education	Below high school level	247 (46.8)	310 (55.7)	557 (51.3)	0.004
	High school or above	281 (53.2)	247 (44.3)	528 (48.7)	
Known diabetes	Yes	16 (3.0)	17 (3.1)	33 (3.0)	1
	No	515 (97.0)	540 (96.9)	1055 (97.0)	
HIV status	Known negative	197 (37.1)	122 (21.9)	319 (29.3)	<0.001
	Known positive	13 (2.4)	8 (1.4)	21 (1.9)	
	Status not known	321 (60.5)	427 (76.7)	748 (68.8)	
Smoking status	Never smoked	482 (90.8)	514 (92.3)	996 (91.5)	0.431
	Ex-smoker	26 (4.9)	27 (4.8)	53 (4.9)	
	Current smoker	23 (4.3)	16 (2.9)	39 (3.6)	
Urgency	Elective (planned)	101 (19.0)	96 (17.2)	197 (18.1)	0.493
	Emergency (unplanned)	430 (81.0)	461 (82.8)	891 (81.9)	
Indication	Malignant disease	28 (5.3)	52 (9.3)	80 (7.4)	<0.001
	Benign disease	388 (73.1)	446 (80.1)	834 (76.7)	
	Trauma	15 (2.8)	25 (4.5)	40 (3.7)	
	Obstetric	100 (18.8)	33 (5.9)	133 (12.2)	
	(Missing)	0 (0.0)	1 (0.2)	1 (0.1)	
Operation location	Foregut	127 (23.9)	145 (26.0)	272 (25.0)	<0.001
	Hindgut	46 (8.7)	60 (10.8)	106 (9.7)	
	Appendix	161 (30.3)	163 (29.3)	324 (29.8)	
	Urogenital	117 (22.0)	45 (8.1)	162 (14.9)	
	Other	77 (14.5)	142 (25.5)	219 (20.1)	
	(Missing)	3 (0.6)	2 (0.4)	5 (0.5)	
ASA grade	Grade I	266 (50.1)	253 (45.4)	519 (47.7)	0.032
	Grade II	204 (38.4)	211 (37.9)	415 (38.1)	
	Grade III	54 (10.2)	88 (15.8)	142 (13.1)	
	Grade IV/V	7 (1.3)	4 (0.7)	11 (1.0)	
	(Missing)	0 (0.0)	1 (0.2)	1 (0.1)	
WHO Checklist	Yes	507 (95.5)	499 (89.6)	1006 (92.5)	<0.001
	No	24 (4.5)	58 (10.4)	82 (7.5)	
Operation grade	Intermediate/Minor	166 (32.0)	184 (33.9)	350 (33.0)	0.568
	Major	352 (68.0)	359 (66.1)	711 (67.0)	
Contamination	Clean/Clean-contaminated	195 (36.7)	127 (22.8)	322 (29.6)	<0.001
	Contaminated	166 (31.3)	233 (41.8)	399 (36.7)	
	Dirty	170 (32.0)	195 (35.0)	365 (33.5)	
	(Missing)	0 (0.0)	2 (0.4)	2 (0.2)	
Approach	Open midline	289 (54.4)	418 (75.0)	707 (65.0)	<0.001
	Open non-midline	237 (44.6)	134 (24.1)	371 (34.1)	
	Laparoscopic attempted	5 (0.9)	4 (0.7)	9 (0.8)	
	(Missing)	0 (0.0)	1 (0.2)	1 (0.1)	
Stoma formation	Yes	23 (4.3)	32 (5.7)	55 (5.1)	0.355
	No	504 (94.9)	521 (93.5)	1025 (94.2)	
	(Missing)	4 (0.8)	4 (0.7)	8 (0.7)	

Figure 5.6 Timing of telephone WHQ administration



Grey line denotes the timing of in-person FALCON follow-up (30-days after surgery).

5.4.4 Data missingness

The level of data missingness overall for all item responses was low 0.1% (13/10089) and similarly for each individual item (range: 0.0-0.1%), so complete case analysis was conducted without imputation.

5.4.5 Diagnostic accuracy

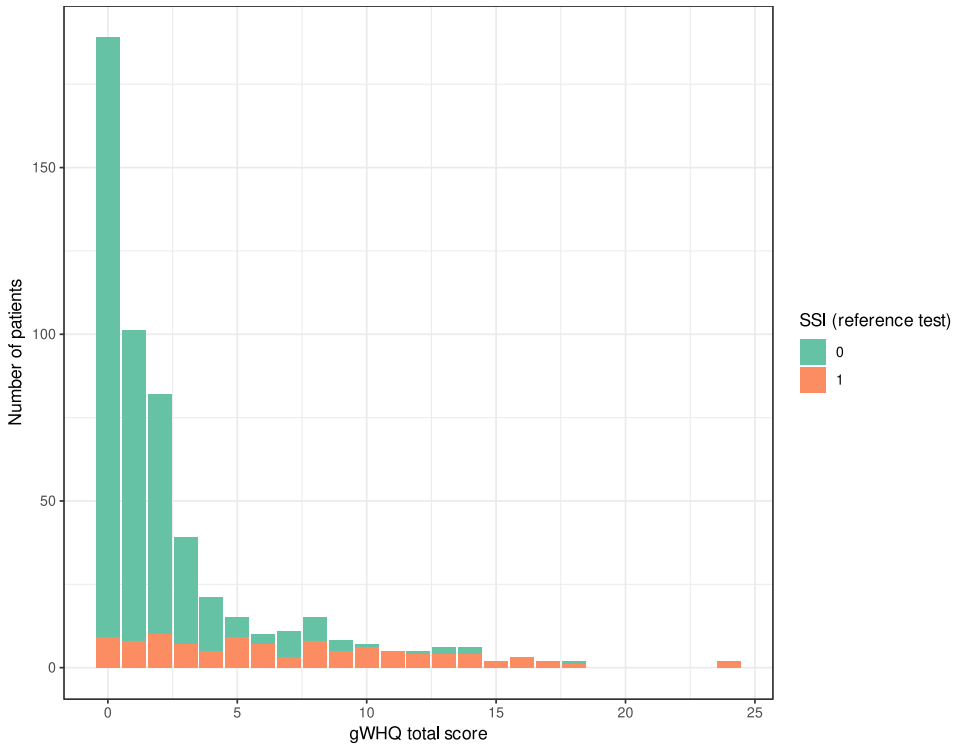
5.4.5.1 Comparison of WHQ scores and reference test

Patients' total WHQ scores in those with and without a diagnosis of SSI made at the FALCON trial assessment 30-days after surgery is presented in *Figure 5.7* and *Table 5.10*. The proportion of patients with SSI at each WHQ point score interval is presented in *Figure 5.8.* As the WHQ point score increased, so the proportion of patients with a reference test diagnosis of 'yes SSI' increased.

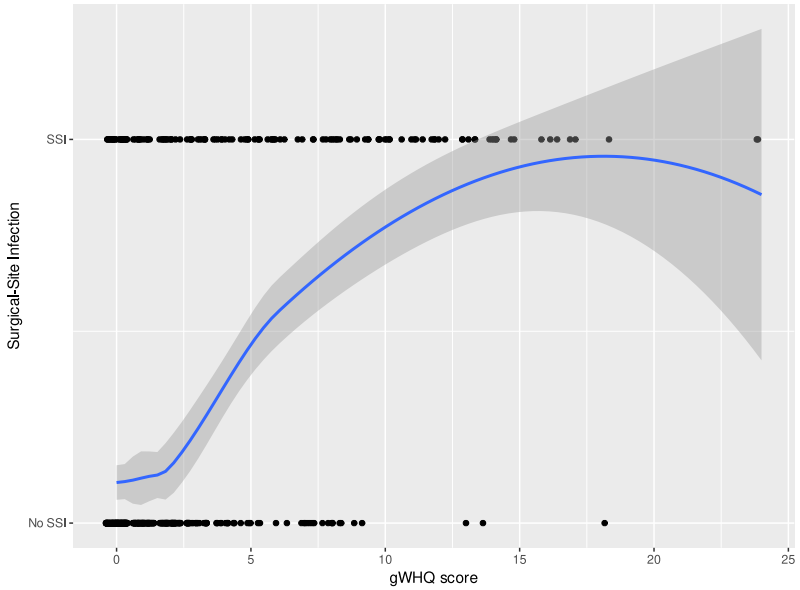
5.4.5.2 Exploration of patients with a WHQ score of zero and reference test 'no SSI'

In patients with a WHQ point score of zero (i.e., did not report any symptoms of SSI over the telephone, N=147) that did go onto have an SSI diagnosis made on 30-day follow-up (N=7), the features that were most commonly detected in person were purulent fluid (6/7), pain at the wound site (6/7) and diagnosis of SSI by a clinician or on imaging (6/7, *Table 5.11*).

Figure 5.7 (Panel) Wound Healing Questionnaire score versus reference test diagnosis of SSI (N=388)



Graph represents per-protocol analysis (N=388)



Spline curve represents proportion of patients with SSI at each WHQ total score value, estimated with generalised additive modelling

Table 5.10 Cross-tabulation of patients WHQ score and whether or not they received a diagnosis of SSI at the in-person assessment 30-days after surgery

(A) Per-protocol analysis (N=388)

	Global Wound Healing Questionnaire score (index test)																							
Reference test	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**	**21**	**22**	**23**	**24**
No SSI	140	62	44	19	11	2	2	2	5	2	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0
Yes SSI	7	8	8	6	5	9	7	3	8	5	4	5	4	4	4	2	3	2	1	0	0	0	0	0	2

(B) Sensitivity analysis including out of protocol (N=531)

	Global Wound Healing Questionnaire score (index test)																							
Reference test	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**	**21**	**22**	**23**	**24**
No SSI	180	93	72	32	16	6	3	8	7	3	1	0	1	2	2	0	0	0	1	0	0	0	0	0	0
Yes SSI	9	8	10	7	5	9	7	3	8	5	6	5	4	4	4	2	3	2	1	0	0	0	0	0	2

Figure 5.8 Proportion of patients with SSI diagnosis in reference test at each WHQ point score level (per-protocol analysis)

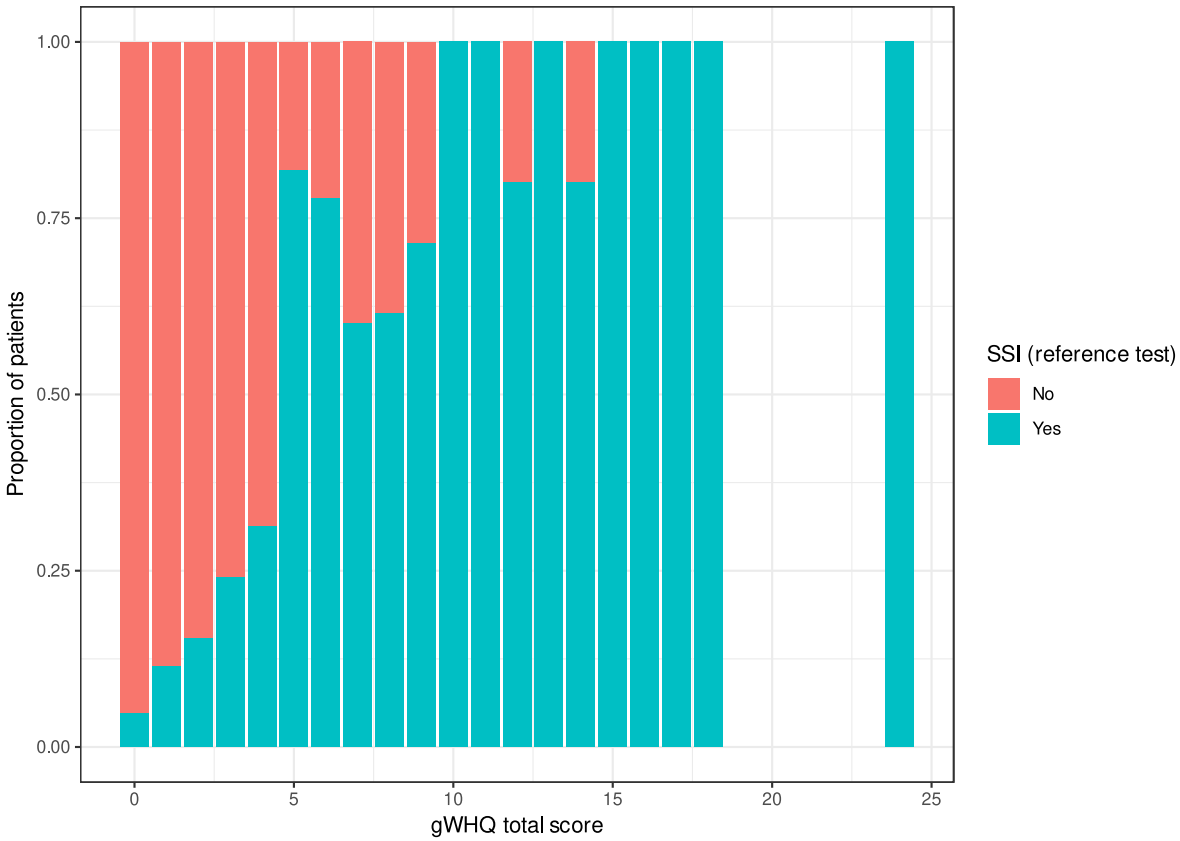


Table 5.11 Misclassification in patients with a WHQ score of zero (N=147)

		Reference test result		
Component of in-person FALCON trial follow-up	**Response levels**	**No SSI N=140**	**Yes SSI N=7**	**Total**
Was there redness of the wound?	Yes	0 (0.0)	***2 (28.6)***	2 (1.4)
	No	140 (100.0)	5 (71.4)	145 (98.6)
Did patient have heat of the wound?	Yes	0 (0.0)	***3 (42.9)***	3 (2.0)
	No	140 (100.0)	4 (57.1)	144 (98.0)
Did patient have purulent drainage from the wound?	Yes	0 (0.0)	***6 (85.7)***	6 (4.1)
	No	140 (100.0)	1 (14.3)	141 (95.9)
Was abdominal wound opening present (spontaneously opened or by clinician)?	Yes	1 (0.7)	***5 (71.4)***	6 (4.1)
	No	139 (99.3)	2 (28.6)	141 (95.9)
Was there localised swelling around the wound?	Yes	0 (0.0)	***4 (57.1)***	4 (2.7)
	No	140 (100.0)	3 (42.9)	143 (97.3)
Was there pain or tenderness at the wound?	Yes	12 (8.6)	***6 (85.7)***	18 (12.2)
	No	128 (91.4)	1 (14.3)	129 (87.8)
Did patient have systemic fever (greater than 38 degrees Celsius)?	Yes	0 (0.0)	***2 (28.6)***	2 (1.4)
	No	140 (100.0)	5 (71.4)	145 (98.6)
Has patient been re-admitted?	Yes, not SSI related	2 (1.4)	*1 (14.3)*	3 (2.0)
	Yes, for treatment of SSI	0 (0.0)	0 (0.0)	0 (0.0)
	No	138 (98.6)	6 (85.7)	144 (98.0)
Has patient been re-operated on?	Yes, not SSI related	1 (0.7)	*1 (14.3)*	2 (1.4)
	Yes, for treatment of SSI	0 (0.0)	0 (0.0)	0 (0.0)
	No	139 (99.3)	6 (85.7)	145 (98.6)
Was SSI diagnosed by clinician or on imaging?	Yes	0 (0.0)	***6 (85.7)***	6 (4.1)
	No	140 (100.0)	1 (14.3)	141 (95.9)

5.4.5.3 Diagnostic test accuracy statistics

A summary of the performance metrics and diagnostic test accuracy statistics is shown in in the *Tables 5.12* and *5.13* and *Figure 5.9* below. In the per-protocol analysis (N=388), the WHQ demonstrated excellent overall discrimination (AUROC 0.869, 95% CI 0.824-0.914). The cut-point identified using Youden's index was 3.5 (WHQ total score ≥4), which diagnosed post-discharge SSI with sensitivity of 0.701 (0.610-0.792), specificity of 0.911 (0.878-0.9430), positive predictive value of 0.723 (0.633-0.814) and negative predictive value of 0.901 (0.867-0.935). Diagnostic test accuracy statistics at different cut-points 'to rule in' or 'rule out' SSI are presented *Table 5.14*.

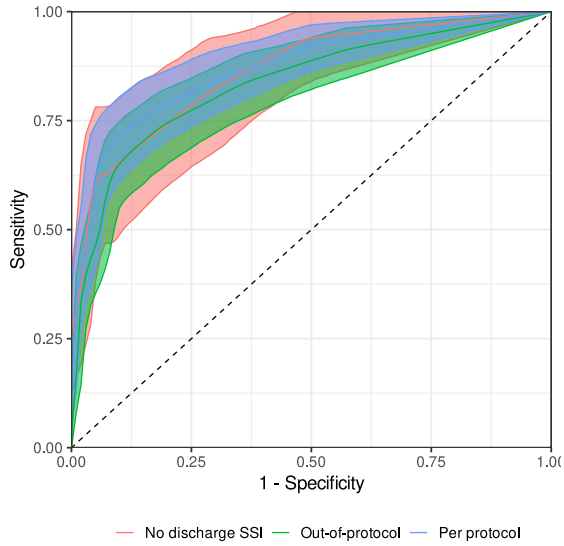
5.4.6 Sensitivity analyses of primary outcome

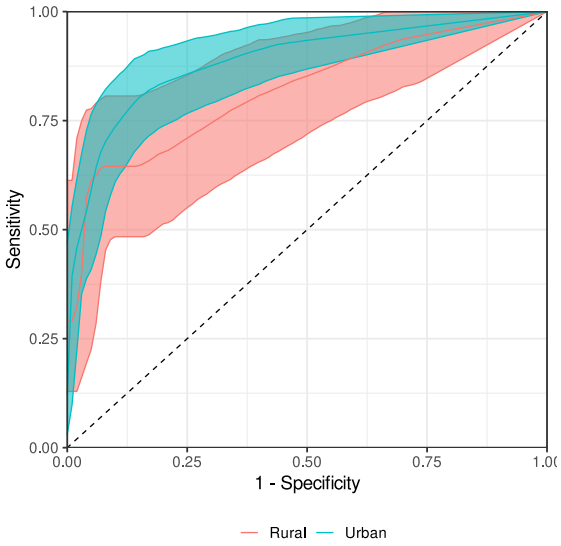
The discrimination was similar in sensitivity analyses including out-of-protocol patients (N=531, AUROC 0.836, 95% CI 0.788-0.883), including post-discharge SSI only (N=300, AUROC 0.863, 95% CI 0.790-0.937) and with inverse probability weighting (AUROC 0.866, 95% CI 0.805-0.927).

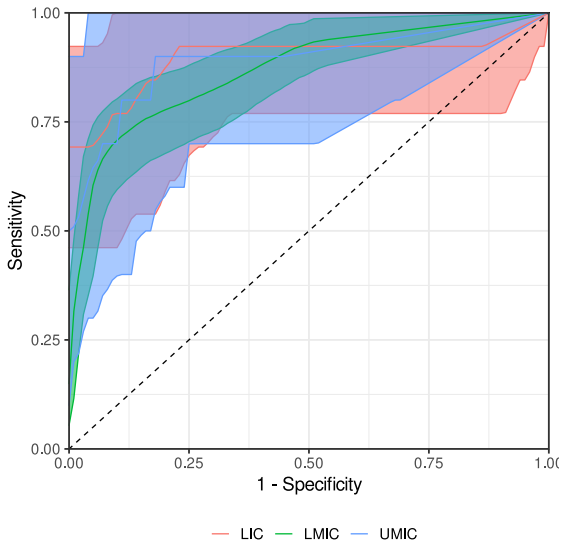
5.4.7 Subgroup analyses of primary outcome

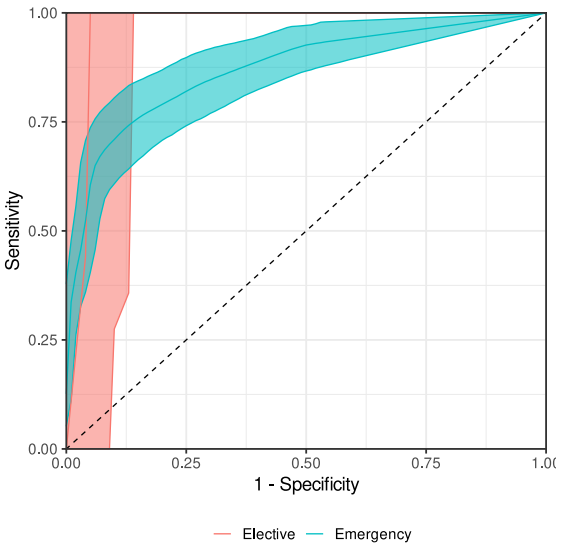
The performance of the WHQ was maintained across key subgroups (*Figure 5.9*). Some differences were observed in reduced overall discrimination in rural (AUROC 0.818, 0.721-0.914) versus urban population (AUROC 0.886, 0.836-0.937) and poorer discrimination after emergency (AUROC 0.871, 0.826-0.916) versus elective surgery (AUROC 0.966, 0.895-1.000), although the 95% confidence intervals overlapped for both comparisons and interpretation of the analysis of elective surgery was limited by a low SSI rate in the elective surgery subgroup.

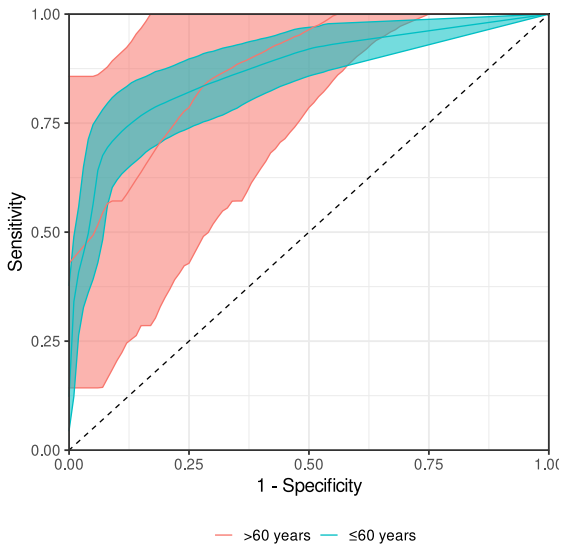
Figure 5.9 (Panel) Receiver operating characteristic curves for the Wound Healing Questionnaire in detecting surgical site infection up to 30-days after surgery











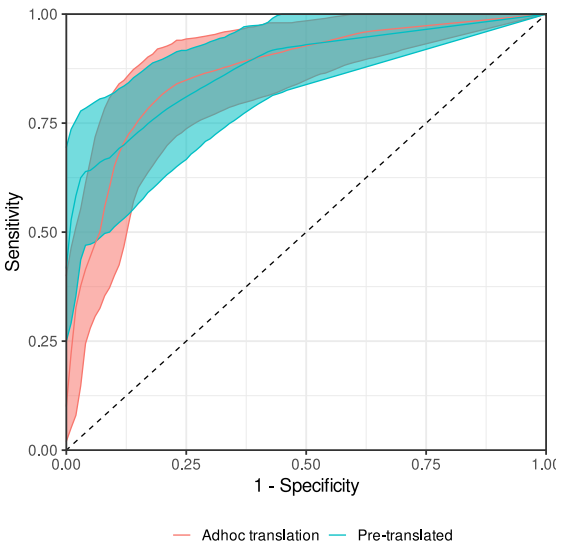


Table 5.12 Summary of diagnostic test accuracy characteristics overall and across subgroups (1)

Patient group	N=	SSI prevalence[$] N=, (%)	AUROC	Accuracy	Sensitivity	Specificity	WHQ score cut-off[€]
Overall*	**388**	**97 (25.0%)**	**0.869 (0.824-0.914)**	**0.858 (0.858-0.859)**	**0.701 (0.610-0.792)**	**0.911 (0.878-0.9430**	**3.5**
Sensitivity analyses							
Including out-of-protocol	531	104 (19.6%)	0.836 (0.788-0.883)	0.842 (0.841-0.842)	0.673 (0.583-0.763)	0.883 (0.852-0.913)	3.5
No discharge SSI	300	32 (10.1%)	0.863 (0.790-0.937)	0.923 (0.923-0.924)	0.625 (0.457-0.793)	0.959 (0.935-0.983)	4.5
Subgroup analyses Patient home location							
Urban	266	66 (24.8%)	0.886 (0.836-0.937)	0.838 (0.837-0.839)	0.818 (0.725-0.911)	0.845 (0.795-0.895)	2.5
Rural	122	31 (25.4%)	0.818 (0.721-0.914)	0.877 (0.875-0.879)	0.613 (0.441-0.784)	0.967 (0.930-1.004)	4.5
Country income							
Upper-middle	39	10 (25.6%)	0.888 (0.741-1.000)	0.846 (0.840-0.853)	0.900 (0.714-1.086)	0.828 (0.690-0.965)	2.5
Lower-middle	314	74 (23.6%)	0.868 (0.817-0.918)	0.866 (0.866-0.867)	0.689 (0.584-0.795)	0.921 (0.887-0.955)	3.5
Low	35	13 (37.1%)	0.892 (0.748-1.000)	0.829 (0.837-0.837)	0.923 (0.778-1.068)	0.773 (0.598-0.948)	1.5
Patient age group							
>60 years	43	7 (16.3%)	0.867 (0.729-1.000)	0.721 (0.712-0.730)	0.857 (0.589-1.116)	0.694 (0.544-0.845)	1.5
≤60 years	345	90 (26.1%)	0.869 (0.821-0.916)	0.852 (0.851-0.853)	0.722 (0.630-0.815)	0.898 (0.861-0.935)	3.5
Urgency of surgery							
Emergency	364	95 (26.1%)	0.871 (0.826-0.916)	0.830 (0.829-0.830)	0.758 (0.672-0.844)	0.855 (0.813-0.897)	2.5
Elective	21	2 (8.3%)	0.966 (0.895-1.000)	0.958 (0.955-0.962)	1.000 (1.000-1.000)	0.955 (0.868-1.042)	4.5
Questionnaire translation							
Formal translation	184	36 (19.6%)	0.875 (0.803-0.946)	0.913 (0.912-0.914)	0.611 (0.452-0.770)	0.986 (0.968-1.005)	4.5
Ad hoc translation[£]	178	50 (28.1%)	0.866 (0.805-0.927)	0.798 (0.796-0.796)	0.840 (0.738-0.942)	0.781 (0.710-0.853)	2.5
Severity of SSI							
No re-operation (mild SSI only)	360	73 (20.3%)	0.855 (0.801-0.908)	0.822 (0.821-0.823)	0.726 (0.624-0.828)	0.847 (0.805-0.888)	2.5

SSI = Surgical Site Infection. AUROC = Area under the receiver operating characteristic curve, used as an overall measure of discrimination. *Overall analysis included only patients with per-protocol WHQ administration. [£]Includes Ad hoc, translated from English by questionnaire administrator and Ad hoc, translated from English with formal translator. [$]events = Surgical site infection recorded using reference test of 30-day in-person FALCON trial follow-up. [€]Cut-off scores defined using Youden's index, in which I maximize the sum of sensitivity and specificity in the cohort of interest. Implementation of the WHQ should be supported by clinical decision making using cut-point scores in *Table.3.14*

Table 5.13 Summary of diagnostic test accuracy characteristics overall and across subgroups

Patient group	N=	PLR	NLR	PPV	NPV	WHQ score cut-off[€]
Overall*	**388**	**7.846 (5.317-11.579)**	**0.328 (0.242-0.446)**	**0.723 (0.633-0.814)**	**0.901 (0.867-0.935)**	**3.5**
Sensitivity analyses						
Including out-of-protocol	531	5.748 (4.289-7.704)	0.370 (0.280-0.489)	0.583 (0.495-0.672)	0.917 (0.891-0.944)	3.5
No discharge SSI	300	15.227 (8.046-28.818)	0.391 (0.250-0.612)	0.645 (0.477-0.814)	0.955 (0.931-0.980)	4.5
Subgroup analyses						
Patient home location						
Urban	266	5.279 (3.746-7.438)	0.215 (0.129-0.360)	0.635 (0.533-0.738)	0.934 (0.897-0.970)	2.5
Rural	122	18.591 (5.902-58.564)	0.400 (0.257-0.624)	0.864 (0.720-1.007)	0.880 (0.816-0.944)	4.5
Country income						
Upper-middle	39	5.220 (2.290-11.896)	0.121 (0.019-0.782)	0.643 (0.392-0.894)	0.960 (0.883-1.037)	2.5
Lower-middle	314	8.706 (5.508-13.760)	0.338 (0.240-0.475)	0.729 (0.624-0.833)	0.906 (0.869-0.942)	3.5
Low	35	4.062 (1.850-8.916)	0.100 (0.015-0.663)	0.706 (0.489-0.922)	0.944 (0.839-1.050)	1.5
Patient age group						
>60 years	43	2.805 (1.574-5.000)	0.206 (0.033-1.279)	0.353 (0.126-0.580)	0.962 (0.888-1.035)	1.5
≤60 years	345	7.083 (4.814-10.422)	0.309 (0.221-0.433)	0.714 (0.621-0.807)	0.902 (0.865-0.938)	3.5
Urgency of surgery						
Emergency	364	5.228 (3.828-7.139)	9.283 (0.198-0.406)	0.649 (0.560-0.737)	0.909 (0.874-0.945)	2.5
Elective	21	22.000 (3.242-149.303)	0.000 (0.000-0.000)	0.667 (0.133-1.200)	1.000 (1.000-1.000)	4.5
Questionnaire translation						
Formal translation	184	45.222 (11.141-183.565)	0.394 (0.262-0.594)	0.917 (0.806-1.027)	0.912 (0.869-0.956)	4.5
Ad hoc translation[£]	178	3.840 (2.709-5.444)	0.205 (0.108-0.389)	0.600 (0.485-0.715)	0.926 (0.877-0.975)	2.5
Severity of SSI						
No re-operation (mild SSI only)	360	4.736 (3.486-6.433)	0.324 (0.222-0.472)	0.546 (0.447-0.645)	0.924 (0.892-0.956)	2.5

SSI = Surgical Site Infection. AUROC = Area under the receiver operating characteristic curve, used as an overall measure of discrimination. *Overall analysis included only patients with per-protocol WHQ administration. [£]Includes Ad hoc, translated from English by questionnaire administrator and Ad hoc, translated from English with formal translator. [$]events = Surgical site infection recorded using reference test of 30-day in-person FALCON trial follow-up. [€]Cut-off scores defined using Youden's index, in which I maximize the sum of sensitivity and specificity in the cohort of interest. Implementation of the WHQ should be supported by clinical decision making using cut-point scores in *Table.3.14.*

Table 5.14 Diagnostic accuracy of the WHQ score across different cut-points to 'rule in' or 'rule out' SSI

WHQ score cut-off	True prevalence*	Diagnostic accuracy	Sensitivity	Specificity
'Rule in' SSI				
≥1	0.621 (0.571-0.670)	0.593 (0.542-0.642)	0.373 (0.312-0.438)	0.952 (0.904-0.981)
≥2	0.441 (0.391-0.492)	0.732 (0.685-0.775)	0.480 (0.403-0.557)	0.931 (0.889-0.961)
≥3	0.307 (0.261-0.355)	0.825 (0.783-0.861)	0.622 (0.528-0.709)	0.914 (0.874-0.945)
≥4	0.242 (0.200-0.288)	0.858 (0.820-0.891)	0.723 (0.622-0.811)	0.901 (0.861-0.933)
≥5	0.201 (0.162-0.244)	0.874 (0.836-0.905)	0.808 (0.703-0.888)	0.890 (0.850-0.923)
≥6	0.173 (0.136-0.214)	0.856 (0.817-0.889)	0.806 (0.691-0.892)	0.866 (0.824-0.901)
≥7	0.149 (0.116-0.189)	0.843 (0.803-0.878)	0.810 (0.686-0.901)	0.848 (0.805-0.885)
≥8	0.137 (0.104-0.175)	0.840 (0.800-0.875)	0.830 (0.702-0.919)	0.842 (0.798-0.879)
≥9	0.103 (0.075-0.138)	0.832 (0.792-0.868)	0.900 (0.763-0.972)	0.825 (0.781-0.863)
≥10	0.085 (0.059-0.117)	0.825 (0.783-0.861)	0.939 (0.798-0.993)	0.814 (0.770-0.853)
'Rule out' SSI				
≤1	0.559 (0.508-0.609)	0.268 (0.225-0.315)	0.069 (0.039-0.111)	0.520 (0.443-0.597)
≤2	0.693 (0.645-0.739)	0.175 (0.139-0.217)	0.086 (0.055-0.126)	0.378 (0.291-0.472)
≤3	0.758 (0.712-0.800)	0.142 (0.109-0.180)	0.099 (0.067-0.139)	0.277 (0.189-0.378)
≤4	0.799 (0.756-0.838)	0.126 (0.095-0.164)	0.110 (0.077-0.150)	0.192 (0.112-0.297)
≤5	0.827 (0.786-0.864)	0.144 (0.111-0.183)	0.134 (0.099-0.176)	0.194 (0.108-0.309)
≤6	0.851 (0.811-0.884)	0.157 (0.122-0.197)	0.152 (0.115-0.195)	0.190 (0.099-0.314)
≤7	0.863 (0.825-0.896)	0.160 (0.125-0.200)	0.158 (0.121-0.202)	0.170 (0.081-0.298)
≤8	0.897 (0.862-0.925)	0.168 (0.132-0.208)	0.175 (0.137-0.219)	0.100 (0.028-0.237)
≤9	0.915 (0.883-0.941)	0.175 (0.139-0.217)	0.186 (0.147-0.230)	0.061 (0.007-0.202)
≤10	0.925 (0.894-0.949)	0.186 (0.148-0.228)	0.195 (0.155-0.240)	0.069 (0.008-0.228)

WHQ score cut-off	Positive likelihood ratio	Negative likelihood ratio	Positive predictive value (PPV)	Negative predictive value (NPV)
'Rule in' SSI				
≥1	7.842 (3.737-16.457)	0.658 (0.593-0.730)	0.928 (0.857-0.970)	0.481 (0.422-0.540)
≥2	6.937 (4.155-11.583)	0.559 (0.482-0.649)	0.845 (0.758-0.911)	0.694 (0.638-0.747)
≥3	7.273 (4.802-11.016)	0.414 (0.327-0.522)	0.763 (0.666-0.843)	0.845 (0.799-0.885)
≥4	7.334 (5.079-10.590)	0.307 (0.221-0.426)	0.701 (0.600-0.790)	0.911 (0.872-0.941)
≥5	7.364 (5.267-10.296)	0.216 (0.137-0.341)	0.649 (0.546-0.744)	0.948 (0.916-0.971)
≥6	6.017 (4.449-8.137)	0.224 (0.137-0.366)	0.557 (0.452-0.658)	0.955 (0.925-0.976)
≥7	5.348 (4.026-7.105)	0.224 (0.131-0.381)	0.485 (0.382-0.588)	0.962 (0.933-0.981)
≥8	5.247 (3.984-6.911)	0.202 (0.111-0.366)	0.454 (0.352-0.558)	0.969 (0.942-0.986)
≥9	5.134 (3.998-6.594)	0.121 (0.048-0.308)	0.371 (0.275-0.475)	0.986 (0.965-0.996)
≥10	5.053 (3.997-6.387)	0.074 (0.019-0.286)	0.320 (0.229-0.422)	0.993 (0.975-0.999)
'Rule out' SSI				
≤1	0.144 (0.086-0.241)	1.789 (1.542-2.075)	0.155 (0.089-0.242)	0.306 (0.253-0.362)
≤2	0.137 (0.091-0.208)	2.418 (1.915-3.054)	0.237 (0.157-0.334)	0.155 (0.115-0.201)
≤3	0.136 (0.094-0.197)	3.259 (2.345-4.529)	0.299 (0.210-0.400)	0.089 (0.059-0.128)
≤4	0.136 (0.097-0.190)	4.630 (2.933-7.308)	0.351 (0.256-0.454)	0.052 (0.029-0.084)
≤5	0.166 (0.123-0.225)	4.463 (2.735-7.285)	0.443 (0.342-0.548)	0.045 (0.024-0.075)
≤6	0.187 (0.141-0.248)	4.474 (2.623-7.631)	0.515 (0.412-0.618)	0.038 (0.019-0.067)
≤7	0.191 (0.145-0.251)	4.957 (2.729-9.006)	0.546 (0.442-0.648)	0.031 (0.014-0.058)
≤8	0.195 (0.152-0.250)	8.247 (3.251-20.922)	0.629 (0.525-0.725)	0.014 (0.004-0.035)
≤9	0.198 (0.157-0.250)	13.432 (3.503-51.513)	0.680 (0.578-0.771)	0.007 (0.001-0.025)
≤10	0.209 (0.166-0.264)	11.673 (3.062-44.500)	0.722 (0.621-0.808)	0.007 (0.001-0.025)

*Proportion of patients that would be classified as having an SSI event with this cut-point. All estimates provided with 95% confidence intervals in brackets.

5.4.8 Community Engagement and Involvement

Patients had a direct impact on study delivery and reporting. Firstly, variables related to acceptability, the number of attempts needed, and time taken were added to the telephone WHQ pathway item set in response to pilot testing, and early exploration of the data during study monitoring. Secondly, several suggestions were provided to iteratively improve the implementation of telephone WHQ administration. To summarise this shared learning, a toolkit was co-produced and provided to sites to share best practice for acceptable and inclusive delivery of a telephone follow-up pathway (*Figure 5.10*). This was presented as a slide presentation (Microsoft Powerpoint®, Microsoft Corporation, Redmond, WA) and infographic poster (Adobe Illustrator®, Adobe, San Jose, CA). Finally, I added a subgroup analysis for mild SSI only, due to concerns that patients with less severe problems may be missed and so delay receiving care.

Figure 5.10 Co-produced toolkit for optimising postoperative telephone follow-up

Download the full toolkit presentation at: https://bit.ly/TALONtips

5.5 Discussion

5.5.1 Summary of key findings

This prospective validation study within a large international pragmatic trial demonstrated high feasibility and validity of telephone assessment for diagnosis of surgical site infection in low-resource environments using the adapted WHQ. The WHQ was demonstrated to be suitable for use across a diverse range of settings, countries, and languages in three continents with high completion and low missing data rates. The diagnostic accuracy of the WHQ score was good when delivered per-protocol and was robust to several sensitivity analyses. However, it was less discriminative in certain subgroups, such as patients living in rural areas. Several cut-points of the WHQ score and their corresponding diagnostic accuracy statistics were presented to facilitate application of the WHQ to different contexts. Co-production of the telephone WHQ administration pathway facilitated cultural and contextually attuned delivery. This tool is now available for global implementation in postoperative surveillance pathways and to optimise efficient trial design and conduct.

5.5.2 Findings in context

5.5.2.1 Comparison to diagnostic accuracy in comparable studies

Few existing high-quality studies have evaluated the diagnostic accuracy of telemedicine methods for remote diagnosis of SSI. A prospective cohort study published in 2022 raised a significant concern for under-detection of SSI using unstandardised methods. On meta-analysis, only four studies were identified with paired in-person and telephone follow-up for which diagnostic test accuracy statistics could be calculated [195]. Three studies were at high risk of bias, and just one, the UK validation of the English language WHQ, was identified as being at low risk of bias [192]. Therefore, this instrument was chosen to update and adapt to use in this international study. These international data therefore play an important role in informing safe upscaling of methods for remote postoperative surveillance. Differences in

performance for patients living in rural versus urban settings may reflect differences in items related to the treatment pathway for wound infection (e.g., seeking advice for a wound problem, readmission to hospital) and patients access to care in rural environments. Whilst the sensitivity may be marginally reduced, remote follow-up methods may improve reach into these communities, improve diversity and representation, and reduce attrition bias.

5.5.2.2 Application of the adapted WHQ

There are several ways in which this WHQ instrument may be applied. First, it may be used in research studies to provide a diagnosis of SSI (i.e., binary outcome of SSI / no SSI) remotely, without the need for in-person review. Choice of cut-off SSI threshold score here would need to consider a balance of sensitivity and specificity, and the consequences of missing or over diagnosing SSI. This has important implications for trial design and conduct. Trials in SSI need to be large and pragmatic, and a validated remote method for assessing SSI will reduce trials costs. Second, the WHQ may be used in clinical practice to triage patients into existing clinical care pathways i.e., with those at very low risk of a SSI diagnosis given reassurance, and those with a moderate or high risk being asked to return for outpatient assessment. This could be adopted by either primary or secondary care depending on the structure of the local health system. Other work in this area has suggested that triage using remote, digital methods is safe, feasible and has cost-savings [131]. Combining remote tools to detect SSI and other common postoperative complications could be an accessible and rapid step towards the digital future of surgery. Provision of large data sets with accurate SSI assessment will inform future data synthesis and decision-making.

5.5.2.3 Loss to telephone contact and risk of attrition bias

This study confirms that digital follow-up pathways in low-resource environments are feasible, and resilient. This supports estimates of high access to mobile communications by

the World Bank [129]. By moving to remote, telephone assessment over 50% additional patients were able to be followed-up that may otherwise have been lost to follow-up, substantially improving trial retention [23, 128, 245, 246]. Intuitively, the time from surgery to attempted follow-up was strongly associated with the likelihood of successful contact. Certain groups were highlighted to be more challenging to reach. Patients with non-midline incisions may represent patients undergoing appendicectomy or cholecystectomy, who are likely to return to work soon after their operation and have limited physical opportunity and/or reflective motivation to complete follow-up [247]. An association between SSI diagnosis and reduced odds of successful contact highlights a potential risk for attrition bias.

Loss to follow-up in randomised trials is rarely random [248, 249]. Attrition is often more likely in patients at highest or lowest risk of a postoperative event [119]. For example, patients with a very severe SSI in the community may be too unwell to return to hospital or receive a phone call, or may have been admitted for care in another hospital and remain uncontactable to telephone follow-up. Patients that have returned to their baseline functional states may be pre-occupied with work or caring responsibilities and so may have competing pressures or down-prioritise continuing research involvement. To safely implement postoperative telesurveillance for SSI in clinical practice and randomised trials, further work is required to explore contextually important reasons for dropout and discontinuation in surgical RCTs [122, 240, 250]. Specific efforts to improve retention in these groups should be co-developed with CEI partners [86].

5.5.2.4 Task shifting to build surgical capacity

Postoperative surveillance is burdensome in high-volume, low-resource settings, both for patients and health systems. Remote follow-up is likely to substantially reduce direct and indirect costs (e.g., time out of work, informal caregivers), who may already be at risk of

catastrophic expenditure because of their surgical episode [9]. Task shifting of wound assessment to more junior or non-clinical staff is likely to significantly improve efficiency and reduce the ‘footprint’ of research studies on local systems. Here, the WHQ was largely delivered by non-expert assessors, both reducing the opportunity cost to the limited surgical workforce and building capacity in research skills and wound evaluation.

5.5.2.5 Video follow-up and future development

Video and photographic assessment of the healing surgical wound is a promising area of innovation that was not evaluated in this study [132, 251]. Assessment using the telephone WHQ was less accurate for ‘mild SSI’ (i.e., not needing reoperation) in a subgroup analysis, and signs such as purulent fluid, wound opening and greater than expected pain on palpation were sometimes missed by the WHQ in an exploratory analysis. The evidence base for adoption of this ‘enhanced’ remote assessment remains scarce, but it has been widely adopted during the SARS-CoV-2 pandemic [125, 252]. Our data shows promise for the feasibility of video and photo assessment in low resource contexts with 64.2% of patients having access to a camera phone (range by country: 11.1% in Rwanda to 80.6% in Mexico). Urgent evidence is required to better understand the safety and potential limitations of this practice.

5.5.3 Limitations

This research was delivered in accordance with a pre-published protocol, integrated into the platform of a randomised trial and in accordance with best-practice guidelines. It represents high-quality evidence to support implementation of postoperative tele-surveillance. However, it has several limitations.

5.5.3.1 Design limitations

It is assumed that the reference test of in-person assessment could correctly detect when a wound infection had or had not occurred. Whilst in the FALCON trial there was a minimum training requirement for those involved in wound evaluation, false positives or false negatives at in-person review would affect the estimates of diagnostic accuracy upon administration of the WHQ. Second, there was a theoretical risk of patients developing a new SSI between their WHQ completion and 30-day follow-up. This is clinically unlikely and not supported by existing cohort data [23]. Third, despite a careful quality assurance and training process, I did not have repeated measures to evaluate inter-rater or intrarater reliability. Fourth, acceptability of telephone follow-up was assessed at the end of telephone WHQ and not anonymised, so was at risk of social acceptability bias.

5.5.3.2 Delivery limitations

First, the WHQ was commonly performed with ad hoc translation by the questionnaire administrator. This may have decreased both the reproducibility and accuracy of the instrument, but reflected the diverse, real-world setting of delivery, and no significant difference was seen in discrimination when translation was performed ad hoc versus with a pre-translated questionnaire. Second, I was also underpowered to explore differences in accuracy between countries or languages. Third, the study excluded patients who died before 30-days (N=29), representing a competing risk when interpreting the generalisability of the results to a highest-risk group of patients. Fourth, there was a risk of partial verification bias in only including patients with in-person FALCON follow-up in the diagnostic accuracy analysis but addressed this inverse probability weighting. Fifth, small changes to the published SAP were made in this report, however these were responsive to CEI and investigator priorities and are described transparently.

5.5.3.3 Interpretation limitations

We are unable to extrapolate from these data that the WHQ pathway is feasible and accurate across all settings globally. However, by testing the telephone pathway across a diverse range of patients, hospitals and languages it is highly likely the findings are generalisable. I am unable to comment on the accuracy of the WHQ in detecting SSI in non-abdominal surgery, highlighting an important area for further research.

5.6 Conclusion

In this chapter, I evaluated whether a novel telephone follow-up pathway could be delivered in low resource settings and whether it was accurate in detecting SSI, the most common complication of abdominal surgery. I demonstrated a high telephone contact rate, particularly when attempted within around the 30th postoperative day. Feasibility diminished over long periods of time from surgery to follow-up. Importantly, patients that had a reference test diagnosis of SSI were less likely to be contactable, raising a risk of attrition bias when applied to follow-up in randomised trials. Patients found telephone follow-up highly acceptable, and the pathway was implemented flexibly across diverse contexts and in a variety of languages. As the WHQ point score increased, so did the likelihood of a patient having an SSI diagnosis recorded in the reference test (in-person follow-up). The WHQ demonstrated excellent discrimination between patients with and without SSI which was robust to several sensitivity and subgroup analyses. TALON-2 also provides a proof-of-concept for international SWATs which can now be used to explore other high-priority methodological challenges in other global health trials, including outcome assessment in other perioperative events.

6 Discussion

6.1 Summary of key findings

Existing methods of post-discharge telesurveillance of the healing surgical wound after abdominal surgery are not fit for purpose for use outside of English speaking, high-income environments. In Chapter 3, I identified significant underdetection of SSI when telemedicine methods were used in comparison to in-person assessment in an international cohort study. Upon meta-analysis of all existing data, 1 in 3 patients with infection were missed using current tools. One potential tool to support high-quality wound assessment is the Bluebelle Wound Healing Questionnaire (WHQ), a universal reporter outcome measure that can be completed over the telephone. However, cross-cultural and cross-language equivalence of the WHQ has not been explored to date. In Chapter 4, I explored the cross-cultural equivalence of the 19 items included in the WHQ in qualitative data, identifying themes relating to comprehension, response mapping, retrieval, and judgement, and translating the WHQ into 5 languages. In quantitative data, I demonstrated that the WHQ was likely to be unidimensional, and was valid for use as with ordinal total score. I identified overlapping category probability thresholds in several items, and 11 highly correlated item pairs. Triangulating these data, I made recommendations for adaptation of 9 items and modified the response structure for 11 items. I then implemented the adapted WHQ in a validation study within a trial in seven LMICs. In Chapter 5, I demonstrated that telephone follow-up was highly feasible across diverse environments and languages, delivered by non-surgeon junior doctors and research nurses. However, patients with a diagnosis of SSI upon face-to-face follow-up were less likely to be contactable identifying an important group to target in future research to avoid attrition bias. The WHQ successfully discriminated patients with and without SSI. I presented several cut-points of the WHQ score with corresponding diagnostic test accuracy statistics allowing future researchers and clinicians to tailor use of the WHQ to their local context. Working closely with CEI partners had impact throughout the design,

delivery and interpretation of these studies. Co-production of a toolkit to optimise implementation of telephone follow-up in global surgery will support onwards adoption.

6.2 Findings in context

6.2.1 Potential applications of the adapted WHQ

Now a high-quality pathway for remote detection of SSI has been developed and validated for use across a variety of settings, this could have wide reaching impacts on global surgical delivery. Firstly, this tool can now be implemented in RCTs to test SSI prevention measures. Face-to-face follow-up is no longer required to ensure accurate diagnosis, reducing costs of travel and time away from work for patients, resource use and staff time for health systems and increasing overall efficiency [69, 113, 253-256]. For centres with high surgical volumes where face-to-face follow-up can present a real burden to already resource-scarce hospitals, this could be transformational, allowing trial recruitment numbers to rise and reducing opportunity cost [257]. Secondly, the tool could be used by local clinicians in their routine clinical practice. As the most common complication of abdominal surgery, integration of the WHQ into care pathways could allow for earlier identification of SSI [24, 25]. Patients with concerning symptoms could be triaged back for review at the hospital or by a local primary care provider or community healthcare worker [258]. This could avoid patients with SSI delaying access to care, reducing risk of sepsis, severe wound complications and need for readmission or reoperation [22, 23, 259-261]. Thirdly, the WHQ could be used by governments in tracking population level outcomes of surgery to improve benchmarking, comparison, and improvement initiatives. The World Bank currently measures and reports six essential indicators covering surgical safety, volume, distance to provider, workforce, and patient expenditure [65, 262, 263]. Having robust, easy to deliver pathways to monitor rates of the most common postoperative complications could enrich these indicator sets in the future, providing higher fidelity insights into care standards and allowing targeted

improvement efforts and advocacy for greater investment in surgical services over time [264, 265].

6.2.2 How should the adapted WHQ be implemented?

In the validation study, I have presented several cut-point scores for the adapted WHQ which allow researchers, clinicians, or governments to prioritise sensitivity versus specificity based on their local priorities. Here, I will present two illustrative example of how the WHQ might be optimally applied in different contexts.

6.2.2.1 Case study 1. Use of the WHQ in a research study

In the ChEETAh cluster randomised controlled trial in SSI prevention, hospitals were randomised to change sterile gloves and instruments before fascial closure versus standard practice [44]. In this trial, patients over 10-years of age undergoing non-caesarean abdominal surgery in pre-defined theatres in each participating hospital were included. Inclusion was monitored in a dedicated ChEETAh trial aggregate register to ensure all eligible patients were included in outcome assessment. Patients provided consent for 30-day outcome assessment but not directly for exposure to the trial intervention as it was both applied at a cluster level and was deemed to be very low risk. A major source of selection bias in RCTs occurs where patients are lost to follow-up or withdraw consent post-randomisation (refusal or attrition bias) [266-268]. The difference between cluster and individually randomised designs lies in the fact that all consecutive eligible patients in a cluster are included after the cluster is randomised, whereas not all eligible patients will be identified, approached, or randomised in an individual patient design. In the ChEETah trial, in high volume centres this led to very rapid patient recruitment with the fastest recruiting centre completing recruitment of 200 patients in less than 3-months. Mandating an additional face-to-face assessment for research purposes for 60-80 additional patients per month would have been impractical and

inefficient, with a very high burden of involvement to local clinical teams. If the adapted WHQ were to be implemented in telemedicine follow-up here, the research team would be likely to select a cut-point that would represent a fair balance of both sensitivity and specificity, for example using the Youden index method (e.g., WHQ total score ≥4, sensitivity: 0.701 (0.610-0.792), specificity: 0.911 (0.878-0.9430)) [269, 270]. Other options also exist including the hypervolume under the manifold or maximum absolute determinant [271]. Whichever method is used, so long as the same WHQ score cut-point was used for all patients in a trial, no measurement bias would be introduced as any misclassification would be non-differential between randomisation arms. This would allow a high-quality RCT to be run efficiently, with a low risk of bias in outcome assessment [96]. There would be no concern over patient safety as the research follow-up represents an additional healthcare interaction without deviation from routine clinical follow-up in the local environment.

6.2.2.2 Case study 2. Use of the WHQ for triage in local care pathways

Where the implications of a 'missed' SSI diagnosis are more severe, for example in clinical surveillance pathways that may replace routine face to face follow-up, a different application of the WHQ score cut-point may be more appropriate. In this instance, clinicians and patients are likely to favour a more conservative cut-point score that favours specificity over sensitivity (i.e., fewer false negatives). Here, the clinical team might select, for example, a WHQ score of ≥1 for diagnosis (62.1% of patients, sensitivity: 0.373 (0.312-0.438), specificity: 0.952 (0.904-0.981)) to identify patients that require clinical review as this would capture most patients with a mild, moderate, or severe SSI. Adopting this cut-point score would bring efficiency to the clinical service by reducing the need for routine in-person follow-up for wound assessment by 37.9%, whilst maintaining patient safety. Other variations of use of these cut-point scores, could be used to identify patients at very high risk of having an SSI (e.g., WHQ score ≥8, 13.7% of patients, sensitivity: 0.830 (0.702-0.919), specificity: 0.842

(0.798-0.879)) and ask them to return to the treating hospital and use a lower threshold (e.g., WHQ between 1 and 8) to be assessed by a local doctor or community worker, to refer on with any concerns [272]. This powerful triage system could increase efficiency and early access to care for patients in the community. As highlighted by our CEI partners, all patients should be provided with a safety net and be directed to access care as soon as possible in the case of any new symptoms or deterioration.

6.3 Potential advantages of telemedicine wound assessment

The use of telemedicine in wound assessment has several potential advantages.

6.3.1 Reducing cost of travel back to hospital for patients and their families

Geospatial mapping of hospital sites around the world has demonstrated that patients in LMICs must often travel further to reach a hospital that provides surgery than in HICs [113, 273]. In some cases, patients may have to travel even further to reach a central hospital (e.g., in a capital city) that is able to provide subspecialised care (for example for neurosurgery or cardiac procedures) [37, 39]. Patients in LMICs are also at high risk of catastrophic expenditure as a result of their initial surgery, with many facing financial ruin [9, 65, 68]. Returning to hospital for additional follow-up therefore may mean incurring additional costs for transport, accommodation, and subsistence that will add to their financial burden [12, 69, 272, 274]. Some patients may also require a family member to travel with them to hospital due to anxiety, disability, or other causes, who may also incur loss of income [274]. Both are largely avoidable using telemedicine follow-up [275].

6.3.2 Reducing time out of work or away from care for patients

Similarly, an unnecessary return to hospital will take postoperative patients away from their normal activities of daily living. Due to the financial pressures of surgery where Universal

Healthcare Coverage is not available, patients may be more likely to return to work early after surgery in LMICs, and a hospital visit may therefore cause loss of income [114, 276-278]. Others may have care responsibilities for children, the elderly or other family members which they will be taken away from whilst visiting hospital and may have emotional, financial, or practical implications [279].

6.3.3 Reducing barriers to access to care

Access to care is one of the key priority areas in global surgical research and policy [14, 40, 280, 281]. Timely access to care requires patients to: (1) seek health intervention; (2) get to a health facility; (3) receive care in a health facility [282]. Whilst difficult to directly measure, this 'first delay' relating to health seeking behaviour leads to substantial harm worldwide and is likely to disproportionately impact those with low health literacy [283]. Early identification of postoperative complications using validated tools such as WHQ, will support patients to seek health interventions and potentially reduce complications of delayed presentation.

6.3.4 Increasing patient satisfaction

In the validation study, patients reported very high satisfaction with telephone follow-up and feeling valued through increased interaction with hospital services. It is likely that integrating routine postoperative surveillance into care pathways in a contextually appropriately manner (e.g., learning from our co-produced toolkit) will improve patient experience during their recovery. I was however, unable to comment on relative satisfaction with telephone versus in-person follow-up, nor attitudes towards the substitution of in-person with telephone follow-up where the former is current practice.

6.3.5. Reducing burden on local clinical teams

Task shifting is a concept which describes the controlled transfer of specific clinical tasks to a less specialised member of the care delivery team [27, 153, 154, 284, 285]. It is commonly performed to improve access to care, reduce burden on highly trained staff and address workforce shortages [285]. The WHQ tool here was administered by non-expert assessors (junior doctors, research nurses or administrators), which task-shifted postoperative follow-up away from trained surgeons to more junior or non-clinician team members. The workforce shortage is one of the most pressing issues facing global health today [286, 287]. Reducing opportunity cost by reducing workload on trained surgery, anaesthesia and obstetric providers may help to relieve some of these issues as part of a broader workforce strategy [288].

6.3.6 Reducing resource use in low-income hospitals

Outpatient clinic services for postoperative follow-up requires administration, waiting room space, appropriately equipped clinic rooms, staffing, and transport networks. Adoption of a telemedicine first follow-up strategy therefore is likely to significantly reduce resource usage.

6.3.7 Reducing risk of nosocomial infection

Surgical patients are at high risk of complications of SARS-CoV-2 infection, particularly in the first 30-days after surgery [46, 47, 49, 141]. It is possible that other common respiratory viruses such as influenza, or other transmissible disease also increases risk in the perioperative setting, although this has been less frequently examined [289]. Telemedicine is likely to reduce the risk of nosocomial transmission of these pathogens to the vulnerable surgical population.

6.3.8 Reducing loss to follow-up

For all of the reasons above, patients may choose not to attend in-person follow-up where offered. In a research setting this may lead to loss to follow-up, and attrition bias if there is differential misclassification between arms [248, 249, 290]. Loss to follow-up is rarely random and may be more likely in patients with severe complications (who may feel too unwell to travel for example) and conversely in those with no complications (who feel further review is unnecessary) [119, 248, 249]. As observed in Chapter 5, telemedicine may offer a simple and accessible method to reduce loss to follow-up in future trials.

6.4 Barriers to wider adoption of telemedicine wound assessment

Despite this promising potential, there are several notable caveats. First, the use of the telephone WHQ only identifies one common postoperative complication [184]. Whilst a telemedicine consultation provides opportunity for patients to self-report other issues during their recovery, the accuracy and safety of this in general postoperative screening was not validated here. Second, there may be some patients that lack access to a telephone for use for a telephone call [129, 291, 292]. In the validation study, a third of patients were able to use a friend or family member's phone for their follow-up with a high level of feasibility demonstrated. Working with CEI partners we optimised the telephone measurement procedures to attempt to overcome problems with connectivity, timing, and privacy concerns. On the rare occurrence that patients report no access to a telephone, they should be offered in-person follow-up as an alternative (if required). Third, for patients that require follow-up, some may prefer a more 'human' face-to-face interaction [293-295]. These patients could be offered in-person or video follow-up as an adjunct if feasible [149, 296].

6.5 Adjuncts to telephone wound assessment

The wide global availability of smart devices that have photo and video capability offers exciting new avenues for exploration in wound healing research [127, 245, 297-299]. Several studies have explored the use of photographs in wound assessment in high-income countries [128, 131, 132, 251]. However, these are typically limited by patients' ability to provide uniform, well-lit photographs of sufficient quality and correct exposure and provide only a single static assessment of the healing wound. Ongoing work is exploring ways to improve standardisation of wound 'selfis' although this is yet to be validated, nor adapted for use across contexts [300]. As we saw in Chapter 4, it may be that visual diagnosis of SSI is harder in patients with darker skin tones for whom more subtle changes such as shining or tightening of the skin may be less apparent in low quality photographs. Wound videos are another exciting area for future exploration, but the field remains immature. Research to date has focussed on the service impact of telesurveillance, for example the number of readmissions or changes to medication schedules, but without exploring the diagnostic accuracy of video assessment [124, 125, 149, 296]. No study to date has explored the intra- or interrater reliability between video and in-person assessments of the healing surgical wound. Self-administered videos that are 'submitted' by patients have several limitations. Firstly, submitted videos are subject to the same limitations on quality and exposure as photographs. Secondly, there is a missed opportunity to ask the patient to correct their positioning or improve the video quality. Third, the patient cannot be asked to interact with their wound (e.g., push on an area of swelling to look for pus or pain, or remove or replace a dressing). Fourth, some health and ethical systems would consider any video submission as a component of the patient record and would require high-security transfer and storage, and even linking to the patient's electronic record [301]. This is impractical in many settings. A 'live' video assessment during a video call may overcome some of these problems, better integrating the desirable components of a wound review (interactivity, palpation, improved exposure) with the efficiency of telemedicine review. As no data are transferred, no storage

of video materials is required, and privacy concerns are overcome. This method has not yet been reported in the literature but is a crucial area of onwards development.

6.6 Artificial intelligence in detection of wound infection

Unstructured machine learning (or 'deep learning') is another area with promise in the field of SSI diagnostics [302-305]. Where high volumes of images or videos of healing surgical wounds can be collated and stored, artificial intelligence platforms could be trained to detect changes consistent with SSI. This has been demonstrated with a high degree of accuracy in other data rich fields such as ophthalmology and dermatology where photograph forms part of routine practice. Prominent examples include the diagnosis of skin cancer and changes related with age-related macular degeneration on optical coherence tomography [306-308]. Whilst these deep learning algorithms require vast amounts of data, they are increasingly accessible in accessible, code-free platforms [309-311]. Collaborative efforts to collate datasets across institutions and multi-country groups would be able to generate sufficient data for successful development and validation and should be the target of future efforts [312]. In the context of LMICs where video or photo capability was sufficient, again this could allow further 'task-shifting' away from low resource clinical teams [27].

6.7 Perspectives on postoperative telemedicine in the literature

6.7.1 Postoperative telesurveillance in high-income settings

Studies exploring the use of telemedicine in post discharge follow-up in high-resource environments have largely evaluated the impact of telemedicine on health resource usage in comparison to standard clinical pathways [313]. The PVC-RAM-1 trial published in *BMJ* in 2021 evaluated a high-intensity remote monitoring programme including daily biophysical measurements, wound photographs, and consultations with a nurse up to 30-days after surgery [149]. They detected no difference in the primary outcome measure of days alive at

home at 30-days after surgery, but observed improvements in pain and reductions in medication errors, particularly in hospitals with early escalation policies. Whilst this provided a useful signal in the utility of telemedicine approaches, it provides no data about detection of specific post-discharge complications, and is likely to be too resource intensive for implementation in LMICs.

The TWIST trial, published in *NPJ Digital Medicine* in 2022 tested a smartphone delivered wound assessment tool which included a SMS-delivered wound screening symptom questionnaire (not previously validated) and wound photograph which could be completed voluntarily, but prompts at days 3, 7 and 15 postoperatively [131]. These data were reviewed directly by a clinician to triage to either reassurance (low-risk), attend community services (moderate-risk), attend hospital (high-risk). The intervention did not improve time to diagnosis of SSI, but did reduce community healthcare attendances and overall patient satisfaction. Whilst this is a helpful example of how telemedicine can be used successful in remote triage to reduce pressure on clinical services, it was not used as a formal diagnostic test that could be substituted for in-person assessment in future trials, nor do we know the feasibility of wound photography in LMICs. Future iterations of automated response platforms such as this could integrate the adapted WHQ to improve diagnostic accuracy [314].

Harkey et al in JAMA Surgery (2021) described a video-based virtual visit for low-risk patients discharged from hospital after minimally-invasive (MIS) appendicectomy or cholecystectomy. They demonstrated non-inferiority of telemedicine versus routine in-person assessment for a primary outcome of hospital encounters (emergency department visits or other unplanned consultations). These data have little generalisability to LMICs for several reasons: (1) the rate of MIS is typically low in most low resource settings [43]; (2) patients more frequently present with delayed presentation of disease in LMICs so are more likely to suffer postoperative complications [23]; (3) the protocol imposed non-pragmatic exclusion

criteria (length of stay ≥3 days, discharge to location other than home, perforated disease) that are likely to exclude large proportions of patients undergoing surgery in LMICs [34]; (4) despite promising signals for smart phone availability in Chapter 5, we remain unsure of the feasibility of video follow-up in low resource contexts.

Several other non-randomised evaluations of wound photography have demonstrated promise in enhancing diagnostic accuracy versus symptoms questions alone in clinical vignettes [132] and reasonable agreement between photographic and in-person SSI diagnosis in two lower limb trauma trials [251] and a vascular surgery cohort [128]. No single solution presented in the high-income country literature provides a high-quality, feasible, reliable, valid method for abdominal SSI diagnosis for research and clinical practice [123, 195, 315].

6.7.2 Emerging examples of telemedicine wound assessment in LMICs

Some promising examples of implementation of telemedicine have been reported from LMICs. In a feasibility study of mHealth-supported SSI diagnosis by Community Health Workers in Rwanda by Kateera *et al.* (2022), there was no difference detected in return to care rates between home visit, phone call and standard of care arms [316] suggesting feasibility of telemedicine follow-up as an alternative to expensive and time-consuming clinic visits. Further data from a cohort of Rwandan patients (N=569, 61 with SSI) undergoing caesarean section described moderate agreement of photograph and in-person SSI diagnosis (Gwet's AC1 agreement estimate of 0.46), reporting challenges with standardisation of photography, image quality and accompanying clinical information that require further exploration [317]. In a diagnostic test accuracy study in a single centre in Tanzania (N=374 patients, 45 with SSI), using a structured questionnaire aligned to the US CDC criteria diagnosed SSI with a sensitivity of 0.72 and specificity of 1.00 [163]. A similar

study in Kenya (N=89, 23 with SSI) also demonstrated telephone CDC assessment to be feasible, with a high contact rate and sensitivity of 0.70 and specificity of 1.00 [318]. Both of these studies show promise for high diagnostic accuracy in telephone wound assessment, but were both single centre, did not use a validated questionnaire, lacked patient involvement in design, limited by their small sample size and were at moderate to high risk of bias. Other data from India [169], Cambodia [319], Haiti [297] and Sudan [320] has demonstrated high rates of telephone contact post discharge and successful detection of SSI, but without formal test accuracy evaluation. Put in the context of the literature, data from this thesis suggests that telephone follow-up pathways are feasible for delivery across a wide variety of settings, with local adaptation to context where required (for example, with the engagement of Community Health Workers in Rwanda). Chapter 4 and 5 in this thesis provides an adapted tool with items that have cross-cultural and cross-language equivalency that is validated for use in global surgical research and practice.

6.8 Telemedicine and loss to follow-up

In the FALCON trial (host trial for the validation study), patients were required to return to hospital for assessment in accordance with the study protocol [43]. During SARS-COV-2, telemedicine follow-up allowed 52% more patients to be followed up that were unable to return in-person. In the validation study, under 10% of patients were uncontactable for telemedicine follow-up. For patients with telephone follow-up attempted per-protocol (i.e., 27-30 days after surgery) the telephone contact rate was almost 96%. This was lower than anticipated during design (15% loss to follow-up include in sample size considerations) and significantly lower than in comparable international surgical RCTs, ranging from 10% to 25% or higher [104, 321-323]. Telemedicine may therefore have a strong potential role in reducing attrition bias in future RCTs in surgery.

6.9 Community engagement and involvement (CEI)

6.9.1 Impact of community engagement and involvement

The impact of CEI in this thesis is reported chapter by chapter. Looking at the overall impact, working with CEI partners from prioritisation to interpretation has fundamentally improved the quality, acceptability, and communication of the research. Whilst designing the thesis research questions, CEI partners from a UK-LMIC advisory group highlighting telephone follow-up as an acceptable and efficient alternative to mandatory in-person follow-up for RCTs. In Chapter 4 and 5, CEI partners supported the co-design of the study protocol and measurement procedures to create a toolkit to optimise telephone outcome assessment in global surgery research. This had a measurable impact on the feasibility of telephone contact, and high data completeness rates. Our CEI partners both generated novel ideas for the toolkit and ratified and refined themes emerging from the qualitative data in Chapter 4. They reviewed the final recommendations for adaptation of the questionnaire, providing further cross-cultural insights. Finally, working with the NIHR Unit CEI lead we co-created a lay summary of the research in Chapter 5 to support public dissemination. This will be translated into local languages to be shared with future participants in research.

6.9.2 Global models of Community Engagement and Involvement

Patient and service user involvement is a fundamental component of high-quality, patient centred research [324-332]. In high-income countries such as the UK, this would typically include patient partnership in grant writing, on a Trial (or study) Management Group and Trial Steering Committee, and with regular patient advisory group meetings throughout a programme [326, 333]. Funding bodies such as the National Institute of Health Research (NIHR) mandate high-level patient partnership in all aspects of publicly funded research, and the quality and depth of this involvement is a major criterion in grant assessment [334]. Researchers are expected to track and report the impact of patient and public involvement in

research transparently, with patients as co-authors on all resultant outputs [225, 325]. However, global models for CEI are still evolving, particularly in global surgery. Whilst participatory action research (where stakeholders are involved in the design of an intervention) is not new, the concept of patients as equal partners in research is still evolving in LMIC research systems. As a result, often CEI in multi-country studies is often performed as 'participation' (patients as the subjects of a research study) or 'consultation' (patients are asked in a one-off interaction to provide feedback on a component of the research) rather than 'co-production' (where patients co-lead the research throughout its life cycle) [335-337]. There are several barriers to the high-income country concept of 'co-production' in lower resource environments. Firstly, for patients who have low levels of literacy, the concepts of research and patient involvement are likely to be challenging to fully understand [338-341]. Secondly, strong power dynamics remain between doctors and their patients in some countries. Rebalancing this can be a real challenge. Thirdly, general awareness of the purpose of research can be lower, and sadly (understandably) often held with high levels of scepticism due to colonial crimes of the past. These problems are not unique to LMICs, but are accentuated in more deprived areas where access to higher or even basic education is not universal, and social divides between professionals such as doctors and the public they serve can be more pronounced. Finally, COVID-19 has exacerbated some of these challenges creating more financial, logistical and safety challenges to cohesive global CEI models [342]. Learning from my experience in this thesis, some recommendations for future CEI in global surgery are made in *Box 6.1* below.

6.9.3 Co-production in telemedicine research in LMICs

Previous authors in telemedicine in LMICs have also discuss the importance of co-design of follow-up pathways. A field development and validation study of telemedicine for patients undergoing caesarean-section in a Rwandan rural district hospital by Cherian et al. (2020)

reported reasonable sensitivity and specificity in wound assessment with task shifting in diagnosis of SSI to general practitioners and community healthcare workers [272]. The study commented on the importance of a pathway that was closely co-designed with local stakeholders. Similarly, a group from Washington made several recommendations for patient empowerment within remote follow-up pathways through a consultation exercise with a Stakeholder Advisory group [343]. Working closely with local partners and care pathways to develop pathways for implementation of the WHQ in the future will be essential in ensuring sustainable and culturally attuned models that benefit both patients and health systems.

Box 6.1 Recommendations for global Community Engagement and Involvement

Recommendation		Description
1	**Start as early as possible**	As soon as you begin to prioritise research topics, think about how you might engage CEI partners that have lived experience of the disease, procedure, or complication of interest.
2	**Aim for diversity & inclusivity**	The more diverse your CEI involvement, the more perspectives you will gain on your research work and the more benefits you will see. Include patients from a range of socioeconomic backgrounds, gender, ethnicities, sexualities, professions, and experiences of the research area of interest.
3	**Don't aim for perfection**	That being said, diversity is very challenging in global CEI. Typically, more educated, wealthier professionals with some previous experience of a medical or related profession are the first to volunteer their expertise. These individuals may be all that is accessible to provide steering and partner with you in the research. Actively keep seeking more diverse representation, but don't let perfection stop you from starting. Some level of CEI is always better than no CEI.
4	**Find locally adapted methods**	There is no single 'best' method for CEI. The best way to identify, contact, engage, and sustain patient and community member involvement will vary from country to country and even from hospital to hospital. For example, in Cape Town, South Africa, there were safety concerns about performing community-based outreach. Instead, a patient champion with strong community ties was identified, was supported for costs of time and transport, and worked closely with the local team to share awareness of the research programme and identify new candidates for future involvement in-hospital. In Rwanda, there was a substantial existing network of Community Health Workers, so the local team worked with these teams to identify potential patient partners, and feedback on key discussion points.
5	**Find patient champions**	A strong patient champion who understands the key concepts of research and patient involvement will be a 'flagship' to encourage other patients and provide mentorship for new members. When you find these individuals, curate your relationship with them carefully, and ensure that they feel engaged and listened too – they are the most essential component of a successful CEI strategy!
6	**Conduct a needs assessment**	Work with local patient champions and research teams to identify the training needs for new patient partners. This might range from simple descriptions of clinical concepts ('jargon-busting') to some basic training on the purpose and principles of research. If possible, learn from experts in PPI/CEI methodology to design simple, targeted training for new patient group members, so that they understand the purpose of their involvement and their terms of engagement very clearly.
7	**Keep in touch regularly**	Once you start to identify high-value individuals that hopefully represent a diverse range of perspectives and understand the key concept of equal partnership in research, make sure you stay in touch with them! Have a plan for regular engagement, even if this is to update on a project's status and receive feedback on progress to date and share learning. This is typically hardest 'mid-way' through a project once the protocol and sites are set-up and data collection is underway. Try hard not to just 'reach out' for consultation at the start and end of a project.
8	**Be flexible with platforms**	There will be easier and more difficult ways to engage with CEI partners and this might vary from individual to individual. Patient advisory group forums can be quite overwhelming for new group members, whilst for others might be empowering. Try to vary your platforms (teleconferencing, WhatsApp calls) and group sizes (one-to-one, small group, larger group) and get a sense of what is working best for your project and partners. Where possible, let you partners take part in choosing the platform and timing to share leadership and ensure you maximise likelihood of active participation.
9	**Ensure their voice is heard**	If you have an international study group meeting with patients and professionals, it can be very intimidating for CEI partners to share their thoughts. Ensure that you give partners adequate time to provide their input in a safe space, including 'offline' via email or WhatsApp if they don't feel confident to do so in a group setting.
10	**Track impact over time**	CEI undoubtedly brings huge benefits to research design and delivery. Don't forget to record the iterative changes that you make over time as a result of CEI steering, as this will be invaluable when you come to report your research in accordance with GRIPP-2 recommendations.
11	**Feedback your results**	When you get to the end of your project it is essential that you share the results of the research with both your CEI partners and the research participants wherever possible. Work closely with your CEI partners to design accessible, mixed-media methods of communicating your study findings. Be sure to consider vulnerable participant groups such as the illiterate, visually impaired, or those with learning difficulties for inclusive and engaging research dissemination.
12	**Report back your experience**	The world is still learning about best methods for global CEI. If you have an experience of engaging CEI partners in research, share it with the global health community. There are likely to be learning points specifically related to global surgery from this research, there will be many cross-cutting themes across procedural specialties, non-communicable disease, and even broader health-related topics. Collaboration will be essential in developing this emerging, but essential area.

6.10 Building resilient perioperative systems

6.10.1 Fragility of surgical systems to external shocks

The SARS-CoV-2 pandemic revealed the fragility of surgical systems around the world [344]. Even for conditions such as cancer where continuation of surgery was prioritised, 1 in 7 patients did not undergo their planned operation during COVID-19 lockdowns [47]. This led to disease progression and death for some patients, and high levels of anxiety for others [146, 151, 182, 345-352]. The shockwave effects of COVID-19 on non-communicable disease are likely to be felt for decades. The status of global surgery as of July 2022 is in crisis. The global waiting lists for surgery are at an all-time high, and systems are struggling to raise capacity during to workforce constraints, increased infection control requirements and ongoing COVID-19 transmission [31]. In addition, there is a substantial but unmeasured 'hidden' waiting list of patients who will require treatment for surgical disease but have not yet been identified due to delayed diagnostics or presentation [346]. In the UK, we estimate that this represents an additional 30% on top of existing waiting lists [353].

6.10.2 Surgical system strengthening and whole health system resilience

However, this is not all new. Since the Lancet Commission on Global Surgery in 2015 there has been global recognition of the essential role of surgery in holistic healthcare delivery and the profound economic, social, and political arguments for strengthening surgery and anaesthesia systems [2, 4, 17]. The commission also recognised sadly, the fragile, unprepared, and under-resourced state of surgical services in many countries around the world. Strengthening surgical systems does not only have direct benefits to patients with surgical treatable disease, but also underpins whole-health system resilience through bolstering critical care capacity, diagnostic capacity, oxygen and medication delivery systems, stable supplies of electricity and water, and pathways for post-discharge surveillance [11, 152, 354].

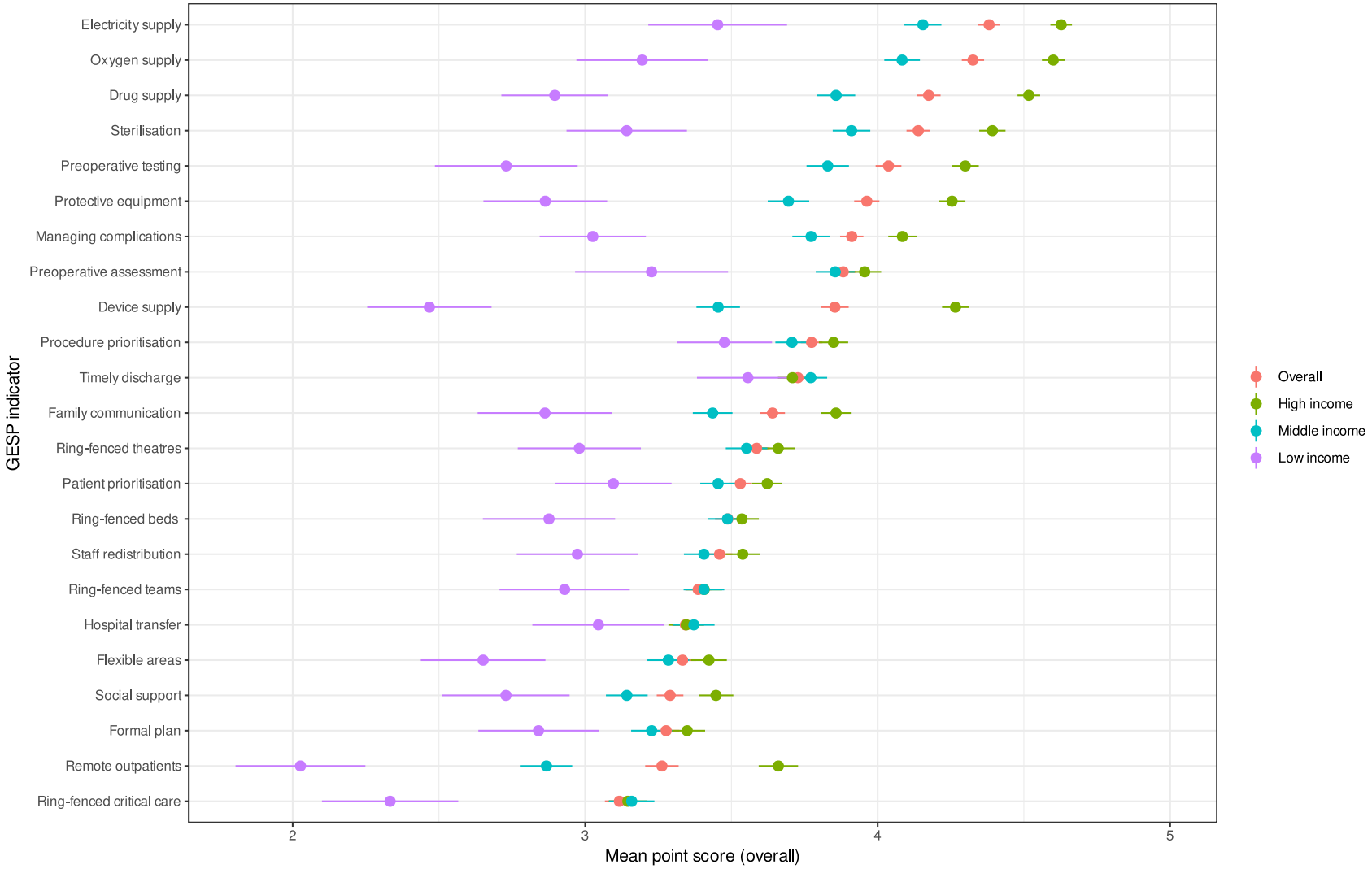
6.10.3 Surgical preparedness index (SPI) for benchmarking and system strengthening

To create a flexible tool that can be applied by clincians, hospital managers and policy makers in hospitals worldwide, we created a Surgical Preparedness Index through a four-stage consensus process with an international writing group. The group prioritised 23 surgical preparedness indicators from 110 candidates longlisted by an international cross-disciplinary stakeholder community. We then undertook an international validation study with 4714 collaborators across 1600 hospitals to explore the association between the SPI score and elective surgical capacity in that hospital. Each item was scored between 1 (very weak) and 5 (very strong), with each participating hospital having a mean score calculated across the 23 indicators between 23 and 115 in total. There was a strong association between the ability of hospitals to continue elective surgery and the SPI total score. This self-assessment process will allow hospitals to benchmark against other hospitals in their country, region, and income category, and identify actionable targets to strengthen their local systems. A live tool is available now at: spi.surgery.

6.10.4 Importance of postoperative telesurveillance in resilient care pathways

Remote methods for post-discharge surveillance were prioritised as one of the key Surgical Preparedness Indicators, reflecting their importance in resilient perioperative care pathways (*Indicator 22.* Capacity to use telephone or video calls for outpatient appointments). Overall, remote outpatients were the second lowest scored SPI (3.26 out of 5) and with the second highest point score difference between mean scores in HIC (3.66 out of 5) and LICs (2.03 out of 5). This highlights the importance of research such as that developed within this thesis in bolstering frugal, accessible and high-quality methods to strengthen surgical systems in LMICs.

Figure 6.1 Mean ratings of hospitals (N=1632) across surgical preparedness indicators



Indicators ordered from highest to lowest mean score (out of 5) 'overall' by indicator

6.11 Limitations

The work described in this thesis has several caveats and limitations. Some specific methodological considerations are described Chapter by Chapter. Here, I will consider in broad terms what the weaknesses of this programme were, how I attempted to mitigate them and considerations for future development. Design limitations were defined as limitations within the study protocol which were anticipated during design. Delivery limitations were defined as those that emerged during study delivery. Interpretation limitations were limits on generalisability of the data to other contexts. Scope limitations were areas that warrant further exploration but were not included within the scope of this thesis.

6.11.1 Design limitations

- Due to limitations of travel and patient contact during the SARS-CoV-2 pandemic, cognitive interviewing was performed over teleconferencing with site investigators. This meant that exploration of cross-cultural equivalence of the WHQ items was limited to investigators impressions of challenges for patients from their experience. Whilst I sampled a broad range of researchers with a variety of perspectives that all had extensive experience of both in-person and telephone follow-up, and proved to be information rich, it is possible that important themes were missed during the WHQ adaptation process.
- Moreover, whilst teleconferencing for qualitative data collection is well described and allowed the study to proceed remotely during a period of travel restrictions, this does have potential limitations [355, 356]. Tele-interviewing limits interpretation of body language and makes it harder to pick up cues from participants and use probes judiciously; effectively ‘disembodying’ the humanistic inquiry of qualitative research [357]. This can lead to fewer codes and statements generated during analysis [358].

It can increase the risk of miscomprehension, particularly for participants where English is not their first language. Connectivity was also a common challenge; many interviews took several attempts to complete, and some focus groups participants were unable to contribute synchronously. However, it was imperative to continue this qualitative research during the pandemic, and virtual methods allowed the study to progress, whilst learning more about some of the potential strengths and weaknesses of this approach [359]. It may also have improved the diversity and inclusivity of the research sample by reducing the burden for busy investigators and financial and logistical barriers to face-to-face approaches in multi-country qualitative research [360].

- Whilst WHQ translation followed *Mapi* recommendations for best practice, structured interviews were not possible in target languages and were all performed in English [134, 215]. Despite baseline translatability assessments and cautious review at each stage by the in-country bilingual consultant it is possible that some meaning of the items was lost across languages. However, the speed of set-up and study progress despite COVID-19 provided a proof of concept for rapid cross-language adaptation of a patient-facing outcome measure within the context of a randomised trial.
- As there was only one WHQ completed per patient, I was unable to examine the inter-rater reliability. Whilst this has been estimated in previously validation work, it is possible that this would vary from country to country or across languages [192].
- The co-produced toolkit for optimising telephone follow-up in global surgery was developed alongside the research packages and integrated into the Chapter 5 follow-up pathways. However, this was performed stepwise, learning over time, and no effectiveness analysis in terms of retention, contact rate or data completeness was planned or completed. The toolkit therefore requires further testing in future work.

- Whilst I put in place several measures to ensure reproducibility in delivery of the WHQ including site investigator training, pilot phases at each site, regular communication through WhatsApp groups, and recording and review of a sample of WHQ administration in two sites, it is possible that there was unmeasured variability in the delivery of the WHQ that have reduced its diagnostic accuracy.

6.11.2 Delivery limitations

- Whilst the overall use of the WHQ total score as a measurement scale was verified on Rasch analysis, there were several potential reasons why the Rasch model misfit the data overall. There was high dependency between 11 item pairs, as might be expected in a diagnostic tool, and there was significant positive skew of the item information distribution [219]. Whilst this mistargeting isn't a problem in itself on this application of Rasch (it follows a clinically intuitive pattern where many patients have no or few symptoms and only those with early to severe manifestations of the disease contribute information to the model) this did mean that the overall power of fit was low, and most patients contributed little to the overall understanding of the psychometric properties of the score. As the Rasch analysis here was really used as an exploratory tool to support or refute findigns from qualitative data during adaptation, I did not perform further manipulation such as subtesting for correlated item pairs, rescoring where threshold probabilities overlapped or adjusting for differential item functioning [219, 361-363]. An exciting area for onwards development would be to attempt to improve model fit, and meet the requirements for a logit-adjusted scoring scale. This could potential both improve the diagnostic accuracy of the scale as a whole, reduce the item set without affecting performance, and better reflect patient-level differences in WHQ scoring based on individual person-factors [364].

- I attempted to estimate the potential 'retention benefit' of telephone follow-up by reporting the proportion of patients from whom follow-up was completed by telephone in the FALCON trial, which mandated in-person follow-up in the study protocol [236]. This proportion was much higher than anticipated due to travel restrictions and safety concerns for patients during the SARS-CoV-2 pandemic, so the 'retention benefit' (as defined) may represent an overestimate [49]. We are unable to state the counterfactual (i.e., that patients would have been lost to follow-up if it weren't for telephone follow-up). However, moving to telephone follow-up did allow patients to have outcome assessment performed safely and with a low rate of loss to follow-up despite pandemic challenges.
- This high proportion of patients unable to return to in-person follow-up also reduced the number of patients with a gold standard reference test result for inclusion in the validation analysis. This reduced the certainty of estimates of the diagnostic test accuracy statistics and introduced a potential for partial verification bias [244, 365, 366]. To address this, I performed an inverse probability weighted (IPW) sensitivity analysis. Whilst there was only a small decrease in the AUROC estimate in the IPW analysis, this is only valid under an assumption of missing data at random; if there was some systematic error in the patients that were and were not asked to return to hospital for in-person follow-up it is possible that the accuracy of the WHQ was overestimated here.
- Whilst there were some interesting signals of differences in accuracy between subgroups such as emergency versus elective care, and urban versus rural patient home location [367, 368], these were exploratory with only small numbers of events per group. Further exploration of possible reasons for any differences including any differential item functioning should be the focus of onwards development of the adapted WHQ.

- The original protocol for the validation study required the WHQ to be completed between 27 and 30-days after surgery (i.e., prospectively, before the date of paired in-person follow-up for the FALCON trial) [186, 241]. Due to constraints of the SARS-CoV-2 pandemic, 27-30d, some centres entered a proportion of patients who had WHQ completion after their face-to-face or telephone follow-up within the FALCON trial. This could have influenced the responses that patients gave in response to each item, as they may have received additional information about the status of their wound at their routine follow-up. If patients completed the WHQ a long time after their operation, they may have also had problems with retrieval and judgement of their wound status up to 30-days after surgery. However, this represented less than a quarter of patients included in the validation analysis, and sensitivity analyses were conducted including all patients and per-protocol timing only, with only small differences in discrimination observed.
- In Chapter 5, we explored the symptom profiles of 'false negative' patients that had a WHQ score of zero (i.e., reported no wound symptoms), but a 30-day FALCON trial diagnosis of SSI (i.e., had signs and symptoms consistent with SSI). On deeper exploration these patients typically had mild symptoms, with a low item location position in the Rasch analysis suggesting those that were 'missed' had early rather than advanced SSI. However, we had not set up procedures to explore differences in responses and clinician assessment with higher fidelity to better understand where misclassification or miscomprehension had occurred. This would be a very useful target for ongoing research.
- Approximately 50% of patients in Chapter 5 had the WHQ administered in non-English language, but without a formal translated WHQ available (i.e., translated by the questionnaire administrator). This occurred either where there was no written version of a regional dialect (common in Northern Ghana), or the patient spoke a

language that was atypical for the area (i.e., had travelled a long distance to receive surgery, commonly in India). Whilst this is not best practice in cross-language use of outcome measures, it reflected the diversity of the delivery network and allowed the pragmatic study to continue with consecutive recruitment without introducing selection bias [134]. In subgroup analyses of patients who received a formal WHQ translation versus adhoc translation by the questionnaire administrator no difference in discrimination was observed.

6.11.3 Interpretation limitations

- This study evaluated feasibility and accuracy of the adapted WHQ in seven LMICs only. Whilst these spanned three continents, and upper-middle, lower-middle and low-income countries and patients it is not possible to definitively state the adapted WHQ will have equivalence across all countries worldwide. It is likely to need further iterative modification and cross-language adaptation during more widespread implementation.
- We have not explored the accuracy of telephone detection of other postoperative complications that can occur after discharge such as pneumonia, venous thromboembolism, urinary tract infection, anastomotic leak, sepsis, cardiac injury, arrhythmias, or cerebral events [25]. The subtlety to SSI measurement is that many wound complications are managed in the community, highlighting the importance of post-discharge surveillance, whereas many major complications would be much more likely to warrant return to hospital for review. Future work should explore how patient reported outcome measures and routinely collected hospital data can be used in complementary ways to create high-quality remote diagnosis systems for a range of common complications.

- We are also unable to comment on the impact of reducing in-person clinical contact on the management on non-clinical issues such as wellbeing, psychological and social support, and delivery of information [369, 370]. Equally, whilst patients reported high levels of satisfaction with telephone follow-up, we are unable to assert the counterfactual: that patients would be happier if they moved to telephone only follow-up rather than face-to-face. Consideration of these important issues should be woven into implementation of remote assessment pathways, informed by CEI partners.
- In the feasibility model in Chapter 5, patients with a reference test diagnosis of SSI in the FALCON trial 30-day follow-up were less likely to be contactable by telephone in the mixed-effects model. Despite likely improvements in retention with the introduction of telephone follow-up in RCTs, this highlights an ongoing risk of attrition bias. Contextually appropriate, behaviourally informed strategies to improve retention to trials in global surgical trials is an area for urgent ongoing exploration [122, 240, 248, 250].

6.11.4 Scope limitations

There are several other interesting areas of enquiry in the field of remote assessment of the healing surgical wound that I have been unable to include in this thesis.

- Video assessment and other digital solutions: As discussed, ‘live’ video assessment by a trained assessor with a patient may be an efficient, high-quality method adjunct to WHQ completion that minimises issues with data storage and linkage to patient records. Data are being collected within the ongoing ChEETAh randomised controlled trial using a similar Study Within A Trial methodology adopted here, and will provide invaluable insight into this application of remote assessment [44]. Mobile applications or text-messaging systems for completion of WHQ items have shown promise in high-income health systems but remain unexplored in LMICs [131].

- Remote physiological monitoring: Wearable technologies for remote physiological monitoring after surgery are being increasingly adopted in high-income countries, but the evidence base remains immature and at high risk of bias [371]. Whilst changes in physiological parameters are non-specific, they may highlight early markers of deterioration to triage patients into formal remote or in-person wound evaluation and can be considered an important part of the data ecosystem.
- Artificial intelligence: As SSI diagnosis is potentially data rich, including physiological parameters, patient reported symptoms, interactions with healthcare practitioners, and signs detectable on video and photographs there is huge potential to build data systems that bring together these disparate sources, and train deep learning algorithms to diagnose SSI with a high degree of precision [307, 372-375]. Governance is likely to be complex, but patient benefit high, and this will be a key area of enquiry in wound research over the next decade.
- Deliverability within routine care pathways: Here the WHQ was delivered by staff within a RCT delivery platform. Different systems, staffing and funding would be required to integrated routine wound screening using the WHQ into clinical pathways, and further infrastructure to aggregate, report and benchmark SSI rates across low-resource hospitals [65, 376]. However, the time requirement for WHQ was low with most patients completing the questionnaire in less than 20 minutes, and a range of non-expert clinicians and non-clinicians administering the questionnaire suggesting that quality can be maintained when delivered flexibly across settings.

6.12 Closing summary

This thesis has aimed to co-produce a new pathway for remote wound assessment after discharge for patients undergoing abdominal surgery in low- and middle-income countries (LMICs). Whilst there were some challenges in delivery during the COVID-19 pandemic, the study methodology was transparent and high-quality, following a pre-published protocol and statistical analysis plan. The Wound Healing Questionnaire (WHQ) was adapted by bringing together qualitative data from frontline researchers involved in wound assessment and Rasch analysis of quantitative data, which allowed recommendations to be made for use across a diverse range of settings. Delivery of the WHQ over the telephone was feasible across 7 LMICs and 22 languages. The WHQ classified patients with or without SSI after discharge with satisfactory accuracy. This adapted tool is now available for implementation in global surgery research and practice, informed by a matrix to select an appropriate WHQ score cut-off and a toolkit to optimise telephone follow-up that was co-produced with patients and community members.

7. References

1. *Early years postgraduate surgical training programmes in the UK are failing to meet national quality standards: An analysis from the ASiT/BOTA Lost Tribe prospective cohort study of 2,569 surgical trainees.* Int J Surg, 2018. **52**: p. 376-382.
2. Meara, J.G., et al., *Global Surgery 2030: evidence and solutions for achieving health, welfare, and economic development.* Lancet, 2015. **386**(9993): p. 569-624.
3. Rose, J., et al., *Estimated need for surgery worldwide based on prevalence of diseases: a modelling strategy for the WHO Global Health Estimate.* Lancet Glob Health, 2015. **3 Suppl 2**: p. S13-20.
4. Meara, J.G., L. Hagander, and A.J. Leather, *Surgery and global health: a Lancet Commission.* Lancet, 2014. **383**(9911): p. 12-3.
5. Shrime, M.G., et al., *Global burden of surgical disease: an estimation from the provider perspective.* Lancet Glob Health, 2015. **3 Suppl 2**: p. S8-9.
6. Roa, L., et al., *Global surgery and the sustainable development goals.* Br J Surg, 2019. **106**(2): p. e44-e52.
7. Alkire, B.C., et al., *Global access to surgical care: a modelling study.* Lancet Glob Health, 2015. **3**(6): p. e316-23.
8. Davies, J.I. and J.G. Meara, *Global surgery--going beyond the Lancet Commission.* Lancet, 2015. **386**(9993): p. 507-9.
9. Shrime, M.G., et al., *Catastrophic expenditure to pay for surgery worldwide: a modelling study.* Lancet Glob Health, 2015. **3 Suppl 2**: p. S38-44.
10. Holmer, H., et al., *Global distribution of surgeons, anaesthesiologists, and obstetricians.* Lancet Glob Health, 2015. **3 Suppl 2**: p. S9-11.
11. Bouchard, M.E., et al., *Investments in surgical systems contribute to pandemic readiness and health system resilience.* J Public Health Policy, 2021. **42**(3): p. 493-500.
12. Hedt-Gauthier, B.L., et al., *Growing research in global surgery with an eye towards equity.* Br J Surg, 2019. **106**(2): p. e151-e155.
13. Esquivel, M.M., et al., *Proposed minimum rates of surgery to support desirable health outcomes: an observational study based on four strategies.* Lancet, 2015. **385 Suppl 2**: p. S12.
14. Meara, J.G., L. Hagander, and A.J.M. Leather, *Surgery and global health: a Lancet Commission.* Lancet, 2014. **383**(9911): p. 12-13.

15. Dare, A.J., et al., *Deaths from acute abdominal conditions and geographical access to surgical care in India: a nationally representative spatial analysis.* Lancet Glob Health, 2015. **3**(10): p. e646-53.
16. Chao, T.E., et al., *Cost-effectiveness of surgery and its policy implications for global health: a systematic review and analysis.* Lancet Glob Health, 2014. **2**(6): p. e334-45.
17. Alkire, B.C., et al., *Global economic consequences of selected surgical diseases: a modelling study.* Lancet Glob Health, 2015. **3 Suppl 2**(Suppl 2): p. S21-7.
18. NCD-Countdown-2030-collaborators, *NCD Countdown 2030: pathways to achieving Sustainable Development Goal target 3.4.* Lancet, 2020. **396**(10255): p. 918-934.
19. Weiser, T.G., et al., *Size and distribution of the global volume of surgery in 2012.* Bull World Health Organ, 2016. **94**(3): p. 201-209f.
20. COVIDSurg-Collaborative, *Elective surgery cancellations due to the COVID-19 pandemic: global predictive modelling to inform surgical recovery plans.* Br J Surg, 2020.
21. Odland, M.L., et al., *Identifying a Basket of Surgical Procedures to Standardize Global Surgical Metrics: An International Delphi Study.* Ann Surg, 2021. **274**(6): p. 1107-1114.
22. GlobalSurg-Collaborative, *Global variation in postoperative mortality and complications after cancer surgery: a multicentre, prospective cohort study in 82 countries.* Lancet, 2021. **397**(10272): p. 387-397.
23. GlobalSurg-Collaborative, *Surgical site infection after gastrointestinal surgery in high-income, middle-income, and low-income countries: a prospective, international, multicentre cohort study.* The Lancet Infectious Diseases, 2018.
24. GlobalSurg-Collaborative, *Mortality of emergency abdominal surgery in high-, middle- and low-income countries.* Br J Surg, 2016. **103**(8): p. 971-88.
25. Biccard, B.M., et al., *Perioperative patient outcomes in the African Surgical Outcomes Study: a 7-day prospective observational cohort study.* Lancet, 2018.
26. Kahan, B.C., et al., *Critical care admission following elective surgery was not associated with survival benefit: prospective analysis of data from 27 countries.* Intensive Care Med, 2017. **43**(7): p. 971-979.
27. Daniels, K.M., J.N. Riesel, and J.G. Meara, *The scale-up of the surgical workforce.* Lancet, 2015. **385 Suppl 2**: p. S41.
28. Grimes, C.E., et al., *Systematic review of met and unmet need of surgical disease in rural sub-Saharan Africa.* World J Surg, 2012. **36**(1): p. 8-23.

29. Verguet, S., et al., *Timing and cost of scaling up surgical services in low-income and middle-income countries from 2012 to 2030: a modelling study.* Lancet Glob Health, 2015. **3 Suppl 2**: p. S28-37.
30. Fowler, A.J., et al., *Resource requirements for reintroducing elective surgery during the COVID-19 pandemic: modelling study.* Br J Surg, 2021. **108**(1): p. 97-103.
31. COVIDSurg-Collaborative, *Projecting COVID-19 disruption to elective surgery.* Lancet, 2022. **399**(10321): p. 233-234.
32. Nepogodiev, D., et al., *Global burden of postoperative death.* Lancet, 2019. **393**(10170): p. 401.
33. NIHR-Global-Health-Research-Unit-on-Global-Surgery, *Global guidelines for emergency general surgery: systematic review and Delphi prioritization process.* BJS Open, 2022. **6**(1).
34. GlobalSurg-Collaborative, *Laparoscopy in management of appendicitis in high-, middle-, and low-income countries: a multicenter, prospective, cohort study.* Surg Endosc, 2018. **32**(8): p. 3450-3466.
35. Patel, J., et al., *Perioperative Care Pathways in Low- and Lower-Middle-Income Countries: Systematic Review and Narrative Synthesis.* World J Surg, 2022.
36. Zilla, P., et al., *Global Unmet Needs in Cardiac Surgery.* Glob Heart, 2018. **13**(4): p. 293-303.
37. Watters, D.A., et al., *Surgical training programmes in the South Pacific, Papua New Guinea and Timor Leste.* Br J Surg, 2019. **106**(2): p. e53-e61.
38. Myles, P.S. and G. Haller, *Global distribution of access to surgical services.* Lancet, 2010. **376**(9746): p. 1027-8.
39. Parvin-Nejad, F.P., et al., *Surgical subspecialists in West Africa: Workforce size, training opportunities, and contributing factors.* Surgery, 2021. **170**(2): p. 478-484.
40. NIHR-Global-Health-Research-Unit-on-Global-Surgery, *Prioritizing research for patients requiring surgery in low- and middle-income countries.* Br J Surg, 2019. **106**(2): p. e113-e120.
41. Kilama, W.L., *The 10/90 gap in sub-Saharan Africa: resolving inequities in health research.* Acta Trop, 2009. **112 Suppl 1**: p. S8-s15.
42. Heneghan, C., et al., *Evidence for non-communicable diseases: analysis of Cochrane reviews and randomised trials by World Bank classification.* BMJ Open, 2013. **3**(7).
43. NIHR-Global-Health-Research-Unit-on-Global-Surgery, *Reducing surgical site infections in low-income and middle-income countries (FALCON): a pragmatic,*

multicentre, stratified, randomised controlled trial. Lancet, 2021. **398**(10312): p. 1687-1699.

44. NIHR-Global-Health-Research-Unit-on-Global-Surgery. *Cheetah - Sterile Glove and Clean Instrument Change at the Time of Wound Closure to Reduce Surgical Site Infection (Cheetah) (NCT03980652)*. 2019 [cited 2021 12th February]; Available from: https://clinicaltrials.gov/ct2/show/NCT03980652.
45. NIHR-Global-Health-Research-Unit-on-Global-Surgery. *Perioperative Respiratory Care and Outcomes for Patients Undergoing High Risk Abdominal Surgery (PENGUIN) (NCT04256798)*. 2020 [cited 2021 12th February]; Available from: https://clinicaltrials.gov/ct2/show/NCT04256798.
46. COVIDSurg-Collaborative, *Timing of surgery following SARS-CoV-2 infection: an international prospective cohort study.* Anaesthesia, 2021. **76**(6): p. 748-758.
47. COVIDSurg-Collaborative, *Effect of COVID-19 pandemic lockdowns on planned cancer surgery for 15 tumour types in 61 countries: an international, prospective, cohort study.* Lancet Oncol, 2021. **22**(11): p. 1507-1517.
48. COVIDSurg-Collaborative, *Elective Cancer Surgery in COVID-19-Free Surgical Pathways During the SARS-CoV-2 Pandemic: An International, Multicenter, Comparative Cohort Study.* J Clin Oncol, 2020: p. Jco2001933.
49. COVIDSurg-Collaborative, *Mortality and pulmonary complications in patients undergoing surgery with perioperative SARS-CoV-2 infection: an international cohort study.* Lancet, 2020. **396**(10243): p. 27-38.
50. NIHR-Global-Health-Research-Unit-on-Global-Surgery, *Delphi prioritization and development of global surgery guidelines for the prevention of surgical-site infection.* Br J Surg, 2020. **107**(8): p. 970-977.
51. COVIDSurg-Collaborative, *Global guidance for surgical care during the COVID-19 pandemic.* Br J Surg, 2020. **107**(9): p. 1097-1103.
52. Mangram, A.J., et al., *Guideline for Prevention of Surgical Site Infection, 1999. Centers for Disease Control and Prevention (CDC) Hospital Infection Control Practices Advisory Committee.* Am J Infect Control, 1999. **27**(2): p. 97-132; quiz 133-4; discussion 96.
53. Owens, C.D. and K. Stoessel, *Surgical site infections: epidemiology, microbiology and prevention.* J Hosp Infect, 2008. **70 Suppl 2**: p. 3-10.
54. Clavien, P.A., et al., *The Clavien-Dindo classification of surgical complications: five-year experience.* Ann Surg, 2009. **250**(2): p. 187-96.

55. Abbas, M. and D. Pittet, *Surgical site infection prevention: a global priority.* J Hosp Infect, 2016. **93**(4): p. 319-22.
56. NIHR-Global-Health-Research-Unit-on-Global-Surgery, *Meta-analysis of high quality randomised controlled trials testing alcoholic chlorhexidine skin preparation or triclosan coated sutures to reduce surgical site infection.* Lancet Infectious Disease, 2022. **[accepted, ahead of print]**.
57. Gheorghe, A., et al., *Health Utility Values Associated with Surgical Site Infection: A Systematic Review.* Value Health, 2015. **18**(8): p. 1126-37.
58. Rosenthal, V.D., et al., *Surgical site infections, International Nosocomial Infection Control Consortium (INICC) report, data summary of 30 countries, 2005-2010.* Infect Control Hosp Epidemiol, 2013. **34**(6): p. 597-604.
59. Shepard, J., et al., *Financial impact of surgical site infections on hospitals: the hospital management perspective.* JAMA Surg, 2013. **148**(10): p. 907-14.
60. Bagheri Nejad, S., et al., *Health-care-associated infection in Africa: a systematic review.* Bull World Health Organ, 2011. **89**(10): p. 757-65.
61. Andersson, A.E., et al., *Patients' experiences of acquiring a deep surgical site infection: an interview study.* Am J Infect Control, 2010. **38**(9): p. 711-7.
62. Coello, R., et al., *Adverse impact of surgical site infections in English hospitals.* J Hosp Infect, 2005. **60**(2): p. 93-103.
63. Leaper, D.J., et al., *Surgical site infection - a European perspective of incidence and economic burden.* Int Wound J, 2004. **1**(4): p. 247-73.
64. Allegranzi, B., et al., *Burden of endemic health-care-associated infection in developing countries: systematic review and meta-analysis.* Lancet, 2011. **377**(9761): p. 228-41.
65. Holmer, H., et al., *Evaluating the collection, comparability and findings of six global surgery indicators.* Br J Surg, 2019. **106**(2): p. e138-e150.
66. Badia, J.M., et al., *Impact of surgical site infection on healthcare costs and patient outcomes: a systematic review in six European countries.* J Hosp Infect, 2017. **96**(1): p. 1-15.
67. Shrime, M.G., et al., *A global country-level comparison of the financial burden of surgery.* Br J Surg, 2016. **103**(11): p. 1453-61.
68. Shrime, M., et al., *Catastrophic Expenditure to Pay for Surgery: A Global Estimate.* Lancet Glob Health, 2016. **3**(38-44).

69. Nkurunziza, T., et al., *Prevalence and predictors of surgical-site infection after caesarean section at a rural district hospital in Rwanda.* Br J Surg, 2019. **106**(2): p. e121-e128.
70. Allegranzi, B., et al., *New WHO recommendations on preoperative measures for surgical site infection prevention: an evidence-based global perspective.* Lancet Infect Dis, 2016. **16**(12): p. e276-e287.
71. Leblebicioglu, H., et al., *Surgical site infection rates in 16 cities in Turkey: findings of the International Nosocomial Infection Control Consortium (INICC).* Am J Infect Control, 2015. **43**(1): p. 48-52.
72. Ramirez-Wong, F.M., et al., *Surgical Site Infections Rates in More Than 13,000 Surgical Procedures in Three Cities in Peru: Findings of the International Nosocomial Infection Control Consortium.* Surg Infect (Larchmt), 2015. **16**(5): p. 572-6.
73. Singh, S., et al., *Surgical site infection rates in six cities of India: findings of the International Nosocomial Infection Control Consortium (INICC).* Int Health, 2015. **7**(5): p. 354-9.
74. Chu, K., R. Maine, and M. Trelles, *Cesarean section surgical site infections in sub-Saharan Africa: a multi-country study from Medecins Sans Frontieres.* World J Surg, 2015. **39**(2): p. 350-5.
75. Itatsu, K., et al., *Risk factors for incisional surgical site infections in elective surgery for colorectal cancer: focus on intraoperative meticulous wound management.* Surg Today, 2014. **44**(7): p. 1242-52.
76. Alvarez-Moreno, C., et al., *Surgical site infection rates in 4 cities in Colombia: findings of the International Nosocomial Infection Control Consortium (INICC).* Am J Infect Control, 2014. **42**(10): p. 1089-92.
77. Portillo-Gallo, J.H., et al., *Surgical site infection rates in four Mexican cities: findings of the International Nosocomial Infection Control Consortium (INICC).* J Infect Public Health, 2014. **7**(6): p. 465-71.
78. Pathak, A., et al., *Incidence and factors associated with surgical site infections in a teaching hospital in Ujjain, India.* Am J Infect Control, 2014. **42**(1): p. e11-5.
79. Mangram, A.J., et al., *Guideline for prevention of surgical site infection, 1999. Hospital Infection Control Practices Advisory Committee.* Infect Control Hosp Epidemiol, 1999. **20**(4): p. 250-78; quiz 279-80.
80. Forrester, J.A., et al., *Clean Cut (adaptive, multimodal surgical infection prevention programme) for low-resource settings: a prospective quality improvement study.* Br J Surg, 2020.

81. Ravi, K., et al., *Systematic analysis of authorship demographics in global surgery.* BMJ Glob Health, 2021. **6**(10).
82. Khan, M., et al., *Decolonising global health in 2021: a roadmap to move from rhetoric to reform*, in *BMJ Glob Health*. 2021.
83. Abimbola, S. and M. Pai, *Will global health survive its decolonisation?* Lancet, 2020. **396**(10263): p. 1627-1628.
84. Nafade, V., P. Sen, and M. Pai, *Global health journals need to address equity, diversity and inclusion.* BMJ Glob Health, 2019. **4**(5): p. e002018.
85. Abimbola, S., et al., *Addressing power asymmetries in global health: Imperatives in the wake of the COVID-19 pandemic.* PLoS Med, 2021. **18**(4): p. e1003604.
86. Tembo, D., et al., *Effective engagement and involvement with community stakeholders in the co-production of global health research.* Bmj, 2021. **372**: p. n178.
87. The-Lancet-Global-Health, *Closing the door on parachutes and parasites.* Lancet Glob Health, 2018. **6**(6): p. e593.
88. Sam-Agudu, N.A. and S. Abimbola, *Using scientific authorship criteria as a tool for equitable inclusion in global health research*, in *BMJ Glob Health*. 2021.
89. National-Research-Collaborative, *Recognising contributions to work in research collaboratives: Guidelines for standardising reporting of authorship in collaborative research.* Int J Surg, 2017.
90. Wilson, A.P., et al., *A scoring method (ASEPSIS) for postoperative wound infections for use in clinical trials of antibiotic prophylaxis.* Lancet, 1986. **1**(8476): p. 311-3.
91. Public-Health-England. *Protocol for the Surveillance of Surgical Site Infection: Surgical Site Infection Surveillance Service*. 2013 [cited 2022 27th September]; 6th Edition:[Available from: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1048707/Protocol_for_the_Surveillance_of_Surgical_Site_Infection.pdf.
92. Matthews, J.H., et al., *Underreporting of Secondary Endpoints in Randomized Trials: Cross-sectional, Observational Study.* Ann Surg, 2016. **264**(6): p. 982-986.
93. Horan, T.C., M. Andrus, and M.A. Dudeck, *CDC/NHSN surveillance definition of health care-associated infection and criteria for specific types of infections in the acute care setting.* Am J Infect Control, 2008. **36**(5): p. 309-32.
94. Antimicrobial-Resistance-Collaborators, *Global burden of bacterial antimicrobial resistance in 2019: a systematic analysis.* Lancet, 2022. **399**(10325): p. 629-655.
95. Laxminarayan, R., et al., *The Lancet Infectious Diseases Commission on antimicrobial resistance: 6 years later.* Lancet Infect Dis, 2020. **20**(4): p. e51-e60.

96. Sterne, J.A.C., et al., *RoB 2: a revised tool for assessing risk of bias in randomised trials.* Bmj, 2019. **366**: p. l4898.
97. Barratt, H., et al., *Mixed methods evaluation of the Getting it Right First Time programme - improvements to NHS orthopaedic care in England: study protocol.* BMC Health Serv Res, 2017. **17**(1): p. 71.
98. Getting-It-Right-First-Time. *GIRFT Surgical Site Infection Audit - Getting It Right First Time - GIRFT*. 2017 [cited 2017 29th August 2017]; Available from: http://gettingitrightfirsttime.co.uk/ssi-audit/.
99. Forrester, J.A., et al., *Development of a Surgical Infection Surveillance Program at a Tertiary Hospital in Ethiopia: Lessons Learned from Two Surveillance Strategies.* Surg Infect (Larchmt), 2017.
100. Richtmann, R., et al., *Surgical Site Infection Rates in Four Cities in Brazil: Findings of the International Nosocomial Infection Control Consortium.* Surg Infect (Larchmt), 2016. **17**(1): p. 53-7.
101. Woelber, E., et al., *Proportion of Surgical Site Infections Occurring after Hospital Discharge: A Systematic Review.* Surg Infect (Larchmt), 2016. **17**(5): p. 510-9.
102. Reeves, N., et al., *Diagnosis of colorectal and emergency surgical site infections in the era of enhanced recovery: an all-Wales prospective study.* Colorectal Dis, 2021.
103. Gundel, O., et al., *Timing of surgical site infection and pulmonary complications after laparotomy.* Int J Surg, 2018. **52**: p. 56-60.
104. Pinkney, T.D., et al., *Impact of wound edge protection devices on surgical site infection after laparotomy: multicentre randomised controlled trial (ROSSINI Trial).* Bmj, 2013. **347**: p. f4305.
105. Ljungqvist, O., M. Scott, and K.C. Fearon, *Enhanced Recovery After Surgery: A Review.* JAMA Surg, 2017. **152**(3): p. 292-298.
106. Chemali, M.E. and G.D. Eslick, *A Meta-Analysis: Postoperative Pain Management in Colorectal Surgical Patients and the Effects on Length of Stay in an Enhanced Recovery After Surgery (ERAS) Setting.* Clin J Pain, 2017. **33**(1): p. 87-92.
107. Ripollés-Melchor, J., et al., *Association Between Use of Enhanced Recovery After Surgery Protocol and Postoperative Complications in Colorectal Surgery: The Postoperative Outcomes Within Enhanced Recovery After Surgery Protocol (POWER) Study.* JAMA Surg, 2019. **154**(8): p. 725-736.
108. Kehlet, H., *Postoperative recovery: DrEaMing as a wake-up call?* Br J Anaesth, 2022. **129**(1): p. 1-3.

109. Oliver, C.M., et al., *Delivery of drinking, eating and mobilising (DrEaMing) and its association with length of hospital stay after major noncardiac surgery: observational cohort study.* Br J Anaesth, 2022. **129**(1): p. 114-126.
110. Lee, Y., et al., *Enhanced recovery after surgery (ERAS) versus standard recovery for elective gastric cancer surgery: A meta-analysis of randomized controlled trials.* Surg Oncol, 2020. **32**: p. 75-87.
111. Małczak, P., et al., *Enhanced Recovery after Bariatric Surgery: Systematic Review and Meta-Analysis.* Obes Surg, 2017. **27**(1): p. 226-235.
112. Hajibandeh, S., V. Bill, and T. Satyadas, *Meta-analysis of Enhanced Recovery After Surgery (ERAS) Protocols in Emergency Abdominal Surgery.* World J Surg, 2020. **44**(5): p. 1336-1348.
113. Raykar, N.P., et al., *Geospatial mapping to estimate timely access to surgical care in nine low-income and middle-income countries.* Lancet, 2015. **385 Suppl 2**: p. S16.
114. Muka, T., et al., *The global impact of non-communicable diseases on healthcare spending and national income: a systematic review.* Eur J Epidemiol, 2015. **30**(4): p. 251-77.
115. Buyse, M., et al., *Statistical Validation of Surrogate Endpoints: Problems and Proposals:.* http://dx.doi.org/10.1177/009286150003400213, 2016.
116. Prentice, R., *Surrogate endpoints in clinical trials: definition and operational criteria.* Statistics in medicine, 1989. **8**(4).
117. Alonso, A. and G. Molenberghs, *Surrogate marker evaluation from an information theory perspective.* Biometrics, 2007. **63**(1): p. 180-6.
118. Bluebelle-Study-Group, Severn-and-Peninsula-Audit-and-Research-Collaborative-for-Surgeons, and West-Midlands-Research-Collaborative, *Bluebelle study (phase A): a mixed-methods feasibility study to inform an RCT of surgical wound dressing strategies.* BMJ Open, 2016. **6**(9): p. e012635.
119. Tin, W., et al., *Outcome of very preterm birth: children reviewed with ease at 2 years differ from those followed up with difficulty.* Arch Dis Child Fetal Neonatal Ed, 1998. **79**(2): p. F83-7.
120. Akl, E., et al., *Potential impact on estimated treatment effects of information lost to follow-up in randomised controlled trials (LOST-IT): systematic review.* Bmj, 2012. **344**: p. e2809.
121. Schulz, K. and D. Grimes, *Sample size slippages in randomised trials: exclusions and the lost and wayward.* Lancet, 2002. **359**(9308): p. 781-5.

122. Brunsdon, D., et al., *What are the most important unanswered research questions in trial retention? A James Lind Alliance Priority Setting Partnership: the PRioRiTy II (Prioritising Retention in Randomised Trials) study.* Trials, 2019. **20**(1): p. 593.
123. Sandberg, C., et al., *Using Telemedicine to Diagnose Surgical Site Infections in Low- and Middle-Income Countries: Systematic Review.* JMIR Mhealth Uhealth, 2019. **7**(8): p. e13309.
124. Harkey, K., et al., *Postdischarge Virtual Visits for Low-risk Surgeries: A Randomized Noninferiority Clinical Trial.* JAMA Surg, 2021. **156**(3): p. 221-228.
125. Barsom, E.Z., et al., *Video consultation during follow up care: effect on quality of care and patient- and provider attitude in patients with colorectal cancer.* Surg Endosc, 2021. **35**(3): p. 1278-1287.
126. Dorsey, E.R. and E.J. Topol, *Telemedicine 2020 and the next decade.* Lancet, 2020. **395**(10227): p. 859.
127. Sonderman, K.A., et al., *Using mobile health technology and community health workers to identify and refer caesarean-related surgical site infections in rural Rwanda: a randomised controlled trial protocol.* BMJ Open, 2018. **8**(5): p. e022214.
128. Totty, J.P., et al., *Use of photograph-based telemedicine in postoperative wound assessment to diagnose or exclude surgical site infection.* J Wound Care, 2018. **27**(3): p. 128-135.
129. Arie, S., *Can mobile phones transform healthcare in low and middle income countries?* Bmj, 2015. **350**: p. h1975.
130. Bastawrous, A. and M.J. Armstrong, *Mobile health use in low- and high-income countries: an overview of the peer-reviewed literature*, in *J R Soc Med*. 2013. p. 130-42.
131. McLean, K.A., et al., *Remote diagnosis of surgical-site infection using a mobile digital intervention: a randomised controlled trial in emergency surgery patients.* NPJ Digit Med, 2021. **4**(1): p. 160.
132. Kummerow Broman, K., et al., *Evaluation of Wound Photography for Remote Postoperative Assessment of Surgical Site Infections.* JAMA Surg, 2019. **154**(2): p. 117-124.
133. Wild, D., et al., *Principles of Good Practice for the Translation and Cultural Adaptation Process for Patient-Reported Outcomes (PRO) Measures: report of the ISPOR Task Force for Translation and Cultural Adaptation.* Value Health, 2005. **8**(2): p. 94-104.

134. Tennant, A., et al., *Assessing and adjusting for cross-cultural validity of impairment and activity limitation scales through differential item functioning within the framework of the Rasch model: the PRO-ESOR project.* Med Care, 2004. **42**(1 Suppl): p. I37-48.
135. Eze, N., C. Mateus, and T. Cravo Oliveira Hashiguchi, *Telemedicine in the OECD: An umbrella review of clinical and cost-effectiveness, patient experience and implementation.* PloS one, 2020. **15**(8).
136. de la Torre-Díez, I., et al., *Cost-utility and cost-effectiveness studies of telemedicine, electronic, and mobile health systems in the literature: a systematic review.* Telemedicine journal and e-health : the official journal of the American Telemedicine Association, 2015. **21**(2).
137. COVIDSurg-Collaborative, *Death following pulmonary complications of surgery before and during the SARS-CoV-2 pandemic.* Br J Surg, 2021. **108**(12): p. 1448-1464.
138. COVIDSurg-Collaborative, *Machine learning risk prediction of mortality for patients undergoing surgery with perioperative SARS-CoV-2: the COVIDSurg mortality score.* Br J Surg, 2021. **108**(11): p. 1274-1292.
139. Simoes, J. and A. Bhangu, *Should we be re-starting elective surgery?* Anaesthesia, 2020. **75**(12): p. 1563-1565.
140. COVIDSurg-Collaborative, *Preoperative nasopharyngeal swab testing and postoperative pulmonary complications in patients undergoing elective surgery during the SARS-CoV-2 pandemic.* Br J Surg, 2021. **108**(1): p. 88-96.
141. Glasbey, J.C. and A. Bhangu, *Elective Cancer Surgery in COVID-19-Free Surgical Pathways During the SARS-CoV-2 Pandemic: An International, Multicenter, Comparative Cohort Study.* J Clin Oncol, 2020: p. Jco2001933.
142. COVIDSurg-Collaborative. *CovidSurg-3: Outcomes of Surgery in COVID-19 Infection (CovidSurg-3)*. 2021 [cited 2022 25th February].
143. Fryer, K., et al., *Implementation of Obstetric Telehealth During COVID-19 and Beyond.* Matern Child Health J, 2020. **24**(9): p. 1104-1110.
144. Hakim, A.A., et al., *Implications for the use of telehealth in surgical patients during the COVID-19 pandemic.* Am J Surg, 2020. **220**(1): p. 48-49.
145. Maaløe, N., et al., *The injustice of unfit clinical practice guidelines in low-resource realities.* Lancet Glob Health, 2021.
146. The-Lancet-Oncology, *Valuing all lives equally: cancer surgery, COVID-19, and the NHS in crisis.* Lancet Oncol, 2021. **22**(2): p. 155.

147. Chakrabarti, S., et al., *Association of Human Mobility Restrictions and Race/Ethnicity-Based, Sex-Based, and Income-Based Factors With Inequities in Well-being During the COVID-19 Pandemic in the United States.* JAMA Netw Open, 2021. **4**(4): p. e217373.
148. ACCCOS-Collaborators, *Patient care and clinical outcomes for patients with COVID-19 infection admitted to African high-care or intensive care units (ACCCOS): a multicentre, prospective, observational cohort study.* Lancet, 2021. **397**(10288): p. 1885-1894.
149. McGillion, M., et al., *Post-discharge after surgery Virtual Care with Remote Automated Monitoring-1 (PVC-RAM-1) technology versus standard care: randomised controlled trial.* BMJ (Clinical research ed.), 2021. **374**.
150. Augestad, K., A. Sneve, and R. Lindsetmo, *Telemedicine in postoperative follow-up of STOMa PAtients: a randomized clinical trial (the STOMPA trial).* Br J Surg, 2020. **107**(5): p. 509-518.
151. Lal, A., et al., *Minimizing COVID-19 disruption: Ensuring the supply of essential health products for health emergencies and routine health services.* Lancet Reg Health Am, 2022. **6**: p. 100129.
152. Pyda, J., et al., *Towards resilient health systems: opportunities to align surgical and disaster planning.* BMJ Glob Health, 2019. **4**(3): p. e001493.
153. Federspiel, F., et al., *Global surgical and anaesthetic task shifting: a systematic literature review and survey.* Lancet, 2015. **385 Suppl 2**: p. S46.
154. Chu, K., et al., *Surgical task shifting in Sub-Saharan Africa.* PLoS Med, 2009. **6**(5): p. e1000078.
155. Kruk, M.E., et al., *Building resilient health systems: a proposal for a resilience index.* Bmj, 2017. **357**: p. j2323.
156. Berrios-Torres, S.I., et al., *Centers for disease control and prevention guideline for the prevention of surgical site infection, 2017.*
157. Ahmad, T., et al., *Use of failure-to-rescue to identify international variation in postoperative care in low-, middle- and high-income countries: a 7-day cohort study of elective surgery.* Br J Anaesth, 2017. **119**(2): p. 258-266.
158. COVIDSurg-Collaborative, *Surgery during the COVID-19 pandemic – Authors' reply.* Lancet, 2020. **396**(10261): p. E79.
159. Chao, G.F., et al., *Use of Telehealth by Surgical Specialties During the COVID-19 Pandemic.* JAMA Surg, 2021.

160. Grenda, T.R., S. Whang, and N.R. Evans, 3rd, *Transitioning a Surgery Practice to Telehealth During COVID-19.* Ann Surg, 2020. **272**(2): p. e168-e169.
161. Carr, A., et al., *Growing backlog of planned surgery due to covid-19.* Bmj, 2021. **372**: p. n339.
162. Barie, P.S., et al., *Surgical Infection Society Guidance for Restoration of Surgical Services during the Coronavirus Disease-2019 Pandemic.* Surg Infect (Larchmt), 2021.
163. Nguhuni, B., et al., *Reliability and validity of using telephone calls for post-discharge surveillance of surgical site infection following caesarean section at a tertiary hospital in Tanzania.* Antimicrob Resist Infect Control, 2017. **6**: p. 43.
164. Collaborative, G., *Determining the worldwide epidemiology of surgical site infections after gastrointestinal resection surgery: protocol for a multicentre, international, prospective cohort study (GlobalSurg 2).* BMJ Open, 2017. **7**(7): p. e012150.
165. Ho, D., et al., *Matching as Nonparametric Preprocessing for Reducing Model Dependence in Parametric Causal Inference.* Political Analysis, 2007. **15**: p. 199–236.
166. Rubin, D.B., *The design versus the analysis of observational studies for causal effects: parallels with the design of randomized trials.* Stat Med, 2007. **26**(1): p. 20-36.
167. Ouzzani, M., et al., *Rayyan-a web and mobile app for systematic reviews.* Systematic reviews, 2016. **5**(1).
168. Abu-Sheasha, G.A., et al., *Cost-effectiveness analysis of three methods of surgical-site infection surveillance: Less is more.*
169. Pathak, A., et al., *Feasibility of a Mobile Phone-Based Surveillance for Surgical Site Infections in Rural India.* Telemedicine journal and e-health : the official journal of the American Telemedicine Association, 2015. **21**(11).
170. McIntyre, L.K., et al., *The incidence of post-discharge surgical site infection in the injured patient.*
171. Reilly, J., et al., *A study of telephone screening and direct observation of surgical wound infections after discharge from hospital.*
172. Bediako-Bowan, A., et al., *Surveillance of surgical site infection in a teaching hospital in Ghana: a prospective cohort study.*
173. Golub, A.V., et al., *Surgical Site Infections after Open Appendectomy and Effectiveness of Complex Approach to Their Prevention. [Russian].*
174. Pham, J.C., et al., *Surgical site infection: Comparing surgeon versus patient self-report.*

175. Petrosillo, N., et al., *Surgical site infections in Italian hospitals: A prospective multicenter study.*
176. A, P., et al., *Feasibility of a Mobile Phone-Based Surveillance for Surgical Site Infections in Rural India.* Telemedicine journal and e-health : the official journal of the American Telemedicine Association, 2015. **21**(11).
177. Burlingame, J., et al., *Assessment of cesarean delivery surgical site infection: A comparison of methods.*
178. Pathak, A., et al., *Feasibility of a Mobile Phone-Based Surveillance for Surgical Site Infections in Rural India.*
179. Raykar, N.P., et al., *A qualitative study exploring contextual challenges to surgical care provision in 21 LMICs.* Lancet, 2015. **385 Suppl 2**(Suppl 2): p. S15.
180. Shaw, R., *UK Head and neck cancer surgical capacity during the second wave of the COVID-19 pandemic: Have we learned the lessons? COVIDSurg collaborative.* Clin Otolaryngol, 2021.
181. Hanna, T.P., et al., *Mortality due to cancer treatment delay: systematic review and meta-analysis.* Bmj, 2020. **371**: p. m4087.
182. Sud, A., et al., *Collateral damage: the impact on outcomes from cancer surgery of the COVID-19 pandemic.* Ann Oncol, 2020. **31**(8): p. 1065-1074.
183. Ghaferi, A.A., J.D. Birkmeyer, and J.B. Dimick, *Variation in hospital mortality associated with inpatient surgery.* N Engl J Med, 2009. **361**(14): p. 1368-75.
184. Macefield, R.C., et al., *Development of a single, practical measure of surgical site infection (SSI) for patient report or observer completion.* J Infect Prev, 2017. **18**(4): p. 170-179.
185. Reeves, B.C., et al., *A mixed-methods feasibility and external pilot study to inform a large pragmatic randomised controlled trial of the effects of surgical wound dressing strategies on surgical site infections (Bluebelle Phase B): study protocol for a randomised controlled trial.* Trials, 2017. **18**(1): p. 401.
186. NIHR-Global-Health-Research-Unit-on-Global-Surgery, *Feasibility and diagnostic accuracy of Telephone Administration of an adapted patient-reported wound heaLing QuestiONnaire for assessment of surgical site infection following abdominal surgery in low and middle-income countries (TALON): Protocol for a validation study within a trial (SWAT).* Trials, 2021. **Accepted ahead of print**.
187. Karzai, F., R.A. Madan, and W.L. Dahut, *The World of Clinical Trial Development Post COVID-19: Lessons Learned from a Global Pandemic.* Clin Cancer Res, 2020. **26**(16): p. 4198-4200.

188. Lorusso, D., et al., *Clinical research disruption in the post-COVID-19 era: will the pandemic lead to change?* ESMO Open, 2020. **5**(5): p. e000924.
189. Imrey, P., *Limitations of Meta-analyses of Studies With High Heterogeneity.* JAMA network open, 2020. **3**(1).
190. Loudon, K., et al., *The PRECIS-2 tool: designing trials that are fit for purpose.* Bmj, 2015. **350**: p. h2147.
191. Macefield, R., et al., *Development of a 'universal-reporter' outcome measure (UROM) for patient and healthcare professional completion: a mixed methods study demonstrating a novel concept for optimal questionnaire design.* BMJ Open, 2019. **9**(8): p. e029741.
192. Bluebelle-Study-Group, *Validation of the Bluebelle Wound Healing Questionnaire for assessment of surgical-site infection in closed primary wounds after hospital discharge.* Br J Surg, 2019. **106**(3): p. 226-235.
193. Elliott, D. and Bluebelle-Study-Group, *Developing outcome measures assessing wound management and patient experience: a mixed methods study.* BMJ Open, 2017. **7**(11): p. e016155.
194. Nepogodiev, D. and A. Bhangu, *Pragmatic multicentre factorial randomized controlled trial testing measures to reduce surgical site infection in low- and middle-income countries: study protocol of the FALCON trial.* Colorectal Dis, 2020.
195. NIHR-Global-Health-Research-Unit-on-Global-Surgery, *Global use of telemedicine for post-discharge assessment of the surgical wound: international cohort study, and systematic review with meta-analysis*. 2022: Annals of Surgery.
196. Tong, A., P. Sainsbury, and J. Craig, *Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups.* Int J Qual Health Care, 2007. **19**(6): p. 349-57.
197. Gaglio, B., et al., *Methodological standards for qualitative and mixed methods patient centered outcomes research.* Bmj, 2020. **371**: p. m4435.
198. Prakash, V., S. Shah, and K. Hariohm, *Cross-cultural adaptation of patient-reported outcome measures: A solution or a problem?* Ann Phys Rehabil Med, 2019. **62**(3): p. 174-177.
199. Creswell, J. and V. Plano Clark, *Designing and Conducting Mixed Methods Research*. 2007, SAGE Publishing.
200. Harris, P.A., et al., *Research electronic data capture (REDCap)--a metadata-driven methodology and workflow process for providing translational research informatics support.* J Biomed Inform, 2009. **42**(2): p. 377-81.

201. Reynolds, J., et al. *Quality assurance of qualitative research: a suggested approach for assessing and strengthening qualitative research within global health trials*. 2019 [cited 2019 22nd November]; Available from: https://globalhealthtrials.tghn.org/articles/quality-assurance-qualitative-research-suggested-approach-assessing-and-strengthening-qualitative-research-within-global-health/.
202. Treweek, S., et al., *Trial Forge Guidance 1: what is a Study Within A Trial (SWAT)?*, in *Trials*. 2018: England. p. 139.
203. Baena, I. and J. Padilla, *Cognitive interviewing in mixed research*, in *Cognitive interviewing methodology*. 2014, John Wiley & Sons: London. p. 133-152.
204. COVIDSurg-Collaborative and STARSurg-Collaborative, *Death following pulmonary complications of surgery before and during the SARS-CoV-2 pandemic.* Br J Surg, 2021. **108**(12): p. 1448-1464.
205. Willis, G., *Cognitive Interviewing, Chapter 7: Analyzing and Documenting Cognitive Interview Results*. 2011, SAGE Publications, Inc.
206. Ames, H., C. Glenton, and S. Lewin, *Purposive sampling in a qualitative evidence synthesis: a worked example from a synthesis on parental perceptions of vaccination communication.* BMC medical research methodology, 2019. **19**(1).
207. Battaglia, M., *Purposive Sample*, in *Encyclopedia of Survey Research Methods*, P. Lavrakas, Editor. 2008, Sage Publications Inc. p. 645-647.
208. Ponterotto, J., *Brief Note on the Origins, Evolution, and Meaning of the Qualitative Research Concept Thick Description.* The Qualitative Report, 2006. **11**.
209. Birt, L., et al., *Member Checking: A Tool to Enhance Trustworthiness or Merely a Nod to Validation?* Qual Health Res, 2016. **26**(13): p. 1802-1811.
210. Clarke, A., *Focus group interviews in health-care research.* Professional nurse (London, England), 1999. **14**(6).
211. Mansell, I., et al., *The learning curve: the advantages and disadvantages in the use of focus groups as a method of data collection.* Nurse researcher, 2004. **11**(4).
212. Malterud, K., V.D. Siersma, and A.D. Guassora, *Sample Size in Qualitative Interview Studies: Guided by Information Power.* Qual Health Res, 2016. **26**(13): p. 1753-1760.
213. Swaine-Verdier, A., et al., *Adapting quality of life instruments.* Value Health, 2004. **7 Suppl 1**: p. S27-30.
214. Hagell, P., et al., *Effects of method of translation of patient-reported health outcome questionnaires: a randomized study of the translation of the Rheumatoid Arthritis Quality of Life (RAQoL) Instrument for Sweden.* Value Health, 2010. **13**(4): p. 424-30.

215. Beaton, D.E., et al., *Guidelines for the process of cross-cultural adaptation of self-report measures.* Spine (Phila Pa 1976), 2000. **25**(24): p. 3186-91.
216. Linacre, M. *Sample Size and Item Calibration(or Person measure) stability: Rasch Measurement Transaction;*. 1994; Available from: https://www.rasch.org/rmt/rmt74m.htm.
217. Tesio, L., et al., *WHY QUESTIONNAIRE SCORES ARE NOT MEASURES: A QUESTION-RAISING ARTICLE.* Am J Phys Med Rehabil, 2022.
218. Hughes, S.E., et al., *Development and validation of the symptom burden questionnaire for long covid (SBQ-LC): Rasch analysis.* Bmj, 2022. **377**: p. e070230.
219. Andrich, D., *Rating scales and Rasch measurement.* Expert Rev Pharmacoecon Outcomes Res, 2011. **11**(5): p. 571-85.
220. Tennant, A. and P.G. Conaghan, *The Rasch measurement model in rheumatology: what is it and why use it? When should it be applied, and what should one look for in a Rasch paper?* Arthritis Rheum, 2007. **57**(8): p. 1358-62.
221. Reeve, B.B., et al., *Comparing Cognitive Interviewing and Psychometric Methods to Evaluate a Racial/Ethnic Discrimination Scale.* Field methods, 2011. **23**(4): p. 397-419.
222. Benítez, I. and J. Padilla, *Analysis of nonequivalent assessments across different linguistic groups using a mixed methods approach: Understanding the causes of differential item functioning by cognitive interviewing.* Journal of Mixed Methods Research, 2014. **8**: p. 52-68.
223. Latcheva, R., *Cognitive interviewing and factor-analytic techniques: a mixed method approach to validity of survey items measuring national identity.* Quality & Quantity, 2011. **45**(6): p. 1175-1199.
224. Noble, H. and R. Heale, *Triangulation in research, with examples*, in *Evid Based Nurs*. 2019: England. p. 67-68.
225. Staniszewska, S., et al., *GRIPP2 reporting checklists: tools to improve reporting of patient and public involvement in research.* Bmj, 2017. **358**: p. j3453.
226. Christensen, K.B., G. Makransky, and M. Horton, *Critical Values for Yen's Q(3): Identification of Local Dependence in the Rasch Model Using Residual Correlations.* Appl Psychol Meas, 2017. **41**(3): p. 178-194.
227. Myles, P.S., et al., *Measurement of quality of recovery after surgery using the 15-item quality of recovery scale: a systematic review and meta-analysis.* Br J Anaesth, 2022. **128**(6): p. 1029-1039.

228. Karumbi, J., et al., *Inclusion of participants from low-income and middle-income countries in core outcome sets development: a systematic review.* BMJ Open, 2021. **11**(10): p. e049981.
229. Masyuko, S., et al., *Patient-reported outcomes for diabetes and hypertension care in low- and middle-income countries: A scoping review.* PLoS One, 2021. **16**(1): p. e0245269.
230. Lal, A., et al., *Fragmented health systems in COVID-19: rectifying the misalignment between global health security and universal health coverage.* Lancet, 2021. **397**(10268): p. 61-67.
231. Leaper, D.J., C.E. Edmiston, Jr., and C.E. Holy, *Meta-analysis of the potential economic impact following introduction of absorbable antimicrobial sutures.* Br J Surg, 2017. **104**(2): p. e134-e144.
232. Kirkland, K.B., et al., *The impact of surgical-site infections in the 1990s: attributable mortality, excess length of hospitalization, and extra costs.* Infect Control Hosp Epidemiol, 1999. **20**(11): p. 725-30.
233. NIHR-Global-Health-Research-Unit-on-Global-Surgery. *Cheetah - Sterile Glove and Clean Instrument Change at the Time of Wound Closure to Reduce Surgical Site Infection (Cheetah)*. 2019 [cited 2021 12th February]; Available from: https://clinicaltrials.gov/ct2/show/NCT03980652.
234. Tanner, J., et al., *Post-discharge surveillance to identify colorectal surgical site infection rates and related costs.* J Hosp Infect, 2009. **72**(3): p. 243-50.
235. Contreras, C.M., et al., *Telemedicine: Patient-Provider Clinical Engagement During the COVID-19 Pandemic and Beyond.* J Gastrointest Surg, 2020. **24**(7): p. 1692-1697.
236. NIHR-Global-Health-Research-Unit-on-Global-Surgery, *Pragmatic multicentre factorial randomized controlled trial testing measures to reduce surgical site infection in low- and middle-income countries: study protocol of the FALCON trial.* Colorectal Dis, 2020.
237. Lang, T.A. and D.G. Altman, *Basic statistical reporting for articles published in biomedical journals: the "Statistical Analyses and Methods in the Published Literature" or the SAMPL Guidelines.* Int J Nurs Stud, 2015. **52**(1): p. 5-9.
238. Cohen, J.F., et al., *STARD 2015 guidelines for reporting diagnostic accuracy studies: explanation and elaboration.* BMJ Open, 2016. **6**(11): p. e012799.
239. Mokkink, L.B., et al., *COSMIN Risk of Bias checklist for systematic reviews of Patient-Reported Outcome Measures.* Qual Life Res, 2018. **27**(5): p. 1171-1179.

240. Gillies, K., et al., *Strategies to improve retention in randomised trials.* Cochrane Database Syst Rev, 2021. **3**: p. Mr000032.
241. NIHR-Global-Health-Research-Unit-on-Global-Surgery. *Statistical analysis plan: Feasibility and diagnostic accuracy of Telephone Administration of an adapted patient-reported Wound HeaLth QuestiONnaire for assessment of surgical site infection following abdominal surgery in low and middle-income countries (TALON)*. 2021 8th March 2021 [cited 2022 9th March]; Available from: https://www.globalsurgeryunit.org/clinical-trials-holding-page/project-talon/.
242. Umemneku Chikere, C.M., et al., *Diagnostic test evaluation methodology: A systematic review of methods employed to evaluate diagnostic tests in the absence of gold standard - An update.* PLoS One, 2019. **14**(10): p. e0223832.
243. Alonzo, T. and M. Pepe, *Assessing accuracy of a continuous screening test in the presence of verification bias.* Series C: Applied Statistics (Royal Statistical Society), 2005. **54**(1).
244. Arifin, W.N. and U.K. Yusof, *Correcting for partial verification bias in diagnostic accuracy studies: A tutorial using R.* Stat Med, 2022.
245. Raza, A., et al., *The efficiency, accuracy and acceptability of smartphone-delivered data collection in a low-resource setting - A prospective study.* Int J Surg, 2017. **44**: p. 252-254.
246. Adepoju, P., *Africa turns to telemedicine to close mental health gap*, in *Lancet Digit Health*. 2020. p. e571-e572.
247. Michie, S., M.M. van Stralen, and R. West, *The behaviour change wheel: a new method for characterising and designing behaviour change interventions.* Implement Sci, 2011. **6**: p. 42.
248. Newlands, R., et al., *Why trials lose participants: a multi-trial investigation of participants' perspectives using the theoretical domains framework.* J Clin Epidemiol, 2021.
249. Skea, Z.C., R. Newlands, and K. Gillies, *Exploring non-retention in clinical trials: a meta-ethnographic synthesis of studies reporting participant reasons for drop out.* BMJ Open, 2019. **9**(6): p. e021959.
250. Elfeky, A., et al., *Non-randomised evaluations of strategies to increase participant retention in randomised controlled trials: a systematic review.* Syst Rev, 2020. **9**(1): p. 224.
251. Bruce, J., et al., *Wound photography for evaluation of surgical site infection and wound healing after lower limb trauma.* Bone Joint J, 2021. **103-b**(12): p. 1802-1808.

252. Oates, E.V., et al., *Are Surgical Patients Satisfied With Remote Consultations? A Comparison of Remote Versus Conventional Outpatient Clinic Follow-Up for Surgical Patients: A Systematic Review and Meta-Analysis of Randomized Controlled Trials.* J Patient Exp, 2021. **8**: p. 23743735211035916.
253. Treweek, S., et al., *Making randomised trials more efficient: report of the first meeting to discuss the Trial Forge platform.* Trials, 2015. **16**.
254. Raykar, N.P., et al., *A qualitative study exploring contextual challenges to surgical care provision in 21 LMICs.* Lancet, 2015. **385 Suppl 2**: p. S15.
255. Sessler, D. and P. Myles, *Novel Clinical Trial Designs to Improve the Efficiency of Research.* Anesthesiology, 2020. **132**(1).
256. Powell, B., T. Nkurunziza, and F. Kateera, *Synthesizing postpartum care pathways, facilitators, and barriers for women after cesarean section: a qualitative analysis from rural Rwanda.* Journal of Global Health Reports, 2021. **5**.
257. Rosala-Hallas, A., et al., *Global health trials methodological research agenda: results from a priority setting exercise.* Trials, 2018. **19**(1).
258. Li, H., et al., *The role of community health workers in the surgical cascade: a scoping review.* Human resources for health, 2021. **19**(1).
259. Oesophago-Gastric-Anastomotic-Audit-Collaborative(OGAA), *Mortality from esophagectomy for esophageal cancer across low, middle, and high-income countries: An international cohort study.* European journal of surgical oncology : the journal of the European Society of Surgical Oncology and the British Association of Surgical Oncology, 2021. **47**(6).
260. Ahmad, T., et al., *Use of failure-to-rescue to identify international variation in postoperative care in low-, middle- and high-income countries: a 7-day cohort study of elective surgery.* British journal of anaesthesia, 2017. **119**(2).
261. International-Surgical-Outcomes-Study-group, *Global patient outcomes after elective surgery: prospective cohort study in 27 low-, middle- and high-income countries.* British journal of anaesthesia, 2016. **117**(5).
262. Davies, J., et al., *Global surgery, obstetric, and anaesthesia indicator definitions and reporting: An Utstein consensus report.* PLoS medicine, 2021. **18**(8).
263. Raykar, N., et al., *New global surgical and anaesthesia indicators in the World Development Indicators dataset.* BMJ global health, 2017. **2**(2).
264. Christie, S., et al., *Strengthening Surgery Strengthens Health Systems: A New Paradigm and Potential Pathway for Horizontal Development in Low- and Middle-Income Countries.* World journal of surgery, 2019. **43**(3).

265. Haider, A., et al., *Development of a Unifying Target and Consensus Indicators for Global Surgical Systems Strengthening: Proposed by the Global Alliance for Surgery, Obstetric, Trauma, and Anaesthesia Care (The G4 Alliance).* World journal of surgery, 2017. **41**(10).
266. Yang, R., et al., *Selection bias and subject refusal in a cluster-randomized controlled trial.* BMC medical research methodology, 2017. **17**(1).
267. Puffer, S., D. Torgerson, and J. Watson, *Cluster randomized controlled trials.* Journal of evaluation in clinical practice, 2005. **11**(5).
268. Easter, C., et al., *Cluster randomized trials of individual-level interventions were at high risk of bias.* Journal of clinical epidemiology, 2021. **138**.
269. Brownie, C. and J. Habicht, *Selecting a screening cut-off point or diagnostic criterion for comparing prevalences of disease.* Biometrics, 1984. **40**(3).
270. Youden, W., *Index for rating diagnostic tests.* Cancer, 1950. **3**(1).
271. Dong, T., et al., *A new diagnostic accuracy measure and cut-point selection criterion.* Statistical methods in medical research, 2017. **26**(6).
272. Cherian, T., et al., *Diagnosing Post-Cesarean Surgical Site Infections in Rural Rwanda: Development, Validation, and Field Testing of a Screening Algorithm for Use by Community Health Workers.* Surgical infections, 2020. **21**(7).
273. Juran, S., et al., *Geospatial mapping of access to timely essential surgery in sub-Saharan Africa.* BMJ global health, 2018. **3**(4).
274. Koch, R., et al. *Financial Risk Protection in Cesarean Section Patients at a Rural District Hospital in Rwanda*. in *14th Annual Academic Surgical Congress (ASC)*. 2019. Houston, TX.
275. Buvik, A., et al., *Cost-Effectiveness of Telemedicine in Remote Orthopedic Consultations: Randomized Controlled Trial.* J Med Internet Res, 2019. **21**(2): p. e11330.
276. Jaspers, L., et al., *The global impact of non-communicable diseases on households and impoverishment: a systematic review.* European journal of epidemiology, 2015. **30**(3).
277. Phull, M., et al., *What is the financial burden to patients of accessing surgical care in Sierra Leone? A cross-sectional survey of catastrophic and impoverishing expenditure.* BMJ open, 2021. **11**(3).
278. Okoroh, J., et al., *Evaluating the impact of the national health insurance scheme of Ghana on out of pocket expenditures: a systematic review.* BMC health services research, 2018. **18**(1).

279. Nwanna-Nzewunwa, O., et al., *The societal cost and economic impact of surgical care on patients' households in rural Uganda; a mixed method study.* BMC health services research, 2021. **21**(1).
280. Odland, M.L., et al., *Identifying, Prioritizing and Visually Mapping Barriers to Injury Care in Rwanda: A Multi-disciplinary Stakeholder Exercise.* World J Surg, 2020. **44**(9): p. 2903-2918.
281. LeBrun, D.G., et al., *Prioritizing essential surgery and safe anesthesia for the Post-2015 Development Agenda: operative capacities of 78 district hospitals in 7 low- and middle-income countries.* Surgery, 2014. **155**(3): p. 365-73.
282. Thaddeus, S. and D. Maine, *Too far to walk: maternal mortality in context.* Soc Sci Med, 1994. **38**(8): p. 1091-110.
283. Heine, M., et al., *Health education interventions to promote health literacy in adults with selected non-communicable diseases living in low-to-middle income countries: A systematic review and meta-analysis.* Journal of evaluation in clinical practice, 2021. **27**(6).
284. Hoyler, M., et al., *Surgical care by non-surgeons in low-income and middle-income countries: a systematic review.* Lancet, 2015. **385 Suppl 2**: p. S42.
285. Beard, J.H., et al., *Outcomes After Inguinal Hernia Repair With Mesh Performed by Medical Doctors and Surgeons in Ghana.* JAMA Surg, 2019. **154**(9): p. 853-859.
286. Gupta, R., et al., *Improving the global health workforce crisis: an evaluation of Global Health Corps.* The Lancet. Global health, 2015. **3**(11).
287. Yudkin, J., et al., *Global health-worker crisis: the UK could learn from Cuba.* Lancet (London, England), 2008. **371**(9622).
288. Britnell, M., *Human: Solving the global workforce crisis in health*. 2019: OUP Oxford.
289. Gögenur, M., et al., *Association of postoperative influenza vaccine on overall mortality in patients undergoing curative surgery for solid tumors.* International journal of cancer, 2021. **148**(8).
290. Nunan, D., J. Aronson, and C. Bankhead, *Catalogue of bias: attrition bias.* BMJ evidence-based medicine, 2018. **23**(1).
291. Ronquillo, C. and L. Currie, *The digital divide: Trends in global mobile and broadband Internet access from 2000-2010.* NI 2012 : 11th International Congress on Nursing Informatics, June 23-27, 2012, Montreal, Canada. International Congress in Nursing Informatics (11th : 2012 : Montreal, Quebec), 2012. **2012**.

292. Greenberg-Worisek, A., et al., *Tracking Healthy People 2020 Internet, Broadband, and Mobile Device Access Goals: An Update Using Data From the Health Information National Trends Survey.* Journal of medical Internet research, 2019. **21**(6).
293. Scott, J.W., et al., *Contextual Challenges to Safe Surgery in a Resource-limited Setting: A Multicenter, Multiprofessional Qualitative Study.* Ann Surg, 2018. **267**(3): p. 461-467.
294. Scott Kruse, C., et al., *Evaluating barriers to adopting telemedicine worldwide: A systematic review.* Journal of telemedicine and telecare, 2018. **24**(1).
295. Bulik, R., *Human factors in primary care telemedicine encounters.* Journal of telemedicine and telecare, 2008. **14**(4).
296. Stommel, W., H. van Goor, and M. Stommel, *The Impact of Video-Mediated Communication on Closed Wound Assessments in Postoperative Consultations: Conversation Analytical Study.* Journal of medical Internet research, 2020. **22**(5).
297. Matousek, A., et al., *Community health workers and smartphones for the detection of surgical site infections in rural Haiti: a pilot study.* Lancet (London, England), 2015. **385 Suppl 2**.
298. Yadav, S., et al., *Smartphone-Based Application for Tele-follow-up of Patients with Endocrine Disorders in Context of a LMIC: A Compliance, Satisfaction, Clinical Safety and Outcome Assessment.* World journal of surgery, 2020. **44**(2).
299. Bunogerane, G.J., et al., *Using Touch Surgery to Improve Surgical Education in Low- and Middle-Income Settings: A Randomized Control Trial.* J Surg Educ, 2018. **75**(1): p. 231-237.
300. Macefield, R. *The self-taken 'Selfi' wound study: Exploring the feasibility of collecting patient/carer of surgical wounds after leaving hospital*. in *Wound Research Network (WReN)*. 2019. Southampton, UK.
301. Edwards, S., *Review of a medical illustration department's data processing system to confirm general data protection regulation (GDPR) compliance.* Journal of visual communication in medicine, 2019. **42**(3).
302. Fletcher, R., et al., *The Use of Mobile Thermal Imaging and Deep Learning for Prediction of Surgical Site Infection.* Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Annual International Conference, 2021. **2021**.
303. Samareh, A., et al., *Artificial Intelligence Methods for Surgical Site Infection: Impacts on Detection, Monitoring, and Decision Making.* Surgical infections, 2019. **20**(7).

304. Chen, W., et al., *Artificial Intelligence-Based Multimodal Risk Assessment Model for Surgical Site Infection (AMRAMS): Development and Validation Study.* JMIR medical informatics, 2020. **8**(6).
305. Scardoni, A., et al., *Artificial intelligence-based tools to control healthcare associated infections: A systematic review of the literature.* Journal of infection and public health, 2020. **13**(8).
306. G, Z., et al., *Clinically relevant deep learning for detection and quantification of geographic atrophy from optical coherence tomography: a model development and external validation study.* The Lancet. Digital health, 2021. **3**(10).
307. OT, J., et al., *Artificial intelligence and machine learning algorithms for early detection of skin cancer in community and primary care settings: a systematic review.* The Lancet. Digital health, 2022. **4**(6).
308. PA, K. and T. EJ, *With an eye to AI and autonomous diagnosis.* NPJ digital medicine, 2018. **1**.
309. Faes, L., et al., *Automated deep learning design for medical image classification by health-care professionals with no coding experience: a feasibility study.* The Lancet. Digital health, 2019. **1**(5).
310. Korot, E., et al., *Predicting sex from retinal fundus photographs using automated deep learning.* Scientific reports, 2021. **11**(1).
311. Korot, E., et al., *Clinician-driven artificial intelligence in ophthalmology: resources enabling democratization.* Current opinion in ophthalmology, 2021. **32**(5).
312. Guinness-World-Records. *Most authors on a single peer-reviewed academic paper: COVIDSurg and GlobalSurg Collaboratives at the University of Birmingham and the University of Edinburgh in the UK, as verified on 24 March 2021.* 2021; Available from: https://www.guinnessworldrecords.com/world-records/653537-most-authors-on-a-single-peer-reviewed-academic-paper.
313. Baniasadi, T., et al., *Surgical Patients Follow-Up by Smartphone-Based Applications: A Systematic Literature Review.* Studies in health technology and informatics, 2020. **271**.
314. Evans, H. and W. Lober, *A Pilot Use of Patient-Generated Wound Data to Improve Postdischarge Surgical Site Infection Monitoring.* JAMA surgery, 2017. **152**(6).
315. Ng, H., D. Huang, and V. Rajaratnam, *Diagnosing surgical site infections using telemedicine: A Systematic Review.* The surgeon : journal of the Royal Colleges of Surgeons of Edinburgh and Ireland, 2021.

316. Kateera, F., et al., *The Effect and Feasibility of mHealth-Supported Surgical Site Infection Diagnosis by Community Health Workers After Cesarean Section in Rural Rwanda: Randomized Controlled Trial.* JMIR mHealth and uHealth, 2022. **10**(6).
317. Hedt-Gauthier, B., et al., *Telemedicine for Surgical Site Infection Diagnosis in Rural Rwanda: Concordance and Accuracy of Image Reviews.* World journal of surgery, 2022.
318. Aiken, A.M., et al., *Evaluation of surveillance for surgical site infections in Thika Hospital, Kenya.* J Hosp Infect, 2013. **83**(2): p. 140-5.
319. Srun, S., et al., *Surveillance of post-caesarean surgical site infections in a hospital with limited resources, Cambodia.* Journal of infection in developing countries, 2013. **7**(8).
320. Elbur, A., et al., *Post-discharge surveillance of wound infections by telephone calls method in a Sudanese Teaching Hospital.* Journal of infection and public health, 2013. **6**(5).
321. Reinforcement-of-Closure-of-Stoma-Site-(ROCSS)-Collaborative and West-Midlands-Research-Collaborative, *Prophylactic biological mesh reinforcement versus standard closure of stoma site (ROCSS): a multicentre, randomised controlled trial.* Lancet (London, England), 2020. **395**(10222).
322. Allegranzi, B., et al., *A multimodal infection control and patient safety intervention to reduce surgical site infections in Africa: a multicentre, before-after, cohort study.* Lancet Infect Dis, 2018. **18**(5): p. 507-515.
323. Lofgren, J., et al., *A Randomized Trial of Low-Cost Mesh in Groin Hernia Repair.* N Engl J Med, 2016. **374**(2): p. 146-53.
324. Crocker, J.C., et al., *Impact of patient and public involvement on enrolment and retention in clinical trials: systematic review and meta-analysis.* Bmj, 2018. **363**: p. k4738.
325. Brett, J., et al., *Reaching consensus on reporting patient and public involvement (PPI) in research: methods and lessons learned from the development of reporting guidelines.* BMJ Open, 2017. **7**(10): p. e016948.
326. Bagley, H.J., et al., *A patient and public involvement (PPI) toolkit for meaningful and flexible involvement in clinical trials - a work in progress.* Res Involv Engagem, 2016. **2**: p. 15.
327. Cook, N., et al., *Patient and public involvement in health research in low and middle-income countries: a systematic review.* BMJ Open, 2019. **9**(5): p. e026514.

328. South, A., et al., *Models and impact of patient and public involvement in studies carried out by the Medical Research Council Clinical Trials Unit at University College London: findings from ten case studies.* Trials, 2016. **17**: p. 376.
329. Absolom, K., et al., *Beyond lip service and box ticking: how effective patient engagement is integral to the development and delivery of patient-reported outcomes.* Qual Life Res, 2015. **24**(5): p. 1077-85.
330. Dudley, L., et al., *What Difference Does Patient and Public Involvement Make and What Are Its Pathways to Impact? Qualitative Study of Patients and Researchers from a Cohort of Randomised Clinical Trials.* PLoS One, 2015. **10**(6): p. e0128817.
331. Pizzo, E., et al., *Patient and public involvement: how much do we spend and what are the benefits?* Health Expect, 2015. **18**(6): p. 1918-26.
332. Gamble, C., et al., *Patient and public involvement in the early stages of clinical trial development: a systematic cohort investigation.* BMJ Open, 2014. **4**(7): p. e005234.
333. Pickering, E., et al., *Formalising the induction of patient and public involvement contributors on trial oversight committees.* Research involvement and engagement, 2021. **7**(1).
334. Staniszewska, S., et al., *Reviewing progress in public involvement in NIHR research: developing and implementing a new vision for the future.* BMJ open, 2018. **8**(7).
335. Price, A., et al., *Patient and Public Involvement in research: A journey to co-production.* Patient education and counseling, 2022. **105**(4).
336. Jackson, T., et al., *Patient and public involvement in research: from tokenistic box ticking to valued team members.* BMC medicine, 2020. **18**(1).
337. Kroese, K., et al., *Community engagement and involvement in Ghana: conversations with community stakeholders to inform surgical research.* Research involvement and engagement, 2021. **7**(1).
338. Lee, H. and J. Lori, *Scoping Review: Valid and Reliable Health Literacy Assessment Tools Used in Low- and Middle-Income Countries.* Journal of nursing measurement, 2020.
339. Pleasant, A., C. O'Leary, and R. Carmona, *Health Literacy: Global Advances with a Focus Upon the Shanghai Declaration on Promoting Health in the 2030 Agenda for Sustainable Development.* Studies in health technology and informatics, 2020. **269**.
340. Paakkari, L. and O. Okan, *COVID-19: health literacy is an underestimated problem.* The Lancet. Public health, 2020. **5**(5).

341. Meherali, S., N. Punjani, and A. Mevawala, *Health Literacy Interventions to Improve Health Outcomes in Low- and Middle-Income Countries.* Health literacy research and practice, 2020. **4**(4).
342. Kroese, K., et al., *Challenges and solutions: surveying researchers on what type of community engagement and involvement activities are feasible in low and middle income countries during the COVID-19 pandemic.* BMJ open, 2021. **11**(10).
343. Lavallee, D., et al., *Engaging Patients in Co-Design of Mobile Health Tools for Surgical Site Infection Surveillance: Implications for Research and Implementation.* Surgical infections, 2019. **20**(7).
344. Mazingi, D., et al., *Exploring the Impact of COVID-19 on Progress Towards Achieving Global Surgery Goals.* World J Surg, 2020. **44**(8): p. 2451-2457.
345. Townsend, E., *COVID-19 policies in the UK and consequences for mental health.* Lancet Psychiatry, 2020. **7**(12): p. 1014-1015.
346. Sud, A., et al., *Effect of delays in the 2-week-wait cancer referral pathway during the COVID-19 pandemic on cancer survival in the UK: a modelling study.* Lancet Oncol, 2020. **21**(8): p. 1035-1044.
347. Ranganathan, P., et al., *Impact of COVID-19 on cancer care in India: a cohort study.* Lancet Oncol, 2021.
348. Maringe, C., et al., *The impact of the COVID-19 pandemic on cancer deaths due to delays in diagnosis in England, UK: a national, population-based, modelling study.* Lancet Oncol, 2020. **21**(8): p. 1023-1034.
349. Mansfield, K.E., et al., *Indirect acute effects of the COVID-19 pandemic on physical and mental health in the UK: a population-based study.* Lancet Digit Health, 2021. **3**(4): p. e217-e230.
350. Lai, A.G., et al., *Estimated impact of the COVID-19 pandemic on cancer services and excess 1-year mortality in people with cancer and multimorbidity: near real-time data on cancer care, cancer deaths and a population-based cohort study.* BMJ Open, 2020. **10**(11): p. e043828.
351. Kansagra, A.P., et al., *Collateral Effect of Covid-19 on Stroke Evaluation in the United States*, in *N Engl J Med*. 2020. p. 400-401.
352. Hamadani, J.D., et al., *Immediate impact of stay-at-home orders to control COVID-19 transmission on socioeconomic conditions, food insecurity, mental health, and intimate partner violence in Bangladeshi women and their families: an interrupted time series.* Lancet Glob Health, 2020. **8**(11): p. e1380-e1389.

353. Nepogodiev, D., et al. *Forecasting waiting lists for elective procedures and surgery in England: a modelling study*. 2022; Available from: https://scholar.google.com/citations?view_op=view_citation&hl=en&user=J_g3yRAAAAAJ&cstart=20&pagesize=80&sortby=pubdate&citation_for_view=J_g3yRAAAAAJ:mlAyqtXpCwEC.
354. Mangipudi, S., et al., *Oxygen availability in sub-Saharan African countries: a call for data to inform service delivery.* Lancet Glob Health, 2020. **8**(9): p. e1123-e1124.
355. Janghorban, R., R. Latifnejad Roudsari, and A. Taghipour, *Skype interviewing: the new generation of online synchronous interview in qualitative research.* International journal of qualitative studies on health and well-being, 2014. **9**.
356. Santhosh, L., J. Rojas, and P. Lyons, *Zooming into Focus Groups: Strategies for Qualitative Research in the Era of Social Distancing.* ATS scholar, 2021. **2**(2).
357. Gregory, K., *Online Communication Settings and the Qualitative Research Process: Acclimating Students and Novice Researchers.* Qualitative health research, 2018. **28**(10).
358. Krouwel, M., K. Jolly, and S. Greenfield, *Comparing Skype (video calling) and in-person qualitative interview modes in a study of people with irritable bowel syndrome - an exploratory comparative analysis.* BMC medical research methodology, 2019. **19**(1).
359. Pocock, T., M. Smith, and J. Wiles, *Recommendations for Virtual Qualitative Health Research During a Pandemic.* Qualitative health research, 2021. **31**(13).
360. Upadhyay, U. and H. Lipkovich, *Using online technologies to improve diversity and inclusion in cognitive interviews with young people.* BMC medical research methodology, 2020. **20**(1).
361. Marais, I. and D. Andrich, *Effects of varying magnitude and patterns of response dependence in the unidimensional Rasch model.* Journal of applied measurement, 2008. **9**(2).
362. Andrich, D. and G. Luo, *Conditional pairwise estimation in the Rasch model for ordered response categories using principal components.* Journal of applied measurement, 2003. **4**(3).
363. Hagquist, C. and D. Andrich, *Recent advances in analysis of differential item functioning in health research using the Rasch model.* Health and quality of life outcomes, 2017. **15**(1).
364. Yorke, J., M. Horton, and P. Jones, *A critique of Rasch analysis using the Dyspnoea-12 as an illustrative example.* Journal of advanced nursing, 2012. **68**(1).

365. de Groot, J., et al., *Correcting for partial verification bias: a comparison of methods.* Annals of epidemiology, 2011. **21**(2).
366. Liu, D. and X. Zhou, *Semiparametric estimation of the covariate-specific ROC curve in presence of ignorable verification bias.* Biometrics, 2011. **67**(3).
367. Association-Of-Rural-Surgeons-Of-India, *The Lancet Commission on Global Surgery - Association of Rural Surgeons of India Karad Consensus Statement on Surgical System Strengthening in Rural India.* Healthcare (Amsterdam, Netherlands), 2019. **7**(2).
368. Kim, E.K., et al., *Rural surgery as global surgery before global surgery.* BMJ Glob Health, 2022. **7**(3).
369. Pinto, A., et al., *Surgical complications and their impact on patients' psychosocial well-being: a systematic review and meta-analysis.* BMJ open, 2016. **6**(2).
370. Powell, R., et al., *Psychological preparation and postoperative outcomes for adults undergoing surgery under general anaesthesia.* The Cochrane database of systematic reviews, 2016. **2016**(5).
371. Knight, S., et al., *Mobile devices and wearable technology for measuring patient outcomes after surgery: a systematic review.* NPJ digital medicine, 2021. **4**(1).
372. E, K., et al., *Clinician-driven artificial intelligence in ophthalmology: resources enabling democratization.* Current opinion in ophthalmology, 2021. **32**(5).
373. W, C., et al., *Artificial Intelligence-Based Multimodal Risk Assessment Model for Surgical Site Infection (AMRAMS): Development and Validation Study.* JMIR medical informatics, 2020. **8**(6).
374. A, S., et al., *Artificial intelligence-based tools to control healthcare associated infections: A systematic review of the literature.* Journal of infection and public health, 2020. **13**(8).
375. A, S., et al., *Artificial Intelligence Methods for Surgical Site Infection: Impacts on Detection, Monitoring, and Decision Making.* Surgical infections, 2019. **20**(7).
376. JI, D., et al., *Global surgery, obstetric, and anaesthesia indicator definitions and reporting: An Utstein consensus report.* PLoS medicine, 2021. **18**(8).

8. Appendix

Appendix A. PCORI reporting checklist in Chapter 4

Standards for qualitative methods		Check
QM-1: State the qualitative approach to research inquiry, design, and conduct	A. Identify and describe evidence gaps that support the need for a qualitative component(s) of the study B. Identify the qualitative approach (eg, ethnography, grounded theory) that will be used, including the purpose, why it is an appropriate approach to answer the research question(s), and how it will be operationalized C. Describe the types of data to be collected, strategies for data collection (eg, focus groups, observations, interviews, documents, audio or video recordings), and when the data will be collected D. Describe how confidentiality will be maintained through data collection, management, analysis, and reporting E. State the computer software program used to assist with analysis	✓ ✓ ✓ ✓ ✓
QM-2: Select and justify appropriate qualitative methods sampling strategy	A. Describe and provide the rationale for the sampling strategy (see RQ-3†, RQ-4†, and PC-2‡), including how the strategy flows logically from the qualitative approach and how it fits the research question(s) B. Explain the anticipated sample size, detail any variation in sampling that may occur over the course of study, and state the criteria for deciding when no further sampling is necessary (eg, thematic saturation) C. Describe how the methods will ensure that the data capture the depth of experiences of the participants or phenomenon of interest (see PC-2‡ and PC-3‡)	✓ ✓ ✓
QM-3: Link the qualitative data analysis, interpretations, and conclusions to the study question	A. State who will be involved in the data analysis and interpretation and describe how their qualifications, training, and expertise equip them to understand and address the complexities and challenges unique to qualitative methods B. Describe data analysis procedures and their link to the study's research questions C. Describe the process by which inferences and themes will be identified and developed as well as how this process is congruent with the chosen qualitative approach and its methodology D. Describe how conclusions will be derived and how they relate to interpretations and content of the original data	✓ ✓ ✓ ✓
QM-4: Establish trustworthiness* and credibility of qualitative research	A. State how documentation regarding all phases of the analysis will be captured. Multiple data collection methods (eg, interviews, focus groups, observations) and/or experts with diverse backgrounds can be used to increase trustworthiness, in addition to an inter-coder reliability process B. To enhance credibility, discuss three distinct elements: rigorous techniques and methods, the role of the qualitative researcher, and the value of participants' perspectives and experiences. Credibility must be explained (see RQ-1†, RQ-2†, and IR-7§) and demonstrated in the analysis in at least one of the following three ways: reflexivity, negative case analysis, and/or member checking	✓ ✓
Standards for mixed methods research		
MM-1: Specify how mixed methods are integrated across design, data sources, and/or data collection phases	A. State which mixed methods approach will be used and describe how it will inform the study procedures B. Describe whether the quantitative and qualitative methods will be sequential, concurrent, or a mixture of both, over time C. Describe how the mixed methods design will integrate qualitative and quantitative approaches at one or more stages of the research process and achieve the intent of the design (eg, by aligning the aims to data collection instruments, procedures and analyses of data, and interpretation of the findings)	✓ ✓ ✓
MM-2: Select and justify appropriate mixed methods sampling strategy	A. Provide a clear description of the relationship between the sampling techniques and the generation of different types of data (eg, numeric or closed ended *v* narrative or open ended; see RQ-3†, RQ-4†, and QM-2¶) B. Describe the sampling strategies and outline the temporality with which they will take place as they relate to selected qualitative and quantitative methodologies (see IR-1§, IR-2§, PC-2‡, PC-3‡, and QM-1¶), including a justification of the emergence of other samples that may arise during the study, as applicable	✓ ✓
MM-3: Integrate data analysis, data interpretation, and conclusions	A. Describe the analytic approaches to integration and demonstrate how the analysis plan is congruent with the study design and aims, and that it has been developed based on the methodological approach (eg, either a priori or emergently; see IR-1§, IR-2§, PC-2‡, PC-3‡, QM-1¶, and QM-3¶) B. Identify the order of study components and the points of integration. State who will conduct the integration; describe how their qualifications, training, and expertise equip them to understand and address the complexities and challenges unique to mixed methods analysis; and state how integrated analyses will proceed in terms of the qualitative and quantitative components C. Describe the approach used to interpret integrated data and how conclusions are supported by the context of original qualitative and quantitative findings. Address divergent findings from both qualitative and quantitative components, as well as method-specific biases across the methods (see QM-4¶)	✓ ✓ ✓

Appendix B. Example topic guide (Site researchers) in Chapter 4

Item 1. *Read the question in full.*

Statement: Please reflect on your experience of completing in-person and telephone follow-up with patients in both clinical practice and randomised trials.

Generic probes

- How would a patient go about answering that question?
- In your experience, how would they arrive at that answer?
- Would it be easy or difficult for them to answer? Why?

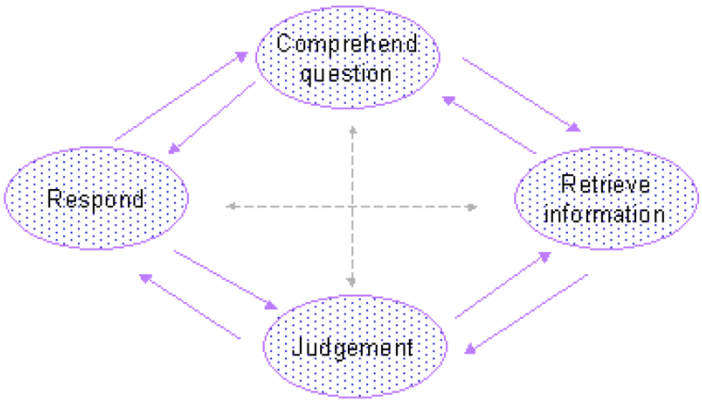


Figure. Four-stage question and answer model

Comprehension probes

- What do you think the term A means to patients?
- Do you think they will understand the question?
- What have patients had trouble understanding in the past?
- Have you ever needed to provide clarification on this question? Why?
- How have you clarified this question when asking it to patients?

Retrieval probes

- Do you think patients typically remember their response to this question?
- How do you think patients you remember that information?

- Do patients ever have problems remembering this information?
- Have you ever made a clarification or prompt to help patients remember?
- What time period do patients talk about when answering this question?

Judgement probes

- How sure are patients of their answers?
- Do they ever change their answer? Why?
- What do they talk about when deciding how to answer this question?
- How accurate would you say their answer is? Why?

Response probes

- How easy or difficult do patients find it to select an answer from the options provided? Why?
- Are there any categories missing from the options provided or do they cover everything? What is missing?
- Why do you think patients choose a particular answer, rather than one of the others?

Appendix C. Summary of qualitative data used in adaptation of 'symptoms' items (1 to 11) and item response categories in Chapter 4

Thematic domain	Original Wound Healing Questionnaire		Item 1	Item 2	Item 3	Item 4	Item 5	Item 6	Item 7	Item 8	Item 9	Item 10	Item 11
	Introductory statement	**Item response categories: 1 = Not at all; 2 = A little; 3 = Quite a bit; 4 = A lot.**	**Was there redness spreading away from the wound?**	**Was the area around the wound warmer than the surrounding skin?**	**Has any part of the wound leaked clear fluid?**	**Has any part of the wound leaked blood-stained fluid?**	**Has any part of the wound leaked thick and yellow or green fluid?**	**Have the edges of any part of the wound separated or gaped open of their accord?**	**If the wound edges opened, did the deeper tissue also separate?**	**Has the area around the wound become swollen?**	**Has the wound been smelly?**	**Has the wound been painful to touch?**	**Have you had, or felt like you have had, a raised temperature or fever (>38oC)?**
Ghana													
Comprehension	None	Four categories hard for questions about symptoms (Interview GH001I). Patients will not understand a four-point scale (Focus group GH001F).	Colour of wound may be challenging for patients with dark skin type (Interview GH002I).	Patients might struggle to answer this question as they are often not health literate (Focus group GH001F).	Patients describe this as 'water' from the wound in Ghana (Interview GH001I).	None	None	None	From FALCON experience, patients will be able to see differences in separation of the deeper part of the wound versus the skin; well reported symptom (Interview GH002I).	Patients should be able to report (Interview GH001I).	None	None	Lots of experience of patients reporting temperature as Ghana is a malarial environment; patients are usually able to report this well (Interview GH002I).
Response mapping	N/A	Patients may be more likely to respond appropriately to yes or no item response categories, or not at all, a little, a lot (Interview GH001I).											None
Retrieval	'Since leaving hospital after having your operation' should be repeated at the top of the second section in order to ensure this is remembered by the assessor and patient.	None	None	The wound may only be open when a nurse/doctor is changing a dressing (Interview GH001I).	None	Family members may also notice leakage from the skin and will be useful collaborators in assessment (Interview GH002I).	None	None	None	None	Patients hate to feel abnormal and they will definitely report something so disturbing (Interview GH002I).	None	None
Judgment	N/A	None	From experience within previous randomised trials, may be challenging for patients to judge (Interview GH001I). May only be if answered correctly if the doctor tells patient their observation (Focus group GH001F).	May be that people only notice heat if doctor has told them this (Interview GH001I).	For children in research studies, parents they might have to answer this question for them Interview GH001I).	None	None	Patients would notice during wound dressing or ward round. May see that the wound is open and see flesh below (Interview GH001I).	Patients will not understand the word tissue, or be able to tell the difference between the superficial and deeper tissues (Focus group GH001F).	Patients may confuse swelling/hardness of the wound related to regular healing (Interview GH002I).	None	May struggle in early phases to differentiate normal pain and pain related to SSI. It may help to use the timing of pain, for example excess pain after the first few days after surgery (if discharged early) (Interview GH001I)	Many commonly report 'feeling hot' as the weather often very warm. May be a poor differentiator of those with and without infection (Focus group GH001F)
Translatability	None	No terms to differentiate a little & quite a bit in Ghanian languages (Interview GH002I). Particular concern about the differentiation between 'a little' and 'quite a bit' which were perceived to be very 'English English' terms (Focus group GH001F).	None	None	None	None	None	None	None	None	None	None	None
Other context	None	None	None	None	None	Patients in Ghana typically stay around the location of the hospital while they're recovering (Interview GH002I). Most of the time wounds are re-dressed in local hospital (Interview GH001I).	None	None	None	None	Patients will be comfortable to report, did not perceive to have stigma attached (Interview GH001I).	None	Patients rarely have a thermometer so quantifying fever is unhelpful here (Interview GH001I).
Nigeria													
Comprehension	None	Patients will not be able to tell the difference between A little and Quite a bit in Nigeria (Interview NG003I).	None	No problems with assessment of this related to CDC criteria to date. Some interviewers may help with understanding in a real world context (Interview NG003).	Colourless, texture may be important (e.g. thin) (Interview NG001I).	None	Memorable & patients will be happy to report this (Interview NG002).	None	None	None	None	None	None
Response mapping	N/A	Difficulty differentiating a little and quite a bit over the telephone, as doesn't appear to be scalar for patients in Nigeria. Might be possible with additional help from person administering the questionnaire if they were able to quantify these levels (Interview NG001I). Quite a bit and a lot not possible to see the difference in. Three point scale would be more appropriate (Focus group NG001F).											Typically patients either feel Y/N, very hot or feel normal (Interview NG002I). Patients unable to differentiate between several response categories for fever ((Interview NG003I).
Retrieval	None	None	None	None	None	None	None	May have opened, then later closed, so may need to ask others/family members. Accuracy may be decreased if family member not available (Interview NG001I). May use having dressing for a longer time than normal or came back to hospital more frequently if wound opened - may help with	Depending on level of education of the patient, they make not understand difference between deeper tissue (beyond the surface of the skin). Unsure whether they will be able to see other layers of below their skin (Interview NG001I). Patients won't understand	None	None	None	None

								retrieval (Interview NG003I).	this, would have to be more can see the deeper layer (fat or flesh or bowel) underneath. Deeper tissue too scientific. (Interview NG002I) May be too complex for some patients to understand. (Focus group NG001F)				
Judgment	N/A	None	Could be helpful to highlight 'on the surface' of the skin. Redness very challenging in patients in dark skin tone (Interview NG002I). Can be difficult to recognise for patients with dark skin types (Interview NG003I). Difficult to identify in dark skin tones. Patients instead report 'shining' or 'irritated' skin, which would be more appropriate (Interview NG001I).	None	Patients should be asked to think of every part of the wound, particularly the main sutured part of the wound (Interview NG001I).	None	None	Patients might sometimes need an assistant for a lower abdominal wound (dependent on incision site) (Interview NG001I). Patients will report this according to disruption of the line of sutures or separately (Interview NG002I).	Some patients will not be aware, or not able to see whether this has occurred (Focus group NG002F).	Patients would be able to gauge intensity & size of swelling (Interview NG001I). May notice that the wound has grown in size. (Interview NG002)	Patients would picture the smell before and while dressing of the wound, whether an odour was perceived & whether that could be traced back to the wound. They would compare this to a usual smell. May realise if a friend or family member around them had made such compliant. Most likely to notice during wound dressing/changing by relative or health workers, or observed herself/himself. Some wound dressings e.g. iodine may come with their own smell which they won't be familiar with (Interview NG001I). Patient will consider any offensive odour from the wound (Interview NG002I).	Patients advised not to frequently touch wound, but likely to notice this symptom (Interview NG001I).	At times patients feel odd, but actually have not no temperature. A little bit/quite a bit not directly related to temperature (Interview NG002I).
Translatability	N/A	Not all all (i.e. completely zero) and a lot easy to answer. Quite a bit compared to a little very hard in Pidgin English (Interview NG002I).	N/A										
Other context	Questionnaire completion in Nigeria is more like a conversation than a very structured interaction in order to keep the patient engaged - the questionnaire can be used to guide this discussion. Patients may 'complete questionnaire' without realising they've answered very structured questions (Interview NG001I). May need skilled administrator. May need to give to relative to help support with responses (Focus group NG001F)	None	None	None	None	None	None	None	Patients with this symptom will want a 'solution', 'next-steps', so their needs to be clear pathway for safety netting. Currently within FALCON, the research nurse asks when is the next clinic appointment, and tells the patient to mention & the doctor will examine, instructed to continue current medications e.g. antibiotics, and eat a balanced diet with sufficient protein (Interview NG001I).	None	Patients may be very reluctant to share this information/ashamed, and only provide the answer if trust is given to the interviewer, especially females and young adults (Interview NG001I).		Patients will not be able to measure their temperature in the community (Focus group NG001F). No patients have access to a thermometer (Focus group NG002F).
India													
Comprehension	None	None	None	None	None	None	None	None	None	None	None	None	None
Response mapping	None	Patients will not be able to differentiate a little and quite a bit. Recommend a three point scale (Interview IN001I). Should be with a three point, rather than a four point scale for responses for patients in India (Interview IN002I).											None
Retrieval	None	None	None	None	None	None	None	None	None	None	None	None	None
Judgment	None	None	None	None	None	None	None	None	Patients will not be able to differentiate between separation of the skin and deeper tissue (Interview IN001I). Can only report Gap in tissues, unless they are medical/been told (Interview IN002I)	Patients will easily be able to report areas of swelling and whether or not this is around the wound (Interview IN002I).	None	The wound may also be painful during daily activities e.g. if patients had chosen not to touch the wound or area around it (Interview IN001I)	None
Translatability	None	Most languages will not facilitate nuance around several middle response categories (Focus group IN001F)	None	None	None	None	None	None	None	None	None	None	None
Other context	A note should be added to confirm verbal consent before proceeding with the questionnaire.	None	None	None	None	None	None	None	None	None	None	None	Patients do not have access to thermometers – so temperature recording would not be accurate. Better to remove 38oC reference value (Focus group IN001F)
Mexico													
Comprehension	None	None	None	None	None	None	None	None	Some patient groups (e.g. lower literacy) will be unable to appreciate differences between superficial and deep tissues of the wound	None	None	None	None
Response mapping	None	Likely to be some problems with patients differentiating between a little & quite a bit (Focus group MX001F).											Concern about several levels in the answer to this item, and would prefer as a Yes or No response (Focus group MX001F).
Retrieval	None	None	None	None	None	None	None	None	None	None	None	None	None
Judgment	None	None	None	None	None	None	None	None	None	None	None	None	None
Translatability	None	Four scalar options work reasonably well in Spanish language (Interview MX001I).	None	None	None	None	None	None	In Spanish language, no good translation for 'deeper tissues' and may	None	None	None	None

									prove a challenge to comprehension (Interview MX001I).				
Other context	None	None	None	None	None	None	None	None	None	None	Useful symptom to collect as commonly reported in patients with SSI in Mexico and not part of current SSI assessment schema (Interview MX002I)	None	Patients do not have access to thermometer, so the number of degrees is unhelpful (Interview MX001I).
Rwanda													
Comprehension	None	None	None	None	None	None	None	Likely to be difficult for patients to understand, and many misunderstand 'opening' as healing (i.e., taking off dressing) or as suture removal. 'Of their own accord' is an important clarification and might need further explanation by the trained assessor (Interview RW001I).	In Rwanda, commonly ask patients 'did you also see the blue sutures' (deep sutures) to help make this clear. Patients do not understand of layered closure, or that there are several layers of the abdominal anatomy (Interview RW002I).	None	None	None	Patients all feel that they have 'temperature rises', as climate very warm and heighten anxiety during recovery, but unless they have a thermometer very difficult to record (Interview RW002I).
Response mapping	None	It will be too challenging for patients in Rwanda to differentiate between 'Quite a bit' and "A little' in local languages - the concepts do not transfer well (Interview RW001I)											None
Retrieval	None	None	None	None	None	None	None	None	None	None	None	None	None
Judgment	None	None	Not very contextually relevant in Rwanda as patients have very dark skin tones. Will be very hard to judge (Interview RW002I).	None	None	None	None	None	None	None	None	None	May not be capturing a useful concept to wound healing in Rwanda. Doesn't adapt well to Rwandan patients (Focus group RW001F).
Translatability	None	Very difficult to translate Quite a Bit versus A little in Kinyarwanda.	Challenging to translate 'redness' - common words in Kinyarwanda relate to 'skin changing from normal' which may be more relevant to patients in Rwanda (Interview RW001I).	None	None	None	None	None	None	None	None	None	None
Benin													
Comprehension	None	None	None	None	None	None	None	None	None	None	None	None	None
Response mapping	None	None	None	None	None	None	None	None	None	None	None	None	None
Retrieval	None	None	None	None	None	None	None	None	None	None	None	None	None
Judgment	None	None	Not possible to identify redness in black skin (Interview BN001I).	None	None	None	None	None	None	None	None	None	None
Translatability	None	None	Concept hard to translate in a way that French speaking patients in Benin will understand (Focus group BN001F).	None	None	None	None	None	Tissue is hard to translate for patients in Benin and they may not understand this (Interview BN001I).	None	None	None	None
Summary of recommendations for adaptation and implementation	No changes to be made. Consent would be confirmed as usual at the start of the interview.	Consensus in qualitative data to reduce the number of response options for symptoms questions to three levels, and to two levels for fever. Decision to triangulate with qualitative data using original response levels in order to make final recommendations for adaptation.	Addition of 'shining of the skin' to support data completion for patients with dark skin tones where redness is difficult to identify.	No changes	Clarification of thin clear fluid	No changes	No changes	No changes. Assessor can ask family member or friend for help to complete questionnaire where needed.	Mixed opinions about ability of patients to tell the difference between the skin and fleshy part beneath. Therefore decision to reword question, but maintain item in adapted Round 1 questionnaire. Assessor ensure safety netting via a formal clinician pathway if any concerns.	No changes	No changes	No changes	Consensus to drop >38oC as patients will not have the ability to record this in a global setting so is redundant. Should be trianqulated with data from pilot study will specifically examine whether this should be considered a Yes / No or three-level item.
Adapted Wound Healing Questionnaire	**Introductory statement**	**1 = Not at all; 2 = A little; 3 = A lot.**	**Was there redness (or shining of the skin) spreading away from the wound?**	**Was the area around the wound warmer than the surrounding skin?**	**Has any part of the wound leaked thin clear fluid?**	**Has any part of the wound leaked blood-stained fluid?**	**Has any part of the wound leaked thick and yellow or green fluid?**	**Have the edges of any part of the wound separated or gaped open of their accord?**	**If the wound edges opened, did the flesh beneath the skin or the inside sutures also separate?**	**Has the area around the wound become swollen?**	**Has the wound been smelly?**	**Has the wound been painful to touch?**	**Have you had, or felt like you have had, a raised temperature or fever?**

Appendix D. Summary of qualitative data used in adaptation of 'treatment' items (12 to 19) in Chapter 4

Thematic domain	Item 12 Have you sought advice because of a problem with your wound, other than at a planned follow-up appointment?	Item 13 Has anything been put on the skin to cover the wound? (dressing)	Item 14 Have you been back into hospital for a problem with your wound?	Item 15 Have you been given antibiotics for a problem with your wound?	Item 16 Have the edges of your wound been deliberately separated by a doctor or nurse?	Item 17 Has your wound been scraped or cut to remove any unwanted tissue?	Item 18 Has your wound been drained? (drainage of pus or an abscess)	Item 19 Have you had an operation under general anaesthetic for treatment of a problem with your wound?
Ghana								
Comprehension	Easy to answer for patients. Will be given a routine review date before discharge, a will be told to come back early if needed (Interview GH001I).	Dressing is not a common term to use in Ghana (Interview GH002I).	None	Patients may not know which medications are antibiotics and which are not. May be an assumption for the interviewer based on other answers (Interview GH002I).	Doctor may have removed a stitch to open the wound – patients should find this easy to answer (Interviewer GH002I).	None	Patients may stay 'water coming from wound', but would be able to identify whether this is bloody, or yellow/green colour 'water' (Interviewer GH001I)	Patients will not understand what a 'general anaesthetic' is. Suggest term 'put to sleep for an operation' (Interviewer GH001I).
Response mapping	None	None	None	None	None	None	None	None
Retrieval	None	None	None	Patients might struggle to remember which medications they were on early after surgery (Focus group 001F).	None	None	None	Patients will definitely remember as this is such a big event. May be difficult to differentiate 'for a wound problem' versus other reason (Interviewer GH002I)
Judgment	None	None	Telling the difference between seeking advice related to a wound problem and going back to hospital difficult (Interview GH001I).	Most patients will be discharged on prophylactic antibiotics from hospital. Unsure whether patients will be able to tell the difference between antibiotics for treatment or prevention (Interview GH001I).	May be some confusion around removal of stitches and opening of a wound (Interview GH001I).	None	Might be hard to differentiate between wound being actively drained and passively draining 'water' (Items 3,4,5) (Interviewer GH002I).	Very likely to be able to answer, as a serious event to return to surgery. Burst abdomen able to report easily (Interviewer GH001I).
Translatability	None	Patients in Ghana use the term 'washing the wound' to describe wound care and dressing, rather than the term dressing (Interview GH001I).	None	None	None	The word tissue is not widely used in Ghana. Suggest the term 'flesh' or 'skin' e.g. the fleshy part underlying the skin (Interview GH001I).	Abscess is not a word that is commonly used or understood by Ghanian patients (Focus group 001F).	None
Other context	It would interesting to explore where the patient had sought advice from, for example from a traditional healer, community healthworker or local health centre (Focus group 001F).	None	Patients will always return to the hospital with any postoperative problems, as traditional healers wont tamper with surgical wounds. Patients will either go back to either district or main hospital (Focus group 001F).	Patients unlikely to be able to read out a medication name over the telephone. Relatives may be able to help answer if more literate (Interview GH001I).	None	None	None	None
Nigeria								
Comprehension	None	Dressings is a term used by patients (Focus group NG001F).	None	Reasonable to assume they will be able to tell whether an antibiotic prescription is due to wound problems (Interview NG003I).	Can happen both on ward & in-theatre, needs to ensure that this is different to Item 19 (Focus group NG002F). Need to be clearly differentiated from reoperation (e.g. in a minors or majors theatre) versus ward/community based wound opening. For example any point in time when doctors/nurses need to separate wound (Interview NG001I).	None	Suggested that yellow or green pus would improve the description (Focus group NG002F). Abscess is not a terminology well known for patients in Nigeria (Focus group NG001F).	May be difficulty in understanding general anaesthetic (Focus group NG001F). Patients won't understand general anaesthesia, a better terminology would be 'put to sleep' (Interview NG002I). Difficult to understand 'general anaesthesia' for Nigerian patients. May be difficult to differentiate re-operation for wound and another cause (Focus group NG002F).
Response mapping	None	None	None	None	None	None	None	None
Retrieval	None	Not many patients would leave hospital with wound dressings still in place, so patients likely to remember if so (Interview NG002I).	Easy for patients to recall as a set event (Interview NG003I).	In general patients will know which drug they are on, and typically will ask what they are for, so likely to be able to answer this (Interview NG001I).	None	None	None	GA will definitely be something a patient remembers (Interview NG001I).
Judgment	None	None	None	However patients may get confused by postoperative prophylactic antibiotics which can be for a prolonged period in Nigeria, but specifically for problem with a wound can be clarified to clear their doubts (Interview NG001I). Patients might find it hard to tell the difference between antibiotics and other medications - this may require furrther explanation or the names of the antibiotics (if literate) (Focus group NG002F).	None	Patients will only be able to answer this if the doctor explains it to them at the time of debridement (Interview NG003I).	None	None
Translatability	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Other context	1st point of contact would be known health worker (doctor/nurse). The question would capture community healers. Other than planned follow-up appointment is useful phrasing to support this (Interview NG001I). This would be a clinic, local hospital, main hospital or community health worker at house (80-90%). <10% of patients would go to traditional healer. (Interview NG002I)	Change of dressing wound be performed by community health worker. Patients may use own dressings (non-standardised) which might include this, of varying materials (Interview NG001I).	Patient may return to hospital where it was performed, or may be sent elsewhere if live a long distance away (Focus group NG002F).	None	Deliberately may have bad connotation in Nigeria - this sounds like a malicious act (Interview NG003I).	Often performed during the process of wound dressing (Interview NG001I).	Pus often releases through pressure on edges or when cut to release pus (Interview NG001I).	None
India								
Comprehension	None	None	None	Patients should know if antibiotic has been provided to them e.g. from pharmacy/doctor (Interview IN002I).	None	None	Patients in general will know the terminology pus but probably not the term 'abscess' (Interview IN002I).	None
Response mapping	None	None	None	None	None	None	None	None
Retrieval	None	None	None	None	None	None	None	None
Judgment	None	None	None	Accuracy may be variable depending on their level of education and the area in which they live (Interview IN001I).	None	None	None	None
Translatability	None	None	None	None	None	None	When translating into Hindi, pus translates as 'bad blood' (Focus group IN001F).	None
Other context	Some patients with complications would travel back to CMC Vellow, others live farther away and would either attend primary or secondary care depending on the area in which they live (Interview IN001I).	Patients often go home with wound care packages, or visit the hospital or healthcare centre for wound care (Interview IN001I).	None	None	None	None	None	Some patients will have spinal anaesthetic for treatment of problesm with their wounds in an operating room but this would not be captured (Focus group IN001F)
Mexico								
Comprehension	None	None	None	None	None	None	Pus an acceptable and comprehensible term (Interview MX002I).	Patients unlikely to understand the term general anaesthesia (Foucs group MX001F).
Response mapping	None	None	None	None	None	None	None	None
Retrieval	None	None	None	Antibiotics are one of the key things they 'like' to receive after an operation (Interview MX002I).	None	None	None	None

Judgment	None	None	None	Patients will be very aware of antibiotics they received & timing of administration (Interview MX001I).	None	None	None	None
Translatability	None	None	None	None	None	None	Patients won't understand the word abscess in Spanish language - concept translates poorly.	None
Other context	Location of help sought will vary depending on the patient's living location; this might be returning to the original hospital or an alternative primary or secondary care location (Interview MX001I)	None	None	None	None	None	None	None
Rwanda								
Comprehension	None	None	None	Confusion between pain medications and antiobiotics for some patients (interview RW002I).	None	None	Concept of pus well understood, and patients can differentiate abnormal fluid. Well differentiated from spontaneous drainage of thick green fluid in item 5 (Interview RW001I).	Second operation' helps with clarification (Interview RW002I). Patients understand the concept of general anaesthesia (Interview RW001).
Response mapping	None	None	None	None	None	None	None	None
Retrieval	None	None	None	None	None	None	None	None
Judgment	None	None	None	When clarifying the type of medication, the form of delivery is a useful addition (e.g., capsule versus tablets). Abx usually take the form of capsules (Focus group RW001).	None	None	None	None
Translatability	None	None	None	None	None	None	Pus translates well into Kinyarwanda (Focus group RW001F).	None
Other context	None	None	None	None	None	This often occurs in a minor operating room or clinical room in Rwanda (Focus group RW001F).	None	None
Benin								
Comprehension	None	None	None	None	Deliberately is easy to understand in French and would not hold negative connotations in Benin (Interview BN001I).	None	None	None
Response mapping	None	None	None	None	None	None	None	None
Retrieval	None	None	None	None	None	None	None	None
Judgment	None	None	None	None	None	None	None	None
Translatability	None	None	None	None	None	None	None	None
Other context	None	None	None	None	None	None	None	None
Adapted Wound Healing Questionnaire	**Have you sought advice because of a problem with your wound, other than at a planned follow-up appointment?**	**Has anything been put on the skin to cover the wound? (dressing)**	**Have you been back into hospital for a problem with your wound?**	**Have you been given medicines (antibiotics) for a problem with your wound?**	**Have the edges of your wound been separated by a doctor or nurse?**	**Has your wound been scraped or cut to remove any unwanted flesh?**	**Has thick, yellow or green fluid (pus) been drained from your wound by a doctor or nurse (abscess)?**	**Have you had to go back to the operating room for treatment of a problem with your wound?**
Summary of recommendations for adaptation and implementation	No changes	No changes	No changes	Change to medicines (antibiotics) in line with other UROM-style items. Assessors recommended to ask patients to read out their medications if possible to double check that they are antibiotics.	Clarification that this question is specifically examining separation of the wound edges, not under general anaesthesia. Word 'deliberately' removed due to perceived negative connotations.	Change to the word tissue to 'flesh' as consensus that this was medical terminology	Limited understanding of the terminology abscess. Concern about crossover with fluid drainage (yellow/green) in Item 5 - decision from working group to highlight drainage 'by doctor or nurse' as an active event.	Consensus that general anaesthesia will not be a concept well understood by LMIC populations. Theme that spinal anaesthesia very common in some LMICs, so would not capture these operations. Terminology adapted to collect information about any procedure carried out in the OR.

Appendix E. Adapted version of Wound Healing Questionnaire (English language)

Wound Healing Questionnaire

For questions with tick boxes, please tick one box per question.

The Wound Healing Questionnaire should be completed between 27-30 days after the patient's operation over the telephone, for patients that will undergo 30-day assessment as part of the trial. The Wound Healing Questionnaire should not be completed by the same person that will complete the standard 30-day Follow-up Form.

	Trial Number	
	Centre name	_____
	Patient name	_____
Patient status		
Has the patient died?	☐ Yes (***please stop at Patient Status***)	☐ No (continue to Follow-up details)
If patient died, date of death	d d m m y y y y	***If patient died, an SAE form must be completed***
If patient died, main cause of death	_____	
Follow-up pathway		
Attempts made to connect with patient	☐ 1 attempt ☐ 4-5 attempts	☐ 2-3 attempts ☐ >5 attempts
Were you able to contact the patient by telephone?	☐ Yes (please continue)	☐ No (***please stop here***)
If telephone contact was made, date of contact	d d m m y y y y	
Ask the patient: What type of phone are you using for this call?	☐ Landline phone ☐ Mobile phone (with a camera)	☐ Mobile phone (without a camera) ☐ Commercial call centre
Ask the patient: Who owns the phone that you're speaking to me on?	☐ Patient themselves ☐ Healthcare worker	☐ Friend or relative ☐ Other (please specify): _____
Ask the patient: Do you live in an urban (mostly city or town) or rural (mostly countryside) area?	☐ Urban	☐ Rural
Ask the patient: What is the highest level of education that you have achieved?	☐ High/secondary school or above	☐ Did not complete first/ primary school or no formal education

	☐ First/primary school level	
What language did the patient use to respond to the questionnaire?	☐ English	☐ Other (please specify): _____
If other: Was the formal translated Questionnaire used?	☐ Yes, formal questionnaire	☐ No, translated by questionnaire administrator
	☐ No, using formal translator	

Wound Healing Questionnaire

Please read the following statement to the patient:

We are interested in knowing how the cut(s) of your skin (called your wound(s)) have healed since you left hospital after your surgery. It is fine to ask someone else to help answer some of the questions, for example if you cannot easily see your wound(s). If you have more than one wound, please answer the questions thinking about just one wound. This should be either the wound with which you have had concerns about how it was healing, or the longest wound if there have been no specific concerns. We would like you to think about the wounds on your skin. Some of the questions I am about to ask you relate to some problems that may occur with wound healing. Please note, many people do not experience these problems after having surgery.

Since you left hospital after having surgery…

Was there redness (or shining of the skin) spreading away from the wound?	☐ Not at all	☐ A little	☐ A lot
Was the area around the wound warmer than the surrounding skin?	☐ Not at all	☐ A little	☐ A lot
Has any part of the wound leaked thin, clear fluid?	☐ Not at all	☐ A little	☐ A lot
Has any part of the wound leaked blood-stained fluid?	☐ Not at all	☐ A little	☐ A lot
Has any part of the wound leaked thick and yellow or green fluid?	☐ Not at all	☐ A little	☐ A lot
Have the edges of any part of the wound separated or gaped open of their accord?	☐ Not at all	☐ A little	☐ A lot
If the wound edges opened: Did the flesh beneath the skin or the inside sutures also separate?	☐ Not at all	☐ A little	☐ A lot
Has the area around the wound become swollen?	☐ Not at all	☐ A little	☐ A lot
Has the wound been smelly?	☐ Not at all	☐ A little	☐ A lot
Has the wound been painful to touch?	☐ Not at all	☐ A little	☐ A lot

Since you left hospital after having surgery…		
Have you had, or felt like you have had, a raised temperature or fever?	☐ Yes	☐ No
Have you sought advice because of a problem with your wound, other than at a planned follow-up appointment?	☐ Yes	☐ No
Has anything been put on the skin to cover the wound? (dressing)	☐ Yes	☐ No
Have you been back into hospital for a problem with your wound?	☐ Yes	☐ No
Have you been given medicines (antibiotics) for a problem with your wound?	☐ Yes	☐ No
Have the edges of your wound been separated by a doctor or nurse?	☐ Yes	☐ No
Has your wound been scraped or cut to remove any unwanted flesh?	☐ Yes	☐ No
Has thick, yellow or green fluid (pus) been drained from your wound by a doctor or nurse (abscess)?	☐ Yes	☐ No
Have you had to go back to the operating room for treatment of a problem with your wound?	☐ Yes	☐ No

Ask the patient: How happy were you with having your follow-up over the telephone?	☐ Very satisfied	☐ Satisfied	☐ Neither satisfied nor dissatisfied
	☐ Very unsatisfied	☐ Unsatisfied	

Before you end the call, inform the patient that **this telephone questionnaire will not replace the 30-day in-person wound assessment** required as part of the trial, and **they must still have their 30-day follow-up appointment**.

Time taken to complete telephone questionnaire	☐ <10 minutes	☐ 11-20 minutes
	☐ 21-30 minutes	☐ >30 minutes

Please add any further comments or details of the telephone follow-up here:

Form completed by

Job role	☐ Surgeon	☐ Other doctor	☐ Nurse	☐ Other (please specify): _____						
Print full name										
Signature		Date form completed	d	d	m	m	y	y	y	y

Please note this questionnaire is licensed from Oxford University Innovation (OUI) outcomes group. Please contact OUI to obtain a free license to use this questionnaire for

Appendix F. Translations of adapted Wound Healing Questionnaire

Translated versions available at:

https://drive.google.com/drive/folders/1RX_HqOS8rKtC43TznhC4cj-PAO0WpFs5?usp=sharing

India	Tamil
	Bengali
	Hindi
	Punjabi
Ghana	Dagbani
	Twi
Mexico	Spanish
Rwanda	Kinyarwanda

Appendix G. STARD Checklist in Chapter 5

Section & Topic	No	Item	Complete
TITLE OR ABSTRACT			
	1	Identification as a study of diagnostic accuracy using at least one measure of accuracy (such as sensitivity, specificity, predictive values, or AUC)	✓
ABSTRACT			
	2	Structured summary of study design, methods, results, and conclusions (for specific guidance, see STARD for Abstracts)	✓
INTRODUCTION			
	3	Scientific and clinical background, including the intended use and clinical role of the index test	✓
	4	Study objectives and hypotheses	✓
METHODS			
Study design	**5**	Whether data collection was planned before the index test and reference standard were performed (prospective study) or after (retrospective study)	✓
Participants	**6**	Eligibility criteria	✓
	7	On what basis potentially eligible participants were identified (such as symptoms, results from previous tests, inclusion in registry)	✓
	8	Where and when potentially eligible participants were identified (setting, location and dates)	✓
	9	Whether participants formed a consecutive, random or convenience series	✓
Test methods	**10a**	Index test, in sufficient detail to allow replication	✓
	10b	Reference standard, in sufficient detail to allow replication	✓
	11	Rationale for choosing the reference standard (if alternatives exist)	✓
	12a	Definition of and rationale for test positivity cut-offs or result categories of the index test, distinguishing pre-specified from exploratory	✓
	12b	Definition of and rationale for test positivity cut-offs or result categories of the reference standard, distinguishing pre-specified from exploratory	✓
	13a	Whether clinical information and reference standard results were available to the performers/readers of the index test	✓
	13b	Whether clinical information and index test results were available to the assessors of the reference standard	✓
Analysis	**14**	Methods for estimating or comparing measures of diagnostic accuracy	✓
	15	How indeterminate index test or reference standard results were handled	✓
	16	How missing data on the index test and reference standard were handled	✓
	17	Any analyses of variability in diagnostic accuracy, distinguishing pre-specified from exploratory	✓
	18	Intended sample size and how it was determined	✓
RESULTS			
Participants	**19**	Flow of participants, using a diagram	✓
	20	Baseline demographic and clinical characteristics of participants	✓
	21a	Distribution of severity of disease in those with the target condition	✓
	21b	Distribution of alternative diagnoses in those without the target condition	✓
	22	Time interval and any clinical interventions between index test and reference standard	✓
Test results	**23**	Cross tabulation of the index test results (or their distribution) by the results of the reference standard	✓
	24	Estimates of diagnostic accuracy and their precision (such as 95% confidence intervals)	✓
	25	Any adverse events from performing the index test or the reference standard	✓
DISCUSSION			
	26	Study limitations, including sources of potential bias, statistical uncertainty, and generalisability	✓
	27	Implications for practice, including the intended use and clinical role of the index test	✓
OTHER INFORMATION			
	28	Registration number and name of registry	✓
	29	Where the full study protocol can be accessed	✓
	30	Sources of funding and other support; role of funders	✓

Appendix H. Collaborating colleagues

TALON study steering group

James Glasbey (UK), Adesoji Ademuyiwa (Nigeria), Alisha Bhatt (India), Bruce Biccard (South Africa), Jane Blazeby (UK), Peter Brocklehurst (UK), Sohini Chakrabortee (UK), JC Allen Ingabire (Rwanda), Francis Moïse Dossou (Benin), Irani Durán (Mexico), Rohini Dutta (India), Dhruva Ghosh (India), Frank Gyamfi (Ghana), Parvez Haque (India), Pollyanna Hardy (UK), Gabriella Hyman (South Africa), Ritu Jain (India), Oluwaseun Ladipo-Ajayi (Nigeria), Ismail Lawani (Benin), Souliath Lawani (Benin), Mwayi Kachapila (UK), Rachel Lillywhite (UK), Rhiannon Macefield (UK), Laura Magill (UK), Janet Martin (Canada), Jonathan Mathers (UK), Kenneth McLean (UK), Punam Mistry (UK), Rohin Mittal (India), Mark Monahan (UK), Rachel Moore (South Africa), Dion Morton (UK), Moyo Ojo (Nigeria), Faustin Ntirenganya (Rwanda), Emmanuel Ofori (Ghana), Rupert Pearse (UK), Alberto Peón (Mexico), Thomas Pinkney (UK), Antonio Ramos de la Medina (Mexico), Tubasiime Ronald (Rwanda), David Roman (Mexico), Emmy Runingamugabo (Rwanda), Alice Sitch (UK), Anita Slade (UK), Stephen Tabiri (Ghana), Donna Smith (UK), Aneel Bhangu (UK).

Statistical advisors

Alice Sitch, Anita Slade, Mike Horton, Duc Khanh To, Aneel Bhangu, Pollyanna Hardy

National hub lead investigators

Adesoji O Ademuyiwa, Lawani Ismail, Dhruva Ghosh, Antonio Ramos de la Medina, Rachel Moore, Faustin Ntirenganya, Stephen Tabiri

Community engagement and involvement partners

Emmy Runingamugabo (Rwanda), Simin Patrawala (India), Angela Prah (Ghana), Christian Oko (Nigeria), Karolin Kroese (UK), Michael Bahrami-Hessari (UK)

TALON collaborators

Benin

Ismaïl Lawani, Francis Moïse Dossou*, Corinne Dzemta, Covalic Melic Bokossa Kandokponou, Souliath LAWANI (Centre Hospitalier Universitaire et Départemental Ouémé-Plateau (CHUDOP), Porto-Novo)

Hulrich Behanzin* (Hopital de Zone de Menontin, Contonou)

Cyrile Kpangon* (Centre Hospitalier Universitaire de Zone de Suru Lere (SLERE), Contonou)

Ghana

Bernard Appiah Ofori, Stephen Tabiri*, Abdul-Hafiz Saba, Gbana Limann, Daniel Kwesi Acquah, Shamudeen Mohammed Alhassan, Sheriff Mohammed, Owusu Abem Emmanuel, Yakubu Musah, Yenli Edwin, Sheba Kunfah, Yakubu Mustapha, Abantanga Atindaana Francis, Emmanuel Ayingayure, Gbana Limann (Tamale Teaching Hospital, Tamale)

Forster Amponsah-Manu, Eric Agyemang, Vera Agyekum, Esther Adjei-Acquah, Emmanuel Yaw Twerefour, Barbra Koomson, Ruby Acheampong Boateng, Ato Oppong Acquah, Richard Ofosu-Akromah, Leslie Issa Adam-Zakariah (Eastern Regional Hospital, Koforidua)

Nii Armah Adu-Aryee, Theodore Wordui (Korle-Bu Teaching Hospital, Accra)

Coomson Christian Larbi, Akosa Appiah Enoch, Mensah Elijah, Kyeremeh Christian, Addo Gyambibi Kwame, Boakye Percy, Kontor Effah Bismark, Gyamfi Brian, Manu Ruth (Techiman Holy Family Hospital (THFH), Techniman)

Romeo Hussey, Samuel Dadzie, Akosua Dwamena Appiah, Grace Yeboah, Cynthia Yeboah, James Amoako, Regina Acquah, Naa Anyekaa Sowah (Berekum Holy Family Hospital (BHFH), Berekum)

Atta Kusiwaa, Esther Asabre (Effia-Nkwanta Regional Hospital (ENRH), Takoradi)

Cletus Ballu, Charles Gyamfi Barimah (Salaga District Hospital, Salaga)

Frank Owusu (St. Patrick's Hospital, Offinso)

Clement Sie-Broni, Vivian Adobea, Prince Yeboah Owusu, Marshall Zume, Abdul-Hamid Labaran, Raphael Adu-Brobbey (Sunyani Regional Hospital (SRH), Sunyan)

Martin Tangnaa Morna, Samuel A. Debrah, Patrick Opoku Manu Maison, Michael Nortey, Donald Enti, Mabel Pokuah Amoako-Boateng, Anthony Baffour Appiah, Emmanuel Owusu Ofori, Richard Kpankpari, Benedict Boakye, Elizabert Mercy Quartson, Patience Koggoh (Cape Coast Teaching Hospital, Cape Coast)

Anita Eseenam Agbeko, Frank Enoch Gyamfi, Joshua Arthur, Joseph Yorke, Christian Kofi Gyasi-Sarpong, Charles Dally, Agbenya Kobla Lovi, Michael Amoah, Boateng Nimako, Robert Sagoe, Anthony Davor, Fareeda Galley, Michael Adinku, Jonathan Boakye-Yiadom, Jane Acquaye, Juliana Appiah, Dorcas Otuo Acheampong, Iddrisu Haruna, Edward Amoah

Boateng, Emmanuel Kafui Ayodeji, Samuel Tuffuor, Naa Kwarley, Yaa Tufuor (Komfo Anokye Teaching Hospital (KATH), Kumasi)

Ramatu Darling Abdulai, Fred Dankwah, Ralph Armah, Doris Ofosuhene, Dorcas Osei-Poku, Arkorful Ebenezer Temitope, Delali Akosua Gakpetor, Victoria Sena Gawu, Christopher Asare, Enoch Tackie ((Greater) Accra Regional Hospital (ARH), Accra)

James Ankomah, Isaac Omane Nyarko, Zelda Robertson, Serbeh Godwin, Appiah Anthony Boakye, Godfred Fosu, Frank Assah-Adjei (Goaso Municipal Hospital, Goaso)

India

Christian Medical College, Ludhiana: Parvez Haque*, Ritu Jain*, Alisha Bhatt, Jyoti Dhiman, Rohini Dutta, Dhruva Ghosh

Christian Medical College, Vellore: Esther Daniel, Priyadarshini K, Latha Madankumar, Rohin Mittal*, Ida Nagomy, Soosan Prasad

Madhipura Christian Hospital, Madhipura: Arpit Jacob Mathew*, Danita Prakash, Priya Jacob, Jeremiah P Anachy, Amy Mathew

Lady Willingdon Hospital, Manali: Josy Thomas, Philip V Alexander*, Pradeep Zechariah, Neerav D Aruldas

Sher-i-Kashmir Institute Of Medical Sciences, Srinagar: Asif Mehraj*,Hafsa Imtiyaz Ahmed, Rauf A Wani, Fazl Q Parray, Nisar A Chowdri

Mexico

Antonio Ramos De la Medina*, Laura Martinez Perez Maldonado, Diana S Gonzalez Vazquez, Iran I Durán Sánchez, Maria J Martínez Lara, Alejandra Nayen Sainz de la Fuente (Hospital Español Veracruz, Veracruz)

Ana O Cortes Flores*, Mariana E Barreto Gallo, Alejandro Gonzalez Ojeda, Monica E Jimenez Velasco (Hospital San Javier, Guadalajara)

Luis Hernández Miguelena*, Reyes J Cervantes Ortiz, Gonzalo I Hernandez Gonzalez, Marco Hurtado Romero, Rosa I Hernandez Krauss (Hospital Regional, Veracruz)

Luis A Dominguez Sansores*, Alejandro Cuevas Avendaño, Celina Cuellar Aguirre, Isaac Baltazar Gomez, Hector Ortiz Mejia (Hospital General de Boca Del Rio, Boca del Rio)

Alejandro González Ojeda*, Oscar E Olvera Flores, Erick A González García de Rojas, Kevin J Pintor Belmontes, Francisco J Barbosa Camacho, Aldo Bernal Hernández, Laura Reyes Aguirre, Rubén E Morán Galaviz, Clotilde Fuentes Orozco, Wenceslao G Ángeles Bueno, Fernando S Ramirez Marbello, Diego E Luna Acevedo, Michel Hernández Valadez, Ana L Bogurin Arellano, Luis R Ramírez-González, Bertha G Guzmán Ramírez, Eduardo Valtierra Robles, Ramona I Rojas García, José V Pérez Navarro, Edgar J Cortes Torres (Hospital De Especialidades, CMNO, Guadalajara)

David R Dominguez Solano* (Hospital Regional de Xalapa, Xalapa)

Alberto N Peón*, Roque D Lincona Menindez, Rozana Reyes Gamez, Maria C Paz Muñoz, (Hospital Espanol Pachuca, Puebla)

Nigeria

Lagos University Teaching Hospital, Lagos (Hub):Orimisan Belie, Victoria Adeleye, Adesoji Ademuyiwa*, Oluwafunmilayo Adeniyi, Opeyemi Akinajo, David Akinboyewa, Felix Alakaloko, Oluwole Atoyebi, Olanrewaju Balogun, Christopher Bode, Olumide Elebute, Francis Ezenwankwo, Adesiyakan Adedotun, George Ihediwa, Jubril Kuku, Oluwaseun Ladipo-Ajayi, Ayomide Makanjuola, Samuel Nwokocha, Olubunmi Ogein, Rufus Ojewola, Abraham Oladimeji, Thomas Olajide, Iyabo Alasi, Oluwaseun Oluseye, Justina Seyi-Olajide, Adaiah Soibi-Harry, Emmanuel Williams, Agbulu Moses Vincent, Nnamdi Jonathan Duru, Kenneth Uche Onyekachi, Christiana Ashley, Chinelo Victoria Mgbemena, Moyosoluwa Ojo, Olowu Oluyemisi, Iyabode Ikuewunmi, Adeoluwa Adebunmi, Edet Glory Bassey, Ephraim Okwudiri Ohazurike, Olayide Michael Amao, Osunwusi Benedetto Oluwaseun, Emily Doris, Olutola Stephen, Christianah Gbenga-Oke, Olawunmi Olayioye, Olowu Oluyemisi, Kayode Oluremi, Esther Abunimye, Christianah Oyegbola, Olayade Kayode, Adeola Ayoola Orowale,

Omolara M Williams*, Olufunmilade A Omisanjo, Omolara M Faboya, Zainab O Imam, Olabode A Oshodi, Yusuf A Oshodi, Ayokunle A Ogunyemi, Olalekan T Ajai, Francisca C Nwaenyi (Lagos State University Teaching Hospital (LASUTH))

Adewale O Adisa*, Adewale A Aderounmu, Funmilola O Wuraola, Oludayo Sowande (Obafemi Awolowo University Teaching Hospital, Ile-Ife (IFE))

Lukman Olajide Abdur-Rahman*, Jibril Oyekunle Bello, HADIJAT OLAIDE RAJI, Nurudeen Abiola Adeleke, Saheed Abolade Lawal, Rafiat Tinuola Afolabi, Abdulwahab Lawal (University of Ilorin Teaching Hospital (UITH), Ilorin)

Okechukwu Hyginus Ekwunife*, Ochomma Amobi Egwuonwu, Chisom Faith Uche (Nnamdi Azikiwe University Teaching Hospital (NAUTH), Nnewi)

Abubakar Bala AB Muhammad, Saminu S Muhammad, Idris Usman IU Takai, Mohammed AS Aliyu Salele, Onyekachi G Ukata, Mahmoud Kawu MK Magashi, Lawal Barau LB Abdullahi, Bello Abodunde BA Muideen, Khadija A Ado, Lofty-John Chukwuemeka LJC Anyawu* (Aminu Kano Teaching Hospital, Kano)

Samson Olori*, Samuel A Sani, Olabisi O Osagie, Ndubuisi Mbajiekwe, Oseremen Aisuodionoe-Shadrach, Godwin O Akaba, Lazarus Ameh, Lazarus Ameh, Francis o Adebayo, Martins Uanikhoba, Felix o Ogbo, Nancy O Tabuanu (University of Abuja Teaching Hospital, Abuja)

Taiwo A Lawal*, Rukiyat A Abdus-Salam, Akinlabi E Ajao, Augustine O Takure, Omobolaji O Ayandipo, Hyginus O Ekwuazi, Olukayode Abayomi, Olatunji O Lawal, Solomon Olagunju, Kelvin I Egbuchulem, Sikiru Adekola Adebayo, Peter Elemile (University College hospital, Ibadan)

Usang E Usang*, Joseph E Udosen, Expo E Edet, Akan W Inyang, Edima M Olory, Gabriel U Udie, Godwin O Chiejina, Adams D Marwa, Faith J Iseh, Sunday A Ogbeche, Mary O Isa (University of Calabar Teaching Hospital, Calabar)

Uchechukwu O Ezomike, Sebastian O Ekenze*, Matthew I Eze, Emmanuel O Izuka, Jude K Ede, Vincent C Enemuo, Okezie M Mbadiwe, Ngozi G Mbah (University of Nigeria Teaching Hospital (UNTH) Enugu)

Rwanda

Alphonsine Imanishimwe, Sosthene Habumuremyi, Faustin Ntirenganya*,JC Allen Ingabire, Isaie Ncogoza, Emmanuel Munyaneza, Jean de Dieu Haragirimana, Christian Jean Urimubabo, Violette Mukanyange, Jeannette Nyirahabimana, Emmanuel Mutabazi (Rwanda University Teaching Hospital of Kigali (CHUK), Kigali)

Christophe Mpirimbanyi*, Olivier Mwenedata, Hope Lydia Maniraguha, Emelyne Izabiriza, Moses Dusabe, Job Zirikana, Francine Uwizeyimana, Josiane Mutuyimana, Elisee Rwagahirima (Kibungo Referral Hospital (KIBUN), Ngoma District)

Alphonsine Imanishimwe, Ronald Tubasiime*, Aphrodis Munyaneza, Sosthene Habumuremyi, Salathiel Kanyarukiko, Gibert Ndegamiye, Francine Mukaneza, Jean Claude Uwimana, Pierrine Nyirangeri, Deborah Mukantibaziyaremye (Kibogora Hospital (KIBO), Kirambo)

Aime Dieudonne Hirwa*, Salomee Mbonimpaye, Piolette Muroruhirwe, Christine Mukakomite, Elysee Kabanda (Ruhengeri Referral Hospital, Musanze District)

South Africa

Rachel Moore, Ncamsile Anthea Nhlabathi, Maria Fourtounas, Mary Augusta Adams, Gabriella Hyman*, Hlengiwe Samkelisiwe Nxumalo, Nnosa Sentholang, Mmule Evelyn Sethoana, Mpho Nosipho Mathe (Chris Hani Baragwanath Hospital, Johannesburg)

Zain Ally* (Helen Joseph Hospital, Johannesburg)

Margot Flint, Bruce Biccard (Groote Schuur Hospital, Cape Town)