Identifying eligible patients for Human Immunodeficiency Virus (HIV) testing in UK primary care setting: derivation and internal validation of HIV risk score using a retrospective cohort study and investigation of issues in implementation.

BY

BENHILDAH NYARADZO RUMBWERE DUBE

A thesis submitted to The University of Birmingham for the degree of

DOCTOR OF PHILOSOPHY

Institute of Applied Health Research

College of Medical and Dental Sciences

The University of Birmingham

June 2022

ABSTRACT

The Acquired Immuno-Deficiency Syndrome (AIDS) and its causative agent, Human Immunodeficiency Virus (HIV) were recognised in the early 1980s. HIV/AIDS is one of the highest contributors to morbidity and a leading cause of mortality, worldwide, making it a public health concern. Diagnosis with HIV during the 1980s and early 1990s was like a death penalty but the life expectancy of the HIV infected individuals has increased over the years and is approaching that for the general population. This is attributed to improvement in management and treatment of people living with HIV/AIDS, the effective use of antiretroviral therapies (ARTs) combined with early HIV diagnosis, because earlier treatment with ARTs is more effective. In addition to enabling individuals to get better health outcomes and optimise their quality of life, early diagnosis of HIV is also important in reducing onward transmission of the disease and reducing healthcare costs. In UK, an action plan was put in plan to eliminate HIV/AIDS by 2025 and this relies on increasing uptake of HIV testing in various settings including primary care, in order to increase early diagnosis.

This thesis developed a prediction model which could be used in primary care to identify patients likely to be HIV positive and to prompt clinicians to offer them HIV testing. It also investigated issues in implementation of a point-of-care alert.

A systematic review was conducted to identify candidate predictor variables to use in the prediction model. Qualitative research was undertaken to find out if GPs used euphemistic terms to record a diagnosis of HIV in primary care records. Using the results from the systematic review and the qualitative research, a cohort study was conducted to derive and internally validate a HIV prediction model. Finally, a systematic review was conducted to elicit clinicians' views on barriers and facilitators of use of point-of-care alerts to ensure that they are considered in the development of a HIV point-of-care alert.

Demographic, socio-economic, lifestyle/behavioural, and clinical and comorbid conditions that predict risk of HIV infection were identified in this research. The demographic and socio-economic predictors associated with a risk of HIV infection were female gender (HR, 1.19 (CI: 1.13-1.25)), 25-34 years age group (HR, 1.29 (CI: 1.21-1.39)), being of black, (HR 10.95 (CI: 10.08-11.89)) and mixed/other ethnicity, deprivation (HR ranging from 1.3 to 1.85, increasing with deprivation) and living in urban areas (HR, 1.12 (CI: 1-1.25)). Lifestyle predictors were current smoker or ex-smoker (HR, 1.01 (CI:1.01-1.02), drug misuse (HR, 2.25 (CI: 2.01-2.52)) and contact abroad (HR, 2.04 (CI: 1.76-2.36)). Clinical and comorbid conditions included Kaposi's sarcoma (HR, 171.01 (CI: 89.06-328.37)), pneumocystis carinii (HR, 71.15 (CI: 10.09-501.98)), progressive multifocal leukoencephalopathy (HR, 55.89 (CI: 14.16-220.66)), syphilis (HR, 10.88 (CI: 6.86-17.27)), non-Hodgkin's lymphoma (HR, 9.31 (CI: 7.04-12.33)), tuberculosis (HR, 2.29 (CI: 1.15-4.55)), cerebral toxoplasmosis abscess (HR, 7.88 (CI: 2.98-20.84)). The other clinical and comorbid predictors were anal cancer or anal intraepithelial dysplasia, aseptic meningitis/encephalitis, oral candidiasis, hepatitis B and C, blood dyscrasia, chronic liver disease, depression and current STI (excluding syphilis) or any previous STI. The C-statistic from the model was 0.74 the optimism adjusted C-slope was 0.990. The sensitivity at 0.25% cut-off was 37% and the specificity was 84%.

The results from the model could develop a risk score to identify patients at high risk of HIV infection in primary care through a point-of-care alert. The identified patients could be offered HIV testing. Barriers and facilitators that affect the use of point-of-care alerts were identified. They fall under the intervention (characteristics of the alert), features pertaining to the setting (GP practice) and person (features related to the clinicians).

This study identified demographic, socio-economic, lifestyle/behavioural, clinical and comorbid conditions that predict risk of HIV infection. These predictors could be used to identify patients at high risk of HIV infection in primary care through a point-of-care alert. The study identified the barriers and facilitators that should be considered to ensure utilisation of a pop-up alert in primary care.

DEDICATION

To my parents Mr Alexius T Rumbwere and my late mother, Mrs Valeria Rumbwere whose guidance, encouragement and support enabled me to be where I am today.

To my children Brandon, Cecil and David, who gave up family time and supported me throughout this programme.

ACKNOWLEDGEMENTS

First and foremost, I am extremely grateful to my PHD supervisory team – Professor Tom Marshall (Lead Supervisor) and Co-Supervisors: Dr Ronan Ryan, Dr Antje Lindenmeyer and Miss Christina Easter. Their guidance, advice, encouragement, patience and support were very valuable to me throughout the whole programme.

My profound gratitude to Ms Susan Bayliss for the support in developing systematic review searching strategy, Dr Linda Nichols for technical support on IMRD data extraction and analysis and to Dr Sarah Price (University of Exeter) and Dr Kym Snell (Keele University) for advice on cohort study data analysis. Thank you goes to Nicola Adderley for providing feedback on the thesis drafts.

Special thank you goes to Dr Modupe Omonijo for her hard work as the second reviewer for the first systematic review and also to Dr Umar Shehu for his patience and hard work as a second reviewer for the last systematic review.

Thank you also goes to Rob Currie for providing me with information on what is happening at local level in reducing late diagnosis of HIV.

Finally, my gratitude goes to IQVIA Medical Research Data (IMRD) for providing the data used in developing the prediction model.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

GLOSSARY OF TERMS

AIDS	Acquired Immuno-deficiency Syndrome
AMR	All-cause Mortality Rate
AMR date	Acceptable Mortality Reporting date
ART	Antiretroviral Therapy
AUC	Area under the curve
BASHH	British Association for Sexual Health
BHIVA	British HIV Association
BIA	British Infection Association
cART	combination Antiretroviral Therapy
CCG	Clinical Commissioning Group
CD4	Cluster of Differentiation 4
CDS	Computerised Decision Support
CDSS	Computerised Decision Support Systems
CI	Confidence Interval
CPRD	Clinical Practice Research Datalink
DCA	Decision Curve Analysis
DH	Department of Health
DHSC	Department of Health and Social Care
EMR	Electronic Medical Records
GP	General Practitioner
GUM	Genitourinary Medicine
HAART	Highly Active Antiretroviral Therapy
HANDD	HIV and AIDS New Diagnoses and Deaths Database
HARS	HIV and AIDS Reporting System
HES	Hospital Episodes Statistics
HIV	Human Immunodeficiency Virus

IDPS	Infectious Diseases in Pregnancy Screening Programme
INB	Incremental Net Benefit
ISOSS	Integrated Screening Outcomes Surveillance Service
IT	Information Technology
IMRD	IQVIA Medical Research Data
KS	Kaposi's sarcoma
MSM	Men who have sex with Men
NAAT	Nucleic Acid Amplification Test
NHS	National Health Service
NICE	National Institute for Health and Care Excellence
NIHR	National Institute for Health Research
NRTI	Nucleoside Reverse Transcriptase Inhibitor
NNRTI	Non-Nucleoside Reverse Transcriptase Inhibitor
OR	Odds ratio
P24 antigen tests	Protein 24 antigen tests
PCP	Pneumocystis carinii pneumonia
PCR	Polymerase chain reaction
PHE	Public Health England
POC	Point-of-care
POCTs	Point-of-care tests
PRISMA	Systematic Reviews and Meta-analyses
PRISMA-P	Systematic Reviews and Meta-analyses for Protocols
PI	Protease Inhibitors
RCT	randomised controlled trial
RDT	Rapid diagnostic tests
RCGP	Royal College of General Practitioners
SIGN	Scottish Intercollegiate Guidelines Network.

THIN	The Health Improvement Network
UKHSA	United Kingdom Health Security Agency
UNAIDS	United Nations Programme on HIV/AIDS
WHO	World Health Organisation

Chapter 1: Introduction

1.1 Introduction to chapter

This chapter will cover the background to HIV, late diagnosis, public health significance, effect of Covid-19 pandemic on HIV, strategy and United Kingdom (UK) policy on HIV/AIDS, the role of primary care in diagnosing and managing HIV and the guidelines relevant to primary care. The last section of the chapter introduces thesis overview, covering Wilson and Jungner principles of screening, a summary of the purpose of this thesis and overview of the chapters.

1.2 Background to HIV

The Acquired Immuno-Deficiency Syndrome (AIDS) is an infectious disease caused by a retrovirus, Human Immunodeficiency Virus (HIV) [1] [2]. AIDS was first recognised in 1981 when rare infections, pneumocystis carinii, pneumonia (PCP) and Kaposi's sarcoma (KS), were diagnosed in injecting drug users and otherwise healthy gay men [3] [4] [5] [6] [7] [8] [9]. These rare infections suggested severe impairment of the immune system due to depletion of T-lymphocytes cells [4] [10]. The patients died, despite treatment of the opportunistic conditions, with a median survival time after diagnosis of less than 20 months [10] [11]. Research identified HIV as the etiologic agent in 1983 [3] [4].

HIV is transmitted through the exchange of a variety of bodily fluids, mainly sexually, perinatal and blood-borne [7] [12] [13]. HIV infected individuals become HIV positive within 3 to 6 months of infection, and high levels of the virus circulate in the body, resulting in rapid replication in infected cells [14]. Following initial infection, some individuals develop acute retroviral syndrome symptoms, followed by up to several years, where they remain asymptomatic after seroconversion [3] [15]. HIV is more concentrated during the acute infection stage and the advanced stage, when viral shedding is at the highest level [16].

The progression from HIV infection to AIDS is characterised by impaired cell-mediated immunity, due to depletion of Cluster of Differentiation 4 (CD4) T-lymphocytes cells, leading to severely damaged immunologic functions [4]. There is a continuum of clinical phases, indicative of the level of immunodeficiency associated with the advancement of HIV infection [17]. The World Health Organisation (WHO) produced a Clinical Staging system. Stage 1 patients are asymptomatic and have persistent generalised lymphadenopathy for longer than 6 months [17] [18]. Stage 2 patients have mild symptoms such as unexplained weight loss of less than 10 percent of total body weight, recurrent respiratory infections and a range of dermatological conditions [17] [18]. Stage 2 may last 10 years or more without HIV treatment [19]. Stage 3 patients are moderately symptomatic with additional clinical conditions including weight loss of over 10 percent of total body weight, prolonged unexplained diarrhoea and severe systemic bacterial infections [17] [18]. Stage 3 lasts about 3 years without treatment [19]. Stage 4 patients are severely symptomatic with all AIDS defining symptoms [17] [18]. At this stage, when CD4 cells decline, patients become more susceptible to opportunistic conditions such as PCP, KS, severe ulcerative herpes simplex infections, oral and oesophageal candidiasis and central nervous system toxoplasmosis [20] [10]. The main cause of death in AIDS patients is opportunistic infections [10].

Initially, the diagnosis with HIV during the 1980s and early 1990s, when no effective treatment was available, was like a death penalty [11] [21] [22]. Improvement in management and treatment of people living with HIV/AIDS, the effective use of antiretroviral therapies (ARTs) combined with early diagnosis, over the years, has resulted in life expectancy of HIV infected individuals almost approaching that for the general population [11] [21]. In a UK study, patients diagnosed with HIV in 1997-2003 had approximately 45 percent higher risk of death compared to those diagnosed in 2008-2012 (hazard ratio=0·55, CI:0·48–0·63) [23]. A study in England and Wales confirmed a decline in All-cause Mortality Rate (AMR) in HIV diagnosed patients aged 15-59 years from 217 per 10,000 people in 1999 to 82 per 10,000 in 2008 [21]. A

Canadian study also discovered a decline of 83 percent in All-cause Standardised Mortality Ratio (≥19 years) from 126.8 (95% CI: 84.9-168.6) per 1000 population in 1996 to 21.3 (95% CI: 17.8-24.8) per 1000 in 2011-2012 [24].

This section provided the background to HIV/AIDS and revealed that it was like a death sentence when it was first diagnosed in the 1980s and early 1990s. The effectiveness of ARTs and improvement in management of care have increased the life expectancy of people living with HIV. This improvement in management of care relied on the use of diagnostic tests to identify infected individuals, covered in the next section.

1.3 Diagnostic tests for HIV

HIV testing is performed through serological tests of CD4 counts and HIV viral load, which detect specific antibodies, antigens, or both in a serum sample, thereby establishing the rate of damage to immune system and the level of immunosuppression [17] [16] [25].The methods used in routine HIV testing either involve use of screening assays on blood for laboratory testing or rapid tests conducted on samples from a finger-prick or mouth swab at point-of-care. The commonly used and recommended first-line assays detect HIV antibodies and the HIV p24 antigens simultaneously [20] [26]. These assays can be utilised within a month of HIV infection [20] [26]. The sensitivity of these assay tests ranges from 99.8 to 100 percent and the specificity ranges from 99.4 to 100 percent [27] [28].

Point-of-care tests (POCTs) are rapid testing devices that diagnose HIV from finger-prick or mouth swap samples within 15 minutes [20] [26] [29]. A diagnostic assessment conducted between 2014 and 2016 found sensitivities for HIV tests of four rapid dual HIV/Syphilis rapid diagnostic tests (RDTs), ranging from 99.5 to 100 percent (CI: 95-100 percent) and specificity ranging from 93.5 to 99.5 percent (CI: 89.1-100 percent) [29]. This was confirmed by a

systematic review that evaluated the performance of RDTs for HIV/Syphilis, which found that the range for sensitivity from selected studies was 96 to 100 percent and specificity was 92 to 100 percent [30]. This shows that rapid HIV tests have as high sensitivity and specificity as the laboratory tests thereby giving significantly acceptable results.

Evidence has shown that the use of rapid HIV tests is feasible and acceptable by patients and the general practitioners. A pilot study conducted in London established that it was feasible to offer Rapid HIV test in primary care and patients likely to accept the offer were Black African and Black Caribbean [31]. A systematic review on acceptability of and preference for rapid point-of-care HIV testing in youth also demonstrated that youth accept the offer for a test in various settings including primary care [32]. A study in Spain investigated attitudes of general practitioners to rapid HIV testing and it revealed that even GPs accept and were willing to use rapid HIV testing, but this should be accompanied by training on the use of rapid tests [33].

1.4 Effects of early HIV diagnosis

Early diagnosis of HIV, when CD4 count is >350/mm^3, reduces onward transmission of the disease, enables individuals to get better health outcomes, optimises quality of life and reduce healthcare costs [34] [35]. On the other hand, in the UK, late diagnosis of HIV (CD4 count <350/mm^3) contributes to HIV/AIDS-related morbidity and mortality, reduces response to antiretroviral therapies, reduces quality of life and increases healthcare costs and undiagnosed infection limits the efforts to reduce transmission [20] [36]. A study in the USA found that 91.5 percent of HIV transmissions in 2009 were attributed to people that were HIV infected but undiagnosed or those diagnosed with HIV but not retained in HIV medical care [37].

Early diagnosis of HIV infection, when plasma viral load is at its highest (during the acute and late stages of HIV infection), allows for implementation of preventative measures aimed at avoiding or reducing further transmission, including use of post-exposure prophylaxis and taking behavioural changes to reduce risk of transmission [16] [25] [38]. Infected individuals unknowingly spread the disease if they are not aware of their HIV status. Research suggests that over 50 percent of new HIV infections are linked to people unaware that they were infected and about a third had not ever taken a HIV test [39] [40]. The research indicated that people with acute HIV infection who are unaware of their HIV status led to considerable HIV transmission [40].

The introduction of ARTs in HIV care and treatment has led to increased life expectancy and improved prognosis of HIV infected individuals, as well as 80 percent reduction in HIV related mortality in industrialised countries [14]. Evidence reveals that ARTs are more effective in reducing disease progression and mortality from HIV associated conditions if they are introduced before severe immunological impairment, which happens when CD4 count<200/mm^3 [41]. A cohort study conducted in outpatient HIV clinics throughout the United Kingdom showed that patients diagnosed late, when their CD4 count had fallen below 350 cells/mm^3, had at least 10 years less life expectancy at age 20 compared to those patients starting ARTs when their CD4 count was above 350 cells/mm^3 [42]. Late diagnosis of HIV is associated with a ten-fold increase in risk of mortality within the first year of diagnosis compared to early diagnosis [43] [44]. Furthermore, patients that started ARTs after early diagnosis benefit from immunologic recovery compared to late presenters [45] [46]. Early initiation of ARTs is associated with restricted viral reservoir establishment and composition [46]. A study conducted in Africa, US and Asia found that early treatment with ARTs led to reduction in onward transmission of HIV to sexual partners [47] [48].

Higher treatment and care costs of HIV infected individuals, in some cases 1.5 to 4 times higher, has been linked to late diagnosis [49] [50]. In addition to high costs of ARTs, patients diagnosed late have a high risk of inpatient episodes and are susceptible to HIV-associated comorbidities [49] [51] [52]. On the contrary, early diagnosis (and early initiation of treatment) reduce costs by minimising hospital attendances and admissions, reducing adverse effects of drugs, preventing HIV-related comorbidities and improving patient care pathways and models [49]. This was confirmed by a UK study which used routinely collected data from HIV units and hospitals to compare costs and cost-effectiveness of routine pre-combination antiretroviral therapy (pre-cART) and first-line combination antiretroviral therapy (cART) between people living with HIV with CD4≤200 cells/mm^3 and those with CD4>200 cells/mm^3 [53]. The study concluded that the annual estimated cost for starting standard first line ARTs was 18 percent lower for patients diagnosed earlier compared to those severely immunocompromised (CD4 count < 200 cells/mm^3) [53].

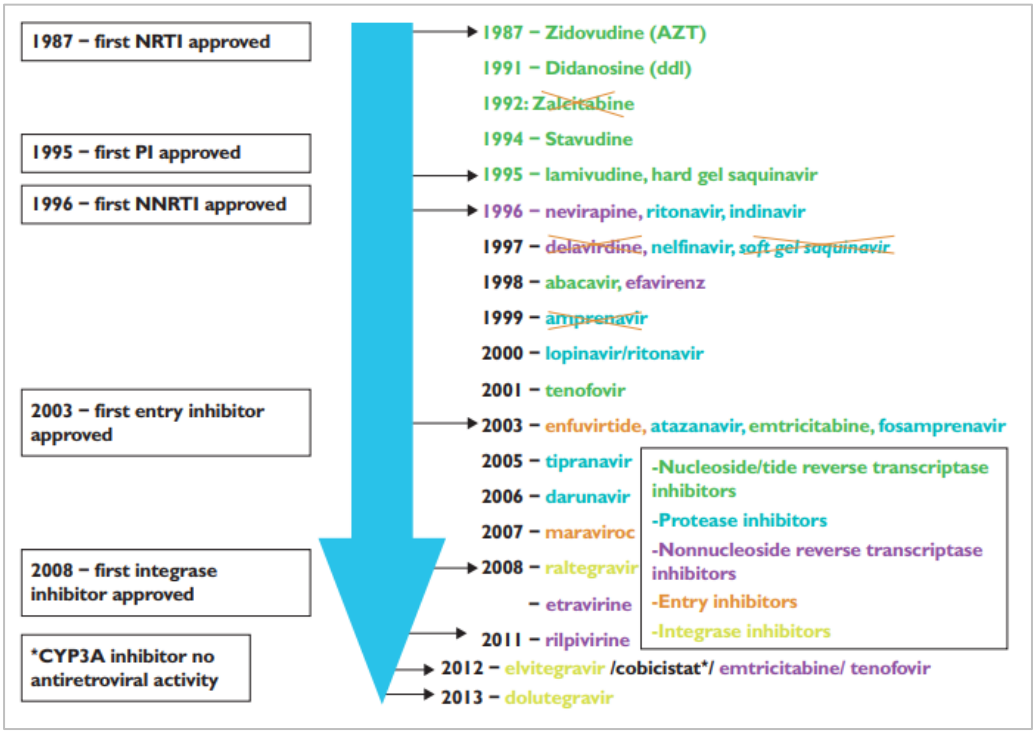
1.5 Treatments for HIV

Treatments of HIV have evolved since the first treatment was approved in 1987. The breakthrough came when researchers realised that Zidovudine, a failed cancer drug discovered in the 1960s, stopped multiplication of the virus thereby helping patients to live longer [54] [55]. Zidovudine and the drugs used prior to 1995 were nucleoside reverse transcriptase inhibitor (NRTI). The challenges of early introduction of these drugs were “high pill burdens, inconvenient dosing, treatment-limiting toxicities and incomplete virological suppression”,

Figure 1.5-1 [54] [55]. The drugs stopped working due to viral resistance and mutations, leading to the development of protease inhibitors (PI) and non-nucleoside reverse transcriptase inhibitor (NNRTI) in 1995, [54] [55]. These drugs led to introduction of Highly Active Antiretroviral Therapy (HAART) regimens (made up of two NRTIs combined with a PI or NNRTI), which were able to suppress the virus [55]. The use of two NRTIs combined with

a third agent (such as PI or NNRTI) is the strategy currently in use and is now referred to as combination antiretroviral therapy (cART) [54] [55]. Additionally, a prevention treatment strategy, pre-exposure prophylaxis (PrEP), was introduced in 2012 and it was administered to healthy individuals to avoid HIV transmission [54] [56].

Figure 1.5-1: Timeline of HIV antiretroviral development of >25 drugs across five different classes, 1987-2013



Adopted from Tseng et al, 2014 [55]
NB: Approval dates refers to US Food and Drug Administration (FDA) approval dates.
Drugs illustrated with an 'X' are those no longer available/used by 2013.

1.6 Global public health significance

The global burden of HIV/AIDS as a public health issue cannot be understated. The World Health Organisation (WHO) estimated that over 79.3 million people have been infected by HIV by 2020 and 36.3 million people had died of HIV/AIDS related diseases, approximately 680,000 in 2020 [57]. In 2020, it was estimated 37.7 million people were living with HIV worldwide and 1.5 million were newly diagnosed [57]. The number of HIV related deaths increased from 290,000 in 1990 to a peak of 1.9 million in 2003, remained constantly high until 2006 and then declined to the estimated 680,000 HIV-related deaths in 2020 [58].

Continental variations in the number of people living with HIV infections are noticeable, with Africa accounting for two-thirds of the global HIV infected population, followed by South-East Asia (9.8 percent) and the Americas (9.8 percent) [59]. In 2020, the prevalence of HIV infected people per 1000 adult population was estimated at 3.6 percent (CI: 2.9–4.2) in Africa, 0.5 percent (CI: [0.3–0.6) in Americas and 0.2 percent (CI: 0.2–0.3) in South-East Asia (SE Asia) [59]. Europe was ranked 4 [h], accounting for 6.9 percent of global HIV/AIDS population with a prevalence rate of 0.4 percent (CI: 0.4–0.5) [59]. Similarly, there were continental variations in incidences of HIV/AIDS in 2020, with Africa recording the highest number of new diagnosis of 880,000 (CI: 590,000–1,300,000), followed by 170,000 (CI: 140,000–200 000) in Europe, 150,000 (CI: 110,000–210,000) in the Americas and 120,000 (CI: 78,000–150,000) in Western Pacific Region [59]. The global total of new cases of HIV infections increased from approximately 1.9 million infected people in 1990 to approximately 3.4 million people in 1995 and declined to the estimated 1.5 million people in 2020 [58].

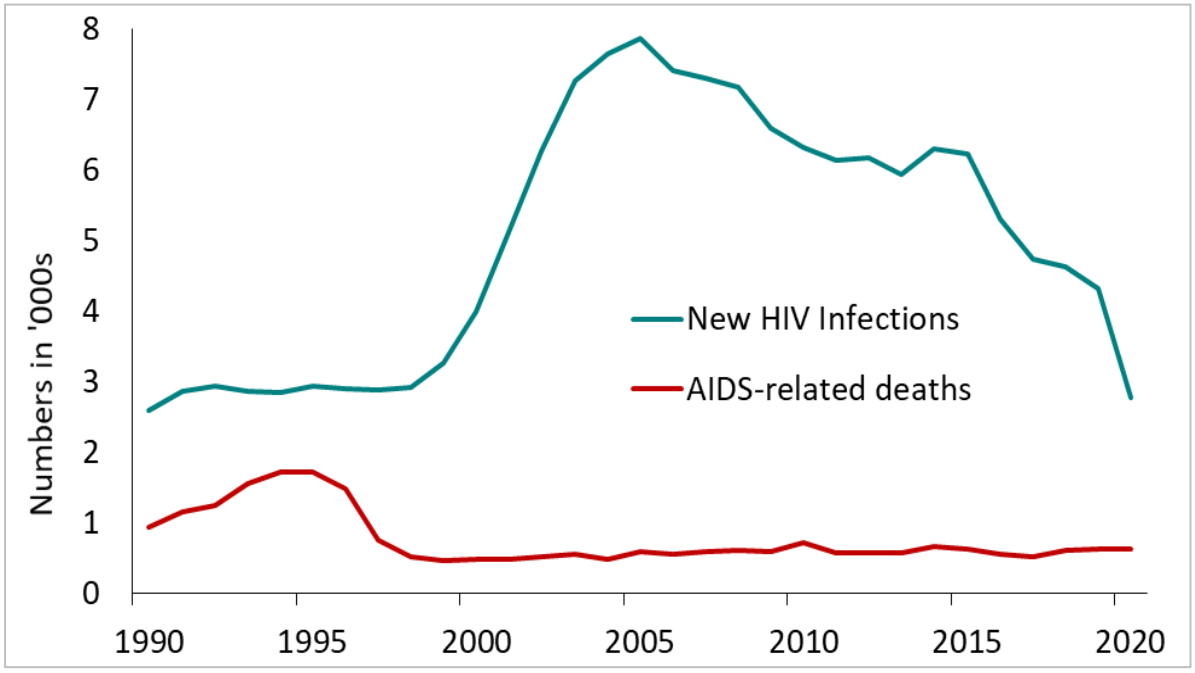
1.7 UK public health significance

The first case of AIDS in UK was a patient referred “with opportunistic infections indicative of immunosuppression” to a London Hospital in December 1981 [60]. Surveillance systems were put in place in 1982, collating data on AIDS related clinical reports and notifications of deaths (56). Data collected through surveillance evolved with improvements in treatments, patient care and laboratory techniques [60]. United Kingdom Health Security Agency (UKHSA), formerly Public Health England (PHE), in collaboration with Department of Health and Social Care (DHSC) and a Clinical Reference Group for HIV, collates and analyses the UK wide data from outpatient HIV services and laboratories via the surveillance systems such as HIV and AIDS Reporting System (HARS) and the National CD4 Surveillance scheme, HIV and AIDS New Diagnoses and Deaths Database (HANDD), Integrated Screening Outcomes Surveillance Service (ISOSS) part of the NHS and Infectious Diseases in Pregnancy Screening Programme (IDPS) [61] [62]. Data published from UKHSA are the official statistics

on HIV. The data on HIV reported in 2020 was complete when disaggregated by gender and age, but data was only 82% complete for ethnicity and 81% complete for country of birth [62].

In 2020, the results from the surveillance data revealed that approximately 106,890 individuals in UK were living with HIV, of whom 2,766 were newly diagnosed and there were 634 deaths [62] [63]. Annual deaths in UK increased from 930 in 1990, to a peak of 1,720 deaths in 1994, followed by a sharp decrease from approximately 1,480 deaths in 1996 to 750 in 1997, a gradual decline to 525 in 2017 and then a slight increase to 634 in 2020, Figure 1.7-1 [58] [62]. The figures for 2020 depict the impact of Covid-19 on access to HIV services with at least 99 deaths attributed to COVID-19 between March and June 2020 [64].

Figure 1.7-1: Number of AIDS related deaths and newly diagnosed HIV infections, 1990 to 2020



Data Source: United Kingdom Health Security Agency, 2021 and Public Health England, 2018.

In UKSHA data, the exposure categories accounting for the highest number of people seen for HIV care in the UK in 2020 were: MSM (46 percent), heterosexual contacts (46 percent), and Black Africans (29 percent) [62] [65]. Similarly, the number of new diagnoses in UK varies

by exposure category as shown by 2020 figures which showed that over a third of the new diagnoses (995) were reported among gay and bisexual and other MSM, 496 among heterosexual men, 571 among heterosexual women and 59 among people who inject drugs [62]. The number of newly diagnosed HIV individuals increased from 1,950 people in 1988 to a peak of 7,870 people in 2005 followed by a gradual decline to 4,328 in 2019 and then a sharp decrease to 2,766 in 2020, Figure 1.7-1 [62]. The sharp decrease in 2019 to 2020 could be a result of the impact of COVID-19 on HIV testing [64] [66].

From 2008 to 2020, approximately 38 percent of newly diagnosed HIV infections in the UK were aged 35-49 years [62]. UKHSA estimated that 23 percent of people living with HIV were 50 years and older in 2008 compared to 39 percent in 2017 [63] [67]. In 2020, males accounted for 71 percent of people newly diagnosed with HIV in the UK and about 62 percent of people living with HIV and aware of their HIV status) [62].

1.8 Effects of Covid-19 pandemic on HIV

As mentioned in the previous section, there is evidence from the UK that HIV diagnoses decreased during Covid-19 pandemic [64]. This may be due to changes in sexual behaviour (lower incidence) but also reduced testing (less diagnosis) in sexual health service and primary care [64]. “Testing rates dropped in 2020 with the steepest decreases in areas of extremely high HIV prevalence in GP and secondary care but increased in emergency departments.” [64] [66] [68]. Analysis of HIV testing in Brighton and Hove found that the Covid-19 pandemic resulted in a decrease in HIV testing by 64% in sexual health services (SHS) and a decrease in testing and diagnoses in primary care, during the “lockdown” from April 2020 [68]. This analysis highlighted that half of HIV testing conducted in 2020 were performed during the first quarter of the calendar year [68]. An observational study conducted in USA (Louisiana, Minnesota, Rhode Island and Washington) found a substantial decrease in trend of HIV testing

and suggested that new HIV infections “may be undiagnosed and not yet linked to clinical care and services” [69].

In addition to a decline in HIV testing, Covid-19 pandemic had an impact on sexual behaviour and access to HIV services (63). A cross-sectional, web-based survey among MSM in UK found that Covid-19 restrictions considerably affected sexual behaviour (25 percent having been sexually active with partners outside their household in a risky behaviour) and mental well-being (such as anxiety and loneliness) of the respondents. Furthermore, a calibrated HIV transmission model during Covid-19 pandemic conducted in Maryland, USA found a 6-month reduction of 25 percent in sexual partners for MSM and the disruptions to preventative measures such as condom use, PrEP initiations/use and ART initiations, which were predicted to result in increased HIV infections [70].

Although there were negative impacts of Covid-19 pandemic, alternative approaches to testing were piloted or intensified during the pandemic. In England, the introduction of “social distance” during Covid-19 pandemic, led to introduction of online or virtual provision of services including online request for tests and self-sampling [71]. A pilot conducted in Washington, during the Covid-19 pandemic, found that mail-out HIV and sexually transmitted infections testing was accepted by high priority populations in urban and diverse areas [72].

Overall, the pandemic may have contributed to late diagnosis and reduced HIV testing in primary care. This, therefore, supports the development of strategies to increase HIV testing in primary care.

1.9 Strategy for elimination of AIDS

The Joint United Nations Programme on HIV/AIDS (UNAIDS), in a strategy to eliminate AIDS by 2030, set a 90-90-90 target for 2020 [73] [74]. This target aims to have 90 percent of HIV infected individuals knowing their status, 90 percent of those diagnosed on ARTs and 90 percent of those on ART having an undetectable viral load. The strategy focuses on drastically reducing new HIV infections, thereby protecting future generations from acquiring the virus [74].

UK met the target of 90:90:90 in 2017. By December 2020, 95 percent of HIV infected individuals knew their status, 99 percent of those diagnosed were on ARTs and 97 percent of those on therapies had an undetectable viral load [64]. Although United Kingdom met the target, it was estimated that 4 to 7 percent of HIV infected individual were still unaware of their HIV status in the UK in 2020 and 906 of the 2,766 new diagnoses (33 percent) were diagnosed late [67]. Hence the need to find alternative ways of increasing uptake of HIV diagnostic testing which could speed up the reduction in HIV transmission [63].

The DHSC published an action plan in 2021, focussing on ending HIV transmission by 2025 [75]. The action plan recommends expansion of settings that offer HIV testing and the use of an opt-out approach to offering diagnostic tests [75]. Opt-out testing is when patients are informed that everyone is routinely tested for HIV, unless they decline [76] [77] [78]. This HIV testing approach is employed in antenatal clinics of UK and has a 99% testing coverage [76]. Opt-in is when HIV testing is offered but patients are not tested unless they explicitly consent to testing [78]. A qualitative study to understand perception towards opt-out testing in primary care discovered that this approach to HIV testing was more acceptable to patients when offered in GP practice than in hospital [77]. An opt-out approach is preferred since it makes no judgement on patients, and it could also help in tackling stigma attached to HIV [79].

However, an HIV test offer might be taken as a judgement of a patient's "sexuality, ethnicity, or behaviour, unless it is clearly explained and understood that the test is offered to all patients" [77]. The ambition of eliminating HIV transmission requires "additional and intensified prevention measures", especially among communities were decline in undiagnosed HIV is slower, even though they have very low prevalence, such as people aged 45-59 years or ethnicity other than Black Africans living outside London [80].

1.10 UK policy on HIV testing

In UK, prior to 2001, HIV testing was mostly restricted to individuals that attended Sexual Health or Genitourinary Medicine (GUM) Clinics and requested a test [20]. The uptake of HIV testing via this route was very low. After 2001, it was recommended that GUM clinics attendees should all be offered HIV testing unless they "opt-out", resulting in an increase in uptake of HIV testing [20]. A universal "opt-out" approach was also introduced in antenatal clinics in 2000, reducing the percentage of women with undiagnosed HIV infection prior to delivery from 18 percent in 2000 to fewer than 10 percent in 2006 [20] [81]. In 2008, another route of HIV testing was introduced, offering tests in a wider range of health settings such as hospitals, termination of pregnancy services, drug treatment centres and primary care (new registrants), especially in areas of high prevalence (greater than two per 1000 population) [20] [82].

A systematic review looking at studies reporting the number of tests offered by providers discovered that adherence to British HIV Association (BHIVA) 2008 guidelines was poor with 27 percent of eligible patients receiving HIV testing [82]. The study concluded that low testing was mostly due to low offers for HIV testing from providers in settings other than the established testing offered in GUM clinics and antenatal care [82]. The systematic review found that low testing coverage was mainly due to low proportion of providers offering HIV testing (40.4 percent) compared to 71.5 percent of patients' acceptance of the tests [82].

Another study on HIV diagnosis in England, Wales and Northern Ireland showed an increase in proportion of patients diagnosed in non-traditional settings (settings other than GUM clinics and antenatal clinics) from 27 percent in 2005 to 32 percent in 2014 [83].

The 2021 action plan in England recommended opt-out HIV testing in high or very high prevalence settings, such as “primary care, emergency departments, prisons and termination of pregnancy services” [75]. It further recommended that additional testing via self-sampling ordered online should be explored, as it has increased since 2018 (65 percent increase from 2018 to 2019, an increase of approximately 10,000 tests in England) [75]. The action plan was in line with 2020 BHIVA guidelines, which recommended that HIV testing should be normalised in healthcare settings, including opt-out and self-testing [84].

1.11 Reasons for late HIV diagnosis in UK

The reasons for delayed HIV testing can be categorised into individual, social, structural and environmental barriers. Individual factors are personal traits and behaviours that hinder or facilitate HIV testing, and these are demographic, migratory, knowledge, lifestyle choices and attitudes [85]. Studies have identified that the personal barriers to HIV diagnosis include: fear of HIV-related stigma and discrimination, misconception that testing positive would lead to deportation (in migrant populations), perceived risk of HIV infection, despondency in relation to lack of treatment, fear of relationships breaking up after testing positive, unawareness on how to access testing centres and fear of changes in life as a result of a positive diagnosis [85] [86] [87] [88] [89] [90] [91] [92]. Results from a Swiss HIV Cohort Study discovered that a high proportion of late presenters of HIV were from Sub-Saharan Africa [93]. The study stated that the reasons for late presentation were: i) lack of knowledge about possibilities of anonymous HIV testing; ii) fear-related issues, including fear of deportation; and iii) having difficulties in accessing testing services [93].

Social barriers and facilitators refer to influences of culture and community, including family, friends and society at large [85]. Social factors include persistence of HIV-related stigma and discrimination, and cultural norms in some communities [88] [89] [90] [94]. Structural and environmental factors refer to barriers and facilitators at a higher level and outside the influence of an individual, community and culture, but affect individuals [85]. These factors are part of the wider social determinants of health that are governmental, systemic, policy decisions and institutional.

Institutional factors are those that affect access to HIV testing services, including clinicians' views and opinions, comfort in discussing HIV testing or giving positive result and in some cases, inability to accept results [95] [96] [97] [98]. Some of the practitioner-based barriers highlighted in literature include lengthy pre-test counselling, anxiety and stigma [99]. A systematic review conducted in USA suggested that provider-related stigma was associated with providers with no or limited training in the past 12 months [100]. This was supported by an online survey conducted in Southern United States, which concluded that offering of HIV testing by primary care providers in high risk areas is affected by providers': 1) opinion of risk factors and non-disclosure of risk behaviours by patients; 2) views on patient obstacles to accessing care; and 3) misconceptions on HIV risk and stigma [101]. However, some of the factors/barriers can be addressed by training, as shown by a training intervention study in a high prevalence area of London where GPs and nurses were trained "on sexual health clinical skills and sexual history", resulting in over 50 percent increase in HIV testing [95]. A systematic review on barriers and facilitators of HIV testing in primary care recommended that providers should have continuous sexual health education including updates on guidelines [102]. A systematic review conducted in 2019 found that facilitators to routine HIV testing included integration of HIV testing into clinical workflows, training of testing champions, removal of formal aspects such as consent and pre/post counselling, and community-level facilitators centred on reducing stigma [103]. Governmental factors included: distrust in government,

policies (such as universal testing) and lack of political will [96] [104]. These barriers to testing should be considered when more testing is recommended in all settings such as primary care.

1.12 Guidelines relevant to primary care.

There has been a notable increase in diagnosis outside the specialised sexual health services (SHS) in UK, in response to national and international recommendations and guidelines produced in the last decade [63]. The guidance documents were produced by WHO, BHIVA, The National Institute for Health and Care Excellence (NICE) and British Association for Sexual Health and HIV (BASHH) and were aimed at increasing uptake of HIV testing for various target groups.

In 2008, BHIVA produced guidance aimed at increasing uptake of HIV testing in primary care [20]. The guidelines in 2020 also recommended increasing uptake of testing in primary care and suggested opt-out testing as the most effective approach [84]. Other guidelines produced by NICE, WHO and BASHH promote increased uptake of HIV diagnostic testing in “at risk” population groups such as people from countries of high prevalence, men having sex with men and sex workers, Table 1.12-1.

Table 1.12-1: Guidance targeting at risk groups

Target group/setting	Guidance source (year)
Adults at risk	BHIVA/BASHH/BIA (2020)
People who may have undiagnosed HIV	NICE (2016)
Men having sex with men (MSM)	NICE (2016)
Black Africans	NICE (2016)
Sex workers	WHO (2016) and BASHH (2013)
Those from countries of high prevalence.	NICE (2016) and BASHH/BHIVA 2008
Routine testing in general practice in high prevalence areas	NICE (2016) and BHIVA 2008
HIV testing in all healthcare settings	BHIVA 2008

Source: NICE (2016), WHO (2016), BASHH/BHIVA (2008 and 2020) and BASHH (2013).

.

1.13 Role of primary care in diagnosing and managing HIV.

There are efforts to increase HIV screening by expanding settings from traditional clinics to other settings such as general practices and emergency rooms [105]. General practices are the first point of contact with health services for nearly all the UK population registered with a GP. They provide opportunities for disease prevention, health promotion and early detection of disease [95] [106]. Evidence shows that a high proportion of patients that are diagnosed with HIV in UK would have seen a General Practitioner (GP) within the previous year [26] [35] [106] [107]. A service evaluation conducted on patients aged 15 years and over, registered with four GPs in Tower Hamlets, London found that 65 percent of patients diagnosed with HIV by their GP had presented at least one HIV related symptom during the three years preceding diagnosis [108]. Additionally, a retrospective case-notes review in City and Hackney, London revealed that 33.3 percent of patients that presented to their GP with at least one HIV indicator were subsequently diagnosed with HIV by their GP [44].

However, a cluster-randomised controlled trial (RCT) implemented in Hackney, London showed that in intervention practices, where HIV testing was offered to newly registered patients, there was a non-significant increase in rate of HIV diagnosis at 0.3 (CI: 0.11-0.85) per 10,000 compared to 0.07 (CI: 0.02-0.20) per 10,000 for the control group (usual care) [109]. This study concluded that using opt-out rapid testing in primary care increase HIV diagnosis in newly registered patients [109]. An interrupted time-series study conducted in 42 general practices in City and Hackney found that promotion of nurse-led HIV screening into routine practice might be linked to an increase in testing and earlier diagnosis of HIV [110]. A cross sectional survey, with convenience sampling, conducted in Spain discovered that offering HIV testing to patients contacting primary care with indicator conditions and lifestyle risk factors was possible and effective [111]. However, it should be noted that the study discovered that testing was cost-effective in population with a prevalence greater or equal to 0.1% [111].

A study in London concluded that promoting screening of HIV in primary care is cost effective in the medium term (13 to 18 years) and it should be promoted, especially in high prevalence areas [112]. On the other hand, another study in the UK concluded that annual testing of high-risk groups in low prevalence areas was cost effective, since the study found that one more HIV test of all adults could identify a large proportion of undiagnosed people living with HIV [113].

Primary care has a role to play in increasing uptake of HIV diagnostic testing, since nearly all the UK population is registered with a GP [114]. HIV testing in general practices can be done by either sending blood samples for laboratory testing or conducting Combined HIV antibody and P24 antigen tests followed by laboratory confirmation [26]. However, among those who visit their GP, a challenge is the fact that many signs and symptoms of HIV/AIDS such as rashes, weight loss and respiratory infections are not specific to HIV/AIDS.

HIV testing in primary care could be improved by having point-of-care/pop-up alerts which identify patients that are most likely to be infected by HIV. An interrupted time series study conducted in Ohio suggested that the use of an alert system in primary care settings could be useful in increasing diagnosis in individuals unlikely to get tested for HIV via the education and awareness campaign alone [115]. The study combined an educational intervention with an Electronic Medical Records (EMR) alert system in health centres and clinics resulting in an increase in HIV diagnosis from 22 previously undiagnosed patients in 2008-09 to 33 patients in 2010-11 [115]. However, the study stated that although there was an increase in testing, many patients remained untested [115]. There is a possibility of participant bias in this study since the patients and providers were those that were willing to participate. Another study in United States established a twofold increase in HIV testing for previously undiagnosed patients aged 18-65 years after the use of EMR reminders [116]. An educational intervention study in Bristol revealed a mean difference in HIV diagnosis rate of 1.9 in the intervention

group but concluded that training should be combined with other behavioural change techniques [117]. Other studies confirmed an increase in the rate of HIV diagnosis from interventions that use clinical alerts in combination with quality improvement, social marketing and incorporating feedback from clinicians [118] [119].

1.14 Electronic primary care records

Electronic primary care records refer to digital documentation used to monitor and manage patients' health care in general practice [120]. The electronic recording, storage and sharing of health information brings the benefit of speedy patient care, reduction of errors required in diagnosis and treatment of patients, increase patients' control of their healthcare and improve the quality and audit of healthcare [121] [122]. The records allow the aggregation of data on performance costs and other standards associated with general practice [123] [124].

In the UK, all general practices have made use of electronic primary care records since the 1990s [125]. Detailed information included in electronic primary care records are name, address, marital status, date of birth, occupation, telephone number, sex, symptoms, allergies, immunisation history, referral information diagnoses, prescriptions and test results [126] [127].

A number of UK general practices contribute anonymised data to well established research databases, these include IQVIA Medical Research Data (IMRD), formerly The Health Improvement Network (THIN), Clinical Practice Research Datalink (CPRD), QResearch and ResearchOne. These databases contain records of patients currently registered with the practices as well as those who died or left the practices. The main differences between the databases are the size (number of general practices) and the GP software system used by the practices (Table 1.14-1).

Table 1.14-1: Common Primary Care databases in United Kingdom

Database	Details
IMRD (FORMERLY THIN)	• Use Vision GP software systems • Data collected from 587 general practices nationwide • Contains data for 18 million patients (3.6m active patients) [128]
CPRD	• Use Vision or EMIS GP software systems • Data collected from 2000 general practices nationwide • Contains data for 16 million active patients [129].
QResearch	• Use EMIS GP software systems • Data collected from 1500 general practices nationwide • Contains data for 35 million patients [130].
ResearchOne	• Use TPP SystmOne software system • Contains data for 6 million patients [131]

The databases are similar because:

1. The data is collected in a way which reflects real-life situations.
2. It is representative of the UK population in terms of demographic, health and socio-economic composition.
3. Ability to select control and subjects from the same population.

The retrospective cohort study (covered in chapter 4) used the IMRD database to develop a prediction model. Apart from easy accessibility of the IMRD database to the University of Birmingham, the database was used because socio-economic status is available at postcode level of the individual's place of residence [132]. Hence, data on socio-economic status is more accurate at lower geographic levels which was not available in the other databases.

1.15 Thesis overview

This thesis aims to develop a strategy of targeted testing of a large number of people, to identify HIV infected individuals. This testing strategy could be considered to be similar to a screening programme, which identifies healthy people with an increased chance of getting a condition or health problem and invites them for diagnostic testing [133].The principles of screening were outlined in 1968 by Wilson and Jungner [134]. Although the Wilson and Jungner principles of screening were designed for a screening programme, and not a testing programme, some of the issues raised provide a useful checklist to consider in targeted testing

strategy, Table 1.15-1. The original checklist had 10 principles but there are some modified versions of the checklist with 12 principles, which were used in this thesis [135].

Table 1.15-1: Principles of screening and application to HIV testing

		Principle	
Disease/condition principles	1)	Epidemiology of HIV should be well understood and the disease is an important health problem.	✓
	2)	Natural history of HIV adequately understood with a recognisable early symptomatic stage.	✓
	3)	Target population for testing should be clearly defined.	✓
Test/intervention principles	4)	Test performance characteristics should be appropriate and specific	✓
	5)	Interpretation of test results should be clearly interpretable and determinate.	✓
	6)	Post-testing options on course of action should be agreed on.	✓
Program/system principles	7)	Facilities and infrastructure for diagnosis and treatment should be available.	✓
	8)	Testing programme should be acceptable to the population and ethical.	✓
	9)	The cost of case-finding should be economically balanced in relation to possible expenditure on medical care as a whole.	✓
	10)	The testing programme should be coordinated and integrated.	
	11)	Testing programme benefits should outweigh harms.	
	12)	Test programme should be quality and performance managed.	

Of the 12 principles, the information provided in this chapter shows that 9 of the principles are already met in relation to HIV testing, Table 1.15-1. The epidemiology of HIV is well understood and HIV is an important national and global health issue, discussed in sections 1.6 and 1.7. The natural history of HIV is well understood with a clearly defined early symptomatic phase, section 1.1. Guidance and policies in place define a target population for testing, as shown in sections 1.10 and 1.12. This thesis aims to improve identification of a target population in primary care settings.

Test performance of diagnostic tests for HIV are appropriate. Point-of-care tests are accepted by the population and have a high sensitivity (99.5-100 percent) and specificity (93.5-99.5 percent), and these are clearly interpretable, section 1.3. There is an agreed course of action after testing positive for HIV, in terms of treatment and follow-up care as discussed in sections

1.4 and 1.5. The treatment for HIV is more effective if it is introduced early and it is also more effective in reducing onward transmission, section 1.4 and 1.5.

There are facilities and infrastructure for diagnosis and treatment available in all settings, but more testing needs to be implemented in some settings such as primary care and emergency departments, section 1.10 and 1.12. This could be made possible since the testing programme for HIV is coordinated and integrated, especially in established settings such as the GUM clinics. The implementation of targeted testing could be integrated into primary care. Moreover, as discussed in section 1.10, the testing programme for HIV is acceptable to the population and it is ethical.

The two principles on benefits/harm of testing and performance management of the testing programme have not yet been met. Furthermore, the cost of case-finding is shown to be economically balanced in relation to possible expenditure on medical care, as a whole, covered in section 1.13. This thesis investigates the possibility of targeted HIV testing in primary care, which could be used to show whether the benefits and costs (indirectly) of finding a HIV patient outweighs the health and social care expenditure if the patient was diagnosed late. It will also attempt to find how HIV testing could be incorporated into primary care using a pop-up alert.

1.15.1 Thesis aims

Given the importance of early diagnosis of HIV and the role that primary care could play in increasing uptake of HIV testing, the aim of this thesis was to derive a prediction model for HIV using a retrospective cohort design. This would provide a predicted probability that an individual might be HIV positive in the United Kingdom (UK) and identify those individuals in whom testing would be worthwhile. It also aims to assess feasibility of using the model in providing a point-of-care alert in general practices. Although UK exceeded the 90:90:90 target

in 2017, it was selected as the setting for this thesis because of the UK's plan to end HIV transmission by 2025. This requires finding alternative ways of increasing uptake of HIV diagnostic testing in order to target those unaware of their HIV status and those diagnosed late.

1.15.2 Thesis objectives and research questions

The objectives of the thesis are:

1 To derive a prediction model for HIV using a retrospective cohort design which provide a predicted probability that an individual might be HIV positive in the United Kingdom.

2 To assess feasibility of using the model in providing a point-of-care alert in general practices.

The objective will be reached through the answers to the following research questions.

1 What are the personal (including demographic and socio-economic) characteristics, symptoms, comorbidities and other factors (such as geographical) available in UK primary care records which might be predictive of a diagnosis of HIV?

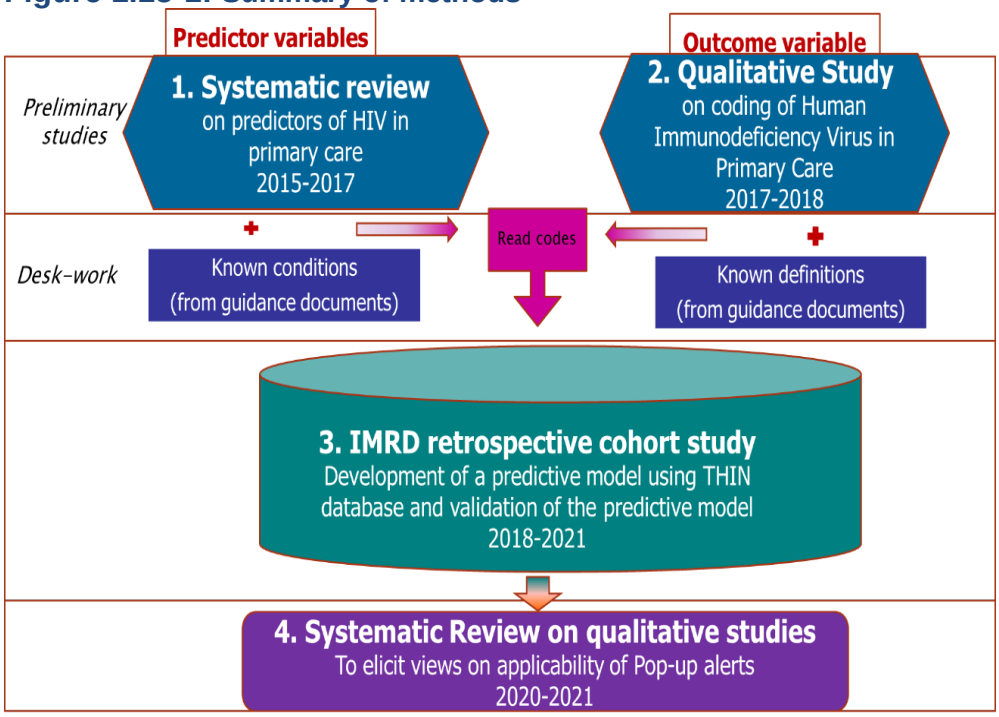
2 Which of these predictors are predictive of HIV in UK electronic primary care records, through derivation and validation of a prediction model for the diagnosis of HIV?

3 Which facilitators and barriers should be considered to ensure the HIV point-of-care alert from the model will be useful in primary care?

1.15.3 Overview of methods

The methods used to answer the questions are detailed in chapters 2-5 and below are the summaries of the methods, Figure 1.15-1.

Figure 1.15-1: Summary of methods



1.15.3.1 Identify predictor variables

A systematic review of risk factors for HIV was conducted to find the predictors of HIV from quantitative studies (chapter 2). This is mainly because the HIV guidelines only list clinical conditions that should be considered as risk factors for HIV, while other studies have also found there are possible socio-economic and lifestyle risk factors. Furthermore, some clinical conditions are not included in the guidelines, while evidence shows that they are predictors of HIV.

Exploration on whether the risk factors identified in the systematic review are found in primary care records was carried out. The reason for the exploration is because not all of the identified risk factors are available in primary care records

1.15.3.2 Recording of outcome variable

Due to stigma attached to HIV, there is a possibility that GPs code HIV using euphemistic terms. This and the fact that HIV is managed in secondary care, means there may be under-reporting or miscoding of HIV in primary care records. A qualitative research was conducted to elicit views from GPs on how HIV is coded in primary care (chapter 3).

Exploration for completeness of HIV in primary care was conducted by finding if there was a difference in incidence rates of HIV in IMRD, compared to incidence rates in published data from UKHSA (whether there was acceptable HIV reporting in IMRD). Furthermore, checking for completeness of "probable HIV" in primary care was performed to show if euphemistic terms were used instead on the recommended HIV codes.

1.15.3.3 Derivation and validation of prediction model in primary care records

Exploration for completeness of candidate predictors in primary care was conducted using descriptive statistics. This identified the predictor variables to be included in model development and internal validation. A prediction model was developed and internally validated, and it will be directly applicable to GPs in primary care (chapter 4).

1.15.3.4 Views on barriers and facilitators of use of pop-up alerts

Systematic review of qualitative research was conducted to elicit views of primary care clinicians on the use of pop-up/point-of-care alerts (chapter 5). This is mainly because there are barriers and facilitators to implementation of point-of-care alerts, which should be considered during the development of HIV pop-up alerts.

1.16 Summary of chapter

Diagnosing HIV earlier has advantages to the patient (better prognosis), to the health service (reduced costs) and to the community (lower transmission). There are acceptable point-of-care tests for HIV which can be used in primary care. HIV testing is underused in primary care while many patients consult primary care practitioners before diagnosis of HIV and many have symptoms of HIV.

Because of the availability of electronic primary care records, it may be possible to identify patients at high risk of HIV from demographic and clinical information (risk factors) recorded in their electronic primary care records. Therefore, in the next chapter, there is need to know which information may be helpful in identifying patients at high risk of HIV and whether this is available in primary care records.

Chapter 2: Predictors of Human Immunodeficiency Virus (HIV) infection in primary care among adults living in developed countries: a systematic review

2.1 Introduction to chapter

This chapter identifies and summarises evidence on possible risk factors for HIV infection in adults (predictor variables) through a systematic review. The last subsection of this chapter identifies which of these risk factors may be available in electronic primary care records and describe how code lists for these predictor variables were developed. The identified characteristics could be used in a prediction model for early detection of HIV in primary care, covered in chapter 4. The systematic review was published in BioMed Central Systematic Reviews Journal; Rumbwere Dube, B.N., Marshall, T.P., Ryan, R.P. et al. Predictors of human immunodeficiency virus (HIV) infection in primary care among adults living in developed countries: a systematic review. Syst Rev 7, 82 (2018). https://doi.org/10.1186/s13643-018-0744-3

https://systematicreviewsjournal.biomedcentral.com/articles/10.1186/s13643-018-0744-3

2.2 Systematic review of HIV predictors

2.2.1 Background

Human Immunodeficiency Virus is a retroviral infection that weakens the immune system and is a subsequent causative agent of Acquired Immuno-Deficiency Syndrome [6] [7]. As mentioned in chapter 1, the virus is transmitted through the exchange of a variety of bodily fluids [7] [12]. HIV/AIDS is one of the highest contributors to morbidity and the sixth leading cause of mortality, worldwide [7] [38]. However, the life expectancy of HIV positive patients

has increased over the years, due to early diagnosis, improved clinical management and Antiretroviral Therapies (ARTs).

The World Health Organisation developed a strategy aimed at reducing new HIV infections, AIDS related mortality and discrimination to zero with one of the HIV strategies being optimisation of “HIV prevention, diagnosis, treatment and care outcomes” [136].

Evidence shows that about 33 percent of patients that are diagnosed with HIV in UK would have seen a General Practitioner (GP) within the previous year [26] [106] [114]. Therefore, primary care has a role to play in increasing uptake of HIV testing since nearly all UK population is registered with a GP [114]. As mentioned in chapter 1, HIV testing in general practices can be done by conducting rapid testing followed by laboratory confirmation [26]. However, among those who visit their GP, a challenge is the fact that HIV/AIDS has many signs and symptoms such as rashes, weight loss and respiratory infections and these are not specific to HIV/AIDS.

Current UK guidelines recommend HIV testing to individuals from high-risk groups, those with symptoms indicative of HIV or where HIV forms part of the diagnosis [20]. However, 74.2 percent of patients consult their GPs in the period prior diagnosis do not present these indicator symptoms and diagnoses [35].

UK primary care clinicians need to identify patients in whom HIV is likely and therefore should be offered HIV testing. A systematic review was therefore necessary to identify demographic, lifestyle, clinical and laboratory characteristics of patients which might be associated with HIV infection in primary care. The identified characteristics were investigated to determine if they are documented in electronic primary care records and whether they can be used to predict which primary care patients are likely to have HIV infection. The results of the analysis can

identify patients in whom HIV is likely and therefore help inform primary care clinicians which patients they should offer HIV testing in the UK.

This systematic review identified, critically evaluated and interpreted available evidence related to the demographic, lifestyle, clinical and laboratory characteristics associated with HIV/AIDS infection in adults in the developed world [137] [138].

2.2.2 Methods

This systematic review conformed to the requirements of the Preferred Reporting Items for Systematic Reviews and Meta-analyses (PRISMA) [139]. The methods were detailed in a published protocol, but a summary was included in this section [140]. The PROSPERO registration number for the protocol is CRD42016042427.

2.2.3 Review question

This systematic review systematically identified and summarised evidence on characteristics of HIV infected adults which could be used in a prediction model for early detection of HIV in primary care. Individual factors in studies with multivariable models of predictors of HIV infection were also identified.

The review question was:

What demographic, lifestyle, clinical and laboratory characteristics are associated with HIV infection in adults aged 18 years and over?

2.2.4 Population, Exposure and Outcome

Studies selected included human participants aged 18 years and over or if both adults and children are included results were reported by age groups. Exposures may be demographic, socio-economic or clinical risk factors or characteristics associated with HIV infection. The comparison group was either people without risk factors or no comparison group. The outcome was laboratory confirmed HIV/AIDS infection. Studies describing frequency, duration and severity of predictive factors were also included. Thus, the search covered clinical indicators, socioeconomic and demographic, lifestyle related risk factors and comorbidities and frequency, duration and severity of predictive factors [20].

2.2.5 Study design

The types of studies most suitable for this aetiological review were mainly observational (analytical) studies that compared groups and produced predictive values or likelihood ratios (case-control and cohort, both retrospective and prospective studies) [141].

2.2.5.1 Search Strategy

Studies were identified via electronic searches of EMBASE (Ovid), MEDLINE (Ovid) and The Cochrane Library (Wiley). Searches for unpublished grey literature were undertaken using Open Grey (SIGLE), Google Scholar and BASE to reduce publication bias [142] [143]. Additional searches were conducted on abstracts or conference proceedings using Web of Science Conference Proceedings Citation Index (CPCI), guidelines (NICE, DH) and examination of reference lists from studies included in the review (reference searching) [143]. The search terms used in this review are outlined below. References were searched and stored using the Refworks referencing programme hosted by the University of Birmingham.

Searches included evidence published in any language. Although research on signs and symptoms on HIV/AIDS started as early as 1984, changes in pattern of infection and patient characteristics had occurred ever since, and more evidence was produced afterwards. Hence the studies covered in this review were those conducted and published from year 1995 onwards.

A. Search terms used

The search terms used in Ovid MEDLINE are shown in Table 2.2-1.

Table 2.2-1: Search terms used in Ovid MEDLINE

1	Human Immunodeficiency Virus.mp.
2	HIV/
3	Acquired Immuno Deficiency Syndrome.mp.
4	AIDS.mp.
5	exp Acquired Immunodeficiency Syndrome/
6	*HIV/
7	aids.mp. or *Acquired Immunodeficiency Syndrome/
8	*Acquired Immunodeficiency Syndrome/
9	sign.mp.
10	signs.mp.
11	symptom Flare Up/ or symptom.mp.
12	symptoms.mp.
13	risk factor$.mp. or Risk Factors/
14	clinical indicat$.mp.
15	clinical feature$.mp.
16	predict$.mp.
17	risk score$.mp.
18	model.mp.
19	1 or 2 or 3 or 4 or 5 or 6 or 7 or 8
20	9 or 10 or 11 or 12 or 13 or 14 or 15 or 16 or 17 or 18
21	primary care.mp. or exp Primary Health Care/
22	Rural Health/ or Nurse Practitioners/ or Family Practice/ or family practic$.mp. or Diagnosis/ or Physicians/
23	Physician-Patient Relations/ or Family Practice/ or Physician's Role/ or general practic$.mp.
24	21 or 22 or 23
25	19 and 20 and 24
26	limit 25 to humans

2.2.5.2 Inclusion/exclusion criteria

To ensure generalisability to a UK setting, only studies undertaken in the following developed countries were included in this review; Europe (European Union and European Free Trade Association nations) and North America (United States of America (USA) and Canada), Australia and New Zealand. The main reason for including these developed countries was the

difficulties in distinguishing the signs and symptoms of HIV/AIDS in developing countries due to i) higher prevalence of the condition, ii) competing causes of ill-health and death, iii) lack of comprehensive primary care means that presentation may be different and iv) different patterns of co-morbidities such as tuberculosis (TB) [7] [12].

Studies which included children only were excluded and non-HIV diagnosis (where HIV diagnosis was not the outcome) were also excluded from this review.

2.2.5.3 Selection procedure

Two reviewers independently pre-screened the articles returned after searching for titles/abstracts to find out if the articles addressed the review question and fulfilled the inclusion and exclusion criteria (Appendix 1-A). Differences between the reviewers were resolved through discussions.

Full text articles were then retrieved either electronically or directly from authors or hard copies from The British Library. Articles were translated using Google Translate, where necessary. Two reviewers independently assessed the suitability of study articles in relation to revised second selection criteria and exclusions were documented (Appendix 1-B).

2.2.5.4 Quality assessment

Quality assessment was done using a checklist for cohort and case-control studies modified from the Scottish Intercollegiate Guidelines Network (SIGN).

2.2.5.5 Data extraction

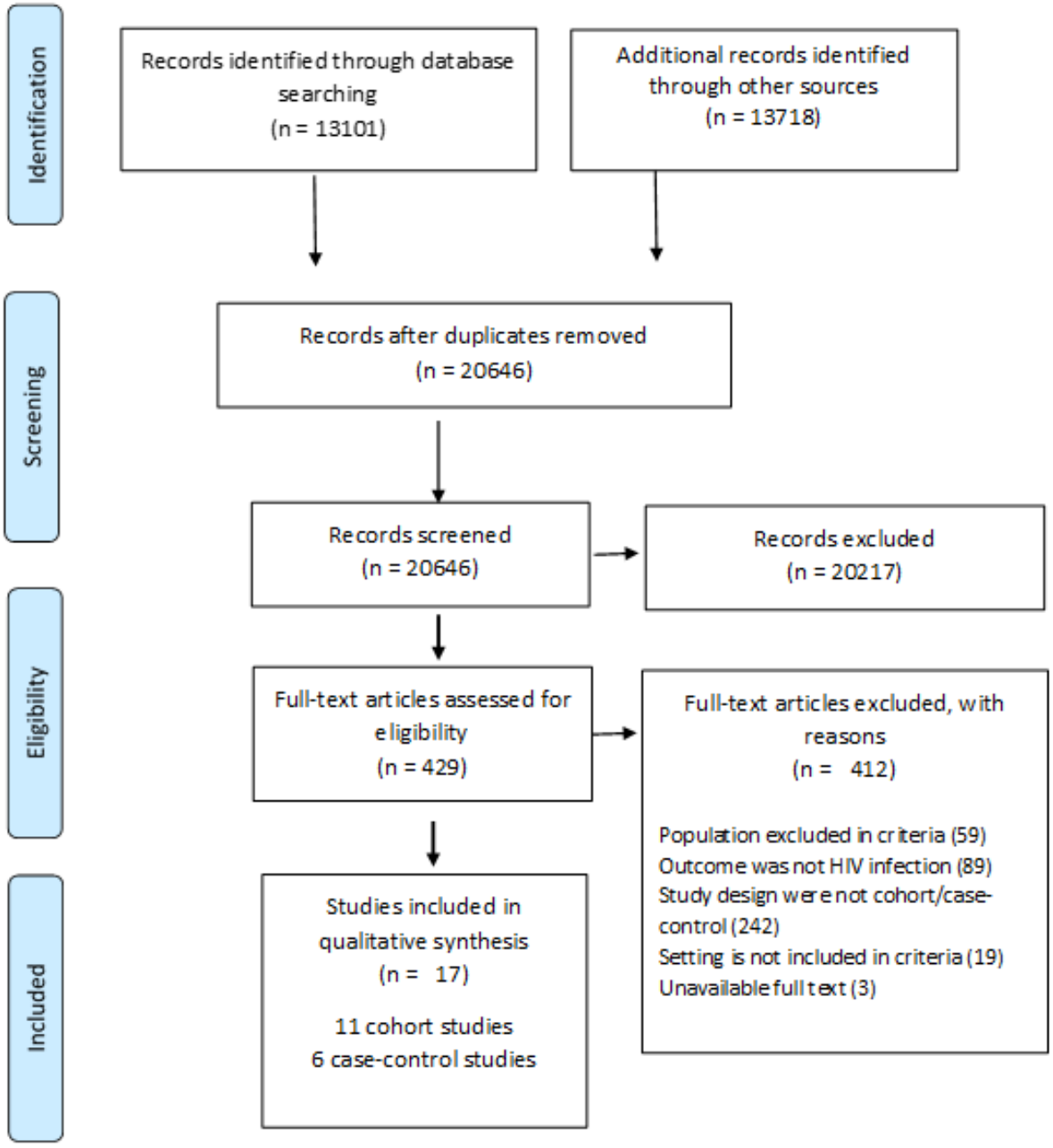
A data extraction form was developed to collate data from selected articles. Synthesis of evidence from systematic reviews can either be narrative or statistically analysed (pooling method or meta-analysis), depending on the homogeneity of the study [144] [145]. Tabulation and narrative of the results were produced in this review since the systematic review was aimed at revealing characteristics that can be used in a prognostic model (in chapter 4) and therefore, required no further analysis. Meta-analysis was not conducted in this systematic review because there was no need to obtain pooled statistics [146]. Tabulation with the description of the articles including the author, publication year, the study design, number of participants, population under study (country, gender and age) and outcome (including relative risks and models produced) was produced. Clarification was requested from the article authors where the information/data was unclear in the study report.

2.2.6 Results

2.2.6.1 Selection procedure

A total of 26,590 hits were returned from the databases (Medline, Cochrane and EMBASE) and the grey literature (SIGLE, BASE, Web of Science and Google scholar); 6,172 duplicates were removed and 20,637 articles were pre-screened, including UK guidelines, summarised in Figure 2.2-1.

Figure 2.2-1: PRISMA 2009 Flow Diagram



The first reviewer selected 438 articles using titles/abstracts, which addressed the research question (including the population, intervention, exposure and setting) and were published in the period under consideration. A second reviewer independently selected articles from a random sample of 100 articles to find out if there were agreements between reviewers on article selection using titles/abstracts. A discussion was held to agree on the articles identified from the sample which confirmed the articles suitable for the second selection phase.

The reviewers independently selected suitable articles using full text. Of the 438 articles, three were non-English and google translate managed to show that the countries of the studies (not stated in the abstract and title) did not meet the inclusion criteria. Only one article was requested from the author. A second discussion was held, and the reviewers agreed on 17 articles: 11 cohort and six case-control studies. The weighted Kappa score for the two independent reviewers on the second article selection (using full text) was 0.42, showing a moderate agreement.

2.2.6.2 Quality assessment

Quality assessment forms were exported into Excel and a column in the workbook identified whether a study was a cohort or case-control study. All the six case control studies were of acceptable standard, and half of them were of high quality, in terms of participant recruitment, number of cases and controls and on how they dealt with bias. However, it was difficult to identify the main confounders accounted for in most of the case controls studies. Form used for quality assessment is in Appendix 2 and the summary of the quality assessment is shown in Table 2.2-2.

All the 11 cohort studies were of acceptable standard, but only two were of high quality, in terms of participant recruitment, number of cases and controls and on how they dealt with bias. About 70 percent of the articles did not clearly identify and account for the main confounders. Details on the quality assessment form are in Appendix 2 and the summary of the quality assessment is shown in Table 2.2-2.

2.2.6.1 Data extraction

Data was extracted from 17 studies using a data extraction form, summarised in Table 2.2-3. Of the six case control studies selected, one study was conducted in UK, one in Netherlands, two in USA and two came from the same study in Canada. The total number of cases were

1,412 and the total number of controls were 3,423. The age range for most of the case-control studies was 18 years and over. The study duration ranged from 1 to 12 years with an average follow-up of 6 months.

The number of cohort studies selected were 11 in total, of which three were from UK, one from Ireland, one from Australia and six from USA. The number of participants ranged from 32 to over 20000 and most of the studies focused on patients aged 18 years and over. The study duration ranged from 1 to 5 years but some of the studies did not state follow-up intervals.

Table 2.2-2: Quality assessment summary: Cohort and Case-control Studies

Study	Design	Study addresses an appropriate and clearly focused question.	Participants being studied are selected from same source populations	Indication of how many people participated (in each group for case-control)	Main potential confounders identified and accounted for.	How well was the study done to minimise the risk of bias or confounding?
1. Joore I.K. et al, (2015)	Case-Control study	Yes	Yes	Yes	Cannot say	+
2. Damery S. et al (2013)	Case-Control study	Yes	Yes	Yes	Yes	++
3. Szerlip M.A. et al (2005)	Case-Control study (retrospective)	Yes	Yes	Yes	Cannot say	+
4. Ellerbrock T.V. (2004)	Case-Control study	Cannot say	Yes	Yes	Cannot say	+
5. Burchell, A.N. (2010)	Case-Control study	Yes	Yes	Yes	Yes	++
6. Burchell, A. N. (2003)	Case-Control study	Yes	Yes	Yes	Yes	++
7. Hodder, S.L. (2013)	Cohort study (prospective)	Yes	Yes	Yes	Yes	+
8. Moran, J. (2012)	Cohort study	Yes	Cannot say	Yes	No	+
9. Desai M. (2012)	Cohort study	Yes	Cannot say	Yes	No	+
10. Guy R.J. (2011)	Cohort study	Yes	Yes	Yes	Cannot say	+
11. Krauskopf K. (2011)	Cohort study	Yes	Yes	Yes	Yes	++
12. Niyonsenga T (2013)	Cohort study	Yes	Cannot say	Yes	Cannot say	+
13. Ross, J. D. (1997)	Cohort study	Yes		Yes	Cannot say	+
14. Gordon S. M. (1995)	Cohort study	Yes	No	Yes	No	+
15. Marder K. (1995)	Cohort study (prospective)	Yes	Yes	Yes	Yes	++
16. Hafner J. W. (1997)	Cohort study (retrospective)	Yes		Yes	Cannot say	+
17. Landau R. (1997)	Cohort study (retrospective)	Yes	Yes	Yes	No	+

Modified from Scottish Intercollegiate Guidelines Network (SIGN)

Minimise risk of bias or cofounding: High quality (++) □ Acceptable (+) □ Unacceptable – reject 0

Table 2.2-3: Data extracted from selected studies

Author	Study design	Setting	Population	Duration and follow-up	Outcome
Joore I.K. (2015)	Case-Control study	Amsterdam, Netherlands	102 cases and 299 controls	2002-2012	HIV infection
Damery S. et al (2013)	Case-Control study	UK	Patients aged ≥18 939 cases and 2576 control	January 1989 – September 2010	HIV/AIDS diagnosis
Burchell, A.N. (2010)	Case-Control study	Ontario, Canada	Gay and bisexual men 123 cases and 240 controls	1998-2006	HIV infection
Burchell, A.N. (2003)	Case-Control study	Ontario, Canada	Adults aged 18 years and over 80 cases (seroconverts) and 106 controls	June 1998 – December 2001	Diagnosed HIV infection
Ellerbrock T.V. (2004)	Case-Control study	Florida, USA	217 cases 395 controls	1998-2000	HIV diagnosis
Szerlip M.A. et al (2005)	Case-Control study (retrospective)	New Orleans, USA	Older patients aged 55 years and over (53 cases and 106 controls)	6 months interval up to 12 months	Diagnosis of HIV infection
Moran J. (2012)	Cohort study	Ireland	N=1,404	2008-2011	HIV infection
Desai M. (2012)	Cohort study (prospective)	UK (Health Protection Agency database)	N=328	September 2010 – December 2011	HIV infection
Guy R.J. (2011)	Cohort study (prospective)	Victoria, Australia	MSM N=7,857	April 2006 – June 2009	HIV positivity
Krauskopf K. (2011)	Cohort study (prospective)	Bronx, New York, USA	HIV infected and at risk men aged 49 years and older N=643	2001-2006 6-month follow-up	HIV infection and hyperlipidemia, hypertension, chronic liver disease and diabetes
Niyonsenga T (2013)	Cohort study	Florida, USA	All cases with HIV/AIDS diagnosis N= 20,528	1998-2002	AIDS/HIV incidence
Ross, J.D. (1997)	Cohort study	Lothian and Glasgow region of Scotland	Population aged 16 and over N = 8466	January 1989 - December 1993	HIV positive results
Gordon S.M. (1995)	Cohort study (retrospective chart review)	Atlanta, Georgia, USA	Patients aged ≥60 N= 32 HIV positive	January 1985 – July 1992	HIV positivity
Marder K. (1995)	Cohort study (prospective)	New York City, USA	Intravenous drug users 99 HIV +ve patients 124 HIV -ve patients	3.5 years	HIV infection
Hafner J.W. (1997)	Cohort study (retrospective)	Albuquerque, New Mexico, USA (Emergency department)	N = 344	19-month period July 1993 – January 1995	HIV diagnosis
Hodder S.L. (2013)	Cohort study (prospective)	USA	Women aged 18-44 with 1 or more personal or partner risk factors	6 months interval up to 12 months	HIV prevalence and incidence
Landau R. (1997)	Cohort study (retrospective)	London UK	A&E patients aware and unaware of HIV status N=133	1991-1994	HIV infection

2.2.6.2 Identified predictors of HIV infection

The predictors of HIV infection identified from the 17 retrieved articles were categorised into: demographic and socioeconomic, behavioural or lifestyle, clinical features and comorbidities. The 5 guidelines used to compare with evidence were three from UK (BHIVA and two NICE guidance), one from USA (Aberg et al (2013)) and one from Australia (Donovan and Ross (2000)) [20] [147] [148] [149] [150]. Statistically significant characteristics or those with highest percentages were included.

A. Demography and socio-economic

The guidelines recommend offering HIV testing to men having sex with men (MSM) and people of Black African ethnicity. Evidence from five studies revealed that significant demographic characteristics associated with HIV infection were i) homosexuals and/or bisexuals, mainly Men who have sex with Men (MSM), ii) black ethnicity and iii) age ranges, mainly between 27 and 40 years. Of the five studies, two revealed that MSM have 1.79 or 2.7 times the risk of HIV infection compared to heterosexuals, detailed in Table 2.2-4 [151] [152]. The other three studies could not provide odds ratio (OR) but revealed that the highest proportion of HIV infected individuals at the end of the study were homosexuals [153] [154] [155]. Only one study by Ellerbrock et al (2004) identified Black ethnicity to have 6.77 times risk associated with HIV infection compared to white ethnicity, detailed in Table 2.2-4 [151]. A study by Guy et al (2011) revealed a 68 percent increase in odds of HIV infection from aboriginal ethnicity though not significant [156].

Evidence showed three more demographic characteristics that predict HIV infection, which are not listed in the guidelines namely: age, gender and country of birth, Table 2.2-4. Three studies identified age groups as a predictor of HIV infection [156] [157] [152]. Age groups at higher risk of HIV infection are between 27 and 40 years with odds ratio of HIV infection as high as 11.54 for the 34 years and over compared to 18-26 years age group.

Table 2.2-4: Demographic characteristics identified in selected studies

Studies		Ellerbrock 2004	Guy 2011	Hodder 2013	Ross 1997	Gordon 1995	Hafner 1997	Moran 2012
		OR	OR	OR	OR	%	%	%
Gender	Male	1.37* (0.98-1.92)			1.00			
	Female	1.00			6.6*			
Age	Reference group		<30 y	18–26 y	21-25 y			
	26-30				1.7 (1.05-2.8)			
	27-33			5.83 (1.22-27.96)				
	30-39		1.91 (1.27- 2.87)					
	31-35				0.3*			
	34+			11.54 (2.71-49.05)				
	36-40				1.6*			
	40+		1.81 (1.19-2.75)					
Ethnicity	Black race (Reference = white)	6.77 (4.17-11)						
	Aboriginal or Torres Strait Islander		1.68* (0.41-6.94)					
Country of birth	Born in USA	1.76 (1.22-2.53)						
	Born in Australia		1.42* (1.00-2.02)					
Sexuality	Homosexual/ Bisexual	1.79* (0.67-4.79)			2.7 (1.5-4.8)	37%	57%	61%
	Heterosexual	1.00			1.00		3%	28%

*Not statistically significant
NB: % do not add up to 100% because they are provided for all risk factors

The two studies that identified gender as a predictor of HIV infection had conflicting and not significant results; one study found males were at 37 percent higher risk of being HIV infected while the other study found females were six times more likely to be HIV infected [151] [152]. The studies conducted in United States identified being born in that country as a significant predictor of HIV infection [151]. However, a study conducted in Australia did not show any significant odds of HIV infection in relation to country of birth [156].

The selected guidelines did not state socioeconomic indicators that should be considered when recommending HIV testing. Socio-economic conditions identified in the selected studies are related to deprivation, covering poverty levels, annual income, employment status, housing problems and education levels, detailed in Table 2.2-5. One study by Niyosenga et al (2013) showed that HIV positivity correlates with poverty in urban areas but not in rural areas [158]. Annual income under $10,000 was shown by Ellerbrock et al (2004) to increase the risk of HIV infection by 13 times. The study by Desai et al (2012) found that 26 percent of people that were HIV positive at the end of the study were unemployed and 17 percent had housing problems. Lower education attainment such as not being a high school graduate was a risk factor for HIV infection as shown by Ellerbrock et al (2004). Hodder et al (2013) discovered that having low education attainment more than double the risk of HIV infection and having education attainment beyond high school reduce (but not significant) the odds by 57 percent. The study by Ross et al (1997) results showed that the region of HIV testing was a predictor of HIV status.

Table 2.2-5: Socio-economic factors identified in selected studies

Studies	Ellerbrock 2004	Hodder 2013	Niyonsenga 2013	Desai 2012
	OR	OR	CC**	%
Housing problems				17%
Poverty index in rural areas			-0.25*	
Poverty index in urban areas			0.58	
Annual income <$10,000	13.2 (7.91-22)			
Farmworker	2.09 (1.47-2.96)			
Unemployed	5.08 (3.18-8.14)			26%
Education beyond high school		0.43* (0.15–1.24)		
Not a high school graduate	2.15 (1.48-3.1)			

*CC=Correlation coefficient

B. Behavioural or lifestyle predictors

Behavioural characteristics identified to increase the risk of HIV infection from literature can be categorised into personal lifestyle, partner lifestyle and effects of life events. The guidelines in BHIVA (2008), Donovan and Ross (2000) and Aberg et al (2013) recommended offering HIV testing to patients with a history of injecting drugs, current drugs use and having HIV partners. There were seven studies that identified injecting drug use as a predictor of HIV infection, Table 2.2-6. The study by Ellerbrock et al (2004) found injecting drug users were 21 times more likely to have HIV infection and studies by Guy et al (2011), Hodder et al (2013) and Ross (1997) found 3 times the risk. The other four studies did not provide the odds ratio (OR) but showed that injecting drug users made 10 to 30 percent of HIV infected patients, at the end of their respective studies. The study by Ellerbrock et al (2004) also showed that patients that smoke crack cocaine are 23 times at risk of HIV infection.

One study by Desai et al (2012) showed that 25 percent of HIV infected individuals were current smokers [159]. Binge drinking was identified by Hodder et al (2013) and Szerlip et al (2005) to increase the risk of HIV infection by 57 percent and by 12.8 times, respectively [157] [160]. The risk of substance misuse, a combination of drug and alcohol misuse was revealed by Hodder et al (2013) to increase the risk of HIV infection by 2.5 times. The study by Desai et al (2012) also established that 22 percent of the HIV infected patients at the end of the study were binge drinking and abusing drugs.

The guidelines in Aberg et al (2013) recommended offering HIV testing to patients that exchange money or drugs for sex. The study by Ellerbrock et al (2004) identified that exchanging money or drugs for sex increase the risk of HIV infection by 19 times. The study by Guy et al (2011) showed that having male anal sex increase the risk of HIV infection by 63 percent and the study by Desai et al (2012) showed that 10 percent of the HIV infected individuals were obese.

Table 2.2-6: Behavioural or lifestyle - Personal choices identified in selected studies

Predictor	Ellerbrock 2004	Gordon 1995	Guy 2011	Hafner 1997	Hodder 2013	Moran 2012	Ross 1997	Desai 2012	Szerlip 2005
	OR	%	OR	%	OR	%	OR	%	OR
Injected drugs users	21.1 (4.89-90.9)	18%	2.97 (1.77 -5.0)	30%	2.71 (1.33-5.53)	10%	2.3 (1.5-3.5)		
Ever smoked crack cocaine	22.8 (12.6-41.5)								
Binge-drinking or alcohol misuse					1.57* (0.74-3.33)				12.8 (1.65-99)
Substance use (combined)**					2.52 (1.22-5.21)			22%	
Current smokers								25%	
Unsafe sex			1.84 (1.6- 3.2)					60%	
HIV positive partner			3.24 (1.47-7.11)						
Sex with drug user	17.2 (7.18-40.9)								
Contact abroad							2*		
Ever exchanged money or drugs for sex	19.3 (11.2-33.2)								
Male anal sex in the last ≥ 6 months			1.63 (1.13-2.35)						
Multiple life partners	M: 5.51 (3.18-9.55) F: 19.8 (8.81-44.2)								
Obesity								10%	

* Not statistically significant.
** includes drug use or binge-drinking

Personal sexual behaviours included in the guidelines are unsafe sex, having sexual contact abroad and having multiple sex partners. Two studies identified unsafe sex as a risk factor for HIV infection; study by Guy et al (2011) showed that unsafe sex (no condom use) increased the risk of HIV infection by 84 percent and the study by Desai et al (2012) found that 60 percent of the HIV infected individuals had unsafe sex.

Guidelines on HIV testing in BHIVA (2008), Donovan and Ross (2000) and Aberg et al (2013) recommended offering HIV testing to patients that had sexual contact abroad or had sex in with a partner who came from countries with high HIV prevalence. Only one study identified that having a contact abroad doubly increase the risk of HIV infection. Three studies identified having multiple sex partners as a risk factor for HIV, Ellerbrock et al (2004) revealed that having 10 or more lifetime partners in men had 5.5 times increase in risk of HIV infection and almost 20 times the risk in women with 3 or more lifetime partners. However, a study by Ross et al (1997) showed that having multiple partners had a 74 percent reduction in risk (OR = 0.26).

Table 2.2-7: Behavioural or lifestyle- number of partners and partner lifestyle identified in selected studies

Predictor		Ellerbrock 2004	Guy 2011	Hodder 2013	Ross 1997
		OR	OR	OR	OR
Multiple partners					0.26
Lifetime partners	(Men—≥10)	5.51			
	(Women—≥3)	19.8			
Partner Characteristics	HIV positive		3.24	8.19	
	Alcohol dependence			1.42	
	Binge-drinking			1.82	
	Illicit drug use	17.2		1.57	

The guidelines in BHIVA (2008), Donovan and Ross (2000) and Aberg et al (2013) recommended offering HIV testing to patients whose partners have some risky behaviours namely: being HIV positive and men having sex with men. Three studies identified partner characteristics as risk factors for HIV infection. The studies by Guy et al (2011) and Hodder et al (2013) showed that having an HIV positive partner increased the risk of HIV infection by three times and eight times, respectively. Two studies identified partner's use of illicit drugs as

a risk factor for HIV infection: increasing the risk by 17 times in the study by Ellerbrock et al (2004) and a 57 percent increase in the study by Hodder et al (2013). The other partner characteristics revealed by Hodder et al (2013) was alcohol dependence (increasing the risk by 42 percent) and Binge drinking (increasing the risk by 82 percent).

No guidelines stated stressful events as a risk factor of HIV. However, stressful events in men having sex with men were identified by Burchell et al (2010) as risk factors of HIV infection, Table 2.2-8 [161]. The odds ratio for stressful events increased with number of events, starting from a 20 percent increase for 1 to 2 events to over double the risk for five events or more. The odds of HIV infection were 7.5 times higher in stressful events for younger ages under 30 years, and it was 2.7 for stressful events for people over 40 years. Any type of stressful event increased the risk of HIV infection by more than double with some stressful events such as bereavement and death of a close friend and financial crisis increasing the risk by three times, and relationship breakdown (romantic and other relations) doubling the risk.

Table 2.2-8: Behavioural or lifestyle – stressful events in men having sex with men identified in selected studies

Predictor		Burchell 2010
Stressful events		OR
Number of events	1–2	1.2
	3–4	1.8
	5+	2.5
Age	<30	7.5
	30–39	2.3
	40+	2.7
Event	Beaten up/attacked	2.5
	Bereavement event	3.1
	Crime event	2.4
	Death of a close friend	3.1
	Financial or security-related	2.5
	Health-related	2.8
	Major financial crisis	2.9
	Problems with drugs/alcohol	3.2
	Relationship-related	2.1
	Romance ends	2.2

C. Clinical features

This section will cover clinical features, including signs and symptoms associated with HIV infection. The guidelines listed the clinical signs and symptoms together with the other comorbid diseases. The features listed include AIDS-defining illness, weight loss and diarrhoea. Four studies identified HIV associated signs and symptoms which included flu-like symptoms (fever/chills, cough), rash, weight loss, abdominal pain, diarrhoea, minor trauma and nausea/vomiting, Table 2.2-9. Three studies identified flu-like symptoms/fever/chills as risk factors for HIV infection. The study by Joore et al (2015) showed that fever/chills were associated with 4.5 times increased risk of HIV infection, while the studies by Burchell et al (2003) and Hafner et al (1997) reported that 76 percent and 13 percent of the HIV positive patients at the end of the study had fever/chills, respectively [161]. Two studies identified weight loss as a predictor of HIV infection; the study by Damery et al (2013) revealed a 13 times likelihood and the study by Joore et al (2015) revealed a 39.6 times likelihood [35] [162].

One study by Hafner et al (1997) revealed that seven percent of the HIV positive patients had cough, six percent minor trauma and five percent abdominal pain. Only one study by Burchell et al (2003) revealed that rash had an odds ratio of 4.5 and the study by Joore et al (2015) revealed that having diarrhoea double the likelihood of being HIV infected.

Table 2.2-9: Clinical features identified in selected studies

Condition		Damery 2013	Joore 2015	Hafner 1997	Burchell 2003
		OR (CI)	OR (CI)	%	% & OR
Weight loss		13.4 (5.15-6.7)	39.6 (6.2-∞)		
Fever or chills			4.5* (0.5-54.3)	13%	
Cough				7%	
Flu like symptoms					76%
Diarrhoea			2* (0.2-17.4)		
Diarrhoea consultation	One only	3.7* (0.9-5.48)			
Diarrhoea consultation	Two	4.4 (2.3-2.81)			
Abdominal pain				5%	
Minor trauma				6%	
Nausea/vomiting				6%	
Rash					4.5
Number of HIV indicator conditions	One		11.7 (6-23.6)		
	Two		77.5 (18.2-700.8)		

D. Comorbidities

The clinical indicator conditions recommended for HIV testing in the BHIVA guidelines are categorised into the following: respiratory, dermatology, neurology, gastroenterology, gynaecology, haematology, ophthalmology, Ear, Nose and Throat (ENT) and other (including mononucleosis-like syndrome, pyrexia of unknown origin, any lymphadenopathy of unknown cause and any sexually transmitted infection). The clinical conditions listed in Donovan and Ross (2000) and Aberg et al (2013) are all included in the BHIVA guidelines.

The respiratory conditions listed in the guidelines include tuberculosis, pneumocystis, bacterial pneumonia and aspergillosis. The respiratory condition identified in the selected studies was pneumonia and pneumocystis, Table 2.2-10. The studies by Damery et al (2013) and Joore et al (2015) revealed that pneumonia is associated with 47.7 times and 8-fold increased risk of HIV infection, respectively. The study by Landau et al (1997) revealed that 52 percent of HIV infected individuals had pneumocystis carinii [163].

Table 2.2-10: Comorbidities – respiratory, dermatological and neurological conditions identified in selected studies

Condition		Damery 2013	Joore 2015	Marder 1995	Landau 1997
		OR (CI)	OR (CI)	OR (CI)	%
Respiratory	Pneumonia	47.7 (3.54-52)	8.3 (2-49.8)		
	Pneumocystis carinii				52%
Dermatology	Psoriasis		2.9* (0.1-∞)		
	Psoriasis — one consultation only	2.6 (1.69-1.5)			
	Psoriasis — two consultations	3.0 (1.38-2.5)			
	Herpes zoster	25.4 (5.76-14.2)	10.9 (2.0-108.9)		
Neurology	Peripheral neuropathy		15.9(2-∞)		
	Neurologic disability in women			2.4	
	Neurologic disability in men			1.9 (1.1-3.2)	

The dermatological conditions listed in the guidelines include Kaposi's sarcoma, severe or recalcitrant seborrhoeic dermatitis, severe or recalcitrant psoriasis and multidermatomal or recurrent herpes zoster. The two dermatological conditions identified by two studies were

psoriasis and herpes zoster, Table 2.2-10. The study by Joore et al (2015) revealed that psoriasis is associated with three times the risk of HIV infection. The study by Damery et al (2013) discovered that the number of psoriasis consultations show the risk of HIV infection; one consultation had odds ratio of 2.6 and two consultations had three times the risk. The risk associated with herpes zoster was identified by Damery et al (2013) and Joore et al (2015), 25.4 times and 10.9 times, respectively.

There were two studies that identified a likely association of HIV infection and neurological condition. The study by Joore et al (2015) found that peripheral neuropathy was 15.9 times likely associated with HIV infection [164]. The study by Marder et al (1995) established that HIV positive women were 2.4 times more likely to have neurologic disabilities (cranial nerve abnormalities and fine limb movement) and men were 90 percent more likely, compared to HIV negative individuals [165].

The neurological conditions listed in the guidelines include:

- cerebral toxoplasmosis, primary cerebral lymphoma,
- cryptococcal meningitis,
- progressive multifocal leukoencephalopathy,
- aseptic meningitis,
- cerebral abscess space occupying lesions of unknown cause
- Guillain-Barré syndrome
- Peripheral neuropathy
- Dementia
- Leukoencephalopathy

The gastroenterological conditions listed in the guidelines include persistent cryptosporidiosis, oral candidiasis, oral hairy leukoplakia, chronic diarrhoea of unknown cause, salmonella, shigella or campylobacter, Hepatitis B infection and Hepatitis C infection. There were 4 studies that identified some of the gastroenterological conditions namely: oral candidiasis and hepatitis or liver diseases, shown in Table 2.2-11. Two studies, by Damery et al (2013) and

Joore et al (2015) revealed that oral candidiasis is associated with an increased risk of HIV infection by 29.4 times and 7.1 times, respectively. Hepatitis B was associated with a risk of HIV infection of 11.5 times and 8.3 times as shown by Joore et al (2015) and Szerlip et al (2005). A study by Krauskopf et al (2011) revealed that 22 percent of HIV infected patients at the end of the study, had chronic liver disease.

Table 2.2-11: Comorbidities – gastroenterological, oncological and gynaecological conditions identified in selected studies

Condition		Damery 2013	Joore 2015	Szerlip 2005	Krauskopf 2011
		OR (CI)	OR (CI)	OR (CI)	%
Gastroenterology	Oral candidiasis	29.4 (4.57-21.8)	7.1 (0.6-∞)		
	Hepatitis B		11.5 (1.2-∞)	8.3 (2.65-26.2)	
	Chronic liver disease				22% (15%-29%)
Oncology	Non-Hodgkin's lymphoma	12.6 (2.13-15)			
	Lymphogranuloma venereum		7.1 (0.6-∞)		
Gynaecology	Cervical dysplasia		2.9* (0.4-232.4)		
	Condyloma acuminata		12.1 (1.2-600.9)		

The oncological conditions listed in the guidelines include non-Hodgkin's lymphoma, anal cancer or anal intraepithelial dysplasia, lung cancer, seminoma, head and neck cancer, Hodgkin's lymphoma and Castleman's disease. Two oncological conditions were identified in selected studies: Non-Hodgkin's lymphoma and lymphogranuloma venereum, shown in Table 2.2-11. A study by Damery et al (2013) revealed that non-Hodgkin's lymphoma was associated with 12.6 increased likelihood of HIV infection. Another study by Joore et al (2015) revealed that lymphogranuloma venereum was associated with a 7-fold likelihood of HIV infection.

The gynaecological conditions listed in the guidelines include cervical cancer, vaginal intraepithelial neoplasia and cervical intraepithelial neoplasia Grade 2 or above. Only one study identified two gynaecological conditions associated with increased risk of HIV diagnosis, shown in Table 2.2-11. The study by Joore et al (2015) established a likelihood of 2.9 times and 12 times more from cervical dysplasia and condyloma acuminata, respectively.

The haematological conditions listed in the guidelines include any unexplained blood dyscrasia such as thrombocytopenia, neutropenia and lymphopenia. The two haematological conditions identified were leucocytopenia and blood dyscrasia, shown in Table 2.2-12. Leucocytopenia was revealed by Joore et al (2015) to have an associated increase in risk of HIV infection by 11.5 times. Blood dyscrasia was revealed by Damery et al (2013) to increase the risk of HIV infection by 5.7 times.

Table 2.2-12: Comorbidities – Haematological, ENT and Other conditions identified in selected studies

Condition		Damery 2013	Hodder 2013	Joore 2015	Krauskopf 2011
		OR (CI)	OR	OR (CI)	% (CI)
Haematology	Leucocytopenia			11.5 (1.2-∞)	
	Blood dyscrasia	5.7 (2.44-4)			
ENT	Lymphadenopathy	11.3 (5.15-5.3)		29.8 (4.4-∞)	
	Parotitis	8.6 (1.68-11)			
Other	Mononucleosis-like illness			6.2 (1.6-29)	
	Pyrexia of unknown origin	7.2 (4.05-3.5)			
	Depressive symptoms		1.45		
	Trichomoniasis			2.9	
	Hyperlipidemia				25% (17%-32%)
	Hypertension				10% (4%-16%)
	Diabetes				10% (5%-14%)

The ophthalmological conditions listed in the guidelines include cytomegalovirus retinitis, infective retinal disease including herpes viruses and toxoplasma, and unexplained retinopathy. No study identified any of the ophthalmological conditions.

The ENT conditions listed in the guidelines include lymphadenopathy of unknown cause, chronic parotitis, lymphoepithelial parotid cysts. Two studies identified ENT conditions associated with a risk of HIV infection: lymphadenopathy and parotitis, shown in Table 2.2-12. Lymphadenopathy was revealed by Damery et al (2013) and Joore et al (2015) to have an associated increase of HIV infection risk of 11.3 and 29.8 times, respectively. The study by Damery et al (2013) established that parotitis increased the risk of HIV infection by 8.6 times.

The other conditions listed in the guidelines include mononucleosis-like syndrome (mainly during primary HIV infection), pyrexia of unknown origin, any lymphadenopathy of unknown cause and any sexually transmitted infection. There were seven other conditions identified in the selected studies, shown in Table 2.2-12. Only one study by Joore et al (2015) revealed that mononucleosis-like illnesses is associated with six times increase in the risk of HIV infection. The study by Damery et al (2013) showed that pyrexia of unknown origin increases the risk of HIV infection by seven times. Depressive symptoms are not listed in the guidelines but were revealed by Hodder et al (2011) to increase the risk of HIV infection by 45 percent. The study by Joore et al (2015) revealed a treble risk of HIV infection associated with trichomoniasis. The other conditions that were present in HIV infected individuals at the end of the study by Krauskopf et al (2011) were hyperlipidemia (25 percent), hypertension (10 percent) and diabetes (10 percent).

Sexually transmitted infections (STI) were identified by seven studies to have an association with HIV infection, shown in Table 2.2-13. Four studies identified the odds of HIV infection from ever having an STI; the study by Szerlip et al (2013) revealed a 7.3 times risk of HIV infection and the study by Ellerbrock et al (2004) and Damery et al (2013) revealed a 10-fold and 10.8 times risk, respectively. The study by Guy et al (2004) revealed the risk of STIs diagnosed within 2 years and within 14 days to increase the risk of HIV infection by 2.72 and 3.19 times, respectively. The risk of HIV infection was shown by Joore et al (2015) to increase 14.6 times by one STI to 37.9 times from two or more STIs. Three studies identified syphilis to increase the risk of HIV infection; the study by Joore et al (2015) revealed an increase of risk by 39 times, while the study by Ellerbrock et al (2004) showed 12.7 times. The study by Guy et al (2004) revealed the risk of syphilis diagnosed within 2 years and within 14 days to increase the risk of HIV infection by 3.86 and 4.9 times, respectively. Two studies identified chlamydia as a risk factor for HIV infection; 11.8 increase in risk revealed by Joore et al (2015) while the study by Guy et al (2004) revealed a diagnosis within 2 years and within 14 days to

increase the risk of HIV infection by 2.3 and 2.6 times, respectively. Two studies, by Joore et al (2015) and Ellerbrock et al (2004) revealed that gonorrhoea increase the risk of HIV infection by 15.9 times and 6.5 times, respectively. Only one study by Joore et al (2015) showed that genital herpes almost trebles the risk of HIV infection.

Table 2.2-13: Comorbidities – Sexually transmitted Infections identified in selected studies

Condition		Damery 2013	Joore 2015	Szerlip 2005	Guy 2011	Ellerbrock 2004
		OR (CI)	OR (CI)	OR (CI)	OR (CI)	OR (CI)
Sexually Transmitted Infection (STI)		10.8 (3.38-7.6)		10.1 (3.39-30.12)		10.1 (6.89-14.9)
STI diagnosis	≤ 2 yrs				2.72 (1.77-4.2)	
	≤14 days				3.19 (2.05-4.96)	
Number of STIs per patient	One		14.6 (5.5-45.6)			
	≥ 2		37.9 (5.6-∞)			
Syphilis			39.3 (5.7-1703.9)			12.7 (7.28-22.3)
Seropositive for syphilis						7.29 (4.15-12.8)
Infectious Syphilis diagnosis	≤ 2 yrs				3.86 (1.99-7.5)	
	≤14 days				4.9 (2.51-9.56)	
Chlamydia			11.8 (3-67.5)			
Chlamydia diagnosis	≤ 2 yrs				2.31 (1.4-3.81)	
	≤14 days				2.62 (1.56-4.39)	
Gonorrhoea			15.9 (2-∞)			6.51 (4.4-9.65)
Genital herpes			2.9 (0.1-∞)			

2.2.7 Discussion

This systematic review was conducted to identify predictors of HIV infection that are available in electronic patient records and could be incorporated in a prediction model to identify primary care patients with undiagnosed HIV. The results identified four main categories: demographic and socio-economic, behavioural, clinical features and comorbidities. Some of these candidate predictors are routinely recorded in electronic primary care records. Others require further investigation to assess if they can be reliably identified and included in a future clinical prediction model.

The demographic and socio-economic predictors identified in this systematic review and available in primary care records are age and gender. Deprivation quintile available in primary

care records can be used as a proxy for some of the socioeconomic predictors such as income and employment status. Studies that identified the demographic predictors such as age, gender, country of birth sometimes had conflicting predictive values, which could be attributed to differences in population characteristics and geography (for example, a study conducted in Australia had contradicting results on country of origin compared to a study conducted in USA). Population characteristics included factors such as sexuality of the population under study could have affected the results. Behavioural predictors that are available in electronic health records are drugs use (injected or smoked), binge-drinking or alcohol misuse, current smokers and obesity. Most of the predictors with an association with HIV were considered in model development. However, some of the studies had small sample sizes which had an influence on possible predictors that were included/excluded in the model.

Some of the clinical features and comorbid diseases identified in literature were listed in the BHIVA guidelines (Appendix 3). Clinical features identified in literature which are most probably available in electronic health records are fever or chills or flu-like symptoms, cough, weight loss, diarrhoea, abdominal pain, minor trauma, nausea/vomiting, rash. The comorbid diseases identified in literature are available in primary care records, summarised in Table 2.2-14.

Table 2.2-14: Summary of comorbidities identified in selected studies

Respiratory
Pneumocystis
Bacterial pneumonia

Dermatology
Kaposi's sarcoma
Severe or recalcitrant psoriasis
Multidermatomal or recurrent herpes zoster

Neurology
Peripheral neuropathy
Neurologic disabilities (cranial nerve and fine limb movement abnormalities)

Oncology
Non-Hodgkin's lymphoma
Lymphogranuloma venereum

Gynaecology
Cervical cancer
Condyloma acuminata

Haematology
Leucocytopenia
Blood dyscrasia

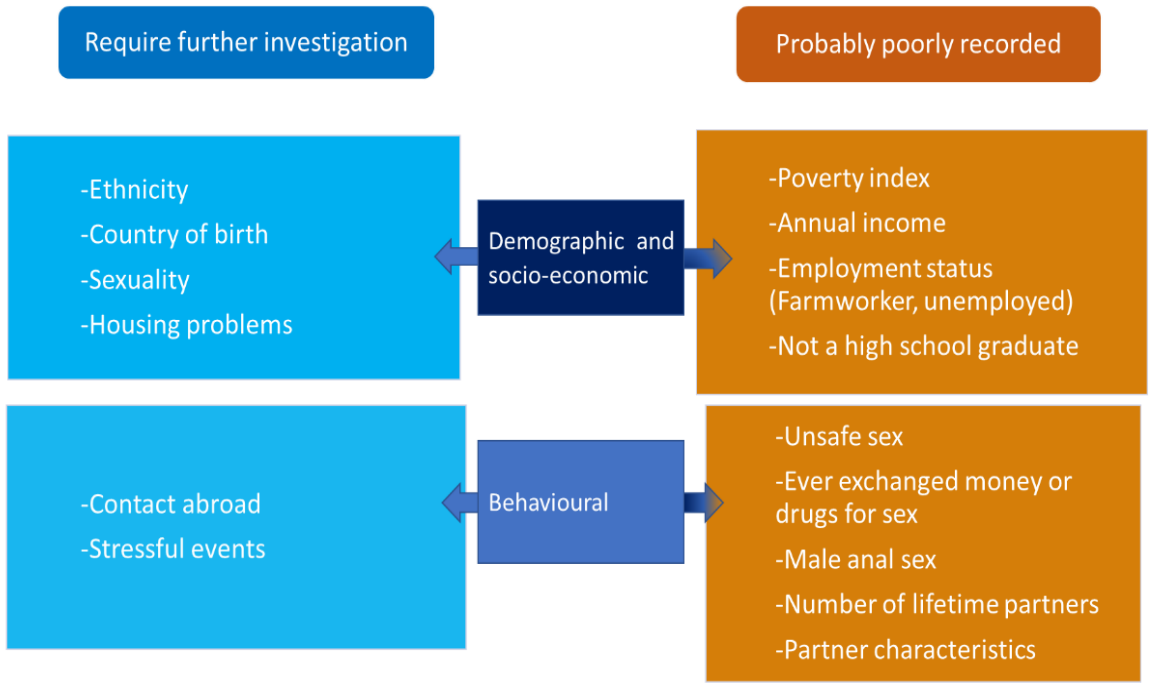
Gastroenterology
Oral candidiasis
Hepatitis B infection
Chronic liver disease

ENT
Lymphadenopathy
Parotitis

Other conditions
Mononucleosis-like syndrome (primary HIV infection)
Pyrexia of unknown origin
Depressive symptoms
Trichomoniasis
Hyperlipidemia
Hypertension
Diabetes
Any sexually transmitted infection

Some of the demographic, socioeconomic and behavioural predictors identified in literature might be available in primary care records and therefore require further investigation on completeness or missingness. Other predictors identified are likely to be missing from primary care electronic records at all and they are likely to be excluded from the model, Figure 2.2-2.

Figure 2.2-2: Predictors requiring further exploration or probably poorly recorded



2.2.8 Limitations

This systematic review focused on studies conducted in developed countries whereas most of the studies on HIV predictors were conducted in developing countries, mostly in Africa. Most of the studies conducted on HIV were case studies, qualitative studies and cross-sectional

studies which are not suitable in identifying risk factors. This might explain lack of evidence on some of the clinical indicator conditions included in the BHIVA guidelines.

Half of the selected studies did not clearly describe how the main potential confounders were identified and catered for, which affected their quality. However, none of the selected studies had an unacceptable quality.

About 40 percent of the cohort and case-control studies did not provide the odds ratio in their results and ended up reporting the percentage of HIV patients who ended up with certain characteristics. This made interpretation of risk association difficult. However, since this study was interested in identifying factors that have an association with HIV infection, the higher percentages were considered.

2.2.9 Conclusion

This systematic review revealed existing scientific evidence on predictors of HIV that can be used to inform decision making in prognostic model development [142]. The predictors identified in the literature falls into four categories: demographic and socioeconomic, lifestyle, clinical features and comorbid conditions. There were 10 distinctive demographic and socioeconomic characteristics identified in literature, of which two are available in electronic primary care records (age and gender), four are sometimes recorded (ethnicity, country of birth, sexuality and housing problems) and four are not available (poverty levels, annual income, employment status and education levels). Predictors that could be available in primary care electronic records require further exploration to determine whether they can be applied in a prediction model.

There were 11 distinctive lifestyle or behavioural characteristics, of which four are frequently included in electronic primary care records (drugs use, binge-drinking or alcohol misuse,

current smokers and obesity), two might be recorded (sexual contact abroad and stressful events) and five are unlikely to be recorded in electronic health records (ever exchanged money or drugs for sex, male anal sex in the last six months or less, personal and/or partner characteristics and unsafe sex).

Of the 51 clinical conditions in BHIVA guidelines, 18 of them were identified as significant predictors in this systematic review. BHIVA guidelines are developed using data from research studies with additional expert input from “writing group made up of BHIVA members – doctors, pharmacists, nurses and researchers” [166]. However, no information on the rationale for including the indicator conditions was provided for the 2008 BHIVA guidelines. Additional predictors clearly identified in literature but not included in the guidelines were fever or chills or flu-like symptoms, cough, abdominal pain, minor trauma, nausea/vomiting, rash, depressive symptoms, trichomoniasis, hyperlipidemia, hypertension and diabetes.

2.3 Developing code lists for the HIV predictors

2.3.1 Predictor variable operational definitions

Operational definitions of HIV predictors in primary care: demographic, socioeconomic, clinical, general practice characteristics were developed. Operational definitions were linked to Read codes which are a hierarchical “coded thesaurus of clinical terms”, named after their inventor, Dr James Read [167] [168]. General Practitioners in the United Kingdom (UK) have used Read codes (which can be broadly mapped to International Classification of Diseases (ICD) codes) since 1985 [169]. Read codes were standard vocabulary used by clinicians in recording patient diagnosis and procedures in primary and secondary health and social care [167]. Versions of Read codes were revised as new classifications emerge. The most recent version of Read Codes was Version 3, which was in use till they stopped to be operational in 2020 [167] [170]

The Read codes are hierarchical Clinical Classification composed of four-digit alpha-numeric codes with numerals of 0–9 and letters A–Z [168]. The characters represent levels with first character relating to level 1 and the four-digit codes increase in detail from left to right [168]. Each next hierarchical level increases the detail of the clinical diagnosis [171]. Chapter 0 contains terms that relates to occupation, chapters 1 to 9 relate to terms associated with history, examinations, procedures and administration while chapters A to U covers conditions, diagnoses and injuries and chapter Z, unspecified conditions [172].

2.3.2 Development of Read code lists for predictor variables.

Recommended Read codes for all predictors identified in the systematic review and from BHIVA/NICE guidelines were extracted from NHS Digital Read code browser (version 3 (CTV3 or v3)) to develop code lists for conditions, symptoms and treatment of interest (list in Appendix 4). All the characters or levels of the read codes were considered in the development of the code lists of the 74 HIV predictors listed, of which seven were demographic and socio-economic, 10 lifestyle or behavioural and 67 were clinical conditions and comorbidities (Appendix 5).

2.4 Summary of chapter

In this chapter, demographic and socioeconomic, lifestyle, clinical features and comorbid conditions predictors of HIV infection were identified through a systematic review. For those that might be recorded in primary care records, appropriate clinical (Read) codes were identified. The next chapter will investigate the possibility that codes used to identify the outcome of HIV may be hidden or disguised. Additionally, the next chapter will also explore completeness of recording of HIV in IMRD database.

Chapter 3: Recording of HIV diagnosis in primary care

3.1 Introduction to chapter

The previous chapter identified the risk factors associated with HIV in adults which will be used in developing a prediction model. This chapter identifies the outcome variable and explores its completeness in the IMRD database. The chapter is divided into three subsections namely i) qualitative exploration of coding practices for HIV/AIDS, ii) development of Read code lists for outcome variables (confirmed and "probable" HIV) and iii) explore completeness of recording of HIV diagnosis in primary care records including finding out the time from "probable HIV" to confirmed HIV using IMRD.

3.2 A qualitative exploration of coding practices for HIV/AIDS

3.2.1 Background

The investment in electronic medical records has improved patient care through increased efficiency and quality of services [114] [173] [174]. Furthermore, improving the use of information technology has provided a cost-effective way to access data for clinical audit, public health and epidemiological research [121] [122]. General Practitioners are recommended to code patient symptoms and diagnoses, and this requires those entering the data and researchers making use of the records to fully understand the code schemes [175]. Thus, although electronic medical records in health institutions are primarily recorded for clinical and routine patient care purposes, their quality and reliability should be reasonably good for use in research and for other purposes [121] [176].

Sharing of electronic patient data recorded by clinicians requires good information governance and privacy, which includes maintaining safe storage of data, patient consent and confidentiality [175]. Confidentiality of patient information has been a cornerstone in medical ethics and is central in building trust for a good doctor-patient relationship [177] [178]. A

narrative review to understand sharing data for people living with long-term health conditions found that the barriers to patients' willingness to share health data via digital technology were trust, security, privacy and identity [179]. The 2015 GP patient survey conducted in England revealed that 66 percent of patients trusted the GP they last had contact with, and 71 percent trusted the nurse they last had conduct with [180]. Furthermore, a survey conducted by PHE in 2017 on about 4,400 people living with HIV found that almost all the participants (98 percent) "were registered with a GP", of which 94 percent stated that their GP knew their HIV status [181]. Another survey using audio computer-assisted self-interview conducted in New York and published in 2011 found that 84 percent of patients living with HIV were willing to share their information with primary care clinicians involved in their care, mainly due to trust and respect from clinicians [182]. However, another study demonstrated that only 40 percent of patients infected with HIV in the UK disclosed their status to their GP because they did not trust staff at their general practice due to stigma and discrimination, lack of confidentiality, and attitude of reception staff [114] [183]. In some instances, patients might not disclose information to their healthcare providers because they want to protect themselves from security and privacy risks associated with electronic records [184] [185].

Similarly, there are examples of sensitive information or conditions known to clinical staff which are not being coded correctly because clinicians are reluctant to record highly confidential information in routinely accessed electronic health records. While some studies have shown that the use of different codes, within a list of standard codes, by GPs in recording non-stigmatised conditions such as heart disease,` does not affect the estimated prevalence of the diseases [122] [186]. This may not be the case with conditions where information disclosure would have a major impact on the patient. A study on 120 clinicians running university-based, outpatient mental health clinics in USA revealed that 63 percent were less willing to record highly sensitive information on electronic records [187]. There is a wide variety of codes used

to recorded concerns of child maltreatment in routine health records including indirect or euphemistic codes [188] [189] [190].

In relation to HIV, GPs are required to use the codes recommended by the Genito-Urinary Medicine Clinics listed in Appendix 6. However, a retrospective cohort study conducted in UK between 1995 and 2005 revealed that cases of HIV disclosed to GPs are sometimes recorded using a code such as ‘‘chronic viral illness’’ or the information is either recorded in free text (where it is inaccessible to systematic searching and therefore “hidden”) or not recorded at all [191]. In an online GP discussion forum, several respondents said that patients sometimes did not reveal a HIV diagnosis to their GP and that GPs sometimes used codes such as “immunosuppressed”, “viral illness” and “additional problem” to record HIV infection in free text of patient records. This, and comparisons with estimates of national prevalence from another study on HIV using GP records, suggest that HIV positive status may be recorded in hidden text or disguised using another code [35].

There is a need for further studies aimed at exploring barriers to recording of HIV positive status in primary care [191]. This study attempts to explore GPs’ perspectives on coding practice in relation to sensitive health conditions, particularly HIV, in primary care. It will establish how to identify HIV status when the coding has been hidden by using free text or alternative clinical codes. The identification of different ways in which HIV may be coded will improve detection of patients with HIV in primary care records. This will improve the predictive characteristics of a prediction model for HIV infection using information recorded in primary care records. A predictive model could inform primary care clinicians which patients they should offer HIV testing. Increased testing for HIV in primary care might lead to increased diagnoses and HIV diagnoses at an earlier stage, which means that it is important to create records which are correct and retrievable.

3.2.2 Methods

A qualitative approach was the most appropriate method of exploring behaviours, views and opinions on how patient data is recorded [192]. This approach refers to procedures or skills and frameworks used in exploring beliefs, norms and values shared by a community [193]. It allowed clinicians to explain their motivations and reasons for coding practice and to provide non-standard codes which might be used instead of standard (recommended) codes. Additionally, the sensitivity of the topic on why different recording was used, as it is better addressed by using a qualitative approach [194].

Semi-structured interviews were conducted either face to face or via Skype, or phone call. Semi-structured interviews are based on loosely structured open-ended questions (topic guide) that define the area under exploration which is used to initiate discussion and the interviewer and interviewee may diverge through follow-up questions [195]. The interviewer probed further to get additional information as the interviews progressed.

3.2.2.1 Recruitment

The sampling method that could have been used in this research was a combination of purposive sampling (recruiting participants that are representative of the population of interest) and snowball sampling (researcher asking participants if they know other participants that can be recruited) [192] [196]. However, the sampling used in this research was not purposive, as very broad invitations were sent out and interviews were held with participants who volunteered. This was done because the research was exploratory, and GPs were very challenging to recruit. The GPs were invited to participate in the study through various methods which included 1) requesting organisations that provided research passport to forward an email to their GPs, 2) asking for interest to participate from attendees of conferences and 3) requesting organisations that work with GPs in research to invite GPs for

expression of interest. The organisations that were approached to provide research passports were the National Institute for Health Research (NIHR) in Thames Valley and South Midlands, East London, West London and South London. The conferences attended were the Second Joint Multidisciplinary Conference of BHIVA and Royal College of General Practitioners (RCGP) in September 2016 and the BHIVA Autumn Conference 2016 in October 2016. GPs were also invited to participate through online GP platforms. The study aimed to recruit participants until the saturation point was reached. The saturation point is "the point at which gathering more data about a theoretical construct reveals no new properties, nor yields any further theoretical insights about the emerging grounded theory" [197]. This was described by Glaser and Strauss (1967) as the point when a researcher gets similar response over and over [198].

The lead researcher approached volunteer GPs and agreed on the venue and time suitable for the interview to take place, or via Skype or phone calls. The participants were informed about the length of the interview (20 minutes). Assurance that the information provided will be confidential and electronically recorded was elucidated before interviewees signed a form giving consent to participate, Appendix 7. The researcher ensured that signed consent was returned before the interview if a Skype/telephone interview was scheduled.

3.2.2.2 Topic guide

The respondents were asked about their experiences of recording sensitive data, if they think GPs ever disguise HIV diagnosis in electronic patient records, their opinion of why they would do that and how the records may be disguised (free text or use and type of another code that can be used).

Topic guide

Questions on sensitive conditions

- Do you think that GPs ever disguise sensitive conditions in the electronic patient records?
- Which conditions do they disguise?
- Would they do this for some types of patients more than others?

HIV specific coding

- With HIV in mind, how do you think GPs might do this?
- What type of other codes might they use?

Additional information was gathered on the participants' interest in HIV issues, years of experience and age ranges.

3.2.2.3 Recording of responses

The interviewer digitally recorded the interview as well as taking notes during and immediately following the interviews.

3.2.2.4 Ethics

Participants were paid for their time as stated in the guidance on "Attributing the cost of health and social care Research and Development" (AcoRD) [199]. The AcoRD guidance was introduced by the Department of Health in 2012 and it offers a consistent and transparent way of attributing costs of research in the National Health Service (NHS) [199]. Ethical approval and permission to conduct the research was obtained from the University of Birmingham. Additionally, a research passport was acquired from the Research and Development departments of the respective Clinical Commissioning Groups (CCGs).

3.2.2.5 Analysis

Data analysis from this study encompassed data organisation, thematic mapping and interpretation [200]. Data was organised and prepared for analysis by transcribing electronic recordings and typing up the field notes [201]. Analysis of data started while data collection was still in progress to enable the researcher to refine questions and fill in gaps as the study continued.

The data was tabulated according to the main topics namely, 1) recording of sensitive conditions, 2) direct Read codes used to record HIV, 3) surrogate or euphemistic terms used to record HIV and 4) other ways of identifying HIV positive patients. Data was manually coded by the researcher by reading through the notes and transcripts [192] [201]. The researcher used the steps of thematic analysis, which are familiarising with the data, generating initial codes, searched for themes, reviewed themes and write up the findings [193] [202]. Themes and codes emerged as data analysis progressed. Interpretation of the meaning of the themes was done as a final stage of the analysis.

3.2.3 Results

A total of six participants were recruited from London and West Midlands regions from July 2017 to September 2018. The age of participants ranged from 38 to 51 years. The number of years of GP experience ranged from 10 to 27 years. A third of the participants had an interest in HIV and sexual health.

3.2.3.1 Recording of sensitive conditions

The first part of the interview was aimed at finding out if sensitive conditions are recorded in electronic primary care records as set out by guidelines/rules. Half of the participants indicated that there are sensitive conditions or circumstances that are sometimes not shown on

electronic patient records. These conditions include symptoms of child sex abuse, mental health (including depression), domestic violence, termination of pregnancy and HIV (in previous years).

Box 1

Yes. I definitely, have been asked to disguise or to use a euphemism to indicate an HIV infection by a patient... So that's the only time when I had to use a different code. I am just thinking, I don't actually think we do this for any other clinical diagnosis or medical condition definitely not at the request of a patient. I mean if conditions stigmatise... like depression and severe mental illness for example, in the past things like abortion and domestic violence. #2

Some things like termination of pregnancy or domestic violence and I think I do recall HIV being one but that's maybe because that's what you are asking me about so that comes to mind, symptoms of child sex abuse or things like that. #3

"Mental health is sometimes disguised in electronic patient records because patients might not want other people to know or patient did not know that their information was recorded in primary care records and once they realised that, they asked the GP not to record it..." #1

.. one which is quite interesting is some mental health but it also relates to... it's because it's a big issue now, related to depression, and that is because I sort of sense that these days GPs Read code what used to be labelled as depression is now Read code as low mood and one of the reason is because there is extra clinical activity or targets if it is coded as depression. #4

I'm thinking about the mental health conditions, conditions that people would object to. And irritable bowel syndrome probably because it's an opinion and there is no text to say this is irritable bowel and there are no consequences for not coding it #5

Yes we use euphemism for HIV in particular, we use … and for Hepatitis B #6

The participants indicated that the disguise of sensitive conditions is not specific to any type of patients. However, the answers to some of the questions revealed that some conditions are not recorded as they should be due to extra work required to meet targets or when patients request for non-disclosure of the conditions due to stigma and discrimination from either practice staff or their community.

3.2.3.2 Recording of HIV in Primary care records

A. Direct HIV Read codes

The second part of the interview was aimed at understanding ways that are used by GPs to code HIV in primary care records. All the participants agreed that the Read codes used in most cases to code HIV are those recommended and permitted by GUM clinics listed in Appendix 6 [203].

Box 2:

We use 43C3-1 as the code for HIV positive patients and a pop-up is activated so that other GPs would know that it is an HIV positive patient. …. the system was there when I joined the GP practice in 2008, maybe for 10 years. …It was the same system used in the other practices I have worked in #5

We do not choose which code to use for our patients because the codes are provided by the GUM clinics #1

..for patients we know have HIV, there is no disguising of this condition. The way the NHS system works is that people can be tested in sexually transmitted disease clinics Genito Urinary clinic and maybe picked up there as having HIV and the rules on NHS confidentiality says that we are not obliged to know the diagnosis. The patient is encouraged to tell us they have this diagnosis, but they are not obliged to tell us. And often we only find it out a long time after the diagnosis is made. So, within the practice we do not prescribe HIV diagnosis but within the NHS we are sometimes thinking of this. #1

Now or certainly in the last 6 years since I've been using electronic records as a doctor, I have never felt it necessary to hide a code in any way. … when I code HIV, I would just code it as any other condition. I would code it as a major condition because it's a lifelong condition which affects health, so I'd code it in a similar way that I would hypertension, diabetes, cancer and things like that. #3

No, I can't think of any ways that people have hidden the diagnosis since I have been a practising GP. I think maybe in the past, potentially there are other ways that things may not be in the clinical system but no, it's not something I have come across recently #3

B. Surrogative terms used in recording HIV

In addition to the direct Read codes for HIV, there are some Read codes or texts that are sometimes used to identify HIV patients. Half of the participants highlighted some of the hidden ways used to identify HIV patients in the early 2000s. Such recording was either written in text notes or using a recording option. The surrogate codes used included immune/immunity problems, viral load, immunodeficiency and using priority 0 (the priority used to hide the code).

Box 3:

Well, there are two ways to try to make it less obvious. Two ways to make it … less … obvious. The traditional one is to use a euphemism of some sort,… obvious immunity problem or general immune problem…. That could cover things like leukaemia, chemotherapy or somebody died, so it doesn't actually specify that, even you can't tell somebody has HIV without the code. #2

a Read code generally used for HIV before [remotely related to immune problems] which all we just agreed that when we use this for the moment... #2

… about 2001… at that time we were provided with the instructions which had "hide some information by using" whether to prioritize the patient problems. You would prioritise problems, so Priority 1 will be the biggest problem and priority 3 the smallest problem… That is when I used the priority coding system so I think all clinical system used that option to hide to diagnosis and that is what I used all the time you could code it as priority zero so it can be hidden from prying eyes. #3

.. direct codes like HIV diagnosis or HIV positive but we also had surrogate names of HIV like immuno deficiency or viral load so… different types of direct and indirect codes for HIV. #4

C. Other ways of identifying HIV positive patients

Some of the participants suggested that they have other ways of identifying a patient's HIV status if the condition is not disclosed in the records. Most of these are associated with medication such as antiretroviral drugs and services provided to HIV infected patients such as

flu jabs every year or more frequent screening services such as smear tests. The other way suggested was the use of blood test results.

Box 4:

…managing people's prescriptions, look at the blood test results and filling things into your notes … flu jab and you clearly are not over 65. #2

if they had a test then clearly you can look at the test reports and that will be quite clear, somewhere in the history, if they have correspondence from the clinic, then it's on there, if they are taking antiretroviral drugs they are recorded …. #2

.. obvious that it was HIV because the medications would be in, you can't hide medications. There will be letters from infectious disease that will be coded. There will be regular blood tests, it will be pretty obvious to a clinician very quickly to diagnose that the patient has HIV. #3

3.2.4 Discussion

This study identified ways in which HIV is recorded in primary care records. It highlighted that generally, there are some instances where sensitive conditions are not disclosed or recorded correctly in electronic patient records. The reasons highlighted by the participants were stigma, patient preferences, extra clinical requirements (such as NICE and QOF) associated with the condition. However, the list of sensitive conditions might have been limited because the participants might have prepared to answer questions related to HIV since they were introduced to the topic of the research and were also provided with the participant information sheet.

The study also revealed that there were definite improvements in the way HIV is currently recorded in electronic records, compared to periods around year 2000. This could be a result of recommendations for primary care to be involved in HIV testing, HIV care and management in the early 2000s [204]. Most of the participants expressed the view that the Read codes used to identify HIV patients are those itemised in the GUMCAD lists, which are used by the GUM clinics. In those instances where HIV diagnosis was hidden, GPs use euphemism such as immune problem, blood-borne viral Infection, viral load and in the past, priority code 0 and retroviral infection.

This study revealed conditions that can be disguised in primary care records, and it is the first study to identify some of the terms used to disguise HIV recording in primary care. A limitation of this research is sampling bias, since the participants were those that were available for interviews and were only from two regions of England. Furthermore, the study did not reach saturation due to difficulties with recruitment of GPs (not available for the interview or stopped replying to correspondence). As with all qualitative research, the information cannot be generalised in the same way as quantitative research, due to a small sample and the fact that participants were not representative of GPs in UK as they were from London and West Midlands. However, the research can be generalised in terms of conceptual generalisation (‘making sense of the world’) and transferability (use of findings in other settings) [192].

3.2.5 Conclusion

The findings of this study revealed how specific codes that identify HIV positive patients from an electronic patient record database, which could be used to inform a study on development of a HIV predictive model, discussed in the next chapter. The study revealed that some GPs were recording HIV as recommended in the GUM clinic list. The additional codes that could be used to identify probable HIV infected patients are those falling in priority 0, or coded immune problem, retroviral infection, blood-borne viral infection and viral load. Some of these

records are more specific to HIV status but some of them were widely used for different conditions, hence more exploration needs to be carried out to confirm that they were recorded for HIV positive patients. The other method of identifying HIV patients, as suggested by the participants, is the Read codes for HIV related drugs.

3.3 Developing code lists for Outcome variables - Confirmed and "Probable" HIV

The primary outcome of the retrospective cohort study was "confirmed HIV" and the secondary outcome was "probable HIV". All the outcome variables were binary where a patient either had the outcome variable or not. Confirmed HIV is defined by the earliest of any of the following: a clinical code indicating HIV infection; a prescription for a HIV treatment (one used exclusively for HIV); a laboratory test indicating HIV infection; a clinical code for an AIDS defining condition.

The list of recommended and permitted Read codes for HIV provided by GUMCAD and any other extracted from NHS Digital Read code browser (version 3 (CTV3 or v3)) were used to develop code lists for confirmed HIV. All the characters or levels of the Read codes were considered in the development of the code lists (Appendix 8).

"Probable HIV" was included as a code because, as discussed in the section 3.2, it was believed the stigma attached to HIV may have led clinicians to use disguised or euphemistic ways of coding HIV infection. This could either mean the diagnosis of HIV was not recorded or that it was recorded much later than the true date of diagnosis. The definition of probable HIV was informed by qualitative research (covered earlier in this chapter), which identified possible disguised ways of coding HIV infection such as a clinical code for chronic viral

infection accompanied by a “high priority” flag. Definitions of probable HIV are shown in Table 3.3-1.

Table 3.3-1: Probable HIV: euphemistic terms for HIV infection identified from qualitative study

Code description	Is the code in the HIV code list
Priority 0	No
Immune/immunity problem	Yes
Blood-borne viral Infection	Yes
Retroviral infection	Yes
Immunodeficiency	No
Viral load (only HIV viral load included)	No

Read codes for these terms were extracted from NHS Digital Read code browser (version 3 (CTV3 or v3)) to develop code lists for conditions, symptoms and treatments of interest. The code lists are included in Appendix 8.

3.4 Completeness of recording of HIV diagnosis in primary care records

3.4.1 Introduction

It was thought that HIV diagnoses might not always be recorded in primary care records, because patients could receive their HIV treatment from hospital specialists and in the earlier years of the HIV epidemic might have been reluctant to inform their GPs of the diagnosis. This would have resulted in the incidence rate being lower than the expected and the data would be incomplete and unsuitable to use in the analysis. Therefore, quality checks were carried out on the data to determine if there was likely to be complete reporting of numbers of HIV infections.

3.4.2 Methods

The study population included all adult patients aged 18 and over registered in a practice which contributed data to IMRD between 2000 and December 2017, after the date of acceptable mortality reporting (the date when recording of patient mortality in each practice was judged to be complete) and one year after their Vision installation date (to ensure that the practice was using their patient record system to its full extent). Because “confirmed HIV” and “probable HIV” were used as the two definitions of the outcome event, separate cohorts were generated where patients with each outcome of interest at baseline were excluded.

If the identified “probable HIV” conditions recorded in IMRD database are disguised or euphemistic codes for HIV, it is likely that a significant number of these cases will later have a confirmed HIV code. This section corroborates the concept of “probable HIV”. It provides evidence to support the use of “probable HIV” as an outcome. This is important because some predictors of HIV may come after a diagnosis of “probable HIV” which would mean that in reality they were not predictors of HIV because they were recorded after the date of diagnosis. Survival analysis was carried out to determine if many cases of “probable HIV” were indeed later diagnosed as confirmed HIV and to describe the time between the recording of “probable HIV” and confirmed HIV status.

3.4.2.1 Acceptable HIV reporting

The reference standard for numbers of HIV infections was the expected incidence of HIV reported by UKHSA. As mentioned in Chapter 1, UKHSA collates and analyses UK wide data from outpatient HIV services and laboratories via surveillance systems and produces publications in December each year [62].

To assess the completeness of recording of HIV infection in primary care records, the incidence of HIV in the cohort was compared to the observed incidence of HIV produced by UKHSA at national (UK), country and regional levels. The incidence rate is defined as the number of new HIV cases in a specified period divided by person-time at risk [143] [205] [206]. The incidence used in this analysis was stratified by age and sex. This analysis determines whether it is possible to identify a date of “acceptable HIV reporting”, after which observed incidence of HIV infection nationally in IMRD is comparable to observed national incidence produced by UKHSA, if recorded incidence in primary care is always lower than regionally reported incidence, a date after which recorded incidence in primary care is stable, suggesting a degree of consistency in HIV reporting even if reporting is incomplete or the true incidence of HIV is lower in these practices. This uses a similar methodology to that used to define the date of acceptable mortality reporting [207]. If a date of “acceptable HIV reporting” could be identified, the analysis would be confined to data collected after that date and the cohort restricted to outcomes recorded after acceptable HIV reporting date.

The incidence rates of HIV by year were calculated from IMRD, stratified by age and sex. The incidence rates were then applied to national and regional populations to calculate the predicted numbers of new HIV cases per year. The number of new HIV cases were then compared to the numbers observed in UKHSA data nationally and regionally.

Data in IMRD dataset was based on 10 Strategic Health Authority regions which do not entirely correspond to regions used by UKHSA. Therefore, those regions which were the same in IMRD dataset and UKHSA were compared directly (East Midlands, North West, North East, West Midlands, Yorkshire and Humber, London, South West, East Midlands, East of England) and all other regions were combined (South East Coast and South Central).

3.4.2.2 HIV recording in relation to date of registration

Recording of a diagnosis of HIV in primary care records just after a patient registers with the practice, may be recording of a previous (historical) diagnosis of HIV and therefore may not be the date on which HIV was diagnosed. In this case recorded symptoms which seem to precede HIV diagnosis may in fact have occurred after diagnosis. To investigate this problem, the pattern of recording of HIV diagnosis in primary care was investigated in relation to the date of registration of the patient with the practice. If there is an excess of HIV diagnoses just after the date of registration this implies some HIV diagnoses recorded at this time may be previous (historical) diagnoses of HIV. A period of time after registration during which all historical HIV diagnoses are likely to have been recorded was identified.

3.4.3 Analysis

Descriptive analysis was carried out to reveal patterns of distribution in confirmed or “probable HIV” including completeness, trend over time and triangulation with other published statistics such as figures produced by UKHSA. Data visualisation in the form of charts and tables was used to explore variability of the data.

3.4.3.1 Descriptive analysis of outcome variables – confirmed HIV status

Data on patients with confirmed HIV from the IMRD database was used to show the trend of numbers and incidence (per 1000 population) of HIV over time at national, country and regional level. Incidence rates of newly diagnosed HIV were calculated by using the number of registered patients in the database to determine person years at risk and the newly

diagnosed HIV cases. Inconsistent trends were investigated to find out whether there was a reason for the pattern such as increase in HIV positive patients in general practices.

3.4.3.2 Quality checks of outcome variables

Completeness of recording of HIV in the dataset was assessed in exploratory analyses using the age-sex specific incidence rates of new cases of HIV to calculate expected numbers of new cases of HIV in the population of England and these were compared to the number of new cases of HIV for England reported in UKHSA.

Further exploratory analysis was conducted to determine the acceptable HIV date at national and regional level. The age-sex incidence rate of HIV was calculated from IMRD dataset, applied to UK population (population used by UKHSA) and the resultant numbers compared with the observed HIV diagnosis published by UKHSA. The acceptable HIV date is the date after which recorded incidence rates of HIV infection in IMRD is comparable to expected incidence produced by UKHSA.

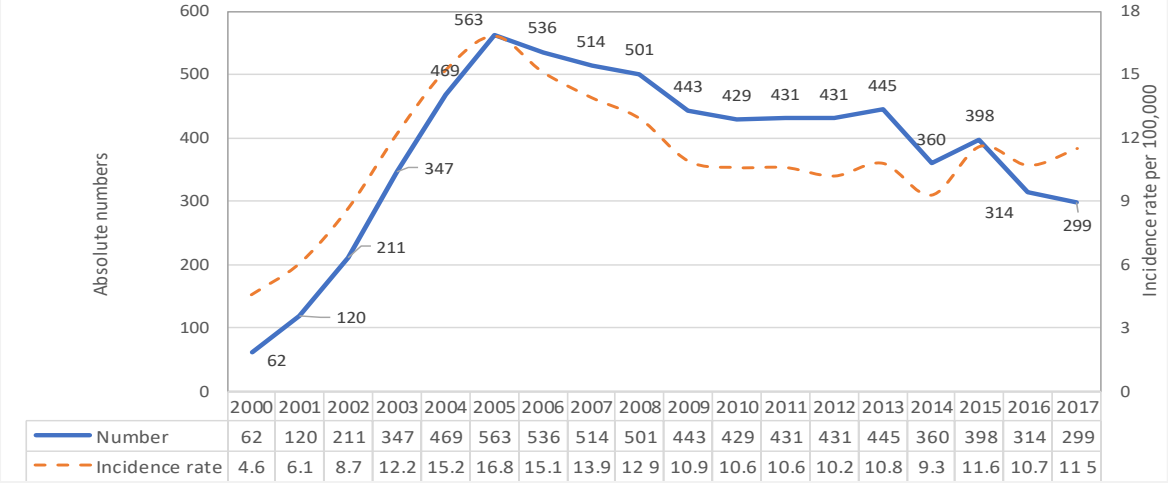
Data was analysed using and Stata SE 17. Preliminary data analysis in this subsection included descriptive analysis, assessment of data quality and completeness of outcome variables. The first part of the analysis was the identification of patients with the primary outcome (confirmed HIV) and assess completeness of its recording in IMRD dataset. The secondary analysis includes exploring the recording of “probable HIV”.

3.4.4 Results

Data for patients aged 18 and over, of both sexes and registered in IMRD between 2000 and 2017 were extracted from the IMRD database. The extracted dataset included a total of 9,415,088 patients, 6,873 (73.0 per 100,000) confirmed HIV positive.

Graphical display of the incidence of HIV in primary care records showed that it rose rapidly until 2005 before becoming relatively stable. This suggests under-recording of HIV status in primary care records prior to 2004, therefore the start of the study period was changed to 2004 (Figure 3.4-1).

Figure 3.4-1: Trend in numbers and incidence rates of newly diagnosed HIV positive patients aged 18 and above, IMRD database, 2000-2017



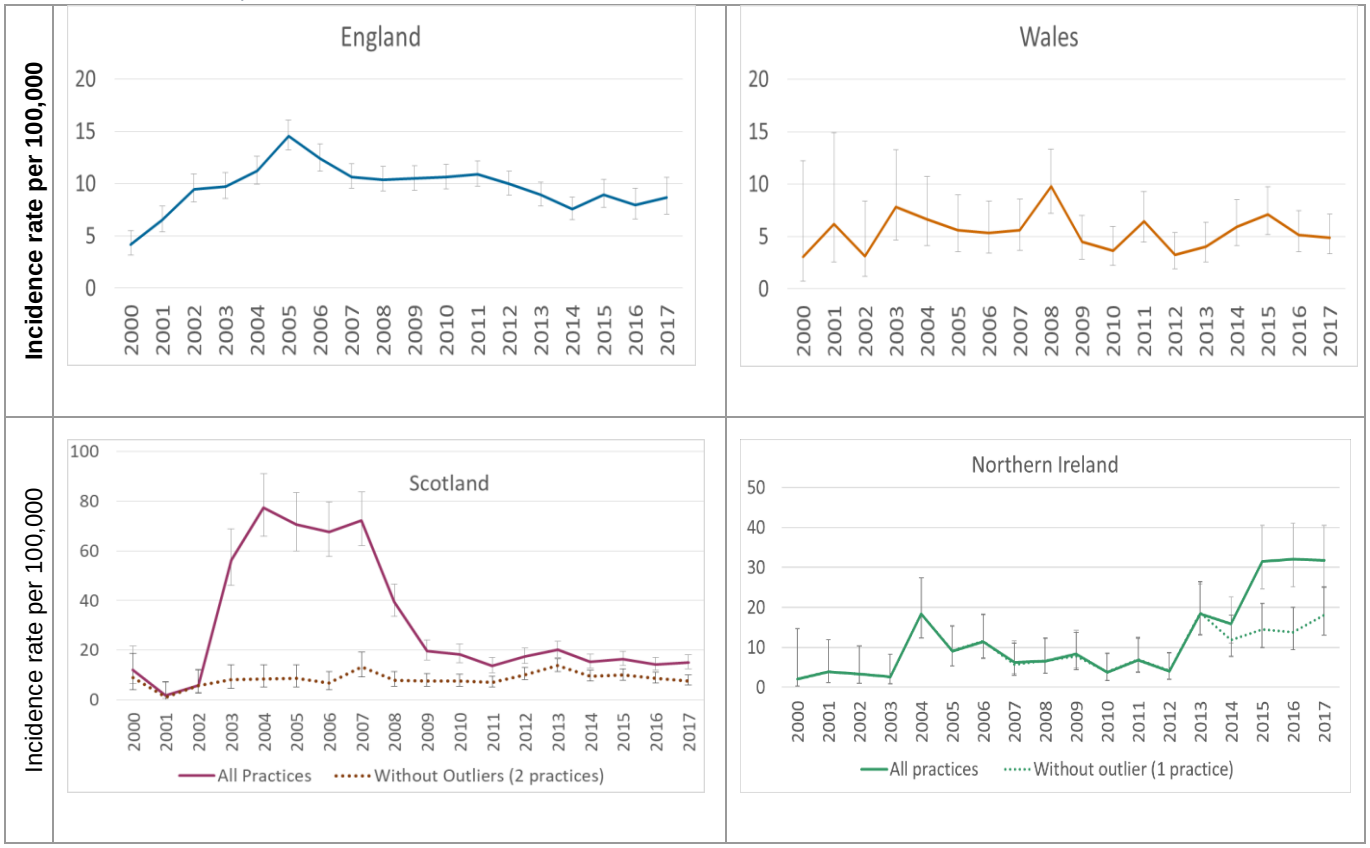
	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017
Number	62	120	211	347	469	563	536	514	501	443	429	431	431	445	360	398	314	299
Incidence rate	4.6	6.1	8.7	12.2	15.2	16.8	15.1	13.9	12 9	10.9	10.6	10.6	10.2	10.8	9.3	11.6	10.7	11 5

3.4.4.1 Descriptive analysis of Outcome variables by country and region – confirmed HIV status.

The IMRD dataset included 4,265 HIV cases in England, 375 in Northern Ireland, 1,872 in Scotland and 361 in Wales. Average incidence per 100,000 person years of new HIV cases from 2000 to 2017 was 10.0 for England, 13.9 for Northern Ireland, 24.2 for Scotland and 5.5 for Wales. Incidence rates for England and Wales were consistent over the study period, while there were some inconsistencies in incidence rates for Scotland and Northern Ireland.

In Scotland between 2003 and 2008, there was a markedly higher incidence of HIV than before and after these dates. Investigation of the high incidence rate in Scotland from 2003 to 2009 revealed a large increase in the number of new HIV cases in two practices with slight increase in registered patients, especially those aged under 25. These practices submitted their patient records to the IMRD database throughout the study period, 2000 to 2017. Approximately 40 percent of the practices' population are in the more deprived Townsend deprivation quintile of 4 and 5. When these practices were excluded, the overall incidence rate for Scotland was 9 per 100,000 population and the trend in incidence rate was more consistent over the study period (Figure 3.4-2).

Figure 3.4-2: Incidence rate of confirmed HIV infection per 1000 population by country – IMRD database, 2000-2017.



In Northern Ireland, there was a markedly higher incidence of HIV from 2013 onwards. A further investigation of the high incidence rate in Northern Ireland revealed a large increase in

the number of new HIV cases in one practice with no reciprocating increase in registered patients. The practice submitted patient records to the IMRD database from 2005 onwards. Over 94 percent of the practice's population are in the most deprived Townsend deprivation quintile of 3 to 5, with 60 percent of the population in the 3rd quintile. The age composition of the practice revealed a 14 percent increase in the under 25 age group and a 12 percent decrease in the over 65 from 2005 to 2017. The under 25s accounted for the increase in HIV cases in the last four years of the study period. When the practice was excluded, the overall incidence rate for Northern Ireland was 10 per 100,000 population and the trend in incidence rate was more consistent over the study period (Figure 3.4-2).

The number of HIV cases for England regions was lowest in Yorkshire and Humber, East Midlands and North East whilst highest numbers were in London (1303, accounting for a third of cases in England),Table 3.4-1.

Table 3.4-1: HIV positive population aged 18 and over by English regions – IMRD database, 2000-2017

Year	East Midlands	East of England	London	North East	North West	South Central	South East Coast	South West	West Midlands	Yorkshire & Humber
2000	3	3	17	0	5	4	8	5	2	1
2001	6	8	35	1	12	11	14	14	7	2
2002	18	17	61	9	24	25	11	13	13	5
2003	9	24	68	19	17	29	22	21	22	2
2004	12	29	87	6	34	39	26	16	26	6
2005	7	89	85	8	28	67	27	27	43	10
2006	11	88	77	7	37	48	26	13	35	8
2007	9	19	93	12	35	55	24	20	40	3
2008	12	24	98	4	29	54	20	28	34	5
2009	9	27	95	6	27	59	26	30	29	7
2010	6	15	110	6	38	75	24	20	16	2
2011	8	17	101	3	25	71	31	19	39	2
2012	1	21	93	1	42	33	31	20	44	3
2013	1	13	102	1	30	22	28	15	31	0
2014	0	17	67	4	20	15	24	10	26	1
2015	0	15	48	1	31	20	24	19	12	2
2016	0	7	40	2	15	9	18	10	10	1
2017	0	6	26	0	9	27	11	4	10	2
Total	112	439	1303	90	458	663	395	304	439	62
% of Total	**2.6%**	**10.3%**	**30.6%**	**2.1%**	**10.7%**	**15.5%**	**9.3%**	**7.1%**	**10.3%**	**1.5%**

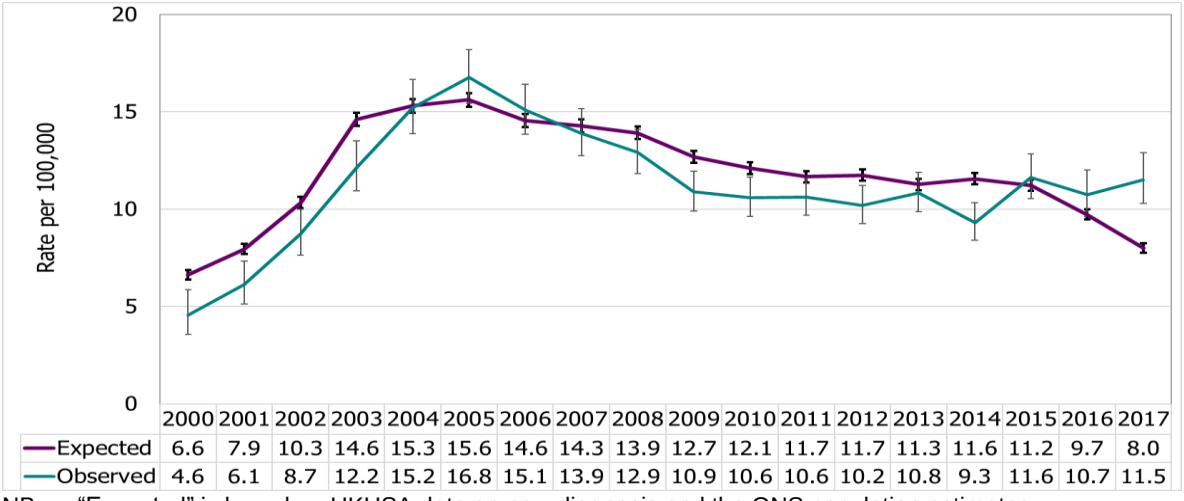
3.4.4.2 Quality checks of Outcome events– Acceptable HIV reporting

Completeness of HIV recording in the IMRD database was assessed by determining the acceptable HIV reporting date, the date after which recorded incidence of HIV infection in IMRD practices is comparable to incidence rates derived from data produced by UKHSA.

A. Incidence of HIV at National (UK) level

Generally, comparison of HIV incidence rates published by UKHSA and the observed incidence rates from IMRD database showed no statistically significant differences for most of the years during the study period, 2000 to 2017, especially from 2004 onwards (shown by overlapping confidence intervals in Figure 3.4-3). Furthermore, the HIV incidence rates through the whole study period for IMRD dataset of 11.55 per 1000 (95% CI:11.23, 11.83) were not significantly different from 11.83 per 1000 (95% CI: 11.73, 11.80) for UKHSA published data. A further analysis to compare the new HIV diagnosis using the incidence rate from IMRD dataset was conducted to assess if the observed numbers from IMRD were significantly different from the UKHSA expected figures.

Figure 3.4-3: Comparison of predicted incidence rates of HIV infected patients aged 18 and above derived from IMRD database and observed incidence in UKHSA, 2000-2017



	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017
Expected	6.6	7.9	10.3	14.6	15.3	15.6	14.6	14.3	13.9	12.7	12.1	11.7	11.7	11.3	11.6	11.2	9.7	8.0
Observed	4.6	6.1	8.7	12.2	15.2	16.8	15.1	13.9	12.9	10.9	10.6	10.6	10.2	10.8	9.3	11.6	10.7	11.5

NB: "Expected" is based on UKHSA data on new diagnosis and the ONS population estimates.
"Observed" are the incidence rates from the IMRD database.

For each year from 2000, the age-sex stratified incidence rates from IMRD dataset were applied to the UK population to generate observed numbers of HIV cases. These were compared with those produced by UKHSA. The results showed a generally lower expected new HIV numbers than the numbers observed and published by UKHSA from 2007 to 2014 (Table 3.4-2). However, the difference was within 20 percent over the period under study, 2001-2017.

Table 3.4-2: Comparison of HIV Incidence from IMRD database and UKHSA, 2001-2017

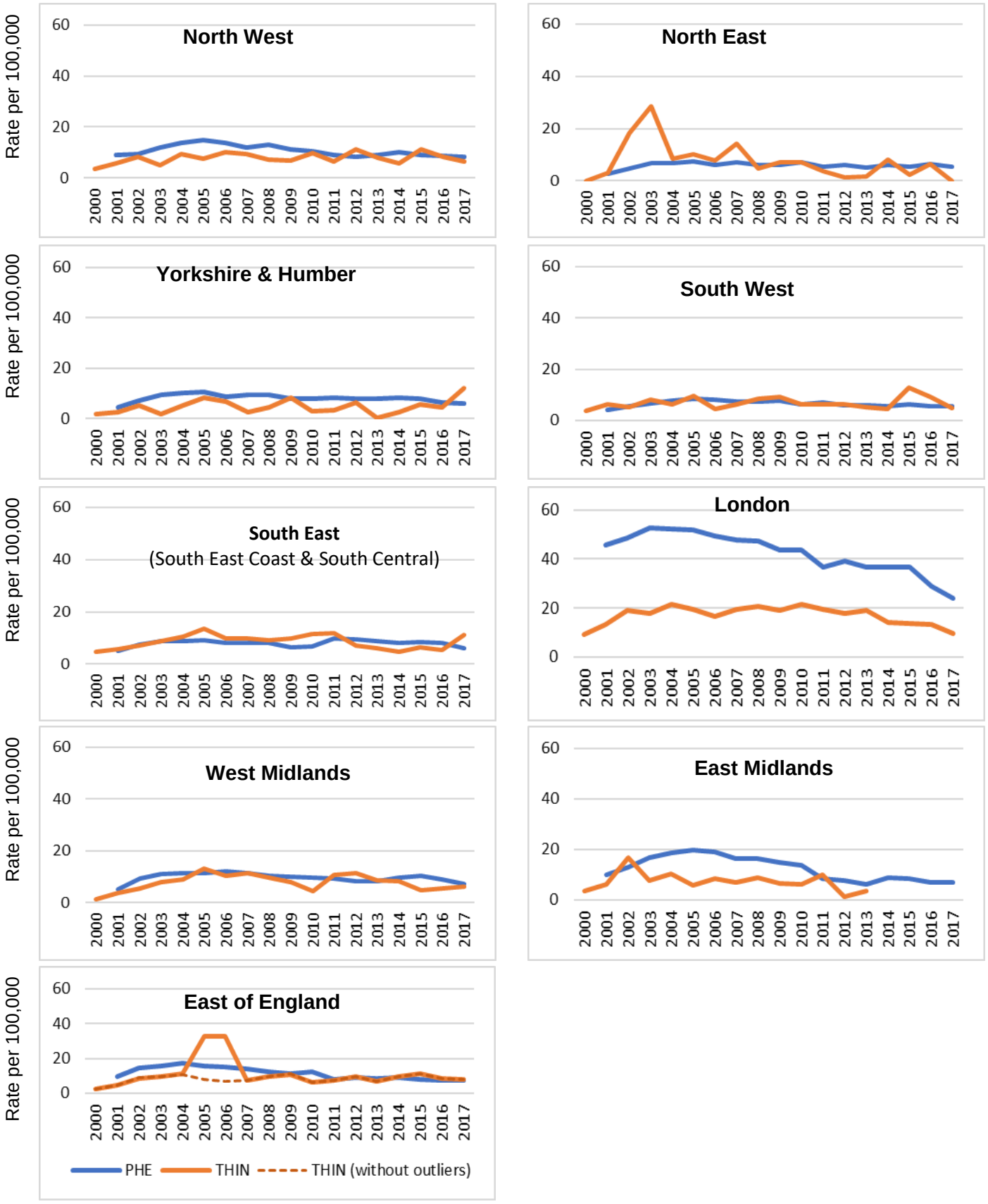
Year	IMRD HIV incidence rate (Expected)	Expected HIV numbers UKHSA*	Observed UKHSA	% Differences from UKHSA (Expected minus observed)
2001	4.6	2180	3249	-33%
2002	6.1	2940	3917	-25%
2003	8.7	4220	5111	-17%
2003	12.2	5914	7110	-17%
2004	15.2	7456	7503	-1%
2005	16.8	8297	7728	7%
2006	15.1	7533	7274	4%
2007	13.9	7012	7194	-3%
2008	12.9	6564	7080	-7%
2009	10.9	5580	6507	-14%
2010	10.6	5471	6255	-13%
2011	10.6	5546	6085	-9%
2012	10.2	5351	6166	-13%
2013	10.8	5716	5950	-4%
2014	9.3	4948	6151	-20%
2015	11.6	6234	6013	4%
2016	10.7	5797	5251	10%
2017	11.5	6248	4334	44%

*Numbers calculated from applying age-sex stratified incidence rates from IMRD on UK population – refers to expected numbers if we apply IMRD incidence rate to population used by UKHSA.

B. Incidence of HIV at English Regional level

Comparing HIV incidence rates by region from 2011 onwards showed no significant difference between incidence rates from UKHSA and IMRD dataset for all regions, except for London and East Midlands where incidence rates from IMRD dataset were lower than UKHSA figures (Figure 3.4-4). The incidence rate for the East Midlands was affected by the decrease in number of practices contributing data to IMRD from 18 in 2010 to two in 2014, followed by no practice contributing to IMRD from 2015 to 2017. The IMRD regions of South East Coast and South Central were mapped to the UKHSA regions of South East and the incidence rates were similar over the entire study period, 2000 to 2017.

Figure 3.4-4: Comparison of HIV Incidence from IMRD database and UKHSA by English regions, 2000-2017



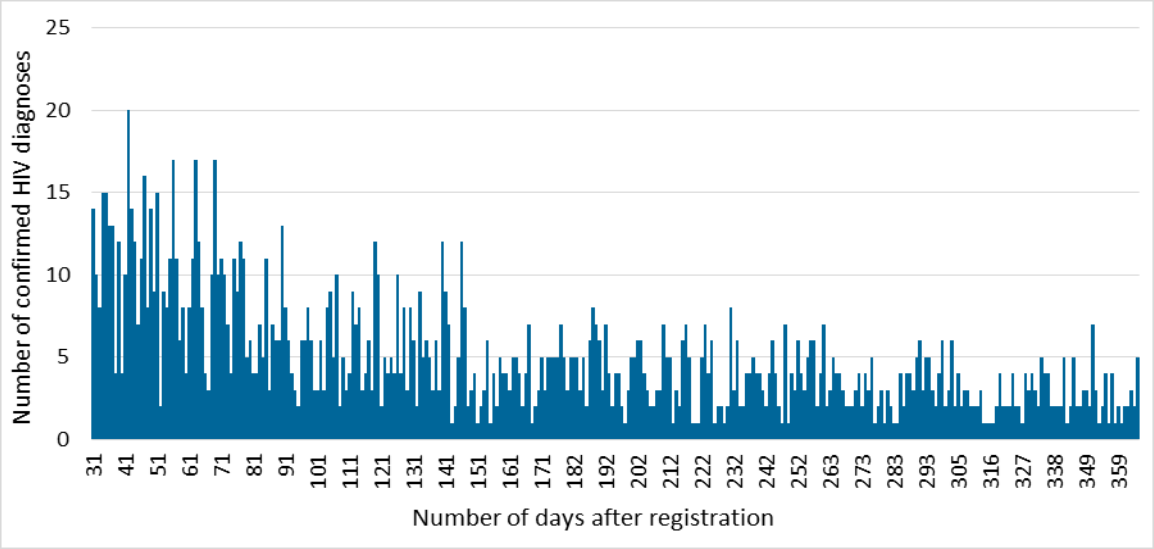
In East of England, there was a markedly higher incidence of HIV in 2005 and 2006 than before and after these dates. Investigation of the high incidence rate revealed a large increase in the number of new HIV cases in one practice which joined IMRD from 2002 to 2011. The practice had a 10 percent increase in young adults aged 18-25 which is the age group which accounted for 90 percent of newly confirmed cases in 2005 and 2006. Over 70 percent of the practice population reside in the least deprived Townsend deprivation quintile of 1 and 2. When the practice was excluded from the overall incidence rate for East of England, the trend in incidence rate was more consistent over the study period, Figure 3.4-4.

.

C. HIV recording in relation to date of registration

The pattern of recording of HIV diagnosis in primary care in relation to the date of registration of the patient with the practice revealed the incidence of HIV recording in the first 150 days after patient registration was approximately double the incidence after 150 days, but there were no sharp spikes in incidence of HIV recording, Figure 3.4-5. This analysis suggested it might be useful to conduct a sensitivity analysis for HIV diagnoses >150 days after registration in the model to investigate whether predictors were different in these two groups.

Figure 3.4-5: Number of HIV cases recorded in relation to date of registration, IMRD database: 2000-2017

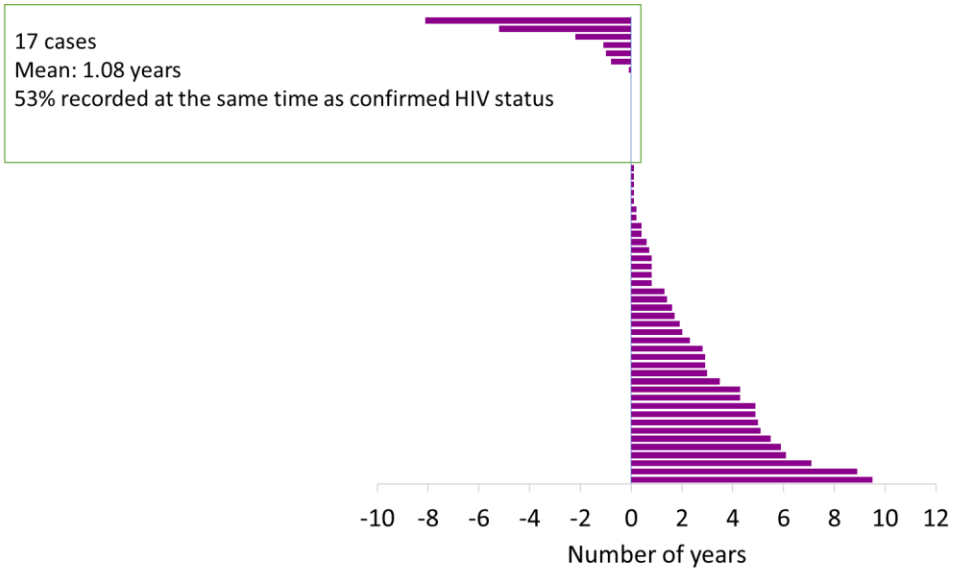


3.4.4.3 Analysis of Outcome variables – "Probable HIV" status

There were 21 patients with "Immune/immunity problems" recorded in IMRD dataset between 2000 and 2017. None had a diagnosis of HIV recorded. There were 152 patients with "Blood-borne viral Infection" recorded in IMRD between 2000 and 2017, none had "Blood-borne viral Infection" recorded before a confirmed HIV diagnosis was recorded and only two (1.3 percent) had "Blood-borne viral Infection" recorded after a confirmed HIV diagnosis was recorded and 98.7 percent had no diagnosis of HIV recorded.

Of 6905 patients with "Immunodeficiency" between 2000 and 2017, only 17 (0.2 percent) had "Immunodeficiency" recorded before a confirmed HIV diagnosis, 50 (0.7 percent) of the patients had "Immunodeficiency" recorded after a confirmed HIV diagnosis and 99 percent had no diagnosis of HIV recorded in the records. Of the 17 patients who had confirmed HIV diagnosed after immunodeficiency, nine (53 percent) were diagnosed on the same day and the average time between "immunodeficiency" and confirmed HIV was 1.08 years (Figure 3.4-6).

Figure 3.4-6: Number of years from confirmed HIV date: Immunodeficiency, IMRD database: 2000-2017



3.4.5 Discussion

There were differences in HIV incidences between geographical areas in UK. England had higher incidence compared to the rest of UK and southern regions of England, mainly London had higher rates than Northern regions. The regional differences could be attributed to varying demographic composition, lifestyles of the population and the urban/rural variations. However, there is decline in HIV incidence in IMRD dataset for East Midlands in 2014 onwards, explained by a decrease in practices contributing to IMRD to two in 2014 followed by no practices contributing to IMRD from 2015 onwards.

Two practices in Scotland had a markedly higher incidence of HIV between 2003 and 2008. This could be because they attracted more HIV patients during this period. The attraction of more HIV patients could be a result of providing good care to populations with a high risk incidence of HIV/associated conditions and/or lifestyle issues (such as drug users) or good HIV patient care and clinical management as recommended by the Medical Foundation for AIDS and Sexual Health in the early 2000s [204]. This could be the same reason for the higher incidence of HIV patients in one practice in Northern Ireland.

The HIV data in IMRD database could be generalised to UK population since the incidence of HIV in the IMRD dataset at UK level was similar to incidence reported by UKHSA from 2004 onwards. At regional level, the incidence of HIV in the IMRD dataset in most regions was similar to incidence reported by UKHSA, except for London (over the entire study period) and East Midlands (from 2014 onwards). The differences in incidence rates in London could be attributed to under-reporting of HIV in GP practices contributing to IMRD compared to what is reported in GUM clinics and other settings (for example, in antenatal clinics). There is a possibility that the prevalence of HIV in primary care could be increased by using databases which are linked to diagnosis recorded in secondary care admissions data, such as using CPRD linked to Hospital Episodes Statistics (HES) [208].

The euphemistic terms mentioned by participants of the exploratory qualitative research were only recorded in less than one percent of observations in IMRD dataset, revealing that GPs entering data in their electronic health records have mostly used the recommended codes for HIV since 2000. Therefore, there was little evidence that codes for immune/immunity problems, blood-borne viral Infection or immunodeficiency were being used as euphemistic codes for HIV in the IMRD dataset. Hence, the "probable HIV" analysis was not likely to yield any useful results in the model and was excluded from further analysis.

The descriptive analysis and quality checks of the outcome variable revealed the following

1. Incidence of HIV in IMRD dataset from 2004 onwards is consistent with incidence in UKHSA. Therefore, it was decided that multivariable modelling should include patients enrolled in all IMRD practices from 2004 onwards. The differences between the IMRD and UKSHA incidence rates from 2000 to 2003 could be credited to low recording of HIV in primary care shown in literature [191]. Improvements of recording of HIV in primary care, from 2004 onwards, could be attributed to standards recommended in the early 2000s, which proposed primary care to be involved in HIV care and testing [204]. These recommendations were published by Medical Foundation for Aids and Sexual Health in collaboration with the Department of Health, BHIVA and National Association of NHS Providers of AIDS Care and Treatment [204].

2. Ascertainment and recording of HIV is therefore likely to be sufficiently good to derive a prediction model

3. From 2003 to 2008 there is unusually high incidence of HIV in Scotland which is accounted for by two general practices. This was the case for a practice in Northern Ireland between 2015 and 2017. Practices with an unusually high incidence of HIV may include patients with different characteristics (for example, a high proportion of drug misusers) which could affect

the multivariable modelling. It was decided that all the practices would be included in the model and a sensitivity analysis would be undertaken to ascertain if exclusion of the outlier practices would have influenced the model.

4. The “probable HIV” analysis found that there were few cases where euphemistic terms were recorded. Therefore, including them in the model would not add value, so they were not used in further analysis.

3.4.6 Conclusion

This section of the chapter identified the acceptable HIV reporting period which should be considered in developing a prediction model. Hence the prediction model and validation used the IMRD dataset from 2004 to 2017. Furthermore, it highlighted that in most cases, GPs use the appropriate Read codes for recording HIV in electronic health records (which are then uploaded to IMRD database) and there is no need to consider the euphemistic terms in the development of a prediction model.

3.5 Summary of chapter

In this chapter, an investigation was conducted to identify euphemistic terms used by GPs in recording HIV in primary care records in a hidden way. The development of code list for the outcome variables was discussed followed by an exploration on the use of the euphemistic terms in IMRD database. Furthermore, this chapter checked the incidence of HIV in IMRD dataset compared to the official data published from UKHSA. The next chapter will develop and validate a model for a risk score of HIV for primary care using the risk factors identified in chapter 2 and informed by the investigation on recording of outcome variables in this chapter.

Chapter 4: Derivation and internal validation of a HIV risk score for primary care in the United Kingdom: a retrospective cohort study using IQVIA Medical Research Data (IMRD) database.

4.1 Introduction to chapter

The second chapter identified the predictors of HIV found in literature (including guidelines) and discussed how the Read codes for the predictor variables were developed. The previous chapter investigated the alternative ways clinicians might record HIV in primary care, discussed how Read codes for outcome variables were developed and explored the quality of outcome variable in electronic primary care records. This chapter will use all the variables identified in these two chapters in deriving and validating a prediction model which could be used in primary care to identify patients for HIV testing.

4.2 Background

The life expectancy of HIV positive patients has increased over the years, almost approaching that of the general population [11] [21] [26]. As mentioned in chapter 1, this is linked to the effectiveness of and adherence to ARTs, early diagnosis (CD4 count above 350/mm^3), and good HIV patient care and clinical management [11] [42].

Late diagnosis of HIV (CD4 count below 350/mm^3) is associated with high morbidity and mortality, onward transmission and increased health and social care costs [114]. Most of AIDS related deaths in England can be attributed to late diagnosis of HIV, making late diagnosis a public health challenge [209]. In England, the Department of Health (DH) set the key strategic priorities on HIV as; reducing the proportion of late HIV diagnoses and increasing the

proportion of HIV infections diagnosed [210]. Hence, the DH found it useful to add an indicator of the estimated proportion of HIV positive patients diagnosed late, with a target of 25 percent or less [210] [211]. Because of the importance of early diagnosis, United Kingdom (UK) national guidelines recommend routine testing for HIV patients in high-risk groups and optimal health care for HIV positive patients [20].

Earlier chapters highlighted that about 33 percent of patients diagnosed with HIV in UK have seen a General Practitioner (GP) within the previous year [26] [35] [106]. A survey conducted in London to identify missed opportunities in health care revealed that as much as 76.4 percent of patients diagnosed with HIV, between April 2004 and February 2006, had presented to their GP [114]. It would therefore be useful to identify patients likely to be HIV positive in primary care in order to encourage GPs to undertake HIV testing in appropriate patients. This could be linked to an alert system in electronic primary care records. It could increase diagnosis of individuals that were unlikely to get tested via the education and awareness campaign alone or through visiting GUM clinics [8]. To identify patients likely to be HIV positive in primary care, there is need to determine predictive factors associated with HIV infection that are recorded in primary care records. These factors were used in the development of a predictive model.

UK guidelines identify 37 clinical indicator conditions for adult HIV infection that should prompt an HIV test [20]. In a previous matched case-control study carried out on data from electronic primary care records, only 12 of these 37 clinical indicator conditions were found to be significantly associated with HIV infection [35]. The study matched cases and controls by age and sex, therefore age and sex could not be considered as predictors. The relationship between indicator conditions recorded in primary care records and risk of HIV is therefore not known. NICE recommended further research into the prevalence of HIV in people with indicator conditions [212].

A cluster randomised controlled trial promoted HIV testing in all newly registered adults in general practices in London between 2010 and 2012 [109]. This was an area with a prevalence of known HIV of 0.8 percent. The intervention significantly increased the rate of HIV diagnosis and cost-effectiveness analysis found HIV testing was likely to be cost-effective in a population with this prevalence [109] [112].

The aim of this research is to generate a prediction model that calculates risk of HIV using information available in primary care records. Calculating a risk of HIV for each patient would enable general practitioners to identify patients with an expected prevalence of 0.8 percent or higher and offer HIV testing [112].

This research used a retrospective cohort design and included additional potential predictors which were not considered in the case control study, and these are: personal characteristics, symptoms and comorbidities. The risk factors included in the prediction model was informed by findings from the systematic review of HIV risk factors, covered in chapter 2. The outcome definition (HIV infection) was informed by a qualitative study and explorative analysis on completeness of recording of HIV diagnoses in primary care records and the use of euphemistic terms, covered in chapter 3.

4.3 Methods

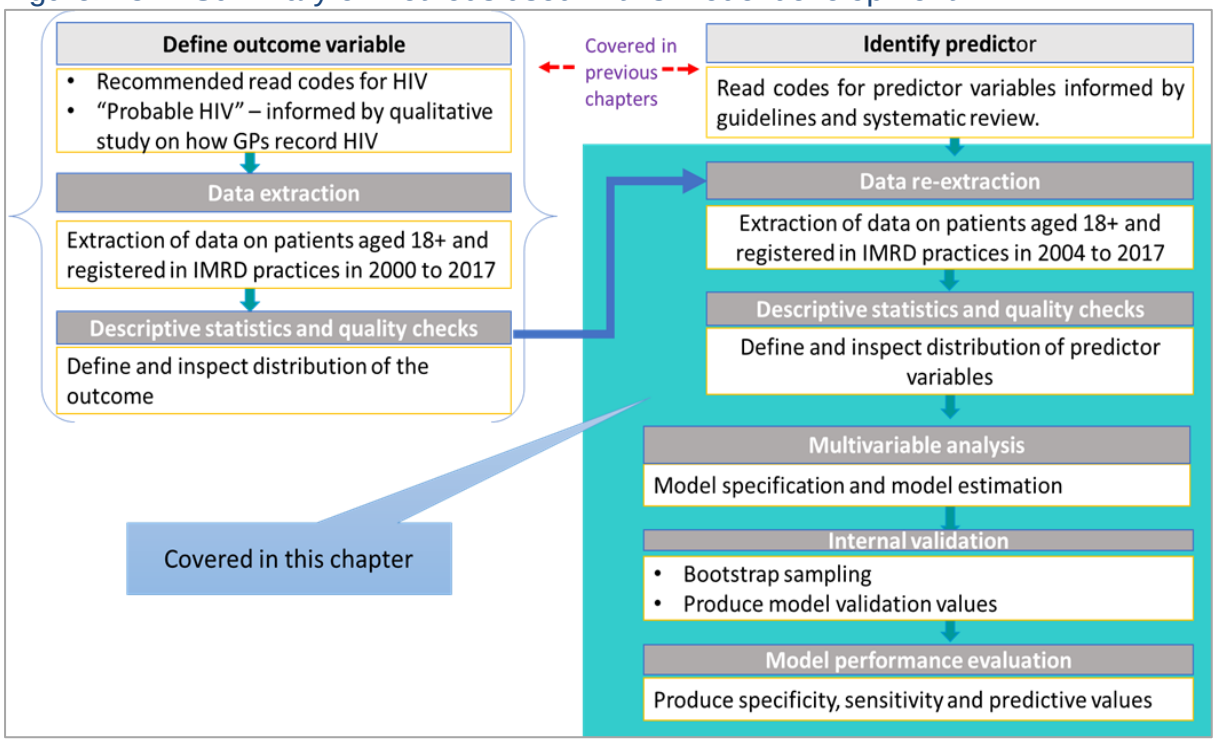
A retrospective cohort study was undertaken using anonymised UK primary care electronic medical records. As shown in chapter one, the IMRD general practice database contains more than 3.6 million active patients and approximately nine million previous patients from approximately 500 practices [128]. Data are regularly updated and the database is broadly

generalisable to the UK population in terms of age, gender, medical conditions and diseases, death rates and socio-economic determinants [213].

The reporting guidelines followed in this study were the Reporting of studies Conducted using Observational Routinely-collected health Data (RECORD) checklist (Appendix 9) and the Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD) [214] [215] [216]. Prediction models are developed to estimate the risk of the presence of a disease or condition (diagnostic) or the risk that an event will occur in future (prognostic) [216] [217] [218]. A diagnostic prediction model starts with subjects' presenting symptoms and explores the cross-sectional relationship (with short follow-up period) between predictors and outcome. A prognostic prediction model starts with healthy subjects and explores relationship between predictors and outcome after a longitudinal follow-up [219] [216]. Although the prediction model developed in this study would be used to identify patients that should be tested for HIV, a prognostic prediction model was developed because most of the signs and symptoms of HIV are not specific to HIV alone and also due to the long-term association between HIV and its predictors during the various clinical stages of HIV/AIDS progression, discussed in section 1.2.

The approach used in this study included i) the definition of problem and data inspection covering the inclusion and exclusion criteria, define and inspect distribution of the predictor variables, and data quality checks, ii) model specification covering assessment of multivariable analysis assumptions, iii) model estimation and performance, and iv) model validation, Figure 4.3-1 [220] [221].

Figure 4.3-1: Summary of methods used in this model development



4.3.1 Inclusion and exclusion criteria

The study population included all adult patients aged 18 and over registered in a practices which contributed data to IMRD between 2004 and December 2017, after the date of acceptable mortality reporting (the date when recording of patient mortality in each practice was judged to be complete) and one year after their Vision installation date (to ensure that the practice was using their patient record system to its full extent) [207]. The included patients were registered with the practice at least 30 days before entry. This ensured that patients who change practice often are included in the study. Consideration on the Acceptable Mortality Reporting (AMR) date, the date from which the general practices recorded date of death which seemed complete and similar to an expected standard, was taken in determining the index date [207] [222]. Therefore, the index date for each patient was the latest of the following dates: 1/1/2000, practice AMR date, Vision installation date plus one year, patient aged 18 years, and patient registration date plus 30 days.

4.3.2 Follow up and censoring events

Patients were followed-up from the index date until death, patient deregistration, last data collection from the practice or when the outcome event occurs. That means patients with HIV at index date were excluded.

4.3.3 Predictor variables

Sociodemographic and geographic predictor variables considered in this study include age, sex, ethnicity, deprivation quintile and urban/rural residence. Lifestyle variables considered were behavioural factors such as smoking, drug and alcohol misuse and unsafe sex. Clinical factors included symptoms such as diarrhoea or lower respiratory infection.

Predictor variables included conditions listed in the guidelines; 2008 British HIV association (BHIVA) and 2011 National Institute for Health and Care Excellence (NICE) guidelines (PH33 and PH34) [20] [223]. Additional candidate predictor variables included in the analysis were derived from a systematic review of predictors of HIV infection, discussed in chapter 2 [140]. Predictor variables included in the study are listed below.

- Socioeconomic and demographic, and geographic:

  Age, gender, ethnicity, country of birth, poverty index (Townsend deprivation), sexual orientation

- Lifestyle/behavioural

  Unsafe sex, smoking status, alcohol misuse, drug misuse, obesity, contact abroad, male anal sex, number of lifetime partners, partner characteristics

- Clinical and co-morbid conditions:

  All conditions identified to have an association with HIV infection listed in Table 4.3-1.

Table 4.3-1: Clinical and co-morbid conditions included in the study

Condition Group	Clinical condition
Respiratory	– Tuberculosis
	– Pneumocystis carinii
	– Bacterial pneumonia
Neurology	– Cerebral toxoplasmosis
	– Primary cerebral lymphoma Cryptococcal meningitis
	– Progressive multifocal leukoencephalopathy
Dermatology	– Kaposi's sarcoma
Gastroenterology	– Abdominal pain
	– Persistent cryptosporidiosis
	– Oral candidiasis
	– Oral hairy leukoplakia
	– Chronic diarrhoea of unknown cause
	– Weight loss of unknown cause
	– Salmonella, shigella or campylobacter
	– Hepatitis B infection
	– Hepatitis C infection
Oncology	– Non-Hodgkin's lymphoma
	– Anal cancer or anal intraepithelial dysplasia
	– Castleman's disease
	– Cervical dysplasia
	– Head and neck cancer
	– Hodgkin lymphoma
	– Space-occupying lesion of unknown cause
	– Lung cancer
Gynaecology	– Cervical cancer
Ophthalmology	– Cytomegalovirus retinitis
	– Any unexplained retinopathy
Haematology	– Blood dyscrasia
Other	– Mononucleosis-like syndrome
	– Pyrexia of unknown origin
	– Hypertension
	– Lymphadenopathy
	– Diabetes
	– Any sexually transmitted infection (includes other STI, chlamydia, gonorrhoea, genital herpes, syphilis)

Socio-demographic predictors recorded in patient records were considered irrespective of time of recording. For example, ethnicity was included whenever it was recorded, even if this was long before HIV diagnosis. All the clinical and comorbid predictors mentioned in the Table 4.3-1 were considered as candidate variables for the model.

4.3.4 Quality checks on candidate predictor variables

Quality checks were carried out on the data to ensure that candidate predictor variables were recorded with sufficient frequency to be useful predictors. Candidate predictor variables with

poor quality issues with completeness, inconsistency and irregularities were dropped since they would be less useful in data analysis [224]. The following criteria was applied during descriptive analysis and stability checks to drop variables with poor quality issues:

1) missing values - the threshold for this was different for each variable depending on whether every patient was supposed to have a value or in some cases no recording is required. For example, the proportion of missing ethnicity could be similar to what is recorded in other published data. Evidence showed that completeness of ethnicity data in IMRD was comparable with other primary care databases such as QResearch and CPRD, and data from CPRD was “comparable to that of the combined censuses for England, Wales, Scotland and Northern Ireland” [225].
2) recorded too infrequently such as a population prevalence of 0.1 percent or less, except for clinical and comorbid conditions which might still have association with HIV even at low population prevalence. The prevalence threshold was decided as a pragmatic method of reducing the number of variables that should be included in a multivariable analysis because variables with a low prevalence are likely to be of little practical value in prediction.
3) inconsistency/variability over time (for example, the recorded frequency changes unpredictably from one year to another). This shows irregularity on how the predictor is recorded over time and this might affect the association between the predictor and HIV.
4) variability between practices which suggest that there may be problems with inconsistent coding. This indicates issues with validity of data for the predictor as different practices might be using different measurements and coding [224] [226] [227].

4.3.5 Ethical approval

Data extracted from the IMRD database was anonymised. Research using the IMRD data collection was approved by the NHS South-East Multi-centre Research Ethics Committee (MREC) in 2003. Under the terms of this ethics approval, studies must undergo scientific

review to help ensure appropriate analysis and interpretation of the data, this study received approval from the Scientific Review Committee (SRC Reference Number: 17THIN009 on 10th February 2017) [228].

4.4 Analysis

The first analysis covered descriptive analysis and quality checks of predictor variables. This was followed by survival analysis between the time when predictors occurred and recording of confirmed HIV, in the multivariable analysis and the internal validation. Sensitivity analysis was conducted to evaluate model performance.

4.4.1 Descriptive analysis and Quality checks of candidate predictor variables

Descriptive analysis and consistency checks were used to decide on the possible predictor variables to include in subsequent analysis. For categorical variables such as ethnicity or Townsend quintile an additional category was created for missing data. Clinical conditions were binary, meaning the presence of a code was taken to indicate that the condition was present. Some variables were categorised or re-categorised at this stage. Conditions with more than one consultation were turned into semi-quantitative variables for example, one consultation for diarrhoea and two or more consultations for diarrhoea.

The frequency of each recorded predictor variable in the cohort was used to determine whether the variable was included in the subsequent analysis. Those with a prevalence of less than 0.2 percent were dropped or further analysis of stability of the variable was undertaken. Stability of predictor variables was assessed by checking whether there was consistency in incidence of the variable over time. Clinical opinion was sought on whether to include clinical and co-morbid conditions which were infrequent but likely to be strongly associated with HIV.

Some of the predictor variables were re-categorised, either in accordance with data published by UKHSA (for example, Black ethnicity and White/Asian ethnic groups) or by combining less frequent conditions to main disease groups that are clinically known to have association with HIV (for example, combining all sexually transmitted infections (excluding syphilis) or grouping cervical conditions if they are individually less frequent).

Several approaches are used in treating missing data is research. The approaches include missing-indicator method, complete or available case analysis and imputation (single and multiple). The missing-indicator method is a method where missing data is replaced by dummy data or a new category representing the missing data [229]. Complete or available case analysis is when researchers exclude cases with missing data and imputation methods is when missing data is replaced by estimated data, either as single or multiple imputations [229] [230]. This study used the missing-indicator method because, in comparison to imputation methods, it is simpler, less computationally intensive and does not involve approximations [229] [231].

4.4.2 Determination of minimum sample size for the model

A minimum sample size was determined to ensure that the number of patients (n) and minimum number of outcome events per predictor satisfied the following conditions:
"(i) small optimism in predictor effect estimates as defined by a global shrinkage factor of ≥0.9, (ii) small absolute difference of ≤ 0.05 in the model's apparent and adjusted Nagelkerke's R^2, and (iii) precise estimation of the overall risk in the population." [232].

4.4.3 Multivariable analysis

A backward stepwise multivariable Cox regression model was used to derive estimate of the coefficients associated with multiple potential risk factors for recorded diagnosis of HIV and generate a risk score or index. Backward elimination, used in this study, is an automated algorithm used in selection of variables to include in a model [233] [234]. The algorithm starts by including all variables into a model and remove the insignificant variables one-by-one, using a p-value threshold [233]. The p-value threshold used for the backward stepwise elimination was 0.157, which could be used as a proxy for subset approaches with Akaike information criterion (AIC) [235] [236] [237]. Prior to multivariable analysis, an investigation was conducted to check which suitable multivariable analysis to use, including if the constant proportional hazards' assumption holds (that is, if the explanatory variable is a predictor of HIV, then the ratio of the hazards is the same at all times) [143]. Examination of the magnitude of violations using Kaplan-Meier curves was used to determine if the variables that violated the proportional hazard assumptions could be included in the model. The model allowed for non-linearity between HIV infection and some of the variables, such as age where the infection is less common under age 20, more common in middle age (20 to 39 years) and less common in older adults. From the Cox regression model, predictions were produced from. baseline survival combined with the linear predictors.

4.4.4 Internal validation and evaluation of model performance

Development of a model requires internal and external validation for calibration and reduction of optimism and overfitting of the model [217] [216]. External validation evaluates the predictive power of the derived model using an independent or external dataset [238] [239]. On the other hand, internal validation estimates how the model performs in a derivation dataset compared to that from a dataset of a similar population, which is either produced through bootstrap sampling or by randomly splitting the original cohort [238] [239]. Split-

sample validation or data splitting is an internal validation approach where the development dataset is divided into one dataset for model development and the other dataset for model validation [219]. However, this method is inefficient as it does not use the available dataset for model development, has two datasets which are closely similar (only vary by chance), resulting in optimistic results and in some cases, splitting of the data reduces sample sizes [219]. Other options of splitting the data, which could improve on the optimistic results include splitting the data by time (temporal validation) or location (geographic validation) [219] [240]. However, temporal and geographic validation also have the same weaknesses as the data splitting method [219] [240]. Bootstrap sampling is an approach that uses all the available data to develop the prediction model and generate bootstrap samples for validation [219]. Bootstrap validation accounts for model overfitting, thus quantifying optimism for the prediction model [219].

Internal validation was undertaken in this study using the resampling technique of random bootstrap samples with replacement [220] [221]. Bootstrap sampling was selected because in addition to a mechanism of quantifying optimism for the final model, it provided the shrinkage factor required to adjust the results from the development model. Stata was used to produce bootstrap samples by drawing subjects at random, with replacement from the original dataset (derivation dataset), meaning the bootstrap sample and model development dataset have same patients with some patients represented many times in the bootstrap samples [221] [220]. Internal validation was evaluated by applying the multivariable analysis from the derivation dataset to bootstrapped samples [220] [241] [242]. This was done by following the steps below:

i. Bootstrap (with replacement) the original dataset.
ii. Fit the multivariable model on each bootstrap and get the linear predictors.
iii. Calculate the Harrell's C-statistics and calibration performance slopes (C-slope) for the bootstraps and original data produced from the bootstrapped model.

iv. Calculate optimism for the Harrell's C-statistics and C-slope for each bootstrapped sample (for each bootstrap sample, find difference between the values in step (iii) above (bootstrap results minus original data results).

v. Using the mean of optimism values, calculate the optimism adjusted c-statistic and C-slope by subtracting mean optimism from Harrell's C-statistic or C-slope of the original multivariable model (before internal validation).

vi. Use the optimism adjusted calibration performance (C-slope) to produce uniform shrinkage values from the original model hazard ratio (log HR) multiplied by optimism adjusted C-slope).

The calibration and clinical usefulness of the model was evaluated by comparing results from the derivation dataset and the validation datasets [220]. The calibration plots showed comparison of results from all predicted risks generated in the validation and derivation datasets. Examination of heterogeneity in performance by general practices was conducted after fitting the model, using the Harrell's C-statistics [243]. The receiver operating curves showed the range of thresholds that quantifies the clinical usefulness of the prediction model [220].

Sensitivity, specificity and predictive values were calculated using several risk score cut off points to show which cut off points would give a better trade-off between sensitivity and specificity while taking into consideration of HIV prevalence in the population. Several cut-off points will be used to inform decision on considering the relative importance of sensitivity and specificity in respect to cost and benefit of testing patients for HIV in primary care [224]. Considering the 2017 prevalence of HIV in UK at 0.15% and that the optimum cut-off point from the case-control study of 0.26%, the cut-off points used in this study were 0.075%, 0.1%, 0.25% 0.5% and 1% [35] [244] [245]. Furthermore, decision curve analysis was conducted to evaluate the possible clinical utility or the benefits of a diagnostic test emanating from the use

of the prediction model [246]. A decision curve is a graphical presentation used to assess the clinical impact of a risk prediction model/biomarkers [247] [248]. Decision curve analysis is a method which evaluates the net benefits at varying levels of threshold predicted probabilities of an event [247] [249] [248] [250]. Net benefit is represented by the following equation:

$$\text{Net benefit} = \frac{True \quad Positives}{n} - \frac{False \quad positives}{n}\left(\frac{p_t}{1-p_t}\right)$$

where, (pt) is a threshold probability
n is the total sample size [249]

The graphic presentation from DCA has the horizontal axis which shows the risk threshold (R), defining high risk and the vertical axis showing the Net Benefit (NB). The lines on the chart are:

- green line for NB for the risk model risk.
- red lines for no intervention in the population (test none),
- blue line for NB when intervention is extended to the whole population.

Decision curve analysis is useful because it is

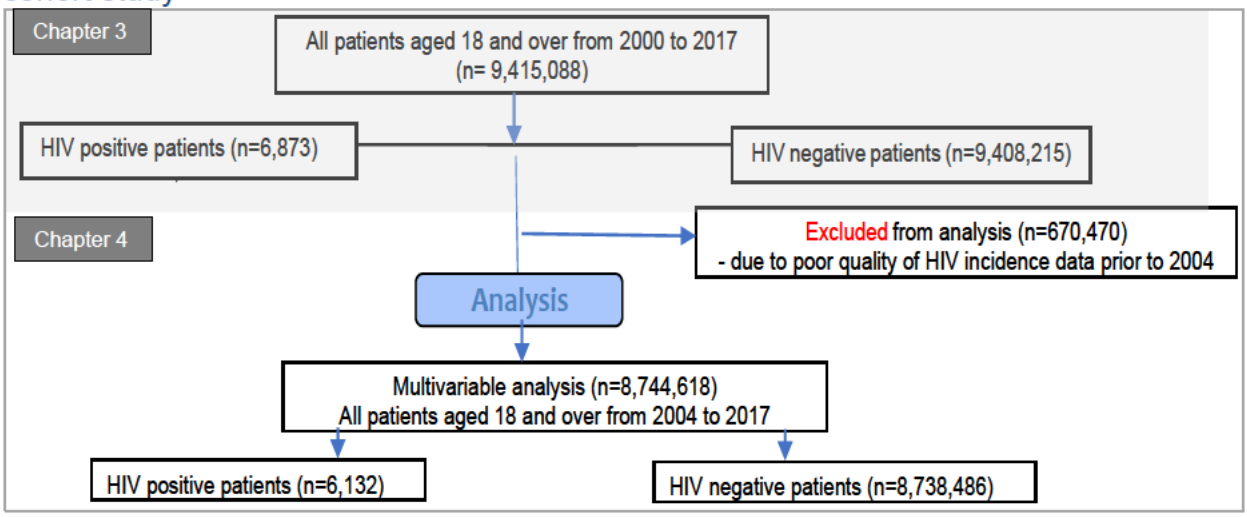
- easier to interpret whether the diagnostic test has a practical clinical value and area under the curve (AUC) may be misleading
- less complex than a full economic evaluation of the costs and benefits of positive and negative test results [248]

A further analysis was conducted to find a diagnostic interval, time between presentations of clinical and comorbid predictors and the recorded date of HIV diagnosis. This established whether using the model would be helpful in early diagnosis of HIV.

4.5 Results

Data for patients aged 18 and over, of both sexes and registered in IMRD between 2004 and 2017 were re-extracted from the IMRD database. The dataset included a total of 8,744,618 patients, 6,132 (70.1 per 100,000 population) confirmed HIV positive patients. Starting from the first data extracted to explore the outcome variable in chapter 2, the number of patients included at each stage are summarised in Figure 4.5-1.

Figure 4.5-1: Flowchart of the inclusion and exclusion of patients in the retrospective cohort study



4.5.1 Descriptive analysis and quality checks

4.5.1.1 Descriptive analysis of Predictor variables

The prevalence of most predictor variables was determined for 8,744,618 patients aged 18 and over, in IMRD databases for the years 2004 to 2017. The prevalence of predictors which are recorded at patient registration were determined from the date of first record.

A. Demographic and socio-economic predictors

Gender and age had complete records and deprivation quintiles had less than 20% missing data, and were all included in the study, Table 4.5-1. Ethnicity was poorly recorded in primary care data extracted from the IMRD database but was included for further explorative work to

check consistency in recording since some of the prevalence by individual sub-categories was 0.1% or more. Sexuality and area of origin (derived from country of birth) were excluded from this study due to a prevalence of less than 0.1% for most of the categories. Triangulation of predictor variables with proportion used in publicly available data formed part of further explorative work, including using census data or UKHSA data to check representativeness of the missing data.

Table 4.5-1: Demographic and socio-economic characteristics of patients aged 18 plus, IMRD database, 2004-2017.

Predictor	Total patients, n (%)	Percentage complete	HIV cases (rate/10,000)	Decision
Gender		100%		
Male	4219484 (48.25)		2931 (6.95)	Include
Female	4525134 (51.75)		3201 (7.07)	Include
Age (years)		100%		
18-24	1859445 (21.26)		1302 (7)	Include
25-34	1886717 (21.58)		2115 (11.21)	Include
35-49	2171913 (24.84)		2016 (9.28)	Include
50-59	1062294 (12.15)		473 (4.45)	Include
60+	1764249 (20.18)		226 (1.28)	Include
Ethnicity		43%		
Asian	266988 (3.05)		179 (6.7)	Include
Black	123818 (1.42)		727 (58.72)	Include
Mixed	43567 (0.5)		61 (14)	Include
Other	61545 (0.7)		86 (13.97)	Include
White	3264557 (37.33)		2007 (6.15)	Include
Not stated	4984143 (57)		3072 (6.16)	Include
Sexuality		0.2%		
Heterosexual	6686 (0.076)		2 (2.99)	Exclude <0.1%
Homosexual	2651 (0.03)		44 (165.98)	Exclude <0.1%
Bisexual	240 (0.003)		240 (10000)	Exclude <0.1%
Trans-sexualism	332 (0.004)		3 (90.36)	Exclude <0.1%
orientation not given - patient	247 (0.003)		247 (10000)	Exclude <0.1%
Other non-specific	7886 (0.09)		8 (10.14)	Exclude <0.1%
Not stated	8726576 (99.8)		5588 (6.4)	
Townsend deprivation quintile		82.8%		
1	1632635 (18.67)		652 (3.99)	Include
2	1498082 (17.13)		845 (5.64)	Include
3	1578757 (18.05)		1131 (7.16)	Include
4	1465381 (16.76)		1279 (8.73)	Include
5	1065129 (12.18)		1274 (11.96)	Include
Not stated	1504634 (17.21)		951 (6.32)	Include
		1.85%		
Area of origin				
Asia	35557 (0.41)		26 (7.31)	Exclude *
Other	109489 (1.25)		75 (6.85)	Exclude *
Sub-Saharan Africa	16264 (0.19)		106 (65.17)	Exclude *
Not stated	8583308 (98.15)		5925 (6.9)	Exclude *
Geographic				
Country / Region of residence (%)				
England: East Midlands	203502 (2.33)	100%	76 (3.73)	Exclude *
England: East of England	542653 (6.21)		387 (7.13)	Exclude *
England: London	1231905 (14.09)		1122 (9.11)	Exclude *
England: North East	151627 (1.73)		61 (4.02)	Exclude *

Predictor	Total patients, n (%)	Percentage complete	HIV cases (rate/10,000)	Decision
England: North West	713085 (8.15)		399 (5.6)	Exclude *
England: South Central	929642 (10.63)		594 (6.39)	Exclude *
England: South East Coast	857728 (9.81)		340 (3.96)	Exclude *
England: South West	634085 (7.25)		251 (3.96)	Exclude *
England: West Midlands	722108 (8.26)		395 (5.47)	Exclude *
England: Yorkshire & Humber	199142 (2.28)		52 (2.61)	Exclude *
Northern Ireland	309147 (3.54)		365 (11.81)	Exclude *
Scotland	1265636 (14.47)		1754 (13.86)	Exclude *
Wales	984358 (11.26)		336 (3.41)	Exclude *
Urban and rural				
Urban/rural				
Rural	1043604 (11.93)	68.24%	376 (3.6)	Include
Urban	4923903 (56.31%)		2895 (5.88)	Include
Not stated	2777111 (31.76%)		2861 (10.3)	Include
Urban/rural type				
Town & Fringe – Less sparse	631762 (7.22)	68.24%	221 (3.5)	Exclude *
Town & Fringe – Sparse	35841 (0.41)		3 (0.84)	Exclude *
Urban >10k - Less sparse	4911644 (56.17)		2893 (5.89)	Exclude *
Urban >10k – Sparse	12259 (0.14)		2 (1.63)	Exclude *
Village, Hamlet & Isolated dwelling – Less sparse	334979 (3.83)		131 (3.91)	Exclude *
Village, Hamlet & Isolated dwellings – Sparse	41022 (0.47)		21 (5.12)	Exclude *
Not stated	2777111 (31.76)		2861 (10.3)	Exclude *

*Excluded due to uneven distribution between categories

B. Geographic predictors

Data was extracted covering all countries in the United Kingdom with the majority of patients (72 percent) residing in England, of which 15 percent were from London, Table 4.5-1. No data was missing on country and regional (England only) distribution. However, region will be excluded from subsequent analysis given that the finding from HIV prevalence by region, from chapter 3, showed no regional differences in prevalence, except for high prevalence in London. The variable indicating whether the postal address was urban or rural was missing for 32 percent of patients, but it was included in subsequent analysis with a category for missing. There was an uneven distribution of population by urban/rural type with half of the types having less than 1 percent each compared to other types, hence the predictor was excluded from further analysis.

C. Lifestyle predictors

Lifestyle or behavioural predictors with prevalence of one percent or more that were included in consistency checks and subsequent analysis were: alcohol misuse, smoking, obesity, having a contact abroad and drug misuse, Table 4.5-2. Behavioural predictors with prevalence less than 0.1 percent such as unsafe sex, having anal sex, having multiple lifetime partners and partner characteristics of high risk were excluded from further analysis.

Table 4.5-2: Lifestyle characteristics of patients aged 18 plus, IMRD database, 2004-2017.

Predictor	Total patients, n (%)	HIV cases (rate per 10,000)	Decision on exclusion
Alcohol misuse	231912 (2.65)	286 (12.33)	Include
Drug misuse	153238 (1.75)	383 (24.99)	Include
Obesity	175048 (2.00)	53 (3.03)	Include
Smoking (current or ex-smoker)	3096282 (35.41)	2355 (7.61)	Include
Contact abroad	119493 (1.37)	187 (15.65)	Include
Stressful events*	600884 (6.87)	453 (7.54)	Include
Unsafe sex	1023 (0.01)	2 (19.55)	Exclude <0.1%
Anal sex	77 (0.00)	3 (389.61)	Exclude <0.1%
Multiple lifetime partners	5115 (0.06)	23 (44.97)	Exclude <0.1%
Partner characteristics	1575 (0.02)	7 (44.44)	Exclude <0.1%

*Stressful events refer to events such as bereavement or death of close friend or financial crisis or relationship breakdown

D. Clinical conditions and Co-morbid predictors

The prevalence of clinical and co-morbid predictors was determined for the years 2004 to 2017. A total of 62 clinical and co-morbid predictors were extracted, of which 57 covered the follow-up period and the other 5 referred to the first time a condition was registered irrespective of the index date, Table 4.5-3. There were 8 indicators with no outcome of HIV case and were not included in subsequent analysis. Of the 54 predictors with at least one outcome of HIV cases, 35 predictors were included in consistency checks and subsequent analysis as they met the inclusion criteria of frequency over 0.2% or were included on clinical grounds. Of these predictors, 21 had a prevalence in the entire patient population of more than 0.2% and 11 predictors were included on clinical grounds, because of the possibility of strong association with HIV. Clinical and co-morbid conditions combined into three new categories and included in subsequent analysis were:

- Chlamydia, gonorrhoea, genital herpes, other STIs and previous STI were combined to sexually Transmitted Infection (excluding Syphilis) and previous infections.
- Cervical dysplasia, cervical cancer and cervical intraepithelial neoplasia were grouped together as cervical conditions.
- Hepatitis B and Hepatitis C were grouped together as Hepatitis conditions.

Table 4.5-3: Clinical and co-morbid predictors of patients aged 18+, IMRD database, 2004-2017.

Condition group	Predictor	Total patients, n (%)	HIV cases (rate/10000)	Decision on exclusion
Respiratory	Aspergillosis	849 (0.01)	1 (11.78)	Exclude <0.2%
Conditions	Cough	2096297 (23.97)	1108 (5.29)	Include
	Pneumocystis carinii	12 (0.00)	1 (833.33)	Include on clinical grounds
	Pneumonia	56300 (0.64)	39 (6.93)	Include
	Tuberculosis	2548 (0.03)	8 (31.4)	Include on clinical grounds
Neurology	Cerebral toxoplasmosis abscess	425 (0.00)	4 (94.12)	Include on clinical grounds
	Cryptococcal meningitis	5 (0.00)	-	Exclude (HIV +ve = 0)
	Dementia	2709 (0.03)	1 (3.69)	Exclude <0.2%
	Guillain–Barré syndrome	980 (0.01)	2 (20.41)	Exclude <0.2%
	Progressive multifocal leukoencephalopathy	40 (0.00)	2 (500)	Include on clinical grounds
	Aseptic meningitis/encephalitis	1737 (0.02)	6 (34.54)	Include on clinical grounds
	Transverse myelitis	693 (0.01)	-	Exclude (HIV +ve = 0)
	Neurologic disability	15869 (0.18)	6 (3.78)	Include on clinical grounds
	Peripheral neuropathy	24789 (0.28)	10 (4.03)	Include
Ophthalmology	Cytomegalovirus retinitis	13142 (0.15)	9 (6.85)	Exclude <0.2%
	Any unexplained retinopathy	386 (0.00)	-	Exclude (HIV +ve = 0)
Dermatology	Herpes zoster	35611 (0.41)	36 (10.11)	Include on clinical grounds
	Psoriasis	158030 (1.81)	123 (7.78)	Include
	Seborrhoeic dermatitis	109699 (1.25)	85 (7.75)	Include
	Kaposi's sarcoma	60 (0.00)	9 (1500)	Include on clinical grounds
	Rash	1164037 (13.31)	935 (8.03)	Include
Gastroenterology	Abdominal pain	712366 (8.15)	405 (5.69)	Include
	Oral candidiasis	48877 (0.56)	62 (12.68)	Include on clinical grounds
	Cryptosporidiosis	416 (0.00)	1 (24.04)	Exclude <0.2%
	Diarrhoea	690671 (7.9)	495 (7.17)	Include
	Hepatitis B	4161 (0.05)	109 (261.96)	Combine hepatitis
	Hepatitis C	7752 (0.09)	48 (61.92)	Combine hepatitis
	Oral hairy leukoplakia	1622 (0.02)	3 (18.5)	Exclude <0.2%
	Nausea/vomiting	509022 (5.82)	386 (7.58)	Include
	Salmonella, shigella or campylobacter	21496 (0.25)	16 (7.44)	Include
	Weight loss	737632 (8.44)	531 (7.20)	Include
	Hepatitis conditions (combined)	11733 (0.13)	152 (129.55)	Include on clinical grounds
Haematology	Blood dyscrasia	28411 (0.32)	59 (20.77)	Include
Oncology	Non-Hodgkin's lymphoma	13244 (0.15)	50 (37.75)	Include on clinical grounds
	Anal cancer or anal intraepithelial dysplasia	993 (0.01)	3 (30.21)	Include on clinical grounds
	Castleman's disease	64 (0.00)	1 (156.25)	Exclude <0.2%
	Cervical dysplasia	3288 (0.04)	9 (27.37)	Combine cervical conditions
	Head and neck cancer	13459 (0.15)	3 (2.23)	Exclude <0.2%
	Hodgkin lymphoma	852 (0.01)	1 (11.74)	Exclude <0.2%
	Space-occupying lesion of unknown cause	135 (0.00)	-	Exclude (HIV +ve = 0)
	Lung cancer	35125 (0.4)	3 (0.85)	Include
	Cervical cancer	2619 (0.03)	2 (7.64)	Combine cervical conditions
Gynaecology	Cervical intraepithelial neoplasia	3891 (0.04)	8 (20.56)	Combine cervical conditions
	Seminoma	990 (0.01)	1 (10.1)	Exclude <0.2%
	Vaginal intraepithelial neoplasia	1019 (0.01)	-	Exclude (HIV +ve = 0)
	Cervical conditions	9463 (0.11)	17 (17.96)	Include on clinical grounds

Condition group	Predictor	Total patients, n (%)	HIV cases (rate/10000)	Decision on exclusion
Ear, nose, and throat	Parotitis	22148 (0.25)	14 (6.32)	Include
	Lymphoepithelial parotid cysts	967 (0.01)	-	Exclude (HIV +ve = 0)
Other Clinical conditions	Mononucleosis-like syndrome	41 (0)	-	Exclude (HIV +ve = 0)
	Pyrexia of unknown origin	298637 (3.42)	277 (9.28)	Include
	Hypertension	1433191 (16.39)	568 (3.96)	Include
	Chronic liver disease	80945 (0.93)	249 (30.76)	Include
	Hyperlipidemia	121514 (1.39)	21 (1.73)	Include
	Depression	1805500 (20.65)	1829 (10.13)	Include
	Lymphogranuloma venereum	11 (0)	-	Exclude (HIV +ve = 0)
	Lymphadenopathy	111373 (1.27)	150 (13.47)	Include
	Diabetes	518855 (5.93)	273 (5.26)	Include
	Minor trauma	369 (0)	-	Exclude (HIV +ve = 0)
	Sexually Transmitted Infections			
	Chlamydia	20947 (0.24)	74 (35.33)	Combine STIs
	Syphilis	818 (0.01)	18 (220.05)	Include on clinical grounds
	Gonorrhoea	1506 (0.02)	11 (73.04)	Combine STIs
	Genital herpes	17081 (0.2)	36 (21.08)	Combine STIs
	Other STIs (excludes above)	31432 (0.36)	53 (16.86)	Combine STIs
	STI	69873 (0.8)	183 (26.19)	Exclude (exclude syphilis)
	Previous STI	162435 (1.86)	207 (12.74)	Combine with STIs
	All STI (excluding Syphilis) and previous STIs	225125 (2.57)	364 (16.17)	Include

4.5.1.2 Quality checks of Predictor variables: Stability of predictor variables over time

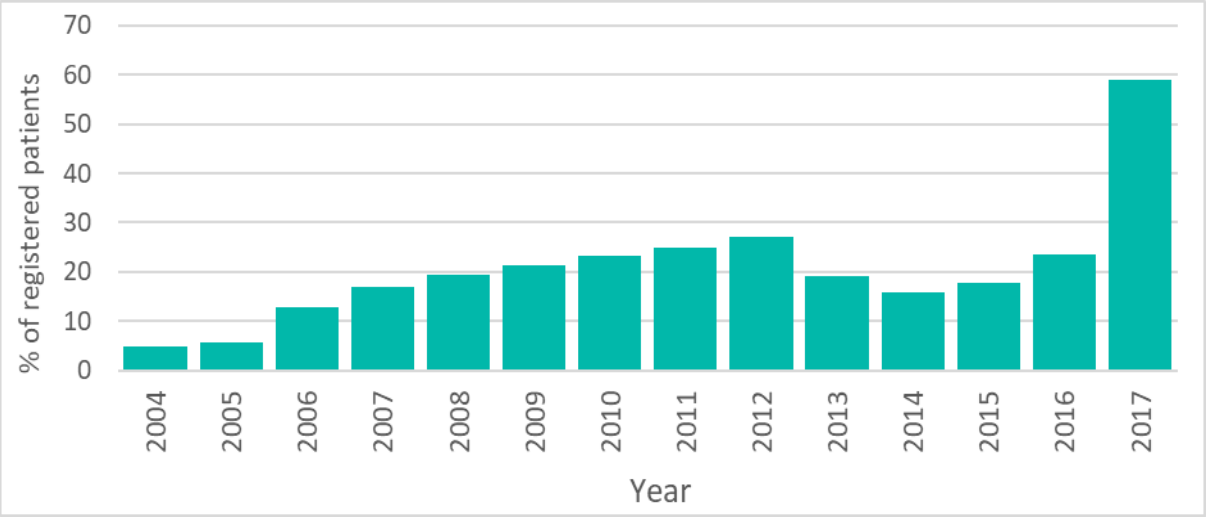
Consistency in the recording of predictor variables over time was determined through trend analysis of the variables from 2004 to 2017, to check if the quality of recording the variables might affect their predictive characteristics.

A. Demographic and Lifestyle predictors

A general increase in proportion of registered patients with recorded demographic and lifestyle predictors in the 1990s and 2000s could be a proportional increase in the predictor or an indication of a general improvement in recording of these characteristics at registration or could be a result of both factors, Appendix 10. The predictors that were further analysed for consistency checks were ethnicity and contact abroad.

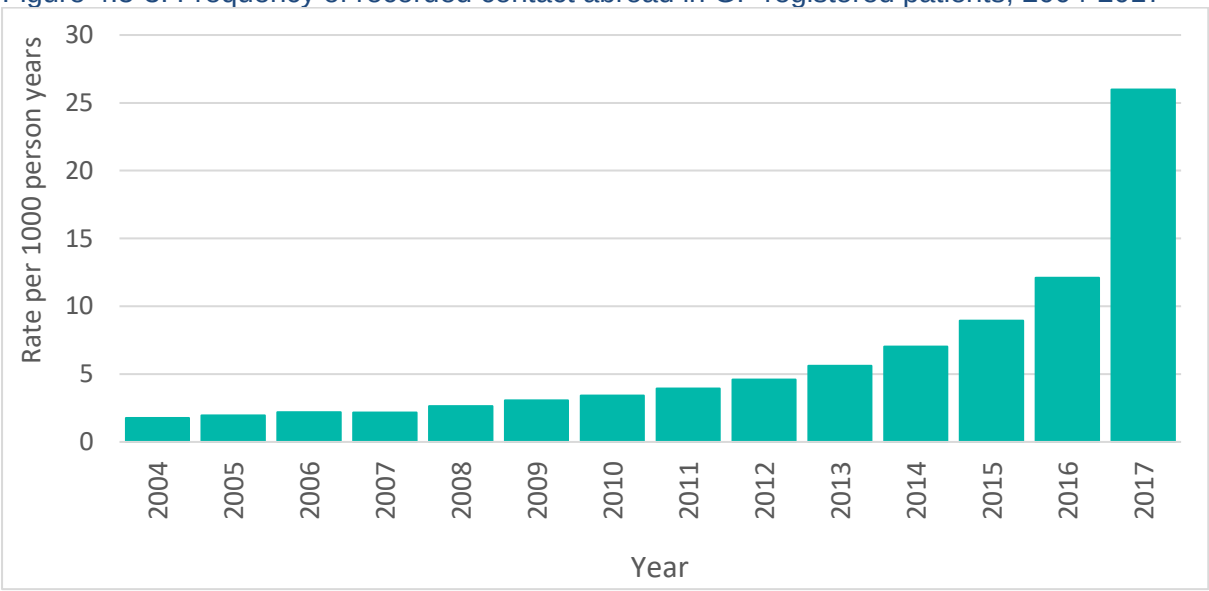
Although ethnicity was recorded for 40% of patients, the proportion of patients with recorded ethnicity consistently increased especially after 2006, hence ethnicity was included in subsequent analysis, Figure 4.5-2.

Figure 4.5-2: Proportion of GP registered patients with ethnicity recorded, 2004-2017



The proportion of those who had contact abroad was 1.37% but there was a consistent increase in frequency of the predictor over the period covered by this study, Figure 4.5-3. Therefore, contact abroad was included in subsequent analysis.

Figure 4.5-3: Frequency of recorded contact abroad in GP registered patients, 2004-2017



B. Clinical conditions and Co-morbid predictors

All clinical and co-morbid predictors which were included from the descriptive analysis were consistently recorded in primary care records and the incidence of these predictors showed some stability over the study period. Therefore, all the 35 clinical and co-morbid predictors were included in subsequent analysis.

4.5.1.3 Implications of the analysis of predictor variables for subsequent analysis

The descriptive analysis and the data quality checks for the predictor variables revealed that the 38 predictor variables will be dropped from subsequent analysis due to population prevalence of 0.2% or less, instability in recording or no HIV positive patient recorded, Appendix 11. All the other variables were included in subsequent analysis.

Predictors listed below were reclassified and included in the multivariable analysis, and the rest of the predictor variables were all included as they were in descriptive analysis.

- Ethnicity; white/Asian and missing ethnicity (reference) versus black and mixed/other. Reclassification was based on nationally published figures which showed that black ethnic group and the mixed/other ethnic groups have higher incidence of HIV compared to white and Asian ethnic groups [67].
- Urban/rural; rural (reference) versus urban and missing.
- Smoking; never smoked (reference) and current smoker/ex-smoker

Of the 81 predictors, 47 of them were included in subsequent analysis after exclusion of 27 individual predictors due to low frequency in population or low HIV prevalence and also after combining seven predictors into three predictors. All predictor variables were categorical and no transformations were done.

4.5.2 Results from determination of minimum sample size and testing Proportional Hazard Assumptions

4.5.2.1 Minimum sample size for the model

Using the formula provided by Riley et al (2019), the minimum sample size required to satisfy the criteria specified in section 4.5.2 was 211,265 with minimum outcome events of 1,230. These figures assumed an events per predictor value of 26.16. This analysis confirmed that the number of patients included for model development (8,744,618 with 6,132 HIV positive patients) were adequate.

4.5.2.2 Testing Proportional Hazard Assumptions

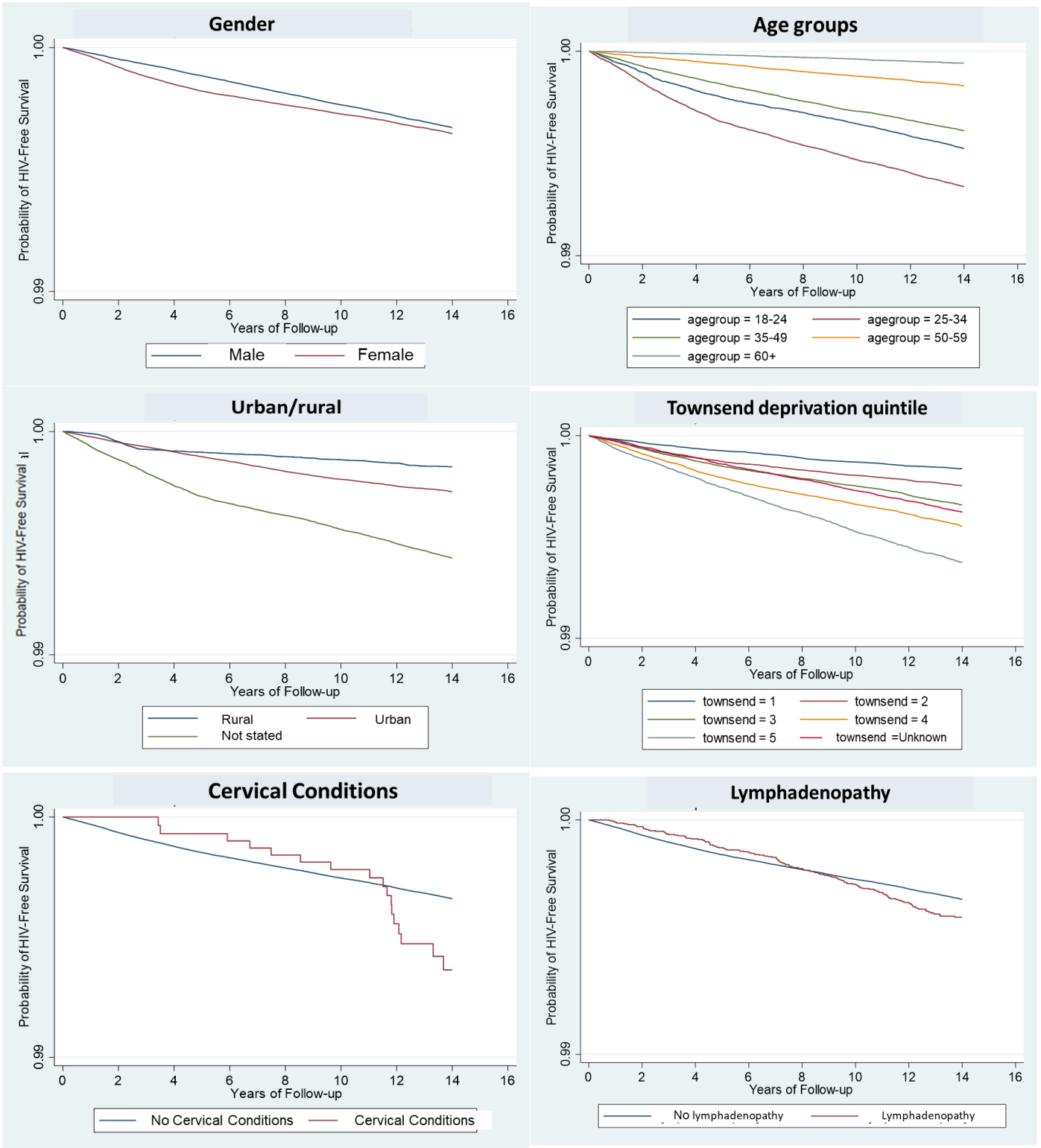
The results from testing of proportional hazard assumptions, using the scaled Schoenfeld residuals, revealed that 13 out of 45 predictors violated the assumptions, shown by a significant p-value, Table 4.5-4. These predictor variables were gender, age groups, Townsend deprivation quintile, urban/rural, abdominal pain, cervical conditions, cough, diarrhoea, nausea/vomiting, pyrexia of unknown origin, rash, weight loss and lymphadenopathy.

Further examination of the magnitude of violations of proportional hazard assumptions was conducted using Kaplan-Meier curves, Figure 4.5 4 (the rest of the Kaplan-Meier curves are in Appendix 12). Hazards can be regarded as proportional if the line for each variable group does not cross each other. The Kaplan-Meier curves showed that all the variables that violated the proportional hazard assumptions using the p-values had proportional hazard except cervical conditions and lymphadenopathy. Although these two variables violated the assumptions, they were included in the multivariable analysis since they were binary.

Table 4.5-4: Results from testing of proportional hazard assumptions

		Predictor	rho	chi^2	df	P-value
Socio-Demographic and geographic	1.	Gender	-0.140	123	1	0.000
	2.	Age group	0.122	65.28	1	0.000
	3.	Ethnic group	-0.032	4.41	1	0.036
	4.	Townsend deprivation quintile	0.127	100.56	1	0.000
	5.	urban	-0.095	57.15	1	0.000
Lifestyle	6.	Alcohol misuse	0.014	1.14	1	0.286
	7.	Drug misuse	0.027	4.32	1	0.038
	8.	Contact abroad	-0.026	4.14	1	0.042
	9.	Smoking status	0.003	0.07	1	0.792
	10.	Stressful events	-0.026	4.11	1	0.043
Clinical and comorbid conditions	11.	Abdominal pain	0.091	51.09	1	0.000
	12.	Anal cancer or anal intraepithelial dysplasia	0.019	2.17	1	0.140
	13.	Aseptic meningitis/encephalitis	0.011	0.72	1	0.398
	14.	Blood dyscrasia	0.026	4.03	1	0.045
	15.	Cerebral toxoplasmosis abscess	-0.010	0.65	1	0.421
	16.	Cervical conditions	0.051	15.95	1	0.000
	17.	Chronic liver disease	-0.012	0.92	1	0.337
	18.	Cough	0.171	182.52	1	0.000
	19.	Cryptosporidiosis	-0.006	0.22	1	0.638
	20.	Cytomegalovirus retinitis	0.032	6.23	1	0.013
	21.	Depression	-0.005	0.17	1	0.684
	22.	Diabetes	0.019	2.31	1	0.128
	23.	Diarrhoea	0.122	93.32	1	0.000
	24.	Head and neck cancer	0.003	0.07	1	0.794
	25.	Hepatitis (B&C)	0.042	10.84	1	0.001
	26.	Herpes zoster	0.015	1.46	1	0.227
	27.	Hyperlipidemia	0.021	2.58	1	0.109
	28.	Hypertension	0.043	11.15	1	0.001
	29.	Kaposi's sarcoma	-0.001	0	1	0.953
	30.	Lung cancer	0.026	4.07	1	0.044
	31.	Lymphadenopathy	0.057	19.92	1	0.000
	32.	Nausea/vomiting	0.057	19.96	1	0.000
	33.	Neurologic disability	0.006	0.24	1	0.624
	34.	Non-Hodgkin's lymphoma	0.016	1.58	1	0.209
	35.	Oral candidiasis	0.023	3.37	1	0.067
	36.	Peripheral neuropathy	0.022	3.02	1	0.082
	37.	Pneumocystis carinii	0.009	0.46	1	0.497
	38.	Pneumonia	0.028	4.84	1	0.028
	39.	Progressive multifocal leukoencephalopathy	0.007	0.33	1	0.568
	40.	Pyrexia of unknown origin	0.071	31.03	1	0.000
	41.	Rash	0.151	140.49	1	0.000
	42.	STI (excludes syphilis) or previous STI	0.004	0.12	1	0.727
	43.	Syphilis	0.013	1.02	1	0.312
	44.	Tuberculosis	-0.014	1.17	1	0.280
	45.	Weight loss	0.145	129.26	1	0.000

Figure 4.5-4: Kaplan-Meier curves for gender, age groups, urban/rural, deprivation, cervical conditions and lymphadenopathy.



4.5.3 Results from the Multivariable model estimation

This section will reveal the results from the multivariable analysis and evaluation of the model.

4.5.3.1 Multivariable (Cox regression) Analysis results

The demographic predictors significantly associated with HIV infection were female gender (HR: 1.19, 95% CI: 1.13-1.25) compared to males, 25-34 age group (HR: 1.3, 95% CI: 1.21-1.39, p-value < 0.001) compared to 18-24 and ethnicity where black ethnicity (HR: 11.22, 95% CI: 10.32 – 12.19, p-value < 0.001) and Mixed and other ethnicity (HR: 2.57, 95% CI: 2.18 – 3.03, p-value < 0.001) compared to reference group of white and Asian.

Deprivation was significantly associated with HIV infection with HR increasing with deprivation levels with most deprived having HR of 1.86 (95% CI: 1.69-2.06, p < 0.001). Living in urban areas was a geographical predictor associated with HIV infection (HR: 1.12, 95% CI: 1.0-1.25, p-value < 0.001).

Lifestyle predictors with significant associated with HIV infection were smoking status of being a current smoker/ex-smoker (HR: 1.14, 95% CI: 1.07-1.21, p-value < 0.001) compared to never smokers, drug misuse (HR: 2.27, 95% CI: 2.02 - 2.54, p-value < 0.001) and contact abroad (HR: 2.05, 95% CI: 1.77 - 2.38, p-value < 0.001)

There were 17 clinical and comorbid conditions significantly associated with HIV infection, the largest hazard ratios were on Kaposi's sarcoma (HR: 180.02, 95% CI: 93.14 – 347.93, p-value < 0.001), pneumocystis carinii (HR: 74.25, 95% CI: 10.32 – 534.14, p-value < 0.001), progressive multifocal leukoencephalopathy (HR: 58.19, 95% CI: 14.54 – 232.87, p-value < 0.001), syphilis (HR: 11.14, 95% CI: 6.99 – 17.76, p-value < 0.001), Non-Hodgkin's lymphoma (HR: 9.52, 95% CI: 7.18 – 12.64, p-value < 0.001), tuberculosis (HR: 2.31, 95% CI: 1.15 -

4.62, p-value < 0.001) and cerebral toxoplasmosis abscess (HR: 8.04, 95% CI: 3.01 – 21.48, p-value < 0.001), Table 4.5-5.

The other clinical and comorbid conditions associated with HIV infection were anal cancer or anal intraepithelial dysplasia, aseptic meningitis/encephalitis, oral candidiasis, hepatitis B and C, blood dyscrasia, chronic liver disease, depression and current STI (excluding syphilis) or any previous STI.

Rash, pyrexia of unknown origin, weight loss, hypertension, nausea/vomiting, abdominal pain peripheral neuropathy, cough, neurologic disability, lung cancer and hyperlipidemia were statistically significant but were associated with decreased risk of HIV infection.

The predictors not significantly associated with HIV infection, after adjusting for other factors were alcohol misuse, cervical conditions (combined cervical cancer, cervical dysplasia and cervical intraepithelial neoplasia), cryptosporidiosis and stressful events.

Table 4.5-5: Multivariable analysis of relationship between predictors and HIV

	Predictor	Hazard Ratio	[95% Confidence Interval]	p-value
Socio-economic, demographic and geographic	Gender			
	Male	1 (reference)		
	Female	1.19	(1.13, 1.25)	<0.001
	Age group			
	18-24	1 (reference)		
	25-34	1.3	(1.21, 1.39)	<0.001
	35-49	0.84	(0.78, 0.9)	<0.001
	50-59	0.43	(0.39, 0.48)	<0.001
	60+	0.16	(0.14, 0.19)	<0.001
	Ethnicity			
	White/Asian/missing	1 (reference)		
	Black	11.22	(10.32, 12.19)	<0.001
	Mixed and other	2.57	(2.18, 3.03)	<0.001
	Deprivation			
	1	1 (reference)		
	2	1.3	(1.17, 1.44)	<0.001
	3	1.55	(1.4, 1.71)	<0.001
	4	1.74	(1.58, 1.91)	<0.001
	5	1.86	(1.69, 2.06)	<0.001
	Not stated	0.72	(0.64, 0.8)	<0.001
	Urban/rural			
	Rural	1 (reference)		
	Urban	1.12	(1, 1.25)	0.045
	Missing	3.19	(2.85, 3.58)	<0.001

	Predictor	Hazard Ratio	[95% Confidence Interval]	p-value
Lifestyle	Drug misuse	2.27	(2.02, 2.54)	<0.001
	Contact abroad	2.05	(1.77, 2.38)	<0.001
	Smoking status (current smoker/ex-smoker)	1.01	(1.01, 1.02)	<0.001
Clinical and comorbid conditions	Kaposi's sarcoma	180.02	(93.14, 347.93)	<0.001
	Pneumocystis carinii	74.25	(10.32, 534.14)	<0.001
	Progressive multifocal leukoencephalopathy	58.19	(14.54, 232.87)	<0.001
	Syphilis	11.14	(6.99, 17.76)	<0.001
	Non-Hodgkin's lymphoma	9.52	(7.18, 12.64)	<0.001
	Cerebral toxoplasmosis abscess	8.04	(3.01, 21.48)	<0.001
	Anal cancer or anal intraepithelial dysplasia	5.23	(1.69, 16.23)	0.004
	Hepatitis	4.62	(3.87, 5.51)	<0.001
	Chronic liver disease	3.69	(3.22, 4.22)	<0.001
	Aseptic meningitis/encephalitis	3.3	(1.48, 7.34)	0.004
	Tuberculosis	2.31	(1.15, 4.62)	0.018
	Blood dyscrasia	1.94	(1.49, 2.51)	<0.001
	Oral candidiasis	1.81	(1.4, 2.33)	<0.001
	STI (excludes syphilis) or previous STI	1.6	(1.44, 1.78)	<0.001
	Herpes zoster	1.45	(1.04, 2.01)	0.027
	Depression	1.43	(1.35, 1.52)	<0.001
	Pneumonia	1.27	(0.93, 1.75)	0.139
	Lymphadenopathy	1.22	(1.03, 1.43)	0.02
	Diabetes	1.1	(0.96, 1.25)	0.157
	Diarrhoea	0.89	(0.81, 0.98)	0.022
	Pyrexia of unknown origin	0.89	(0.79, 1)	0.06
	Hypertension	0.87	(0.79, 0.96)	0.005
	Rash	0.86	(0.8, 0.92)	<0.001
	Weight loss	0.84	(0.77, 0.92)	<0.001
	Nausea/vomiting	0.83	(0.74, 0.92)	0.001
	Abdominal pain	0.54	(0.49, 0.6)	<0.001
	Peripheral neuropathy	0.51	(0.28, 0.96)	0.036
	Cough	0.5	(0.47, 0.53)	<0.001
	Neurologic disability	0.44	(0.2, 0.99)	0.046
	Lung cancer	0.29	(0.09, 0.91)	0.035
	Hyperlipidemia	0.24	(0.16, 0.37)	<0.001

The model had a C-statistic of 0.743 (95% CI: 0.736, 0.749) and a calibration performance slope (C-slope) of 1.000 before validation. The C-slope of one showed that the expected and observed values from the model were equal.

There were four practices with unusually high incidence rates of HIV (two general practices in Scotland, one in Northern Ireland and one in East of England), shown in chapter 3. Using all the variables used on all the practices in UK, the multivariable analysis was re-run on the dataset without the four outlier practices. The demographic predictors that were no longer significantly associated with HIV infection with outlier practices excluded were gender and deprivation level 2, Appendix 13. However, age group of 35-49 years was significantly associated with HIV (HR: 1.46, 95% CI: 1.33 - 1.59). Smoking status and contact abroad were

no longer significantly associated with HIV once the outlier practices were excluded in the model.

There were 18 clinical and comorbid conditions significantly associated with HIV infection after exclusion of outlier practices. The largest hazard ratios were still on Kaposi's sarcoma (HR: 141.13, 95% 95% CI: 72.75 - 273.78, p-value < 0.001), pneumocystis carinii (HR: 73.23, 95% CI: 10.16 - 527.66, p-value < 0.001), progressive multifocal leukoencephalopathy (HR: 65.95, 95% CI: 16.47 - 264.16, p-value < 0.001) and syphilis (11.68, 95% CI: 7.3 - 18.69, p-value < 0.001). The additional comorbid condition which was significantly associated with HIV was pneumonia (HR: 1.5 9% CI: 1.09 - 2.06, p-value < 0.001). The C-statistics for the model without the 4 outlier practices was 0.753 (95% CI: 0.746 - 0.760).

4.5.4 Results from Internal Validation

4.5.4.1 Bootstrapping results

Bootstrapping of 250 samples with replacement was conducted on the extracted IMRD dataset. For each bootstrap, optimism for the C-statistic and a calibration performance slope (C-slope) were calculated. The mean optimism for the C-statistic and C-slope were 0.00048 and 0.0099, respectively. The optimism adjusted C-statistic was 0.746 (95% CI: 0.736 - 0.749) and the optimism adjusted C-slope was 0.990.

After applying the uniform shrinkage to the original hazard ratio, the predictors which were significantly associated with an increased risk of HIV infection in the original (non-bootstrapped model) were the same predictors with HR greater than 1 from the bootstrapped model, Table 4.5-6. For example, the HR still remained significant and slightly lower for black ethnicity, from 11.22 (95% CI: 10.32 - 12.19) to 10.95 (95% CI: 10.08 - 11.89, p-value < 0.001) and for pneumocystis carinii, from 74.25 (95% CI: 10.32 - 534.14, p-value < 0.001) to 71.15 (95% CI: 10.09 - 501.98, p-value < 0.001).

Table 4.5-6: Optimism adjusted Hazard ratios and 95% Confidence Intervals

	Predictor	Hazard Ratio	Lower 95% Confidence Interval	Upper 95% Confidence Interval
Socio-economic, demographic and geographic	Gender			
	Males	1 (reference)		
	Females	1.19	1.13	1.25
	Age group			
	18-24	1 (reference)		
	25-34	1.29	1.21	1.39
	35-49	0.84	0.78	0.90
	50-59	0.44	0.39	0.49
	60+	0.16	0.14	0.19
	Ethnicity			
	White/Asian/missing	1 (reference)		
	Black	10.95	10.08	11.89
	Mixed and other	2.55	2.16	3.00
	Deprivation			
	1	1 (reference)		
	2	1.30	1.17	1.44
	3	1.54	1.40	1.70
	4	1.73	1.57	1.90
	5	1.85	1.68	2.04
	Not stated	0.72	0.64	0.80
	Urban/rural			
	Rural	1 (reference)		
	Urban	1.12	1.00	1.25
	Missing	3.15	2.82	3.53
Lifestyle	Drug misuse	2.25	2.01	2.52
	Contact abroad	2.04	1.76	2.36
	Smoking status (current smoker/ex-smoker)	1.01	1.01	1.02
Clinical and comorbid conditions	Kaposi's sarcoma	171.01	89.06	328.37
	Pneumocystis carinii	71.15	10.09	501.98
	Progressive multifocal leukoencephalopathy	55.89	14.16	220.66
	Syphilis	10.88	6.86	17.27
	Non-Hodgkin's lymphoma	9.31	7.04	12.33
	Cerebral toxoplasmosis abscess	7.88	2.98	20.84
	Anal cancer or anal intraepithelial dysplasia	5.15	1.68	15.79
	Hepatitis	4.55	3.82	5.42
	Chronic liver disease	3.64	3.18	4.17
	Aseptic meningitis/encephalitis	3.26	1.47	7.20
	Tuberculosis	2.29	1.15	4.55
	Blood dyscrasia	1.92	1.49	2.49
	Oral candidiasis	1.80	1.40	2.31
	STI (excludes syphilis) or previous STI	1.60	1.43	1.77
	Herpes zoster	1.44	1.04	2.00
	Depression	1.43	1.35	1.51
	Pneumonia	1.27	0.93	1.74
	Lymphadenopathy	1.21	1.03	1.43
	Diabetes	1.10	0.97	1.25
	Diarrhoea	0.90	0.82	0.98
	Pyrexia of unknown origin	0.89	0.79	1.00
	Hypertension	0.87	0.79	0.96
	Rash	0.86	0.80	0.92
	Weight loss	0.84	0.77	0.92
	Nausea/vomiting	0.83	0.74	0.92
	Abdominal pain	0.54	0.49	0.60
	Peripheral neuropathy	0.52	0.28	0.96
	Cough	0.50	0.47	0.54
	Neurologic disability	0.45	0.20	0.99
	Lung cancer	0.30	0.10	0.92
	Hyperlipidemia	0.25	0.16	0.38

The baseline survival for the optimism adjusted model at 1 year was 0.999895, at 5 years was 0.9995294 and 0.9992635 at 10 years.

The final equation from the prediction model (at 1 year) was:

h(t=1) = 0.999895 × exp(1.19 (Females) + 1.29 (Age group 25-34) + 10.95 (Black) + 2.55 (Mixed and other) + 1.3 (Deprivation 2) + 1.54 (Deprivation 3) + 1.73 (Deprivation 4) + 1.85 (Deprivation 5) + 1.12 (Urban) + 3.15 (Missing Urban/rural) + 2.25 (Drug misuse) + 2.04 (Contact abroad) + 1.01 (Smoking status (current smoker/ex-smoker)) + 171.01 (Kaposi's sarcoma) + 71.15 (Pneumocystis carinii) + 55.89 (Progressive multifocal leukoencephalopathy) + 10.88 (Syphilis) + 9.31 (Non-Hodgkin's lymphoma) + 7.88 (Cerebral toxoplasmosis abscess) + 5.15 (Anal cancer or anal intraepithelial dysplasia) + 4.55 (Hepatitis) + 3.64 (Chronic liver disease) + 3.26 (Aseptic meningitis/encephalitis) + 2.29 (Tuberculosis) + 1.92 (Blood dyscrasia) + 1.8 (Oral candidiasis) + 1.6 (STI (excludes syphilis) or previous STI) + 1.44 (Herpes zoster) + 1.43 (Depression) + 1.21 (Lymphadenopathy) + 1.27 (Pneumonia) + 1.1 (Diabetes) + 0.9 (Diarrhoea) + 0.89 (Pyrexia of unknown origin) + 0.87 (Hypertension) + 0.86 (Rash) + 0.84 (Weight loss) + 0.83 (Nausea/vomiting) + 0.54 (Abdominal pain) + 0.52 (Peripheral neuropathy) + 0.5 (Cough) + 0.45 (Neurologic disability) + 0.3 (Lung cancer) + 0.25 (Hyperlipidemia))

The final equation from the prediction model (at 5 years) was:

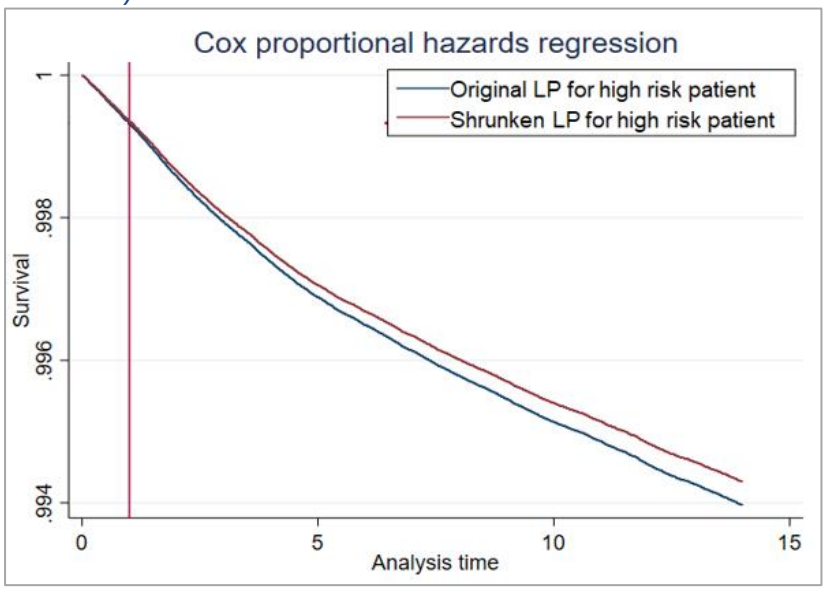
h(t=5) = 0.9995294 × exp(1.19 (Females) + 1.29 (Age group 25-34) + 10.95 (Black) + 2.55 (Mixed and other) + 1.3 (Deprivation 2) + 1.54 (Deprivation 3) + 1.73 (Deprivation 4) + 1.85 (Deprivation 5) + 1.12 (Urban) + 3.15 (Missing Urban/rural) + 2.25 (Drug misuse) + 2.04 (Contact abroad) + 1.01 (Smoking status (current smoker/ex-smoker)) + 171.01 (Kaposi's sarcoma) + 71.15 (Pneumocystis carinii) + 55.89 (Progressive multifocal leukoencephalopathy) + 10.88 (Syphilis) + 9.31 (Non-Hodgkin's lymphoma) + 7.88 (Cerebral toxoplasmosis abscess) + 5.15 (Anal cancer or anal intraepithelial dysplasia) + 4.55 (Hepatitis) + 3.64 (Chronic liver disease) + 3.26 (Aseptic meningitis/encephalitis) + 2.29 (Tuberculosis) + 1.92 (Blood dyscrasia) + 1.8 (Oral candidiasis) + 1.6 (STI (excludes syphilis) or previous STI) + 1.44 (Herpes zoster) + 1.43 (Depression) + 1.21 (Lymphadenopathy) + 1.27 (Pneumonia) + 1.1 (Diabetes) + 0.9 (Diarrhoea) + 0.89 (Pyrexia of unknown origin) + 0.87 (Hypertension) + 0.86 (Rash) + 0.84 (Weight loss) + 0.83 (Nausea/vomiting) + 0.54 (Abdominal pain) + 0.52 (Peripheral neuropathy) + 0.5 (Cough) + 0.45 (Neurologic disability) + 0.3 (Lung cancer) + 0.25 (Hyperlipidaemia))

The final equation from the prediction model (at 10 years) was:

h(t=5) = 0.9992635 × exp(1.19 (Females) + 1.29 (Age group 25-34) + 10.95 (Black) + 2.55 (Mixed and other) + 1.3 (Deprivation 2) + 1.54 (Deprivation 3) + 1.73 (Deprivation 4) + 1.85 (Deprivation 5) + 1.12 (Urban) + 3.15 (Missing Urban/rural) + 2.25 (Drug misuse) + 2.04 (Contact abroad) + 1.01 (Smoking status (current smoker/ex-smoker)) + 171.01 (Kaposi's sarcoma) + 71.15 (Pneumocystis carinii) + 55.89 (Progressive multifocal leukoencephalopathy) + 10.88 (Syphilis) + 9.31 (Non-Hodgkin's lymphoma) + 7.88 (Cerebral toxoplasmosis abscess) + 5.15 (Anal cancer or anal intraepithelial dysplasia) + 4.55 (Hepatitis) + 3.64 (Chronic liver disease) + 3.26 (Aseptic meningitis/encephalitis) + 2.29 (Tuberculosis) + 1.92 (Blood dyscrasia) + 1.8 (Oral candidiasis) + 1.6 (STI (excludes syphilis) or previous STI) + 1.44 (Herpes zoster) + 1.43 (Depression) + 1.21 (Lymphadenopathy) + 1.27 (Pneumonia) + 1.1 (Diabetes) + 0.9 (Diarrhoea) + 0.89 (Pyrexia of unknown origin) + 0.87 (Hypertension) + 0.86 (Rash) + 0.84 (Weight loss) + 0.83 (Nausea/vomiting) + 0.54 (Abdominal pain) + 0.52 (Peripheral neuropathy) + 0.5 (Cough) + 0.45 (Neurologic disability) + 0.3 (Lung cancer) + 0.25 (Hyperlipidaemia))

The closeness of the two models was shown by plotting survival curves of a patient with high risk using the two models (original likelihood probability (lp) and the shrunken lp) on lines corresponding to the patient's survival probability at 1 year, Figure 4.5-5.

Figure 4.5-5 Survival curves of a patient with high risk using the two models (original and shrunken)

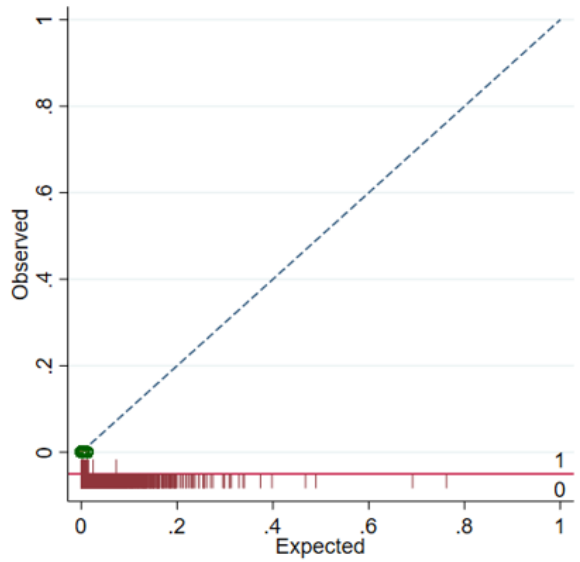


Calibration plots were produced at 1 year, 5 years and 10 years using *pmcalplot* method written by Ensor et al (2018) [251]. The calibration plots showed that the model overpredicts risk a little, Figure 4.5-6. The plots also show outliers with high predictions. This was the same pattern shown by the 5-years and 10-years calibration plots.

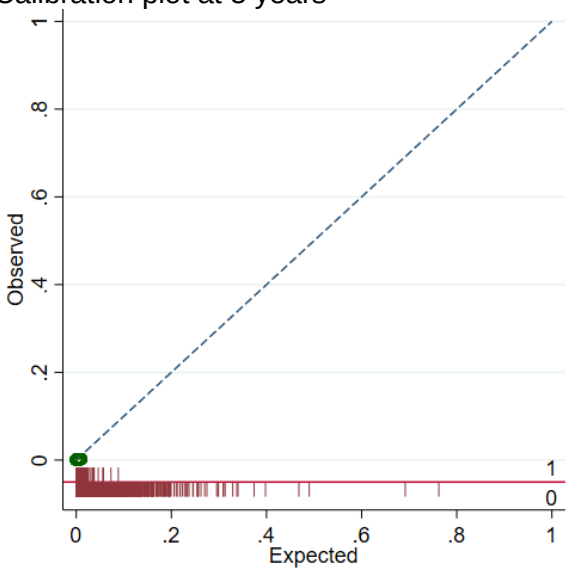
Further analysis on the significant clinical and comorbid predictors established that predictors recorded within the year prior to HIV diagnosis were pneumocystis carinii, progressive multifocal leukoencephalopathy, cerebral toxoplasmosis abscess, anal cancer or anal intraepithelial dysplasia, aseptic meningitis/encephalitis, and blood dyscrasia, Appendix 14

Figure 4.5-6: Calibration plots at 1 year, 5 years and 10 years

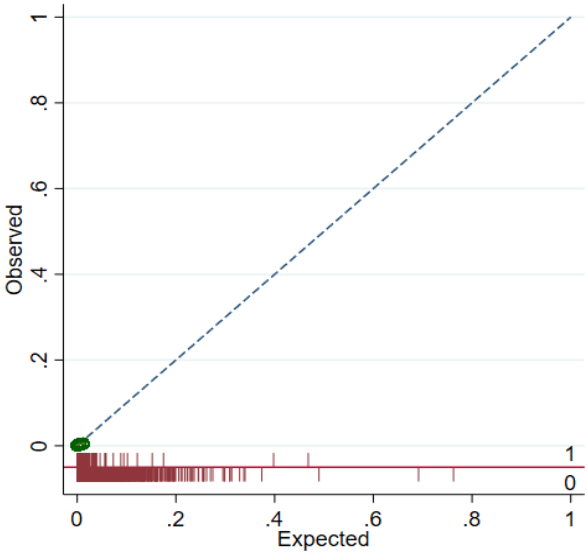
Calibration plot at 1 year

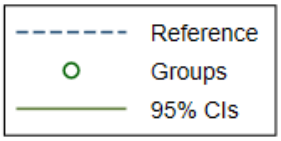


Calibration plot at 5 years



Calibration plot at 10 years

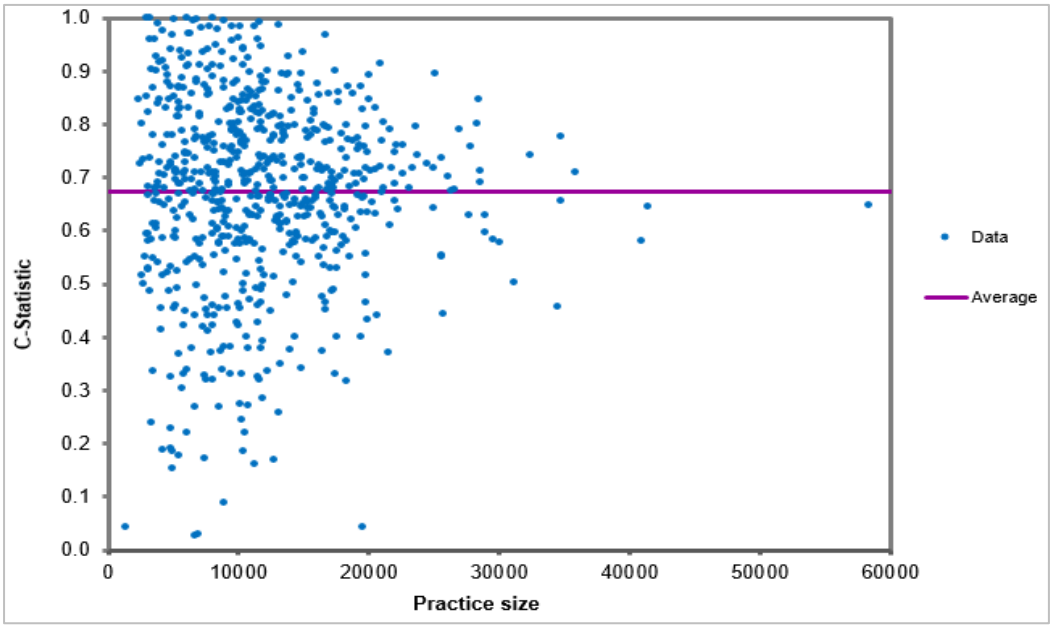




4.5.4.2 Model checks of heterogeneity in performance

Examining the model performance across the 784 practices in IMRD database revealed a wider variation of C-statistics in smaller practices compared to large practices, Figure 4.5-7. The practices with no prevalence of HIV (103 practices, 13%) were excluded from this analysis and had a C-statistic of 0.5. The analysis showed that practices with a patient register under 10,000 (2004-2017) had a variation of C-statistics from almost 0 to 1.00. However, the variation of C-statistics tapered towards 0.60 as practice population increases.

Figure 4.5-7: C-statistics across practices in IMRD dataset, 2004-2017



4.5.5 Further evaluation of model performance

Sensitivity analyses and diagnosis interval were carried out to further evaluate the performance of the model.

4.5.5.1 Sensitivity analysis

The sensitivity analyses were conducted for different geographies considering the findings from incidence of HIV by region or country. For each geography, the values were calculated at 5 cut off points (0.075%, 0.1%, 0.25%, 0.5% and 1%). More results for different geographic breakdowns are included in Appendix 15.

Using the whole dataset (all practices included), the values at 0.075% cut-off points for the model were: a sensitivity of 70%, a specificity of 43%, PPV of 0.09% and NPV of 99.95%, Table 4.5-7. The values for sensitivity and negative predictive value (NPV) decrease with increase in cut-off points and vice versa for specificity and positive predictive value (PPV).

Table 4.5-7: Sensitivity, specificity and predictive values at selected cut-off points for all UK practices in IMRD, 2004-2017.

	Cut-off points				
	0.075%	0.1%	0.25%	0.5%	1%
Sensitivity (%)	70%	62%	37%	21%	10%
Specificity (%)	43%	52%	84%	94%	98%
PPV (%)	0.09%	0.09%	0.16%	0.27%	0.45%
NPV (%)	99.95%	99.95%	99.95%	99.94%	99.94%
Total	8744618	8744618	8744618	8744618	8744618
With HIV	6132	6132	6132	6132	6132
Likelihood Ratio +	1.2	1.3	2.3	3.8	6.4
Likelihood Ratio -	0.7	0.7	0.8	0.8	0.9

A. UK practices, excluding 4 Outlier Practices.

There were four practices with unusually high incidence rates of HIV (two general practices in Scotland, one in Northern Ireland and one in East of England). The four practices were dropped in the second sensitivity analysis with the outlier practices excluded. The values at 0.075% cut-off points for the model were: a sensitivity of 60%, a specificity of 54%, PPV of 0.07% and NPV of 99.96%, Table 4.5-8.

Table 4.5-8. Sensitivity, specificity and predictive values for UK excluding four Outlier Practices at selected cut-off points in IMRD, 2004-2017.

	Cut-off points				
	0.075%	0.1%	0.25%	0.5%	1%
Sensitivity (%)	60%	52%	27%	18%	9%
Specificity (%)	54%	65%	93%	98%	99%
PPV (%)	0.07%	0.08%	0.22%	0.43%	0.47%
NPV (%)	99.96%	99.96%	99.96%	99.95%	99.95%
Total	8693655	8693655	8693655	8693655	8693655
With HIV	4778	4778	4778	4778	4778
Likelihood Ratio +	1.3	1.5	4.1	7.9	8.6
Likelihood Ratio -	0.7	0.7	0.8	0.8	0.9

B. UK practices, excluding London, East Midlands and outlier practices

The third sensitivity analysis excluded London and East Midlands since their HIV incidence from the IMRD dataset were different from UKHSA HIV incidences. With London and the East Midlands removed the values at 0.075% cut-off points were: a sensitivity of 56%, a specificity of 56%, PPV of 0.06% and NPV of 99.96%, Table 4.5-9.

Table 4.5-9. Sensitivity, specificity and predictive values for UK (excluding London and East Midlands) at selected cut-off points.

	Cut-off points				
	0.075%	0.1%	0.25%	0.5%	1%
Sensitivity (%)	56%	47%	23%	15%	8%
Specificity (%)	56%	68%	95%	98%	99%
PPV (%)	0.06%	0.07%	0.24%	0.45%	0.56%
NPV (%)	99.96%	99.96%	99.96%	99.96%	99.95%
Total	7258248	7258248	7258248	7258248	7258248
With HIV	3580	3580	3580	3580	3580
Likelihood Ratio +	1.3	1.5	4.8	9.2	11.4
Likelihood Ratio -	0.8	0.8	0.8	0.9	0.9

C. London Practices

HIV incidence is highest in London; hence sensitivity analysis was conducted for London on its own. The values for London at 0.075% cut-off points were: a sensitivity of 75%, a specificity of 38%, PPV of 0.11% and NPV of 99.94%, Table 4.5-10. The values for sensitivity and negative predictive value (NPV) decrease with increase in cut-off points and vice versa for specificity and positive predictive value (PPV).

Table 4.5-10. Sensitivity, specificity and predictive values for London at selected cut-off points.

	Cut-off points				
	0.075%	0.1%	0.25%	0.5%	1%
Sensitivity (%)	75%	69%	43%	26%	13%
Specificity (%)	38%	48%	83%	94%	97%
PPV (%)	0.11%	0.12%	0.22%	0.40%	0.44%
NPV (%)	99.94%	99.94%	99.94%	99.93%	99.92%
Total	1231905	1231905	1231905	1231905	1231905
With HIV	1122	1122	1122	1122	1122
Likelihood Ratio +	1.2	1.3	2.4	4.4	4.9
Likelihood Ratio -	0.6	0.6	0.7	0.8	0.9

D. High prevalence practices.

Sensitivity analysis was conducted for high prevalence practices with HIV prevalence of 0.07% and above. The values for high prevalence practices at 0.075% cut-off points were: a sensitivity of 84%, a specificity of 21%, PPV of 0.16% and NPV of 99.88%, Table 4.5-11.

Table 4.5-11. Sensitivity, specificity and predictive values for High prevalence Practices at selected cut-off points in IMRD, 2004-2017.

	Cut-off points				
	0.075%	0.1%	0.25%	0.5%	1%
Sensitivity (%)	84%	78%	53%	32%	18%
Specificity (%)	21%	27%	57%	85%	96%
PPV (%)	0.16%	0.16%	0.18%	0.33%	0.62%
NPV (%)	99.88%	99.88%	99.87%	99.88%	99.87%
Total	1626993	1626993	1626993	1626993	1626993
With HIV	2456	2456	2456	2456	2456
Likelihood Ratio +	1.1	1.1	1.2	2.2	4.1
Likelihood Ratio -	0.8	0.8	0.8	0.8	0.9

E. Patients diagnosed with HIV more than 150 days from registration with GP

Sensitivity analysis was conducted for patients diagnosed with HIV more than 150 days from registration with a GP. The values for high prevalence practices at 0.075% cut-off points were: a sensitivity of 55%, a specificity of 59%, PPV of 0.06% and NPV of 99.96%, Table 4.5-12.

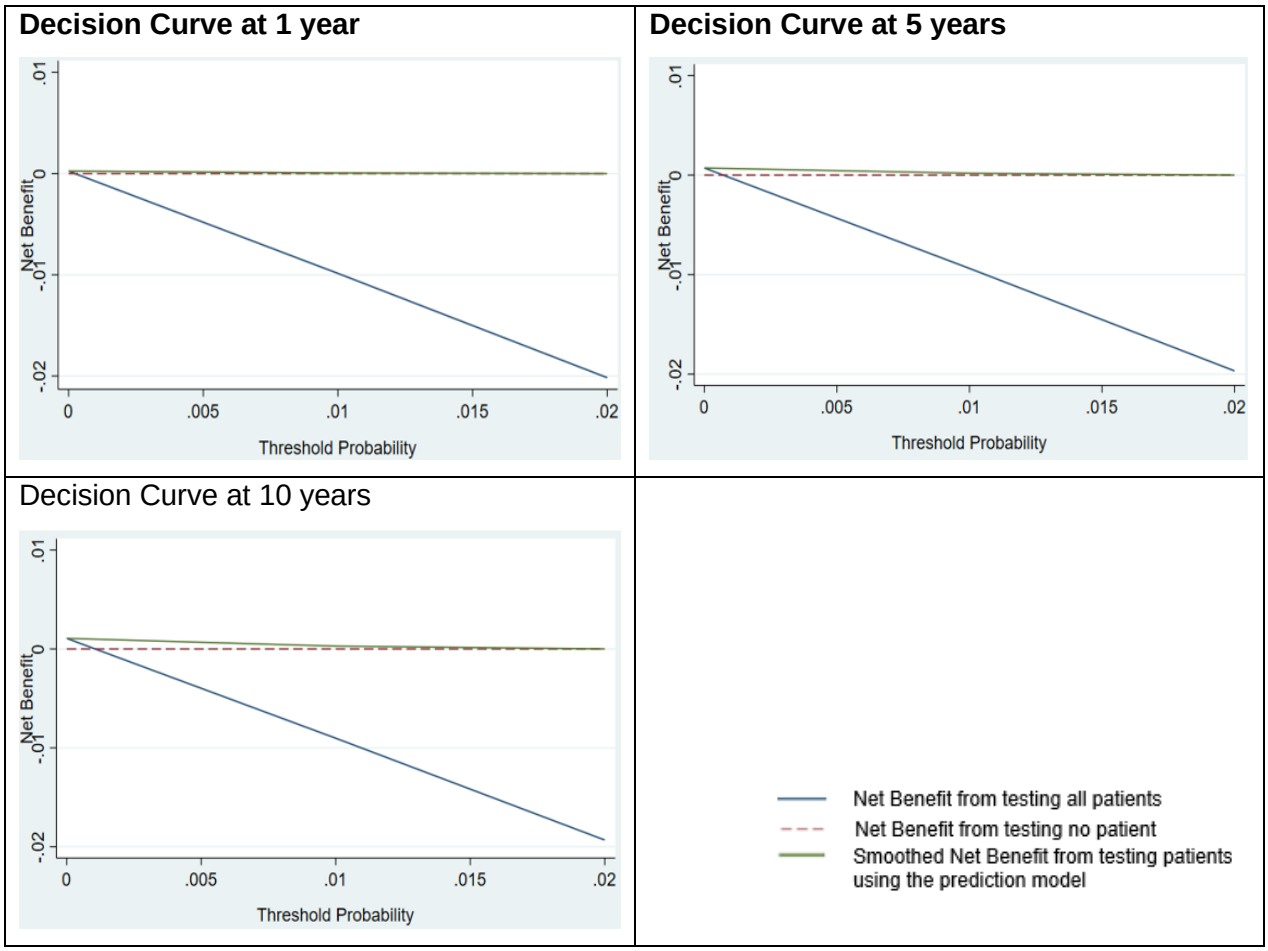
Table 4.5-12: Sensitivity, specificity and predictive values at selected cut-off points for patients diagnosed with HIV more than 150 days from registration with GP.

	CUT-OFF POINTS				
	0.075%	0.1%	0.25%	0.5%	1%
Sensitivity (%)	55%	45%	23%	14%	6%
Specificity (%)	59%	71%	96%	98%	99%
PPV (%)	0.06%	0.08%	0.24%	0.38%	0.40%
NPV (%)	99.96%	99.96%	99.96%	99.96%	99.96%
Total	8692993	8692993	8692993	8692993	8692993
With HIV	4119	4119	4119	4119	4119
Likelihood Ratio +	1.3	1.6	5.2	8.0	8.4
Likelihood Ratio -	0.8	0.8	0.8	0.9	0.9

4.5.5.2 Decision Curve Analysis (DCA)

Decision curve analysis, at 1 year, 5 years and 10 years were produced to show net benefit if all patients were tested (blue line), if no patients were tested (red dash line) and if the prediction model was used to target patients for testing (green line), Figure 4.5-8. Net benefit from using the prediction model was close to zero and higher that "testing all" for 1 year, 5 years and 10 years.

Figure 4.5-8: Decision Curve Analysis at 1 year, 5 years and 10 years.



4.5.5.3 Diagnostic interval from significant predictors (recorded with a date after GP registration) to HIV diagnosis

An analysis conducted to find the diagnostic interval (time between the date when HIV was recorded and the date of the most recent predictor, which had dates) found that two thirds of the patients with HIV either had none of the significant predictors or they were recorded on the same date as HIV diagnosis. Almost a third of the patients (27%) were not diagnosed for 6 months and above Table 4.5-13.

Table 4.5-13: Time from presentation of predictors to HIV diagnosis

Diagnosis interval	Number of patients	% of patients	Cumulative %
Same day	4083	66.6%	100%
<1 month	137	2.2%	33%
2-3 months	129	2.1%	31%
3-6 months	153	2.5%	29%
6 months - 1 year	150	2.5%	27%
Between 1 and 2years	261	4.3%	24%
Between 2 and 3years	183	3.0%	20%
Between 3 and 5 years	279	4.6%	17%
>5 years	753	12.3%	12%

4.6 Discussions

The non-recording of predictor variables in primary care is one of the reasons why some of the risk factors, such as sexuality (only 0.2% recorded) were not included in this study while they are important factors included in official data from UKHSA. This study also showed an improvement in recording of socio-demographic factors in primary care over the years, for example, ethnicity was recorded for 5% of patients in 2004 and the proportion increased to 59% in 2017. Some of the improvements could be a result of incentives provided under the Quality and Outcomes Framework (QOF) from 2006/7 and 2011/12 [225].

A case-control study conducted on the IMRD database found out that 10 clinical conditions were significantly associated with HIV, of which 5 were confirmed in this study, namely: oral candidiasis, blood dyscrasia, non-Hodgkin's lymphoma and STIs [35]. This study identified demographic and lifestyle predictors which were not identified in previous studies, such as black and mixed ethnicity, sexual contact abroad, drug misuse and smoking. However, some of the clinical conditions identified as significant predictors in the case-control study, such as weight loss and diarrhoea were not statistically significant in this study. Furthermore, unlike the case control study which confirmed 12 statistically significant clinical and comorbid conditions listed in the BHIVA guidelines, this study confirmed 17 predictors. Depression is the only predictor which was a statistically significant predictor in this study but not listed in the BHIVA guidelines.

The C-statistic measures the discrimination of the prediction model, and it provided the probability that for randomly selected pair of individuals (with and without HIV), the model assigns a higher probability to the individual with HIV [243]. The C-statistic from the model was 0.74 indicating that the model discriminates better than chance. The C-statistic compared well with the area under the curve of 0.66 from the case-control study and 0.666 (95% CI: 0.619, 0.712) from a similar model developed in North Carolina [35] [252]. The optimism adjusted C-slope of 0.990 was close to the value 1, meaning there was very little overfitting of the model to the data. This means there was no situation where probability of HIV would be overestimated in high-risk patients or underestimated in low-risk patients. This could be attributed to the large IMRD dataset used in this study. Although the predictions are close to zero, the calibration plots also show little over-prediction at 1 year, 5 years and 10 years.

An observation from the sensitivity analysis of the model is that, at cut-off points of 0.5% to 1%, the PPV ranges between 0.4 to 0.5%. Using a threshold of 0.1%, the positive predictive value ranged from 0.27% for all the practices under study to 0.45% for all practices (excluding outliers, London and East Midlands practices). The positive likelihood value for 0.1% threshold ranged from 3.8 to 9.2. The sensitivity analysis for patients diagnosed 150 days after registration had percentages in the ranges above. The sensitivity analysis for UK at 0.25% cut-off almost tallies with results from the case-control study at optimum cut-off of 0.26, sensitivity of 42% and specificity of 80%.

Clinicians choose the cut-off points for specificity and sensitivity of a condition depending on the cost of testing, prevalence of the disease and the consequences of false-negative/false-positive results [253]. It was discussed in chapter 1 that i) late diagnosis of HIV is associated with increased healthcare costs, ii) it is cost effective to promote HIV testing in primary care and iii) the prevalence of HIV is very low (around 1%) in UK. Hence, the cut-off point for

specificity and sensitivity for screening of HIV should be the point where sensitivity is high (that is, where testing correctly detect HIV positive patients).

Decision curve analysis is useful in combining the accuracy measurement, similar to sensitivity and specificity, with the clinical applicability [250]. The decision curves showed no net benefit of testing patients identified from the prediction model and a negative net benefit if mass testing (testing all patients) is conducted. However, these results should be treated with caution because they could be attributed to a very low prevalence of HIV (0.07%), coupled with the c-statistic of 0.743 (95% confidence interval: 0.736 - 0.749) and calibration plot. Furthermore, DCA did not consider the costs/savings associated with the diagnostic tests, such as monetary costs and reduced healthcare costs resulting from early diagnosis, discussed in chapter 1 [246]. Hence, the decision on determining the threshold requires further clinical input, given the low prevalence of HIV and the importance of early diagnosis.

4.7 Study strengths and limitations

This study had the advantage of using routinely collected data from a generalizable sample of practices distributed throughout the UK. However, it was susceptible to the weaknesses of routinely collected data such as incomplete outcome ascertainment and missing or incomplete data on predictor variables [143]. Furthermore, this study is subject to bias since it relies on patients visiting their GP when they have HIV infection or any of the candidate predictors.

This study showed that some predictor variables may not be recorded completely (for example, drug misuse), and important known risk factors, such as country of birth are not routinely recorded in UK general practice. The prediction model is also dependent on unbiased recording of the primary outcome (HIV status). This may not be the case if certain types of

patients are less likely to have their HIV status recorded (for example, a married man who has unprotected sex with other men).

As mentioned in section 3.4.5, the prevalence of HIV could be improved by using databases with linkages to hospital data. Additionally, there is a possibility that other databases could have higher prevalence of some of the predictors excluded from this model development due low prevalence. Hence, although IQVIA data was large enough in development and validation of this model, use of other databases such as CPRD could have produced a model with better performance if they have higher prevalence of the outcome variable and predictors.

4.8 Conclusion

This study developed and internally validated a prediction model for identification of potential HIV patients in primary care. Analysis of predictor variables revealed poor recording of variables such as sexuality, which are identified in published figures to be risk factors of HIV. The model confirmed some of the significant clinical and demographic predictors of HIV which were identified by a case-control study conducted in 2013 [35]. It also confirmed some of the clinical conditions listed in BHIVA guidelines [20]. Additionally, this study identified lifestyle predictors such as drug misuse and contact abroad which were not identified earlier.

4.9 Summary of chapter

In this chapter, a model for a risk score of HIV for primary care was developed and internally validated, using the Read code lists developed in chapter 2 and 3. Further evaluation of the model was also carried out. The risk score will be used in development of a point-of-care/pop-up alert in primary care. The next chapter will elicit views on the barriers and facilitators of using these pop-up alerts in primary care so that the findings could be used to ensure good utilisation of a HIV pop-up alert.

Chapter 5: Systematic review of qualitative studies on views on facilitators and barriers of use of pop-up alerts

5.1 Introduction to chapter

The previous chapter outlined the results from derivation and validation of a prediction model for risk of HIV infection. The predictive model is intended to enable general practitioners to identify patients likely to be HIV infected (through the use a point-of-care alert) and offer HIV testing. This chapter shows the evidence on the barriers and facilitators on utilisation of point-of-care alerts to prompt use of diagnostic tests. The views of primary care clinicians are paramount in the successful implementation of point-of-care alerts as they are the main users. This work will show whether the pop-up alerts from the prediction model developed in chapter 4 will be useful in primary care and what should be considered to increase the adoption of the tool.

5.2 Background

The usage of computer based alerts and prompts in healthcare has been increasing [254] [255]. Alerts range from simple reminder alerts for tasks or prescriptions to more sophisticated support for decision-making including risk scores for clinical conditions [256]. Computer based prompts or alerts have been used in primary care for screening of conditions such as breast cancer, osteoporosis, abdominal aortic aneurysms, and obesity [116].

Point-of-care ("pop-up") alerts and prompts are valueless unless they change clinicians' behaviour. Evidence shows that point-of-care (POC) alerts have mixed effects or lead to small or modest improvements in provider behaviour, processes of care and patient outcome [255] [257] [258] [259] [260]. Some studies established that systems where clinicians are required to respond were more likely to have a positive effect [257] [259]. In the USA, prompts in

primary care were associated with a five-fold increase in hepatitis C virus screening rates among adults born between 1945 and 1965 from 28 percent in the 3 years before to 72 percent in the year after implementation [261]. However, this study only provided the before and after information and did not model trends in testing. A cluster RCT in primary care patients in Spain found that use of alerts resulted in a non-statistically significant increase in participation in a colorectal cancer screening programme in primary care patients [262]. Furthermore, a retrospective cohort study to evaluate the impact of prompts for HIV testing in primary care in the USA found an increase in HIV tests from 15.3 percent to 30.7 percent after the introduction of a reminder (RR 2.02, 95% CI 1.95–2.09, $p < 0.0001$) [116]. A prospective interventional study in Barcelona revealed an increase in requests for HIV test from 12.6 percent in 2013 before the introduction of the prompts to 35.6 percent in 2015 followed by a decrease to 17.9 percent after removal of the prompt in 2016 [263]. A cluster randomised clinical trial in Catalonia showed a 20.6 percent improvement (OR 1.29, 95% CI: 1.25–1.34) in reminder resolution rates with the use of electronic point-of-care reminders in primary care versus monthly feedback [264]. A prospective interventional study conducted in Catalonia on 51 primary healthcare centres discovered that point-of-care alerts increased HIV testing in men (OR 1.26, 95% CI: 1.04–1.52), in patients aged under 50 years (OR 1.77, 95% CI: 1.33–2.38) and OR of 1.51 (95% CI: 1.20–1.92) in patients diagnosed of indicator conditions that were not sexually transmitted infections nor AIDS-defining illness [265].

This evidence showed variations in effectiveness of reminders in primary care. These variations could be due to facilitators and barriers affecting how healthcare providers act on electronic alerts or prompts [266] [267]. Factors highlighted as the barriers to adherence and outcomes include design features, clinicians' views, characteristics of the reminder, whether prompts were part of the clinician's workflow and the healthcare setting [257] [258] [268]. An analysis of data from an electronic prompt system in paediatric clinics in the USA found out that some clinics and physicians were more likely to address prompts than others, more likely

for younger children and for more serious health issues [269]. This study also found that physicians were likely to respond when prompts were at the top of the page [269]. Another study on electronic prompts in English cardiology wards found “that targeting, timing and additional features of alerts are critical factors in determining whether they are acted on or overridden.” [270]. Additionally, other studies showed a decrease in utilisation of alerts with increased number of reminders, overall patient complexity and increased number of repeated reminders [271]. Most of these studies were quantifying the magnitude of the effectiveness of POC alerts (ranging from 10 percent to 5-fold) [261] [262].

The interactions between the alerts and their environment that form the barriers and facilitators are likely to be complex. Clinicians are the main players that determine the success in utilisation of POC alert. This study aims to explore clinicians’ views on factors that facilitate or impede them from acting on clinical decision support (point-of-care alerts) in primary care. Addressing some of the concerns could increase uptake of HIV testing in primary care through the use a point-of-care alert, which prompts HIV testing [210]. A scoping search identified no studies specifically addressing the use of prompts in relation to HIV testing. However, the barriers and facilitators to point-of-care alerts for diagnostic tests for a chronic condition are likely to be similar, irrespective of the condition. The differences between HIV and other chronic conditions that should be considered pertains to sensitivity of HIV and the institutional factors affecting access to HIV testing discussed in section 1.11. Therefore, the aim of this review was to understand barriers and facilitators to use of point-of-care alerts or prompts for diagnostic testing for chronic conditions, as this is likely to be applicable to diagnostic tests for HIV.

5.3 Methods

The steps of this systematic review for qualitative studies included search for articles, selection of studies that address the review question and met the criteria, and extraction and synthesis of data using thematic mapping. The protocol for this review adhered to the requirements of Preferred Reporting Items for Systematic Reviews and Meta-analyses Protocol (PRISMA-P) [139] [272]. The protocol was registered with PROSPERO (ref CRD42021230295). This systematic review conformed to the requirements of the Enhancing transparency in reporting the synthesis of qualitative research (ENTREQ) statement.

5.3.1 Research question

This systematic review aimed to identify and summarise evidence on barriers and facilitators to primary care clinicians acting on point-of-care alerts for diagnostic testing intended to change clinician behaviour.

5.3.2 Population, phenomenon of interest and context

Studies were eligible if the participants were primary healthcare clinicians. Both General Practitioners (GPs) and nurses, were included irrespective of specialty, number of training years and professional level. In cases where professional roles are not specified, the reviewers contacted the authors for clarification. To ensure applicability of findings to primary care settings in the UK, the only studies included were those undertaken in developed countries in Europe (European Union and European Free Trade Association nations) and North America (USA and Canada), Australia and New Zealand. The phenomenon of interest was point-of-care alerts intended to change clinical decision-making. Alerts or prompts included reminders for recommended care, preventative communications and chronic condition management. The POC alerts included reminders to undertake diagnostic tests or laboratory tests for

diseases such Hepatitis B, Hepatitis C, thyroid disease and diabetes. These diagnostic tests are all blood tests which are similar the diagnostic test for HIV. The outcomes were themes on barriers and facilitators to utilisation of alerts or reminders including the characteristics of the provider and the setting.

5.3.3 Study design

This review included qualitative studies, which were more suitable because they can explore how clinicians view the application of point-of-care alerts [192]. This included those studies that used methods of data collection such as:

- interviews (semi-structured or structured),
- focus group discussions,
- surveys that include free text comments,
- observation
- mixed methods studies if they include qualitative methods

These studies had qualitative methods of data analysis such as thematic and/or narrative analysis. Previous systematic reviews of qualitative studies were also included in this review.

5.3.4 Search strategy

Studies were identified and reviewed via electronic searches of online research databases of EMBASE (Ovid), MEDLINE (Ovid), CINAHL Database and The Cochrane Library (Wiley). To reduce possible publication bias, additional searches were conducted on unpublished grey literature search engine of Open Grey (SIGLE) and Google Scholar [142] [143]. Additional searches were conducted on abstracts or conference proceedings using Web of Science Conference Proceedings Citation Index (CPCI) and examination of reference lists from studies included in the review (reference searching) [143]. References were searched and stored using the Refworks referencing software programme hosted by the University of Birmingham.

The review included studies published in any language (English and non-English) to minimise publication bias. Studies covered in this review were those conducted and published in the past 20 years when the use of electronic patient records (EPRs) in primary care practices gathered momentum [273].

5.3.5 Inclusion/exclusion criteria

The studies included in this review were those that explore use of point-of-care alerts for diagnostic testing in primary care. Those focusing on alerts for routine monitoring testing were excluded.

5.3.6 Data analysis and synthesis

Titles or abstracts of the identified studies were assessed and pre-screened by two independent reviewers to check if they addressed the review question. A meeting was conducted by the two reviewers to discuss and resolve differences in selected articles. In case of unresolved disagreements, a third reviewer was asked for their view.

The next stage was the retrieval of full articles included from selection by titles/abstracts. Two independent reviewers assessed the suitability of the studies in relation to the inclusion/exclusion criteria, using the full article. Another meeting was held to check if the reviewers agreed on the final list of the selected studies.

Quality assessment of articles was conducted using Critical Appraisal Skills Programme (CASP) checklist for qualitative research. Studies were graded as high quality, acceptable or poor quality and those of low quality were included if the results were valuable to this review.

The results from the extraction process were collated using a data extraction form. The data

extraction form was pilot tested on a few of the studies to determine suitability and reliability and it was revised to include findings from the selected studies. Strengths and weaknesses of the study, including sampling and non-sampling biases were documented. Any exclusion of studies at this stage was documented.

There is a range of methods used in synthesising qualitative research namely: meta-ethnography, grounded theory, thematic synthesis, textual narrative synthesis, meta-study, meta-narrative, critical interpretive synthesis, ecological triangulation, framework synthesis and 'Fledgling' approaches [274]. Some of the approaches are interpretive and are aimed at developing theory from conceptual literature, for example grounded theory, critical interpretive synthesis and meta-ethnography. On the other hand, other approaches are aggregative where individual primary findings are summarised or they are treated as parts that can be put together to make a whole, for example, meta-summary [274] [275]. Synthesis of evidence for this systematic review employed thematic mapping to reveal characteristics that needs to be addressed to increase utilisation of point-of-care alerts [145] [276]. This study used thematic mapping, which is a hybrid approach between grounded theory and meta-ethnography that can be used to address questions related to interventions [274].

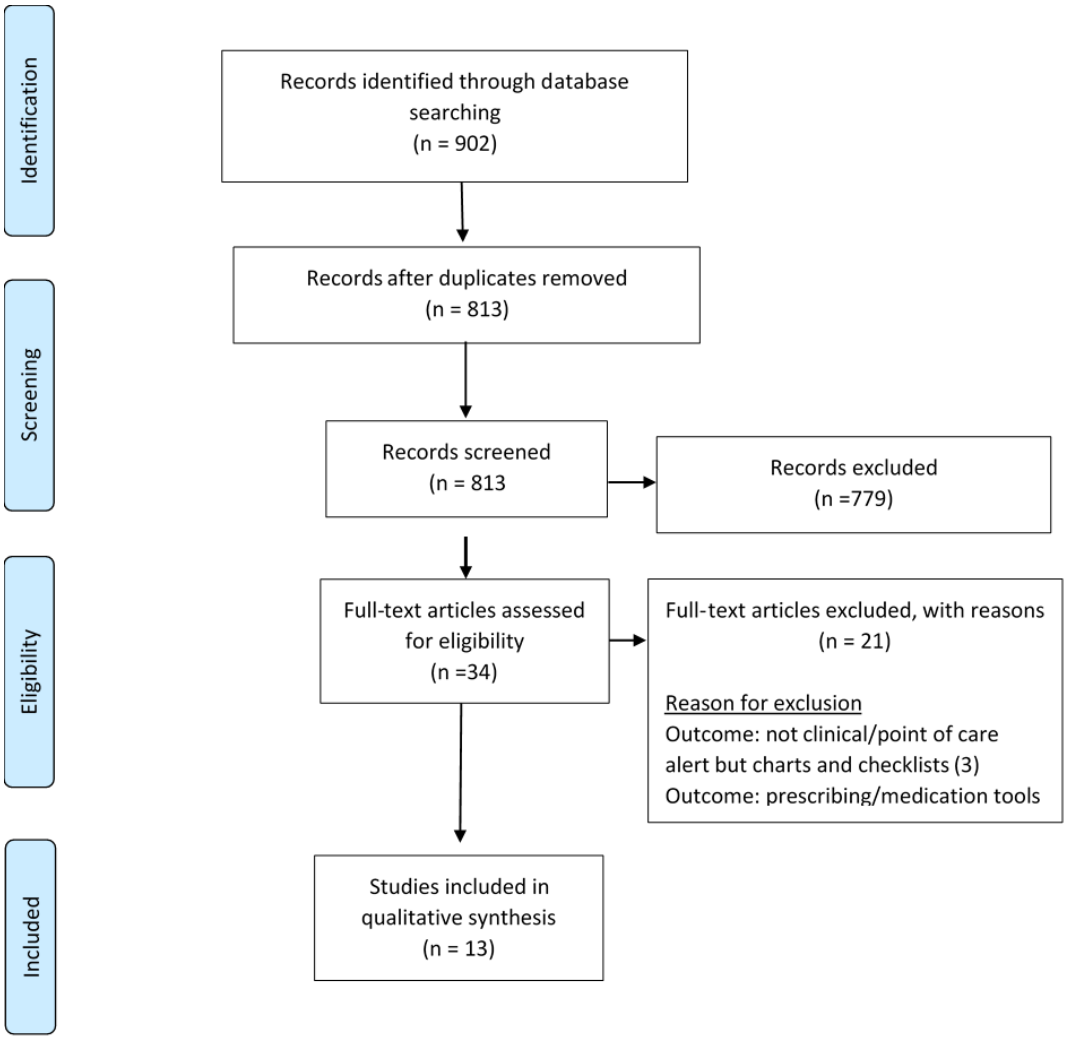
Thematic mapping was done in three stages, and this involved: i) free line-by-line coding of the findings of studies, ii) organising 'free codes' into 'descriptive' themes (related areas) and iii) development of 'analytical' themes [276]. The analysis adhered to the requirements of the PRISMA approach and a PRISMA diagram is included, showing the number of studies considered at each level of screening and assessment [139].

5.4 Results

5.4.1 Selection procedure

A search from the databases (Medline, Cochrane CINAHL and EMBASE) and the grey literature (SIGLE, Web of Science and Google scholar) returned 902 hits, of which 89 duplicates were removed and 813 articles were pre-screened, summarised in Figure 5.4-1.

Figure 5.4-1: PRISMA 2009 Flow Diagram



After review of titles and abstracts the reviewers selected 34 articles which considered the research question (including the population, phenomenon of interest and setting) and the period of publication, Appendix 16-A. There were no unresolved disagreements following discussion.

Full text articles were retrieved. All articles were in English; hence no translation was required. After reviewing the full text using Appendix 16-B selection criteria, the reviewers agreed on 13 articles and there were no unresolved disagreements.

5.4.2 Quality of studies

All 13 qualitative studies were of acceptable standard, of which eight were of good quality, in terms of participant recruitment, detailing of data collection and analysis, consideration on ethical issues (including consent from participants) and how they dealt with interviewer bias. The other five articles were not clear about the aim of the study, did not justify the use of a qualitative methodology and some did not include the ethical approval and participant consent (Appendix 17). However, all 13 articles were included in the review since they were all valuable to the research.

5.4.3 Study characteristics

The 11 qualitative and two mixed method studies were conducted in UK (four studies), Finland (two studies), Australia (two studies), Canada (one study) and USA (four studies). The number of study participants (GPs and Practice Nurses) ranged from 9 to 39. Of the 13 studies, eight employed semi-structured interviews as the data collection approach, Table 5.4-1. Most of the studies (11 of them) analysed data using thematic analysis and the other two used directed content analysis or deductive content analysis.

Table 5.4-1: Data extracted from selected studies

Author (Year)	Title	Setting	Number of participants	Study design	Data collection	Data analysis	Findings highlighted in Results	Findings highlighted in Discussion	Strengths/weaknesses
Peiris (2009)	An electronic clinical decision support tool to assist primary care providers in cardiovascular disease risk management: development and mixed methods evaluation	Australia	21	Mixed methods	In depth interviews	Thematic analysis	Systematic provision of care Risk Communication by clinicians improves Challenges for Implementation in Routine care	Scientific design and functionality of the tool Integrating the tool in routine workflow, automation of decision support, Offer recommendations, not assessment Limit data entry	Self-selective sampling method of GPs introduce bias since the participants are those that might actively contribute to the future tool development. Findings were similar to results from other Australian literature.
Dikomitis (2015)	Embedding electronic decision-support tools for suspected cancer in primary care: a qualitative study of GPs' experiences	UK	23	Qualitative	Semi-structured telephone interviews	Thematic analysis	Interactional workability: integration into GPs' computer systems, facilitators and barriers in the use of the tools and influence on referral behaviour Relational integration: GPs' understanding of the CDS implications and compatibility with existing guidelines Skill-set workability: training on new tools and use of CDS in everyday practice	Make CDS fit for purpose	Self-selecting sampling: interviewed GPs were 'keen' and/or interested in cancer diagnosis therefore were not representative.
Short (2004)	Barriers to the adoption of computerised decision support systems in general practice consultations: a qualitative study of GPs' perspectives	West Midlands, England	15	Qualitative	Semi-structured interviews	Thematic analysis	Limited skills and confidence in IT GP understanding of risk Time pressures in primary care Barriers arising from infrequent use Concerns about patient reactions to use of a support system	The ability to print information for both patient notes and for patients to take away would further enhance the value of any aid.	Not representative of GPs with less IT exposure because the participants were IT enthusiasts.

Author (Year)	Title	Setting	Number of participants	Study design	Data collection	Data analysis	Findings highlighted in		Strengths/weaknesses
							Results	Discussion	
Chase (2017)	Clinical Decision Support and Primary Care Acceptance of Genomic Medicine	USA	35	Qualitative	Semi-structured interviews	Thematic analysis	Priority for health issue in question Perceived potential Clinician lack of knowledge Need for evidence Cost concerns Work flow issues Current implementations and priorities	Evidence that CDS improves outcomes, costs, workflow integration	No statistically significant results but the study was a starting point for future research
Harle (2018)	Information Needs and Requirements for Decision Support in Primary Care: An Analysis of Chronic Pain Care	USA	10	Qualitative	Semi-structured individual interviews & a workshop	Thematic analysis	Information Needs Ideas representing potential decision support concepts and guide detailed design of decision support solutions	Need for reference information or knowledge resources Desire for context related to many clinical information elements needed	The study is relevant to the broader literature exploring clinical information needs. Increases representativeness by using a heterogeneous sampling approach thereby interviewing clinicians from three health systems in urban and rural areas.
Dryden (2012)	Provider perspectives on electronic decision supports for obesity prevention	Cambridge, England	9	Qualitative	Semi-structured interviews	Thematic analysis	Logistical Issues such as inappropriate alerts Appropriateness of use of HIT for Obesity Care	Providers should believe that their actions have an impact on their patients' health outcome	Purposeful sampling may not fully represent all the paediatric providers
Wan (2012)	Qualitative evaluation of a diabetes electronic decision support tool: views of users	Australia	22	Qualitative	Semi-structured interviews	Thematic Mapping	Provide a good reminder of patients' risk factor information Use of the tool in providing better quality of care to patients. Problems with the functionality of the tool software cost time. Users' poor knowledge with the tool's functions. Lengthen consultation time	Automatic provision of CDS as part of clinicians' workflow. Provide recommendations, not assessments. Location of decision making; and computer–based	Not representative since only willing participants were included

Author (Year)	Title	Setting	Number of participants	Study design	Data collection	Data analysis	Findings highlighted in Results	Findings highlighted in Discussion	Strengths/weaknesses
Varonen (2008)	What may help or hinder the implementation of computerised decision support systems (CDSSs): a focus group study with physicians	Finland	39	Qualitative	Focus group discussions	Thematic analysis	Facilitators Flexibility of the system; possibility to tailor for patients and possibility to switch off the system reliable knowledge base and having trusted peers developing the system Simplicity and ease of use Concise reminders Concise and tailored education on the CDSS Barriers experience of imperfect health care information systems loss of own reasoning and clinical autonomy erroneous information: garbage in-garbage out Resistance towards change Issues from poorly interacting computer programme	The younger physicians had the most enthusiastic attitudes and older physicians presented more criticism.	Interviewed physicians came from different parts of the country, thereby increasing representativeness However, computer or technology enthusiasts may have been overrepresented.
Kortteisto (2012)	Clinical decision support must be useful, functional is not enough: a qualitative study of computer-based clinical decision support in primary care	Finland	28	Mixed methods	Focus group discussions	Deductive content analysis	Content of the eCDS guidance Functionality of the system Features related to professionals Environmental factors	• Perceived usefulness of eCDS guidance. • Inaccurate data in eCDS system, • Usability of the EPR itself.	Generalization (as with most qualitative studies) is not possible due to the sample which was not representative

Author (Year)	Title	Setting	Number of participants	Study design	Data collection	Data analysis	Findings highlighted in		Strengths/weaknesses
							Results	Discussion	
Nair (2015)	A Clinical Decision Support System for Chronic Pain Management in Primary Care: Usability testing and its relevance	Canada	13	Sequential exploratory mixed method approach (2 rounds)	Observations	Directed content analysis	Graphical presentation of content i.e. layout, easy navigation and data entry fields. Clinical aspects of the CDS e.g. quality of medication list Software functions i.e. ability to save and having broken links	Use of 'Think aloud' protocols provided specific critique and suggestion	Small sample size Research done during implementation of the tool
Baron (2017)	Recognition of the Relationship Between Patients' Work and Health: A Qualitative Evaluation of the Need for Clinical Decision Support (CDS) for Worker Health in Five Primary Care Practices	USA	76	Qualitative	Semi-structured interviews and observations	Thematic analysis	The Knowledge Resources - Promoted Consistency and Standardization of Care More Evidence to support how the Clinical Recommendations in the Knowledge Resources Would Improve Care Whether the CDS Tool developed saves time is an important factor in clinician acceptance	Perceived ineffectiveness in provision of information	Limited representativeness due to the number and range of practice sites and clinicians included.
McParland (2019)	Differential Diagnosis Decision Support Systems in Primary and Out-of-Hours Care: A Qualitative Analysis of the Needs of Key Stakeholders in Scotland	Glasgow, Scotland	22	Qualitative	Focus group discussions	Thematic analysis	Current Practice Clinician's current practice: Attitudes to DDDSS Implementation Considerations Provide access to the evidence base	Systems could improve history taking and documentation, integrate with their workflow, and provide point-of-care access to a trusted evidence base. Concerned about the deskilling of future doctors through the use of DDDSS	Examined the attitudes of a variety of stakeholders in relation to DDDSS Bias as focus groups were only conducted in one area in the central belt of Scotland, hence only individuals able to travel to venues participated.

Author (Year)	Title	Setting	Number of participants	Study design	Data collection	Data analysis	Findings highlighted in		Strengths/weaknesses
							Results	Discussion	
Harry (2019)	Barriers and facilitators to implementing cancer prevention clinical decision support in primary care: a qualitative study	USA (Minnesota, North Dakota, and Wisconsin)	28	Qualitative	Semi-structured interviews	Thematic analysis	Intervention characteristics: Evidence strength and quality, Relative advantage, Adaptability, and Design quality and packaging Outer setting: Patient needs and resources, external policy and incentives Inner setting: - Culture, Implementation Climate - Tension for change, Compatibility, relative priority, organizational incentives and rewards, Readiness for Implementation - Available Resources, access to Knowledge and Information Characteristics of individuals - Knowledge and beliefs about the intervention	Computerised automation, making the CDS more patient-specific, CDS content; the CDS system; and implementation of the CDS into practice.	Bias because key informants were employed by the healthcare system.

5.4.4 Thematic Analysis

Thematic mapping included free line-by-line coding of the findings of the studies, grouping the 'free codes' into related areas and development of 'analytical' themes [276]. In some cases, themes were further merged if the findings had associated process. In the end, the results from the 13 studies fell under three overarching themes which looked at characteristics relating to the intervention, person and the setting. The themes were.

i) Characteristics of the decision support tool

ii) Features of the primary care setting

iii) Features related to users of the tool

5.4.4.1 Characteristics of the decision support tool

A. Content and functionality of the decision tool

The content of the tool and its accompanying guidance significantly affected the utilisation of the decision support tools [277] [278]. The other determinant was the range of computer operations that should be run before the decision support tool comes up since issues with the functionality of the tool software increases the time required to complete a consultation [277] [279] [280]. A tool with good functionality improves ease of use and the performance speed of the computer [277].

"'*There was generally a problem with the functioning of the electronic patient record system . . . yes, big problems with the computer*.'" [277].

"'*It may be problematic if the current system does not function adequately and then CDSS is added on … We would need one standard, one functional programme that would serve adequately …' (male, primary care*)" [281].

Clinicians with previous experience with poorly performing computer programmes or dysfunctional computer systems increased resistance towards the use of new tools [282]. Usability of the tool is affected by the software functions and the whole system used, including the ability to save time and whether there are broken links during implementation [278] [283].

B. Design quality and packaging

The way the CDS was designed affected its use [278] [280]. This also covered how the CDS was graphically presented in terms of its layout, how it was navigated and whether manual data entry fields were required [283].

> "'*Surely, what we would want would be a user friendly system where you have the information you need with just a few clicks . . .*' (female, secondary care)" [282].

Some clinicians highlighted that some decision support tools end up causing duplication of work with what the clinic staff would have done anyway whether the CDS was implemented or not.

> ""*Is it going to be duplicating everything that they're [clinic staff who identify patients due for screenings before a visit] doing already, I wonder?" (ID 132*)" [278]

Packaging of the CDS and any guidance that accompanies the tool had an impact on its adoption [277]. Furthermore, the location of the tool on the screen also affected its utilisation with some clinicians preferring the reminders to be in a visible position or on the left side of the screen [277] [282].

> *'Reminders' position on the left side of the screen.'* [277]

"The reminders should be behind one button, visible if you want them . . ." (female, primary care)" [282].

C. Relative advantage and Adaptability

Simplicity and easy navigation positively or negatively affected utilisation of CDS [277] [282]. The CDS should be flexible such that it can be tailor made for individual patients and should have an option to switch it off [282] [278]. The CDS should have accurate data and should be appropriate for the health issue in question for it to be fit for purpose [277] [281] [284].

"'What is the most important thing is that the system is simple, not complex as the current ones that are developed by engineers ...' (male, primary care)" [282].

Appropriateness of the CDS ensured that the reminders were concise, thereby improving good patient's risk information and providing better quality of care to patients [279] [282]. The other facilitator to use of CDS was the ability to provide recommendations [279]. Additionally, providers should believe that their actions in using the CDS have an impact on their patients' health outcome [284].

D. Confidence in CDS development

Clinicians wanted to understand why the tool was developed and its purpose, and some of the participants said this information should be in the accompanying documents and guidance [281].

"'*You wouldn't really use it without knowing what or how it was developed, why it was developed, and what it was for'*. (GP/20)" [281].

Additionally, availability and access to the concepts and guide documents showing the design of the CDS increase its acceptability by clinicians [285].

E. Evidence on effect of tool

The need for evidence, in relation to potential in improving care and the effect of the CDS on costs, were raised as a barrier to the utilisation of the CDS [286] [287]. The quality and strength of the evidence improved utilisation of the CDS [278]. This included inaccurate information or other similar alerts being provided to the clinician [278].

"I'm a little wary of it but I'm also excited because I think it will help us get closer to having an individualized plan for people that really is about who they are." [287].

"Because if it's pulling inaccurate information or there's three [EHR alerts] that are firing about the same thing and there's so much overlap, that's going to be a point of frustration and ultimately, people will just wash their hands of it." [278].

Furthermore, the clinician's perception on the potential and usefulness of the CDS and the accompanying guidance facilitated or was a barrier to the utilisation of the CDS [277] [286]. This can be addressed by providing education on how the tool was developed and providing access to evidence on its potential [282] [279].

"*I think it does require a fairly good introduction, introductory session to show where the effectiveness of the tool can, can come in, where you can benefit from this and how to fit it into a consultation. –GP0101*" [279]

F. Save on time and cost

CDS which reduce costs and save time are easily accepted by clinicians [286] [287, 288]. Time was also saved if the CDS had minimal data entry and considered the time pressure in

primary care [280]. Lengthening of consultation time due to the implementation of the CDS was found to be a barrier to use [279].

"Do we make it part of the appointment and give ourselves a little more time so we can discuss it? That's the tough part, just trying to work it in, because you're already going over on most of your appointment slots anyway." [278].

"'*Most GPs are wary of these types of decision supports in terms of the time they take. It seems like a relatively trivial amount of extra time . . . (but it) becomes a major barrier unless it is going to make a dramatic difference.*'' GP03" [288]

5.4.4.2 Features of the primary care setting

A. Primary care priorities

The relative priority of the health issue and how the implementation of the CDS is valued affect the use of the tool by clinicians [278] [286]. Additionally, the relevance of CDS needs to align to organisational and external policies, rewards and incentives. [278].

"We can't do everything…right now the priority is to make the system work for us I think we are working too hard for it." [286].

Some of the priorities for general practices were not well planned at system level, thereby affecting the implementation of the tool at ground level [278].

"*There's so many different initiatives and so many different changes that it's hard to keep them all straight if you're doing four or five changes at once."* [278].

B. Communication and Knowledge

Some of the CDS were seen as contributing to knowledge and communication. The literature highlighted that the CDS provided access to knowledge and information which was a vital resource in improving patient care [278] [285]. Some participants highlighted that the CDS and its accompanying guidance provided knowledge resources as an evidence base [289]. Furthermore, CDS promoted consistent and standardised care to patients [280] [287]. Implementing some of the tools improved communication of risks among clinicians and also improved sharing of patient needs [280] [278].

> *"I think it's useful to us.... It's basically like a mini audit. So, anything that makes you look a little bit deeper at the person sitting in front of you is always worthwhile.... [Interview 19: Male AMS GP 40-49 years]"* [280].

Additionally, beliefs and attitudes towards the CDS and its functions affected its utilisation [278] [279] [289]. The other factors were resistance to change, clinicians fearing loss of own reasoning or clinical autonomy due to reliance on the CDS and clinicians feeling that the CDS was taking attention away from the patient [282] [281].

> *"'Does it mean that you are even more closely tied with the computer, and the patient sits quietly behind your back? There is also risk that you stop using your 'clinical eye', touching and listening to the patient' (male, secondary care)".* [282].

> *"'There's a dichotomy between the very useful information that's on the computer, and actually, you know, sort of, looking at the patients, and giving them, you know, proper attention, as they perceive it, you know'. (GP/15)*" [281] .

However, some clinicians raised concerns that the use of CDS might deskill future doctors [289]. Studies also mentioned the effect of the tool on referral behaviour and how the GPs

were concerned by the patient's reaction to them using a support tool [281] [288]. Other studies found that GPs had to understand the risk and implications of the CDS [281] [288]

C. Compatibility

The CDS must be compatible with the practice's Information Technology (IT) system and existing guidelines [278] [281] [289]. This also involved how the CDS was implemented in practice [278]. Other factors were unwillingness to change and the culture of the practice and the healthcare system [278].

D. Integration and automation into routine work

An easily adopted CDS should not add more into the clinician's workload or should not interrupt in any way with clinical work, rather it should be integrated with clinician's routine work [280] [281] [286] [289]. The computer system used by the CDS should also be integrated into the GPs' computer systems [281]. In addition to integration of the CDS with the clinician's computer system, the automation of the CDS in the clinicians' workflow increased utilisation of the CDS [278] [279].

> *"'I think this issue of the computer resources is important. It's got to be something that becomes a regular part of your practice, really. So, from that point of view, certainly the IT people ought to sort that side of things out'."* [281].

5.4.4.3 Features related to users of the tool

A. Limited skills and confidence

Ability and confidence in using IT was highlighted as a factor in the adoption of a computer-based tool [288]. Some clinicians struggle with their computer system and a CDS would be an additional challenge [288].

"*It doesn't matter how useful it might be, because it has a computer attached to it makes it unusable. That applies to computerised systems full-stop*.'' GP05" [288]

"*It's maybe another skill set in terms of just, you know, getting used to using the apps, isn't it. Ask everything we would ask normally. Transfer it on . . . You just need to get really slick at it*" [Amanda, Trainee ANP, HCP Group 1]" [289]

The other skill-based factor was clinicians' lack of knowledge on the CDS which can be facilitated by concise, accessible and tailored education or training on the new CDS and how it can be used in everyday practice [282] [281] [286].

"'It [the WebEx] was really long and drawn out is the honest answer; I think it was half an hour or an hour, I can't remember. Yes, but actually there was a good eight minute slot that was brilliant that just explained it all, so I would be tempted I think from watching that thing it made a big, eh, it was really useful'. (GP/6)".

Infrequent use of the CDS was also found to be a barrier to its use [288]. One study highlighted that enthusiasm towards the CDS came from younger physicians while the older physicians presented criticism [282].

B. Accuracy

Some clinicians felt that having a system in place might improve history taking and documentation [289]. The clinician's use of the CDS might be affected by the experience they had with an imperfect and an ineffective healthcare information system [282] [287].

However, a CDS developed by peers was easily accepted by clinicians [282]. This would also allow the use of 'Think aloud' protocols that provided specific critique and suggestion [283].

C. Training on the tool

Concise and tailored training on the CDS was a facilitator of its utilisation [282]. Some of the participants felt that e-learning was not an effective method of training as in-person training [278].

> *"I think your best way to approach it personally would be, in a family practice setting at least, is for you guys or whoever is bringing it out to come to both a section meeting [PCPs] and a staff meeting [nurses and rooming staff]."* [278].

5.5 Discussion

This study established that the characteristics of a CDS tool affect its utilisation in relation to ease of use, position, functionality and whether it will add value in terms of relative advantage as well as saving time and reducing costs.

How the CDS was developed and how it will be used influences its utilisation. It is easily adopted if it was developed by peers, can be integrated into the users' workflow and compatible with the users' computer system so that it does not interfere with everyday work.

Furthermore, the CDS should fit with the local priorities of the general practice as well as external policies and incentives.

The study also showed that a CDS could be more easily adopted if it has accompanying guidance and accessible training. The method of training depends on users' preferences and time required, with some participants not preferring e-learning.

5.1 Strengths and limitations

The findings from this systematic review showed a wide range of factors that should be considered when developing a POC alert for HIV. Some of the findings such as design features, clinicians' views and characteristics of the reminder were identified in literature on quantitative studies discussed earlier in the background [257] [258] [268]. However, the findings from this systematic review could be generalised in relation to transferability, which means they can be extrapolated to other settings or conditions [290].

A limitation of this systematic review was that there might be further considerations required due to differences in sensitivities between HIV and other conditions, including extra consent needed for HIV testing as more discussion is required before consent is requested. Furthermore, as with all qualitative studies, findings from this study cannot be generalised in relation to extrapolation from selected participants to views from all primary care clinicians. Further research is needed to test the effects of HIV or other sensitive POC alerts in primary care.

5.2 Conclusion

BHIVA highlighted the need to normalise HIV testing by integrating it into routine work so that no consent would be required for blood tests and no additional costs are incurred [84]. Recommendation from this study was that in order to normalise a CDS for HIV testing, the characteristics of the tool, the primary care setting and the characteristics of the users should be considered during the development of the CDS.

The CDS tool for HIV should be straightforward and easy to use. It should not have many technical features which affects performance of the IT system. Involvement of peers in the development of the tool, especially those with interest in HIV increases confidence in use of the tool. Additionally, confidence improves with availability of publication of prediction model and possible benefits. Furthermore, decision makers are concerned that the tool does not add more consultation time and it should save costs, including pathology cost. However, consideration should be made on the cost of using the CDS and the total cost incurred by the whole system when HIV patients are not diagnosed early.

In low prevalence primary care settings, HIV testing may not be a priority and HIV prevalence in UK varies by country and region [291]. Therefore, the acceptance of the CDS tool could be affected by the prevalence of HIV. However, HIV prompts could enhance knowledge by raising awareness of HIV risk factors and make it easier for clinicians to raise questions on HIV testing. This would be made easier if the HIV testing prompts are integrated into the general practice's routine work. Also, how users view the tool and their confidence and skills in using the tool should be considered. This should include how confident primary care staff are in raising questions on HIV testing and undertaking point-of-care testing. Attention should be taken on additional training required by the staff on HIV tool.

Users (clinicians) should be involved in the development of a point-of-care alert. The POC alert should integrate well into primary care workflow and computer system, should not add more consultation time and should save costs. This showed that the use of the POC/pop-up alerts should be normalised into the routine work of primary care.

In summary, the development of the pop-up/POC alert should consider ease of use, integration into routine work and increasing the confidence of the primary care on how to use the CDS as well as how to do point-of-care testing.

5.3 Summary of chapter

In this chapter, evidence on clinicians' views on the barriers and facilitators on utilisation of POC alerts to prompt use of diagnostic tests were identified. It showed what should be addressed to increase utilisation of HIV pop-up alerts (developed from the prediction model) in primary care.

Chapter 6: Discussion and Conclusion

6.1 Introduction to chapter

This chapter covers the implications of the findings from all the previous chapters on public health and clinical practice. It also covers how the findings from the prediction models could be implemented in clinical practice and recommendations for future research.

6.2 Summary of findings

6.2.1 Background to HIV in UK

HIV/AIDS has been declining in UK in relation to newly diagnosed patients (from 7,870 in 2005 to 2,766 in 2020). The effectiveness of ARTs has contributed to reduction of deaths from 1,720 in 1994 to 634 in 2020. As a way of reducing transmission and increasing the number of people living with HIV under treatment, guidelines are recommending the expansion of HIV testing from established settings (GUM clinics and antenatal clinics) to other settings including primary care. Evidence shows that the involvement of primary care would increase HIV testing, since almost all patients in UK are registered with a GP.

6.2.2 Predictors of HIV

The systematic review revealed that most of the predictors identified in the literature are included in HIV guidelines in UK, USA and Australia. The predictors that were identified in literature but not in the guidelines were, fever or chills or flu-like symptoms, cough, abdominal pain, minor trauma, nausea/vomiting, rash, depressive symptoms, hyperlipidemia, hypertension and diabetes. All the clinical conditions and comorbidities identified in the systematic review are available in primary care records. However, some of the demographic

and socio-economic predictors were not available in primary care records. Those available or partly recorded in primary care include gender, age, sexuality, ethnicity and country of birth. Deprivation, available from primary care records, could be used as a proxy for employment status, poverty levels, education attainment and income. Lifestyle or behavioural factors identified in the systematic review and available in primary care records were smoking status, binge drinking or alcohol misuse, drug misuse, obesity, having contact abroad, personal and sexual partners' characteristics (multiple partners and/or partners misusing drugs) and having gone through stressful events.

Descriptive analysis showed that most of the 8 demographic, geographic and socio-economic predictors are poorly recorded in primary care records, apart from age, gender and region of residence. Ethnicity and sexuality are poorly recorded in IMRD primary care records with completeness of 43 percent and 0.2 percent, respectively. Completeness of deprivation is an issue even in official data and it was 82.8 percent in IMRD. The other predictors were area or country of birth (1.85 percent complete) and urban/rural classification of place of residence (68.2 percent complete).

Descriptive analysis revealed that 10 lifestyle predictors were available in IMRD. However, the frequency of some of the predictors could be a result of poor recording of such predictors in primary care records. The factors include number of lifetime partners, patients having anal sex and partner characteristics.

6.2.3 Recording of HIV

Qualitative research was conducted to find out if GPs disguise recording of HIV in primary care records. It revealed that euphemistic terms potentially used by GPs are priority 0 (used

around year 2000), immune/immunity problem, blood-borne viral Infection, retroviral infection, immunodeficiency and viral load. An exploration of IMRD dataset showed that only 0.2 percent of patients in the dataset had the euphemistic terms recorded in primary care before HIV diagnosis, hence euphemistic terms were not considered in the prediction model. To ensure that the prediction model used HIV data consistent with official figures, an acceptable HIV reporting was established by finding the period when the incidence of HIV from IMRD was similar to the officially published data from the UK Health Security Agency. At national level, the acceptable HIV reporting was in 2004, hence the prediction model used data from 2004 to 2017. The analysis of IMRD dataset at country level revealed that two practices in Scotland and one in Northern Ireland had inconsistently higher incidence in some years, hence these outlier practices were considered in the sensitivity analysis. At regional level, most of the English regions had HIV incidence similar to the UKHSA published data except for London and East Midlands, where incidence from IMRD was lower that the incidence published by UKHSA. There was one practice in East of England which inflated the HIV incidence from 2004 to 2007 and this outlier practice was excluded is the sensitivity analysis.

The results from the qualitative study unveiled that, apart from HIV, there are conditions disguised in primary care records. The conditions include symptoms of child sex abuse, mental health including depression, domestic violence, termination of pregnancy and Hepatitis B.

6.2.4 Development and internal validation of HIV prediction model

The prediction model revealed that the demographic, geographic and socio-economic predictors with significant association with HIV infection were gender, 25-34 years age group and being of black and mixed/other ethnicity, deprivation and living in urban areas. Lifestyle predictors were being a current smoker or ex-smoker, drug misuse and sexual contact abroad. The clinical and comorbid conditions were Kaposi's sarcoma, pneumocystis carinii,

progressive multifocal leukoencephalopathy, syphilis, Non-Hodgkin's lymphoma, tuberculosis, cerebral toxoplasmosis abscess, anal cancer or anal intraepithelial dysplasia, aseptic meningitis/encephalitis, oral candidiasis, hepatitis B and C, blood dyscrasia, chronic liver disease, depression and current STI (excluding syphilis) or any previous STI. The additional significant predictors after exclusion of outlier practices (two in Scotland, one in Northern Ireland and one in East of England) were 35-49 years age group and pneumonia. Pneumonia was a predictor in the previous case-control study and the model results from this study tend to agree once outlier practices were removed [35]. The predictors that were no longer significantly associated with HIV infection, after excluding outlier practices, were deprivation level 2, gender, smoking status and having sexual contact abroad.

Internal validation confirmed all the statistically significant predictors of HIV that were identified in the original model. The results from the validation model were close to the results from the original model, mainly due to robustness of the data used in the models. Sensitivity analysis of the results showed that using a threshold less than 1%, more patients need to be tested to get a reasonable number of HIV positive patients. For example, in high prevalence practices, the sensitivity at a threshold of 0.5% was 32 percent, specificity at 85 percent, positive predictive value of 0.33 percent (meaning testing 300 people result in one positive result), negative predictive value of 99.9 percent, positive likelihood ratio of 2.2 and negative likelihood ratio of 0.8.

6.2.5 Views on facilitators and barriers to use of pop-up alerts

If the prediction model is used and a point-of-care (POC) or pop-up alert needs to be developed in primary care, views of clinicians on the use of the pop-up alert should be considered. The main issues fall under three main categories namely, i) characteristics of a POC alert, ii) features of the primary care setting and iii) features related to users of the tool. The characteristics of the pop-up alert include the content and functionality, design quality and

packaging, relative advantage and adaptability, confidence in CDS development, evidence on effect of tool and saving on time and cost.

Features of the primary care setting include primary care priorities, communication and knowledge, compatibility and integration and automation into routine work. Features related to users of the tool include skills and confidence of the users, perception on accuracy of the tool and availability of training on the tool.

6.2.6 Overall Summary

This study derived and internally validated a prediction model for HIV using a retrospective cohort design and assessed the feasibility of using the model in providing a point-of-care alert in primary care. The model developed and validated in this study could provide a risk score which provides a predicted probability that an individual might be HIV positive in United Kingdom and other countries with patients that experience the same signs and symptoms of HIV/AIDS as those in UK (mainly developed countries). In addition to model development, the study revealed factors that should be considered in implementation of the pop-up alerts in primary care.

6.3 Contribution of this research

The case-control study conducted by Damery et al (2013) was the only HIV prediction model developed in the past 15 years [35]. The case-control study only considered predictor variables listed in the 2008 BHIVA guidelines but this research considered all demographic, socioeconomic and clinical and comorbid conditions found in literature and in BHIVA guidelines. The systematic review revealed eight clinical predictor variables found in literature

which were not included in BHIVA guidelines. Of the 51 clinical predictors of HIV listed in BHIVA guidelines, only 18 were found in literature.

The model developed and internally validated in this study identified 17 clinical and comorbid conditions statistically associated with HIV as opposed to the 12 clinical conditions identified by the case-control study [35]. Furthermore, the model identified demographic, socioeconomic and behavioural predictors which were not included in previous prediction models.

The systematic review on barriers and facilitators to use of pop-up alerts identified the factors that should be considered in developing HIV pop-up alerts. The factors identified fell under characteristics of the decision support tool, features of the primary care setting and features related to users of the tool.

6.4 Public health implications

Late diagnosis and transmission of HIV/AIDS is a public health issue worldwide and in the UK. The number of newly diagnosed individuals and those diagnosed late, has been declining in the UK but there are still pockets of high prevalence and late diagnosis (57) (61). People living with HIV who are unaware of their status and those diagnosed late are at high risk of transmitting the disease to others, making this a public health concern.

Public health teams in local authorities and in UKHSA are responsible for commissioning of HIV prevention services since 2013 [292]. HIV prevention campaigns are aimed at preventing HIV infection and transmission via programmes such as “education and information, promoting safer sexual behaviour, facilitating access to HIV testing, increasing early diagnosis and treatment, and a pre-exposure prophylaxis (PrEP)” [293]. This study identified some

demographic, socio-economic, geographic and lifestyle or behavioural characteristics that could be used to target these campaigns. Demographic and socio-economic characteristics include being in the 25-34 age group, black and mixed/other ethnicity and deprived communities. Lifestyle characteristics are substance misuse such as smoking and drug misuse. Although some of the predictors of HIV infection revealed in the systematic review (chapter 2) were not used in model development, public health should continue to use these risk factors in prevention campaigns and facilitation of access to HIV testing. Such predictors include sexuality, unsafe sex, partner characteristics and number of lifetime partners.

This research developed a prediction model that can be used to identify patients that are most likely to be HIV positive in primary care. This approach would augment other alternative testing routes recommended in BHIVA guidelines. This targeted testing approach is, however, less resource intensive compared to mass testing.

6.5 Implications for clinical practice

6.5.1 Recording of predictors or risk factors of HIV

The official statistics categorise risk groups by sexuality, country of birth and ethnicity. This means that more should be done in improving recording of such risk factors in primary care, which would in turn improve the results of the prediction model. However, it should be noted that recording of ethnicity has improved in recent years, and this could be due to incentives, which were put in place between 2006/7 and 2011/12, with recording of ethnicity for new GP registrants while ethnicity for older patients is not updated. On the other hand, recording of sexuality is not incentivised in primary care and might not be acceptable to patients if they consider it to be personal.

Some of the lifestyle factors included in the prediction model were recorded once and not updated in recent years. Although drug misuse was a risk factor of HIV, it could have been more useful to find out if the association was dependent on whether drug misuse was recent or not recent since some of the HIV positive patients had drug misuse recorded in the early 1980s. Thus, this research shows that primary care needs to consistently update lifestyle characteristics since patients' change behaviours over time.

6.5.2 Recording of HIV

It was believed that GPs might disguise recording of HIV in primary care records by using euphemistic terms. One of the reasons for that was stigma and discrimination. This study showed that there has been an improvement in recording of HIV in primary care, especially compared to early 2000s and HIV incidence is similar to the incidence in UKHSA data. Most GPs are using the recommended and permitted clinical codes listed by the sexual health clinics.

The fact that about two thirds of the risk factors included in the model were recorded on the same day as HIV diagnosis suggests recording issues of either the predictor variable or HIV diagnosis, Table 4.5-13. This could mean that GPs were either recording a risk factor soon after HIV diagnosis or they identify a risk factor and test for HIV on the same day. If it is the later, then it means the GPs were aware of the need to test for HIV. However, this shows that GPs should improve the recording of the conditions and other characteristics in the electronic records.

6.5.3 Predictors of HIV identified in the prediction model

The prediction model revealed risk factors which could identify patients with high probability of HIV infection, listed in Table 4.5-6. The predictors with strongest association with HIV

infection were Kaposi's sarcoma (hazard ratio (HR) of 171) and pneumocystis carinii (HR of 71). These were two of the three conditions which led to recognition of AIDS patients in the early 1980s and they are still the predictors with highest probabilities. The other predictors in the top 5 are progressive multifocal leukoencephalopathy with a HR of approximately 56, being of black ethnicity with a HR of 10 and syphilis with a HR of approximately 11.

The predictors associated with risk of HIV infection with hazard ratio of 2 to 9 were Non-Hodgkin's lymphoma (NHL) with HR of 9, cerebral toxoplasmosis abscess with HR of almost 8, anal cancer or anal intraepithelial dysplasia with HR of 5, hepatitis with HR of about 5, chronic liver disease with HR of about 4, aseptic meningitis/encephalitis with HR of 3, being of mixed and other ethnicity with HR of 2.6, tuberculosis with HR of 2.3, drug misuse with HR of 2.3 and contact abroad with HR of 2.

The rest of the predictors significantly associated with HIV infection were blood dyscrasia, oral candidiasis, STI (excluding syphilis) or previous STI, herpes zoster, depression, deprivation (high deprivation quintiles), lymphadenopathy, age group of 25-34 and being a female.

Patients are likely to accept an offer of HIV if the risk factor is clinical diagnosis such as Kaposi's sarcoma, pneumocystis carinii, pneumonia, multifocal leukoencephalopathy, syphilis, NHL, cerebral toxoplasmosis. A qualitative study on opt-out testing conducted in Brighton revealed that patients felt judged if they are targeted due to their ethnicity or sexuality unless it was clear that the tests were offered to all patients [77]. That means patients would probably accept an offer for HIV test if the risk factors were behavioural such as drug misuse and current smoker/ex-smoker. However, patients might not accept offer for HIV tests if the risk factor is ethnicity.

6.5.4 Utilisation of prediction model

The prediction model developed in this thesis is the second study to develop a prediction model for HIV in UK primary care. The previous case-control study only considered clinical conditions identified in HIV guidelines [35]. This prediction model will be useful once it is packaged in a computer-based tool, developed with consideration to some of the barriers and facilitators mentioned in chapter 5. It is also fundamental to increase awareness and training of GPs on HIV testing [95].

As discussed in the first chapter, this thesis aimed at developing a model which would assist in targeting of patients that might be offered HIV testing, based on their characteristics (socio-economic, demographic, lifestyle and clinical conditions). This would help in meeting the principle of screening which states that the target population for testing should be clearly defined.

The discussion on the principles of screening in chapter 1 identified that 9 of the 12 principles of screening are already met. The prediction model revealed that it is possible to test many people for HIV and get an acceptable number of those that are positive. This approach is less resource intensive compared to mass testing of all registered patients attending an appointment.

6.6 Strengths and Limitations

6.6.1 Model development

This is the first retrospective cohort study which developed a prediction model using candidate predictor variables found in literature (including guidelines). The prediction model used the routinely collected IMRD primary care databases which is representative of the UK population.

In addition, the strength of using IMRD, like all electronic primary care records is that it contains great data with demographic, socioeconomic and clinical characteristics which were required for prognostic research [126] [127]. The dataset was large and robust enough to provide reliable findings. The closeness of performance results from the derivation and internal validation dataset confirms that the data used was a large dataset. However, the trend of HIV incidence in some UK countries and English regions were not consistent, which could be an issue on unexplainable changes in patients' demography for some practices.

The developed model was the first to consider all predictor variables found in literature and HIV guidelines, which were available in primary care database or their proxies. However, some of the variables (potentially important) were excluded from model development due to poor recording in IMRD database. Inclusion of these variables if they are robust enough could improve the results from the model. On the other hand, some of the studies selected in the systematic review had small sample sizes which could result in inclusion of some predictors that were supposed to be excluded. Results from studies with small sample size are unlikely to be differentiated from variation by chance [294].

In addition to a wider list of predictor variables, the research considered a broader definition of outcome variable, which included both confirmed and probable HIV identified through euphemistic terms. The euphemistic terms identified were not included in model development since they had a very minimal effect, with less than 1% of observations recorded before HIV diagnosis. This could be a result of the fact that the qualitative study conducted in chapter 3 was not exhaustive enough as GPs were difficult to recruit. Hence, there was a possibility that some of the HIV patients were excluded as a result of poor participation in the explorative research. However, there is a possibility that increasing the participants for the qualitative study could still provide no difference in euphemistic terms identified.

Model development treated missing data as new categories while there are other methods suggested in literature, such as multiple imputations [230]. These approaches could reduce bias associated with the use of missing-indicator method, which was used in this study [229] [230].

6.6.2 Views on the use of pop-up alerts in primary care

The systematic review on views of clinicians in the use of pop-ups revealed that there are barriers and facilitators that should be considered when developing a pop-up alert. However, there are no qualitative studies that were specific to HIV, and this could have a limitation on professionals' views in relation to institutional factors, which affect access to HIV testing services, discussed in section 1.11. These include the ease at which clinicians feel about discussing HIV testing with their patients.

Although the factors that should be considered in developing pop-up alerts were identified in the systematic review, there was a limitation that they were non-HIV specific. The study would have been more valuable and HIV specific if the study was a primary qualitative research on use of HIV pop-up alerts.

6.7 Recommendations for implementation of model in a clinical setting

Prior to implementation of the model in clinical practice, the prediction model should be externally validated to evaluate how it fits in a different and independent population. This should be followed by setting up of a threshold risk score which will be used in the pop-up or POC alert. Alternative tools used by clinicians to evaluate the possible clinical utility or the benefits of a diagnostic test, apart from the decision curve analysis should be considered [246]. The alternative tools include Incremental Net Benefit (INB) and cost-benefit analysis

[295]. Incremental Net Benefit (INB) is calculated as a function of costs for tests and effectiveness (measured by life-years gained) [295].

A feasibility study to find out if the risk score is viable to use in primary care should be conducted [296]. This should include studying how the risk score will impact on patient outcome, in addition to the effects on clinicians' decision making [297]. A risk analysis should be performed before implementation of the risk score in primary care. This could improve the implementation process, if possible, risks or malfunction are identified early.

To be effective, computer decision tools must be used, and the use of the tool is affected by views of clinicians on the potential of the tool and how it was developed. This requires training on the functionality of the tool and how it was developed. Furthermore, the development of a pop-up alert should involve a multidisciplinary team with relevant knowledge on HIV or model development or IT, such as model developers, clinicians and IT specialists [298].

6.8 Work undertaken to have model incorporated in GP systems.

The HIV model could be combined with some of the pilot studies and trials conducted in UK primary care. An example is the Rapid HIV Assessment (RHIVA) trial where nurse-led screening of HIV testing in primary care was promoted [110]. Although this study did not use a prediction model, hence incorporating the prediction model as a POC alert could improve targeting of patients that are offered HIV testing. Other pilot studies conducted in other countries could be adopted in UK. One example of such studies that could be adopted into a pilot study in UK is a prospective interventional study conducted in Spain, which involved 51

primary healthcare centres [265]. This study incorporated HIV POC alerts using HIV indicator conditions, similar to the POC alerts that could be produced from the prediction model developed in this research.

Alternatives to requesting primary care clinicians to directly offer HIV testing to patients should be investigated and if possible, successful adopted. Such alternatives include mail-out HIV testing and self-collection/self-testing. These options would be improved by using the prediction model for targeting of patients with above the threshold risk score. In UK, a strategy for sexually transmitted infections to self-test is already in place, requesting patients to self-test using a home self-testing kit [299]. This strategy could be expanded with the incorporation of the HIV POC alert in targeting patients. One example of the alternatives to HIV testing is the mail-out pilot study conducted in USA where patients were allowed to order a rapid HIV self-testing through a request form on a secure website [72]. Patients received “In-Home HIV self-test kit, prevention materials, and documentation of local resources” and were requested to self-report results as a response to an email [72]. A similar example is the intervention which was scaled up in USA during Covid-19 pandemic, as part of telemedicine models [300]. In this self-testing model, the patient decides and order a test online, collect specimen using accompanying instructions to self-test and report results when they contact a clinician [300]. Another model of “home sampling for testing, combined with treatment and sexual health counselling” was piloted in Maastricht, Netherlands [301].

6.9 Recommendations for future research.

This thesis developed and internally validated a prediction model for HIV using primary care records. Further research is required to externally validate the model, to ensure that it is robust enough for implementation in practice. External validation requires using another primary care database to find how the model fits in compared to findings from IMRD database. External

validation could be temporal (using different time periods) or geographical (different locations, including primary care data from other countries or other UK databases covering different GP practices, such as QResearch, ResearchOne, CPRD Aurum and SAIL (Wales) [302]. External validation of this model is more suited to developed countries mentioned in chapter 2, those countries with HIV infected population with same predictors as UK due to reasons mentioned in section 2.2.5.2 [7] [12].

More research is required if this model is adopted in other developing countries, such as those in Africa and Asia. Further research includes a review of predictors of HIV in these countries and understanding available databases (including how predictors are coded/recorded). Some of the countries have initiated electronic health records, such as Resource and Patient Management System (RPMS) in India while other countries such as South Africa do not have electronic patient records due to "complex infrastructure, network requirements and user resistance" [303] [304].

Prior to implementation of pop-up alerts, more research should be conducted to determine if GPs are interested in using HIV pop-up alerts. This could be followed by a pilot study on implementation of pop-up alerts to find whether HIV pop-up alerts are acceptable.

The data not well recorded in IMRD database included sexuality, country of birth and deprivation. Sexuality is a personal characteristic, and more research can be done to determine acceptability of recording such information in primary care records. The research could look at the clinicians' views and patients' perspective as well as understanding how such data is collected in other data collection programmes. Information on country of birth is another characteristic which is poorly recorded in IMRD database. More research is needed to understand why the data is not recorded and how the situation could be improved. From the

analysis on deprivation in section 4.5.1.2, it seems there is an improvement in recording of deprivation in primary care records over the years. Further research could explore how recording of deprivation in primary care records could be improved.

This research looked at using the prediction score in primary care to increase HIV testing. Apart from improving HIV testing in primary care, BHIVA guidelines recommended alternatives including increasing HIV testing via self-sampling ordered online. This could be improved if patients have a tool that assist them in self-identification. Hence, it would be useful to understand if patients are willing to use a prediction tool that aid them in self-identifying themselves for HIV testing. Therefore, further research is required to ascertain whether a prediction model could be developed for use in a website or App for patients to self-identify if they are at high risk and require to self-sample. There are at least two options that can be researched on including 1) use primary care data to identify patients at risk and then send them a self-testing kit, 2) give patients access to an app or website to enter their characteristics (age, sex, ethnic group, other information including clinical data) and if they meet certain criteria, they can be offered a self-testing kit for free and 3) combine the two options above, where primary care data is used to identify patients, send a link so that the patients can anonymously enter their own risk factors to a website, ask them to order test kits and then enter HIV test result.

The first option has the following advantages and disadvantages:

Advantages

1) Using existing primary care data and the model from this research that is already developed. This option utilises existing primary care databases such as IMRD or CPRD, and the model

developed in this research (after external validation). The second option would require developing a new model that can be used by patients.

2) Not relying on patients being proactive. The GPs identify patients and recommend HIV testing while the second and third options require patient involvement in the identification process.

3) Almost complete coverage of UK population. The model developed in this research could be applied to primary care in the UK because it was developed using UK population. Additionally, it would cover almost all patients in UK since nearly all the UK population is registered with a GP [114].

Disadvantages

1) Missing or under-recorded important information on risk factors such as sexual behaviour, drug misuse, ethnicity and country of birth. As shown in section 4.6.1.1, some risk factors are not always recorded in primary care records and might affect the performance of the model.

2) Increased staff time/cost. Implementation of this option will increase the consultation time and cost since clinicians have to discuss with patients about HIV testing in addition to booked consultation time. Additionally, there are pathology costs involved with HIV testing.

The advantages and disadvantages of second options are:

Advantages

1) Patients can enter complete information on important risk factors which they may not want to disclose. A study on use of computers to enter alcohol intake and another on antenatal problems revealed that patients shared honest information with a computer more than they did with clinicians [305] [306].

2) No need for consultation in primary care prior to test. Patients enter symptoms in the App and act on the results without consultation with their GP.

Disadvantages

1) Data in primary care is not included (might be relevant but patient may be unaware). The App will not have access to primary care records and patients will enter the symptoms they can recall.

2) Patient needs to be proactive and go to website or app to enter data. This option relies on patients' willingness to find out if they have HIV or not and might not be as successful if patients are not proactive.

3) There is no validated prediction model for this option. There is no model developed and validated for use in a patient App. Hence, more research would be required to develop and validate the model and in developing the App.

Development of a patient facing tool should consider the findings from a qualitative study on efficacy of HIV-related mHealth interventions, which include "avoid an exclusive focus on HIV, be tailored and personalised, come from a trusted source, allay fears and focus on support and health benefits" [90].

6.10 Conclusion

The aim of this thesis was to develop and internally validate a prediction model which could be used in targeting patients likely to be HIV positive in primary care that should be offered a test. The research identified predictors of HIV recorded in electronic primary care records, which could be used in developing point-of-care or pop-up alerts. The top five predictors associated with a risk of HIV infection identified in the prediction model were Kaposi's

sarcoma, pneumocystis carinii, progressive multifocal leukoencephalopathy, being of black ethnicity and syphilis. The use of pop-up alerts from the prediction model satisfies the principle of screening for HIV testing in primary care, which requires that the target population for testing should be clearly defined. The thesis further identified factors that should be considered during the development of the pop-up alert. This would assist in satisfying the principle of screening which requires coordination and integration of the testing programme in primary care.

References

[1] McKnight A, “Quick guide: HIV,” *Current Biology,* vol. 9, no. 7, 1999.

[2] LaBrie M, “HIV/AIDS,” in *Palliative Medicine: A case-based manual*, Oxford, Oxford Scholarship Online, 2012.

[3] Masur H, Michelis MA, Greene JB, “An outbreak of community-acquired Pneumocystis carinii pneumonia: initial manifestation of cellular immune dysfunction,” *N. Engl. J. Med.,* vol. 305, no. 24, p. 1431–8, December 1981.

[4] Begovac J, Lepej SZ, Kniewald T, Lisić M, “Biological, Epidemiological and Clinical Basis of Understanding Human Immunodeficiency Virus Infection,” *Coll. Antropol.,* vol. 25, no. 1, p. 111–126, 2001.

[5] UNAIDS and WHO, “Quality and Coverage of HIV Sentinel Surveillance With a brief History of the HIV/AIDS Epidemic,” in *Workshop on Hiv/Aids and Adult Mortality in Developing Countries*, New York, 2003.

[6] Sharp PM and Hahn BH, “Origins of HIV and the AIDS Pandemic,” *Cold Spring Harb Perspect Med 2011,* vol. 1, no. 1, 2011.

[7] World Health Organisation, “HIV/AIDS factsheet,” World Health Organisation, Geneva, Switzerland, 2015.

[8] Halkitis PN, “The AIDS Generation: Stories of Survival and Resilience,” *Social History of Medicine,* vol. 28, no. 1, 2014.

[9] Jones J and Salazar L, “A Historical Overview of the Epidemiology of HIV/AIDS in the United States,” in *Understanding the HIV/AIDS Epidemic in the United States The Role of Syndemics in the Production of Health Disparities*, © Springer International Publishing, 2016.

[10] Gold JWM, “The Diagnosis and Management of Hiv Infection Management of The Hiv-Infected Patient,” *Medical Clinics of North America,* vol. 80, no. 6, pp. 1283-1307, November 1996 .

[11] Nakagawa F, May M and Phillips A, “Life expectancy living with HIV: recent estimates and future implications,” *Current opinion Infectious diseases,* pp. 17-25, 2013.

[12] Hessol NA, Gandhi M and Greenblatt RM, “Epidemiology and natural history of HIV infection in women,” U.S. Department of Health and Human Services, Health Resources and Services Administration, HIV/AIDS Bureau, 2005.

[13] Baumann RJ and Espinosa PS, “Neuroepidemiology of HIV/AIDS,” in *Handbook of Clinical Neurology: HIV/AIDS and the Nervous System*, Oxford, Elsevier, 2007.

[14] O’Cofaigh E and Lewthwaite P, “Fundamentals of HIV Natural history of HIV and AIDS,” *Medicine,* vol. 41, no. 8, pp. 411-16, 2013.

[15] Romanelli F and Matheny SC, “HIV Infection: The Role of Primary Care,” *Am Fam Physician,* vol. 80, no. 9, pp. 946-952, 2009.

[16] Simon V, Ho DD and Karim QA, “HIV/AIDS epidemiology, pathogenesis, prevention, and treatment,” *Lancet,* vol. 368, no. 9534, p. 489–504, 2006.

[17] Weinberg JL and Kovarik CL, “The WHO Clinical Staging System for HIV/AIDS,” *American Medical Association Journal of Ethics,* vol. 12, no. 3, pp. 202-206, March 2010.

[18] World Health Organization (WHO), Consolidated Guidelines on the Use of Antiretroviral Drugs for Treating and Preventing HIV Infection: Recommendations for a Public Health Approach, 2nd ed., Geneva: World Health Organization, 2016.

[19] Centers for disease control and prevention (CDC), “About HIV,” Centers for disease control and prevention (CDC), 2022. [Online]. Available: https://www.cdc.gov/hiv/basics/whatishiv.html.

[20] British HIV Association (BHIVA), “UK National Guidelines for HIV Testing 2008,” British HIV Association (BHIVA), 2008. [Online]. Available: https://www.bhiva.org/HIV-testing-guidelines. [Accessed 28 11 2016].

[21] Simmons RD, Ciancio BC, Kall MM, Rice BD and Delpech VC, “Ten-year mortality trends among persons diagnosed with HIV infection in England and Wales in the era of antiretroviral therapy: AIDS remains a silent killer,” *HIV Medicine,* vol. 14, p. 596–604, 2013.

[22] Melhuish A and Lewthwaite P, “Natural history of HIV and AIDS,” *Medicine,* vol. 46, no. 6, pp. 356 - 361, 2018.

[23] Croxford S, Kitching A, Desai S, Kall M, Edelstein M, Skingsley A, Burns F, Copas A, Brown AE, Sullivan AK and Delpech V, “Mortality and causes of death in people diagnosed with HIV in the era of highly active antiretroviral therapy compared with the general population: an analysis of a national observational cohort,” *Lancet Public Health,* vol. 2, p. e35–46, 2017.

[24] Eyawo O, Franco-Villalobos C, Hull MW, Nohpal A, Samji H, Sereda P, Lima VD, Shoveller J, Moore D, Montaner JSG, Hogg RS and for the Comparative Outcomes And Service Utilization Trends (COAST) study, “Changes in mortality rates and causes of death in a population-based cohort of persons

living with and without HIV from 1996 to 2012," *BMC Infectious Diseases,* vol. 17, no. 174, 2017.

[25] Handy P, "HIV: diagnosis and assessment," in *Oxford Handbook of Genitourinary Medicine, HIV, and Sexual Health*, 3rd ed., Oxford, Oxford University Press, 2011.

[26] Madge S, Matthews P, Singh S and Theobald N, HIV in Primary Care, London: Medical Foundation for AIDS & Sexual Health (MedFASH), 2011.

[27] Ly TD, Ebel A, Faucher V, Fihman V and Laperche S, "Could the new HIV combined p24 antigen and antibody assays replace p24 antigen specific assays?," *Journal of Virological Methods,* vol. 143, p. 86–94, 2007.

[28] Fanmi AN, Ramière C, Tardy JC and André P, "Real-life evaluation of a human immunodeficiency virus screening algorithm using a single combined p24 antigen–antibody assay.," *Eur J Clin Microbiol Infect Dis,* vol. 32, p. 425–430, 2013.

[29] Heuvel Van Den A, Smet H , Prat I , Sands A, Urassa W, Fransen K and Crucitti T, "Laboratory evaluation of four HIV/syphilis rapid diagnostic tests," *BMC Infectious Diseases,* vol. 19, no. 1, 2019.

[30] Gliddon HD, Peeling RW, Kamb ML, Toskin I, Wi TE and Taylor MM, "A systematic review and meta-analysis of studies evaluating the performance and operational characteristics of dual point-of-care tests for HIV and syphilis," *Sex Transm Infect,* vol. 93, p. S3–S15, 2017.

[31] Prost A, Griffiths CJ, Anderson J, Wight D and Hart G J, "Feasibility and acceptability of offering rapid HIV tests to patients registering with primary care in London (UK): a pilot study," *Sex Transm Infect . doi:10.1136/sti.2008.033233,* vol. 85, p. 326–329, 2009.

[32] Turner SD, Anderson K, Slater M, Quigley L, Dyck M, and Guiang CB, "Rapid Point-of-Care HIV Testing in Youth: A Systematic Review," *Journal of Adolescent Health,* vol. 53, p. 683e691, 2013.

[33] Agustí C, Fernàndez-López L, Mascort J, Carrillo R, Aguado C, Montoliu A, Puigdangolas X, De la Poza M, Rifà B and Casabona J, "Attitudes to rapid HIV testing among Spanish General Practitioners," *HIV Medicine ,* vol. 14 , no. Suppl. 3, p. 53–56, 2013.

[34] Yin Z, Brown AE, Hughes G, Nardone A, Gill ON, Delpech VC and contributors, "HIV in the United Kingdom 2014 Report: data to end 2013," Public Health England, London, November 2014.

[35] Damery S, Nichols L, Holder R, Ryan R, Wilson S, Warmington S, Stokes-Lampard H and Manavi K, "Assessing the predictive value of HIV indicator

conditions in general practice: a case–control study using the THIN database," *British Journal of General Practice,* pp. 370-377, June 2013.

[36] Wohlgemut J, Lawes T and Laing RBS, "Trends in missed presentations and late HIV diagnosis in a UK teaching hospital: a retrospective comparative cohort study.," *BMC Infectious Diseases,* vol. 12, no. 72, 2012.

[37] Skarbinski J, Rosenberg E, Paz-Bailey G, Hall HI, Rose CE, Viall AH, Fagan JL, Lansky A and Mermin JH, "Human Immunodeficiency Virus Transmission at Each Step of the Care Continuum in the United States.," *JAMA Intern Med,* vol. 175, no. 4, p. 588–596, 2015.

[38] European Centre for Disease Prevention and Control (ECDC), "HIV testing: increasing uptake and effectiveness in the European Union," European Centre for Disease Prevention and Control (ECDC), Stockholm, 2010.

[39] Fuster-RuizdeApodaca MJ, Laguia A, Molero F, Toledo J, Arrillaga A and Jaen A, "Psychosocial determinants of HIV testing across stages of change in Spanish population: a cross-sectional national survey," *BMC Public Health (2017) 17:234,* vol. 17, no. 234, 2017.

[40] Cohen MS, Gay C, Kashuba ADM, Blower S and Paxton L, "Narrative review: Antiretroviral therapy to prevent the sexual transmission of HIV-1," *Ann Intern Med,* vol. 146, no. 8, p. 591–601, 2007.

[41] Matic S, "Twenty-five years of HIV /AIDS in Europe," in *HIV/AIDS in Europe: Moving from death sentence to chronic disease management* , Geneva, World Health Organization, 2006.

[42] May M, Gompels M, Delpech V, Porter K, Post F, Johnson M, Dunn D, Palfreeman A, Gilson R, Gazzard B, Hill T, Walsh J, Fisher M, Orkin C, Ainsworth J, Bansi L, Phillips A, Leen C, Nelson M, Anderson J and Sabin C, "Impact of late diagnosis and treatment on life expectancy in people with HIV-1: UK Collaborative HIV Cohort (UK CHIC) Study. BMJ. 2011 Oct 11;343:d6016. doi: 10.1136/bmj.d6016," *BMJ.,* vol. 343, Oct 2011.

[43] Brown AE, Kall MM, Smith RD, Yin Z, Hunter A and Delpech VC, "Auditing National HIV Guidelines and Policies: The United Kingdom CD4 Surveillance Scheme," *The Open AIDS Journal,* vol. 6, no. Suppl 1: M12, pp. 149-155, 2012.

[44] Wellesley R, Whittle A, Figueroa J, Anderson J, Castles R, Boomla K, Griffiths C and Leber W, "Does general practice deliver safe primary care to people living with HIV? A case-notes review," *British Journal of General Practice,* pp. e655-e661, 2015.

[45] Robbins GK, Spritzler JG, Chan ES, Asmuth DM, Gandhi RT, Rodriguez BA, Skowron G, Skolnik PR, Shafer RW, Pollard RB and AIDS Clinical Trials Group 384 Team, "Incomplete reconstitution of T cell subsets on combination

antiretroviral therapy in the AIDS Clinical Trials Group protocol 384.," *Clinical infectious diseases: an official publication of the Infectious Diseases Society of America,* vol. 48, no. 3, pp. 350-61, 2009.

[46] Darcis G, Lambert I, Sauvage A, Frippiat F, Meuris C, Uurlings F, Lecomte M, Léonard P, Giot J, Fombellida K, Vaira D and Moutschen M, "Factors associated with late presentation for HIV care in a single Belgian reference center: 2006–2017," *Scientific Reports,* vol. 8, no. 8594 , 2018.

[47] Cohen MS, Chen YQ, McCauley M, Gamble T, Hosseinipour MC, Kumarasamy N, Hakim JG, Kumwenda J, Grinsztejn B, Pilotto JH, Godbole SV, Mehendale S, Chariyalertsak S, Santos BR, Mayer KH, Hoffman IF, Eshleman SH, Piwowar-Manning E, Wang L, Makhema J, et al, "Prevention of HIV-1 infection with early antiretroviral therapy," *N Engl J Med,* vol. 365, no. 6, pp. 493-505, 11 Aug 2011.

[48] Cohen MS, Chen YQ, McCauley M, Gamble T, Hosseinipour MC, Kumarasamy N, Hakim JG, Kumwenda J, Grinsztejn B, Pilotto JH, Godbole SV, Chariyalertsak S, Santos BR, Mayer KH, Hoffman IF, Eshleman SH, Piwowar-Manning E, Cottle L, Zhang XC, et al , "Antiretroviral Therapy for the Prevention of HIV-1 Transmission," *N Engl J Med,* vol. 375, no. 9, pp. 830-9, 1 Sep 2016.

[49] British HIV Association (BHIVA), "BHIVA guidelines for the treatment of HIV-1-positive adults with antiretroviral therapy 2015 (2016 interim update)," 2016. [Online]. Available: https://www.bhiva.org/HIV-1-treatment-guidelines. [Accessed 14 Apr 2019].

[50] Fleishman JA, Yehia BR, Moore RD, Gebo KA and HIV Research Network, "The economic burden of late entry into medical care for patients with HIV infection," *Med Care,* vol. 48, no. 12, pp. 1071-9, Dec 2010.

[51] Krentz HB, Auld MC, Gill MJ, "The high cost of medical care for patients who present late (CD4 <200 cells/microL) with HIV infection.," *HIV Med,* vol. 5, p. 93–8, 2004.

[52] Halperin J, Katz M, Pathmanathan I, Myers L, Van Sickels N, Seal PS and Richey LE, "Early HIV Diagnosis Leads to Significantly Decreased Costs in the First 2 Years of HIV Care in an Urban Charity Hospital in New Orleans," *Journal of the International Association of Providers of AIDS Care (JIAPAC),* vol. 16, no. 6, pp. 527-530, 2017.

[53] Beck EJ, Mandalia S, Sangha R, Sharott P, Youle M, Baily G, Brettle R, Gompels M, Johnson M, McCarron B, Ong E, Pozniak A, Schwenk A, Taylor S, Walsh J, Wilkins E, Williams I and Gazzard B, "The Cost Effectiveness of Early Access to HIV Services and Starting cART in the UK 1996-2008," *PLoS One,* vol. 6, no. 12, p. e27830, 2011.

[54] Watson S, “The History of HIV Treatment: Antiretroviral Therapy and More”.

[55] Tseng A, Seet J and Phillips EJ, “The evolution of three decades of antiretroviral therapy: challenges, triumphs and the promise of the future,” *British Journal of Clinical Pharmacology,* vol. 79, no. 2, p. 182–194, 2004.

[56] Phanuphak N and Gulick RM, “HIV treatment and prevention 2019: current standards of care.,” *Curr Opin HIV AIDS,* vol. 15, no. 1, pp. 4-12, Jan 2020.

[57] World Health Organisation [WHO), “Global Health Observatory (GHO) data: HIV/AIDS,” July 2021. [Online]. Available: https://www.who.int/data/gho/data/themes/hiv-aids. [Accessed 4 October 2021].

[58] Joint United Nations Programme on HIV/AIDS (UNAIDS), “AIDSinfo country fact sheets,” UNAIDS, 2021. [Online]. Available: http://aidsinfo.unaids.org/. [Accessed 10 October 2021].

[59] World Health Organisation (WHO), “Global progress report on HIV, viral hepatitis and sexually transmitted infections, 2021,” 2021. [Online]. Available: http://apps.who.int/iris/bitstream/handle/10665/342808/9789240030985-eng.pdf. [Accessed 15 October 2021].

[60] Dougan S, Evans BG, Macdonald N, Goldberg DJ, Gill ON, Fenton KA, Elford J, “HIV in gay and bisexual men in the United Kingdom: 25 years of public health surveillance.,” *Epidemiology and infection,* vol. 136, no. 2, pp. 145-156, 2008.

[61] Public Health England, “HIV surveillance systems,” Public Health England, January 2018. [Online]. Available: https://www.gov.uk/guidance/hiv-surveillance-systems. [Accessed 11 March 2019].

[62] United Kingdom Health Security Agency (UKHSA), “United Kingdom, Country and region HIV data tables, No. 2: 2021,” December 2021. [Online]. Available: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1037993/hiv-country-region-data-tables-2021.ods. [Accessed 2 December 2021].

[63] Nash S, Desai S, Croxford S, Guerra L, Lowndes C, Connor N, Gill ON., “Progress towards ending the HIV epidemic in the United Kingdom: 2018 Report - Data to end of December 2017.,” Public Health England, London, 2018.

[64] Martin V, Shah A, Mackay N, Lester J, Newbigging-Lister A, Connor N, Brown AE, Sullivan AK, Delpech VC and contributors, “HIV testing, new HIV diagnoses, outcomes and quality of care for people accessing HIV services: 2021 report. The annual official stat”.

[65] Public Health England (PHE), “HIV: annual data tables,” [Online]. Available: https://www.gov.uk/government/statistics/hiv-annual-data-tables.

[66] Public Health England (PHE), “The impact of the COVID-19 pandemic on prevention, testing, diagnosis and care for sexually transmitted infections, HIV and viral hepatitis in England Provisional data: January to September 2020,” Public Health England, London, 2020.

[67] United Kingdom Health Security Agency (UKHSA), “HIV in the UK 2021 slide set,” December 2021. [Online]. Available: https://khub.net/documents/135939561/561828170/hiv+in+united+kingdom+2021+slide+set.odp/bd35891b-42b2-2888-2c22-14405c35480f. [Accessed 2 December 2021].

[68] Wenlock RD, Shillingford C, Mear J, Churchill D, Vera JH and Dean G, “The impact of COVID- 19 on HIV testing in the UK’s first Fast-Track HIV city,” *HIV Medicine,* vol. 00, p. 1–7, 2022.

[69] Moitra E, Tao J, Olsen J, Shearer RD, Wood BR, Busch AM, LaPlante A, Baker JV and Chan PA, “Impact of the COVID-19 pandemic on HIV testing rates across four geographically diverse urban centres in the United States: An observational study. Lancet Reg Health A,” *Lancet Reg Health Am,* vol. 7, Mar 2022.

[70] Mitchell KM, Dimitrov D, Silhol R, Geidelberg L, Moore M, Liu A, Beyrer C, Mayer KH, Baral S and Boily MC, “Estimating the potential impact of COVID-19-related disruptions on HIV incidence and mortality among men who have sex with men in the United States: a modelling study,” *Lancet HIV,* 19 Feb 2021.

[71] Mitchell HD, Vilaplana TG, Mandal S, Ratna N, Glancy M, Shah A, Simmons R, Penman C, Kirsebom F, Costella A, Brown AE, Mohammed H, Delpech V, Sinka K, Hughes G and UK Health Security Agency, “Effects of COVID-19 Pandemic Response on Service Provision for Sexually Transmitted Infections, HIV, and Viral Hepatitis, England,” *Emerging infectious diseases,* vol. 28, no. 3, p. 739–742..

[72] Fistonich GM, Troutman KM and Visconti AJ, “A Pilot of Mail-Out HIV and Sexually Transmitted Infection Testing in Washington, District of Columbia During the COVID-19 Pandemic,” *Am J Prev Med,* vol. 61, no. 5 Suppl 1, pp. S16-S25, Nov 2021.

[73] Joint United Nations Programme on HIV/AIDS (UNAIDS), 2016–2021 Strategy On the Fast-Track to end AIDS, Geneva: Joint United Nations Programme on HIV/AIDS (UNAIDS), 2015.

[74] Joint United Nations Programme on HIV/AIDS (UNAIDS), 90-90-90 An ambitious treatment target to help end the AIDS epidemic, Geneva: Joint United Nations Programme on HIV/AIDS (UNAIDS), 2014.

[75] Department of Health and Social Care (DHSC), “Towards Zero: the HIV Action Plan for England - 2022 to 2025,” 2021.

[76] HIV Commission, “Transformation,” 2022. [Online]. Available: https://www.hivcommission.org.uk/final-report-and-recommendations/transformation/. [Accessed 9 March 2022].

[77] Glew S, Pollard A, Hughes L and Llewellyn C, “Public attitudes towards opt-out testing for HIV in primary care: a qualitative study,” *British Journal of General Practice,* vol. 64, no. 619, pp. e60-e66., 2014.

[78] White DAE, Scribner AN, Martin ME and Tsai S, “A Comparison of Patient Satisfaction with Emergency Department Opt-In and Opt-Out Rapid HIV Screening,” *AIDS Research and Treatment,* vol. 2012.

[79] Parliamentary Office of Science and Technology, “HIV in the UK,” Parliamentary Office of Science and Technology, Londone, 2007.

[80] Presanis AM, Harris RJ, Kirwan PD, Miltz A, Croxford S, Heinsbroek E, Jackson CH, Mohammed H, Brown AE, Delpech VC, Gill ON and Angelis DD, “Trends in undiagnosed HIV prevalence in England and implications for eliminating HIV transmission by 2030: an evidence synthesis model,” *Lancet Public Health,* vol. 6, p. e739–51, 2021.

[81] The UK Collaborative Group for HIV and STI Surveillance, “Testing Times. HIV and other Sexually Transmitted Infections in the United Kingdom,” Health Protection Agency, Centre for Infections, London, 2007.

[82] Elmahdi R, Gerver SM, Gomez Guillen G, Fidler S, Cooke, G and Ward H, “Low levels of HIV test coverage in clinical settings in the UK: a systematic review of adherence to 2008 guidelines,” *Sex Transm Infect,* vol. 90, p. 119–124, 2014.

[83] Croxford S Yin Z Kall M, Burns F, Simmons R, Copas A, Ireland G, Kirwan P, Chau C and Delpech V, “Where do we diagnose HIV infection? Monitoring new diagnoses made in nontraditional settings in England, Wales and Northern Ireland,” *HIV Medicine,* vol. 19, pp. 465-474, 7 August 2018.

[84] British HIV Association (BHIVA), “British HIV Association/British Association for Sexual Health and HIV/British Infection Association Adult HIV Testing guidelines 2020,” [Online]. Available: https://www.bhiva.org/HIV-testing-guidelines.

[85] Blondell S, Kitter B, Griffin MP and Durham J, “Barriers and Facilitators to HIV Testing in Migrants in High-Income Countries: A Systematic Review,” *AIDS Behav,* vol. 19, no. 2012, 2015.

[86] Deblonde J, De Koker P, Hamers FF, Fontaine J, Luchters S and Temmerman M, “Barriers to HIV testing in Europe: a systematic review,” *European Journal of Public Health,,* vol. 20, no. 4, p. 422–432, 2010.

[87] Dodds C, Hickson F, Weatherburn P, Reid D, Hammond G, Jessup C and Adegbite G, “BASS Line 2007 Survey: Assessing the sexual and HIV prevention needs of African people in England,” Sigma Research., London, 2008.

[88] John O, *HIV prevention among black Africans in England: a complex challenge. A Race Equality Foundation Briefing Paper.,* 2009.

[89] Prost A, “A review of research among black African communities affected by HIV in the UK and Europe.,” MRC Social and Public Health Sciences Unit Occasional paper no. 15, Glasgow, 2006.

[90] Evans C, Turner K, Suggs LS, Occa A, Juma A and Blake H, “Developing a mHealth intervention to promote uptake of HIV testing among African communities in the UK: a qualitative study,” *BMC Public Health,* vol. 16, no. 656, 2016.

[91] Deblonde J, Hamers FF, Callens S, Raquel Lucas R, Barros H, Rüütel K, Hemminki E and Temmerman M, “HIV testing practices as reported by HIV-infected patients in four European countries,” *AIDS Care,* vol. 26, no. 4, pp. 487-496.

[92] Bedert M, · Davidovich U, · de Bree G, van Bilsen W, van Sighem A, Zuilhof W, Brinkman K,· van der Valk M and de Wit J, “Understanding Reasons for HIV Late Diagnosis: A Qualitative Study Among HIV-Positive Individuals in Amsterdam, The Netherlands,” *AIDS and Behavior,* vol. 25, p. 2898–2906, 2021.

[93] Hachfeld A, Darling K, Calmy A, Ledergerber B, Weber R, Battegay M, Wissel K, Di Benedetto C, Fux CA, Tarr PE, Kouyos R, Ruggia LS, Furrer HJ and Wandeler G for the Swiss HIV Cohort Study, “Why do sub-Saharan Africans present late for HIV care in Switzerland?,” *HIV Medicine, ,* vol. 20, pp. 418--423, 2019.

[94] Fakoya I, Prost A, Burns F and Hart G, “Increasing the uptake of HIV testing to reduce undiagnosed infection and prevent transmission among black African communities living in England: Barriers to HIV testing-Full report”.

[95] Deblonde J, Van Beckhoven D, Loos J, Boffin N, Sasse A, Nöstlinger C, Supervie V and HERMETIC Study Group, “HIV testing within general practices

in Europe: a mixed-methods systematic review," *BMC Public Health,* vol. 18, no. 1191 , 2018.

[96] Youssef E, Cooper V, Delpech V, Davies K and Wright J, "Barriers and facilitators to HIV testing in people age 50," *Clinical Medicine,* vol. 17, no. 6, p. 508–20, 2017 .

[97] Elgalib A, Fidler S and Sabapathy K, "Hospital-based routine HIV testing in high-income countries: a systematic literature review," *HIV Medicine,* vol. 19, p. 195—205, 018.

[98] Westrop SJ, Hartney T, Gill N, Delpech V, Crook P and Nardone A, "Addressing Late HIV Diagnosis through Screening and Testing: An Evidence Summary," Public Health England, London, 2014.

[99] Leber W, Anderson J and Griffith C, "HIV testing in Europe: how can primary care contribute," *Sex Transm Infect,* vol. 91, no. 7, 2015.

[100] Geter A, Herron AR and Sutton MY, "HIV-Related Stigma by Healthcare Providers in the United States: A Systematic Review," *AIDS Patient Care STDS,* vol. 32, no. 10, p. 418–424, October 2018.

[101] Drumhiller K, Geter A, Elmore K, Gaul Z and Sutton MY, "Perceptions of Patient HIV Risk by Primary Care Providers in High-HIV Prevalence Areas in the Southern United States," *AIDS PATIENT CARE and STDs,* vol. 34, no. 3, 2020.

[102] Tan K and Black BP, "A systematic review of healthcare provider perceived barriers and facilitators to routine HIV testing in primary care settings in the Southeastern United States," *J Assoc Nurses AIDS Care,* vol. 29, no. 3, p. 357–370, 2018.

[103] Bagchi AD and Davis T, "Clinician Barriers and Facilitators to Routine HIV Testing: A Systematic Review of the Literature," *Journal of the International Association of Providers of AIDS Care,* vol. 19, pp. 1-9, 2020.

[104] Fakoya I, Reynolds R, Caswell G, Shiripinda I., "Barriers to HIV testing for migrant black Africans in Western Europe.," *HIV Med,* vol. 9, no. Suppl 2, p. 23–25, 2008.

[105] Leber W, Beresford L, Nightingale C, Barbosa EC, Morris S, El-Shogri F, McMullen H, Boomla K, Delpech V, Brown A, Hutchinson J, Apea V, Symonds M, Gilliham S, Creighton S, Shahmanesh M, Fulop N, Estcourt C, Anderson J, Figueroa J and Griffiths C, "Effectiveness and cost-effectiveness of implementing HIV testing in primary care in East London: protocol for an interrupted time series analysis.," *BMJ Open,* vol. 7, no. 12, p. e018163., 2017.

[106] Sudarshi D, Pao D, Murphy G, Parry J, Dean G and Fisher M, “Missed opportunities for diagnosing primary HIV infection.,” *Sex Transm Infect,* vol. 84, no. 1, p. 14–16, 2008.

[107] Sullivan AK, Curtis H, Sabin CA and Johnson MA, “Newly diagnosed HIV infections: review in UK and Ireland,” *BMJ,* vol. 330, 2005.

[108] Dorward J, Chinnaraj A, Garrett N, Apea V and Leber W, “Opportunities for earlier diagnosis of HIV in general practice,” *Sex Transm Infect,* vol. 88, no. 7, p. 524, 2012.

[109] Leber W, McMullen H, Anderson J, Marlin N, Santos AC, Bremner S, Boomla K, Kerry S, Millett D, Mguni S, Creighton S, Figueroa J, Ashcroft R, Hart G, Delpech V, Brown A, Rooney G, Sampson M9, Martineau A, Terris-Prestholt F and Griffiths C, “Promotion of rapid testing for HIV in primary care (RHIVA2): a cluster-randomised controlled trial,” *Lancet HIV,* vol. 2, no. 6, pp. e229-35, Jun 2015.

[110] Leber W, Panovska-Griffiths J, Martin P, Morris S, Estela Capelas Barbosab , Estcourt C, Hutchinson J, Shahmanes M, El-Shogria F and Boomla K, “Evaluating the impact of post-trial implementation of RHIVA nurse-led HIV screening on HIV testing, diagnosis and earlier diagnosis in general practice in London, UK,” *EClinicalMedicine ,* vol. 19 , 2020.

[111] Agustí C, Martín-Rabadán M, Zarco J, Aguado C, Carrillo R, Codinachs R, Carmona JM and Casabona J, “Early diagnosis of HIV in Primary Care in Spain. Results of a pilot study based on targeted screening based on indicator conditions, behavioral criteria and region of origin,” *Aten Primaria,* vol. 50, no. 3, pp. 159-165, Mar 2018.

[112] Baggaley RF, Irvine MA, Leber W, Cambiano V, Figueroa J, McMullen H, Anderson J, Santos AC, Terris-Prestholt F, Miners A, Hollingsworth TD and Griffiths CJ, “Cost-effectiveness of screening for HIV in primary care: a health economics modelling analysis,” *Lancet HIV.,* vol. 4, no. 10, pp. e465-e474, October 2017.

[113] Long EF, Mandalia R, Mandalia S, Alistar SS, Beck EJ and Brandeau ML, “Expanded HIV Testing in Low-Prevalence, High-Income Countries: A Cost-Effectiveness Analysis for the United Kingdom,” *PLOS One,* vol. 9, no. 4, p. e95735, 2014.

[114] Burns FM, Johnson AM, Nazroo J, Ainsworth J, Anderson J, Fakoya A, Fakoya I, Hughes A, Jungmann E, Sadiq ST, Sullivan AK and Fenton KA, “Missed opportunities for earlier HIV diagnosis within primary and secondary healthcare settings in the UK,” *AIDS,* vol. 22, no. 1, p. 115–122, 2008.

[115] Avery AK, Del Toro M and Caron A, “Increases in HIV screening in primary care clinics through an electronic reminder: an interrupted time series.,” *BMJ Quality & Safety,* vol. 23, no. 3, pp. 250-6, Mar 2014.

[116] Kershaw C, Taylor JL, Horowitz G, Brockmeyer D, Libman H, Kriegel G and Ngo L, “Use of an electronic medical record reminder improves HIV screening,” *BMC Health Services Research,* vol. 18, no. 14, 2018.

[117] Davies CF, Kesten JM, Gompels M, Horwood J, Crofts M, Billing A, Chick C and May MT, “Evaluation of an educational intervention to increase HIV-testing in high HIV prevalence general practices: a pilot feasibility stepped-wedged randomised controlled trial,” *BMC Fam Pract,* vol. 19, no. 1, p. 195, 13 Dec 2018.

[118] Goetz MB, Hoang T, Bowman C, Knapp H, Rossman B, Smith R, Anaya H, Osborn T, Gifford AL and Asch SM, “A System-wide Intervention to Improve HIV Testing in the Veterans Health Administration,” *Gen Intern Med,* vol. 23, no. 8, p. 1200–7, 2008.

[119] Bokhour BG, Saifu1 H, Goetz MB, Fix GM, Burgess J, Fletcher MD, Knapp H and Asch SM, “The role of evidence and context for implementing a multimodal intervention to increase HIV testing,” *Implementation Science,* vol. 10, no. 22, 2015.

[120] Grandy A, “Electronic health records: how the United States can learn from the French Dossier Médical Personnel,” *Wisconsin International Law Journal,* vol. 30, no. 2, pp. 374-400, 2012.

[121] Khan ND, Harrison SE and Rose PW, “Validity of diagnostic coding within the General Practice Research Database: a systematic review,” *British Journal of General Practice,* 2010.

[122] Gray K, Orr D and Majeed A, “Use of Read codes in diabetes management in a south London primary care group: implications for establishing disease registers,” *BMJ,* vol. 326, p. 1130, 2003.

[123] Berg M, “Implementing information systems in health careorganizations: myths and challenges,” *International Journal of Medical Informatics,* vol. 64, pp. 143-156, 2001.

[124] Swinglehurst D, “Displays of authority in the clinical consultation: A linguistic ethnographic study of the electronic patient record,” *Social science and medicine,* vol. 118, pp. 17-26, 2014.

[125] Millman A, Lee N and Brooke A, “ABC of medical computing. Computers in general practice-1,” *BMJ,* vol. 311, pp. 800-2, 1995.

[126] Watson J, Nicholson BD and Hamilton W, “Identifying clinical features in primary care electronic health record studies: methods for codelist development,” *BMJ Open,* vol. 7, p. e019637, 2017.

[127] Carlisle G, “Overview of the national laws on electronic health records in the EU Member States: National Report for United Kingdom (England),” Milieu Ltd and Time.lex, 2013. [Online]. Available: https://ec.europa.eu/health/sites/health/files/ehealth/docs/laws_united_kingdom_en.pdf. [Accessed 14 January 2019].

[128] UK EMR - IQVIA Medical Research Data, “UK EMR - IQVIA Medical Research Data,” [Online]. Available: https://www.iqvia.com/library/fact-sheets/uk-emr-iqvia-medical-research-data.

[129] Clinical Practice Research Datalink (CPRD), “Clinical Practice Research Datalink,” 2018. [Online]. Available: https://www.cprd.com/home.

[130] Qresearch, “Generating new knowledge to improve patient care,” [Online]. Available: https://www.qresearch.org/.

[131] TPP, “Data and research,” [Online]. Available: https://tpp-uk.com/products/data-and-research/.

[132] Chaudhry Z, Mannan F, Gibson-White A, Syed U, Ahmed S, Kousoulis A, Majeed A, “Outputs and Growth of Primary Care Databases in the United Kingdom: Bibliometric Analysis,” *J Innov Health Inform,* vol. 24, no. 3, p. 942, 2017.

[133] World Health Organisation (WHO), “Screening programmes: a short guide. Increase effectiveness, maximize benefits and minimize harm. :,” WHO, Copenhagen, 2020.

[134] Wilson JMG and Jungner G, “Principles and practice of screening for diseases,” *Public Health Paper Number 34,* 1968.

[135] Dobrow MJ, Hagens V, Chafe R, Sullivan T and Rabeneck L, “Consolidated principles for screening based on a systematic review and consensus process,” *Canadian Medical Association Journal,* vol. 190, no. 14, pp. E422-E429, 2018.

[136] World Health Organisation (WHO), “Global update on the health sector response to HIV, 2014,” Geneva, Switzerland, 2014.

[137] Norman I and Griffiths P, “The rise and rise of the systematic review,” *International Journal of Nursing Studies,* vol. 51, no. 1, pp. 1-3, 2014.

[138] Trivedi D, Goodman C, Dickinson A. Gage H, McLaughlin J, Manthorpe J, Ashaye, K and Iliffe S, “A protocol for a systematic review of research on

managing behavioural and psychological symptoms in dementia for community-dwelling older people: evidence mapping and syntheses," vol. 2, no. 70, 2013.

[139] Moher D, Liberati A, Tetzlaff J and Altman DG, "Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement," *BMJ,* vol. 339, pp. 25-35, 2009.

[140] Rumbwere Dube BN, Marshall TM and Ryan RP, "Predictors of human immunodeficiency virus (HIV) infection in primary care: a systematic review protocol," *Systematic Reviews,* vol. 5, p. 158, 2016.

[141] Glasziou P, Irwig L, Bain C and Colditz G, Systematic reviews in health care: A practical guide, Cambridge, United Kingdom: Cambridge University Press., 2001.

[142] Wright RW, Brand RA, Dunn W, and Spindler KP, "How to Write a Systematic Review," *Clinical Orthopaedics and Related Research,* vol. 455, p. 23–29, 2007.

[143] Bruce N, Pope D and Stanistreet D, Quantitative Research Methods for Health Research: a practical interactive guide to epidemiology and statistics, Chichester, UK: J Wiley & Sons Ltd, 2008.

[144] Akobeng AK, "Understanding systematic reviews and meta-analysis," *Arch Dis Child,* vol. 90, p. 845–848, 2004.

[145] Popay J, Roberts H, Sowden A, Petticrew M, Arai L, Rodgers M, Britten N, Roen K and Duffy S, "Guidance on the conduct of narrative synthesis in systematic reviews," 2006.

[146] Fu R, "Quantitative Analysis," in *Systematic Reviews to Answer Health Care Questions*, Wolters Kluwer Health, 2014, pp. 139-166.

[147] National Institute for Health and Clinical Excellence (NICE), "HIV testing:increasing uptake in black Africans," March 2011. [Online]. Available: https://www.nice.org.uk/guidance/ph33. [Accessed 30 June 2016].

[148] National Institute for Health and Clinical Excellence (NICE), "HIV testing: increasing uptake in men who have sex with men," March 2011. [Online]. Available: https://www.nice.org.uk/guidance/ph34. [Accessed 30 June 2016].

[149] Aberg JA, Gallant JE, Anderson J, Oleske JM, Libman H, Currier JS, Stone VE and Kaplan JE, "Primary care guidelines for the management of persons infected with human immunodeficiency virus: Recommendations of the HIV Medicine Association of the Infectious Diseases Society of America," *Clinical Infectious Diseases,* vol. 39, no. 5, pp. 609-629, 2004.

[150] Donovan B and Ross MW, “HIV risk evaluation,” *Australian Family Physician,* vol. 7, pp. 646-650, 2000.

[151] Ellerbrock TV, Chamblee S, Bush TJ, Johnson JW, Marsh BJ, Lowell P, Trenschel RJ, Von Reyn CF, Johnson LS and Horsburgh CR, “Human immunodeficiency virus infection in a rural community in the United States,” *American Journal of Epidemiology,* vol. 160, no. 6, pp. 582-588, 2004.

[152] Ross JD and Goldberg DJ, “Patterns of HIV testing in Scotland: a general practitioner perspective.,” *Scottish medical journal,* vol. 42, no. 4, pp. 108-110, 1997.

[153] Gordon SM and Thompson S, “The Changing Epidemiology of Human-Immunodeficiency-Virus Infection in Older Persons.,” *Journal of the American Geriatrics Society,* vol. 43, no. 1, pp. 7-9, 1995.

[154] Hafner JW and Brillman JC, “Symptomatology of HIV-related illness and community-acquired illness in an HIV-infected emergency department population,” *Annals of Emergency Medicine,* vol. 29, no. 1, pp. 151-157, 1997.

[155] Moran J, Connell J, Burke D and Hall W., “Early HIV infection: Demographic features and risk factors among individuals with early HIV infection in Ireland 2008-2011.,” *Irish journal of medical science,* vol. 181, p. s127, 2012.

[156] Guy RJ, Spelman T, Stoove M, ElHayek C, Goller J, Fairley CK, Leslie D, Tee BK, Roth N, Grulich AE and Hellard ME, “Risk factors for HIV seroconversion in men who have sex with men in Victoria, Australia: Results from a sentinel surveillance system.,” *Sexual Health,* vol. 8, no. 3, pp. 319-329, 2011.

[157] Hodder SL. Justman J, Hughes JP, Wang J, Haley DF, Adimora AA, Del Rio C, Golin CE, Kuo I, Rompalo A, Soto-Torres L, Mannheimer SB, Johnson-Lewis LT, Eshleman SH and El-Sadr WM, “Womens HIV SeroIncidence Study HIV Acquisition Among Women From Selected Areas of the United States A Cohort Study,” *Annals of Internal Medicine,* vol. 158, no. 1, 2013.

[158] Niyonsenga T, Trepka MJ, Lieb S and Maddox LM, “Measuring socioeconomic inequality in the incidence of AIDS: rural-urban considerations.,” *AIDS & Behavior,* vol. 17, no. 2, pp. 700-709, 2013.

[159] Desai M, Bevin MA, Holland H, Darling D, Mukela A, Menon Johansson A and Fox J, “Understanding the health needs and risk behaviours of new HIV patients.,” *HIV Medicine,* vol. 13, p. 22, 2012.

[160] Szerlip MA, Desalvo KB and Szerlip HM, “Predictors of HIV-infection in older adults,” vol. 17, no. 3, pp. 293-304, 2005.

[161] Burchell AN, Allen V, Grewal R, Moravan V, Remis RS, Gardner S, Bayoumi AM, Kaul R, Mazzulli T, McGee F, Millson P, Raboud J and Rourke SB,

"Gonorrhea co-infection among gay and bisexual men in HIV care in Ontario : Trends in testing and diagnosis, 2008-2011.," *22nd Annual Canadian Conference on HIV/AIDS Research, CAHR 2013 Vancouver, BC Canada,* vol. 24, pp. 60A-61A, 2013.

[162] Joore IK, Arts DL, Kruijer MJP, Van Charante EPM,Geerlings SE, Prins JM and Van Bergen JE, "HIV indicator condition-guided testing to reduce the number of undiagnosed patients and prevent late presentation in a high-p," *Sexually transmitted infections,* vol. 91, no. 7, pp. 467-472, 2015.

[163] Landau R, Coker R, Vermeulen E, Touquet R, Fothergill J and Poznansky MC, "Patients unaware of their HIV status present to an inner city accident and emergency department with respiratory complications," vol. 14, no. 5, 1997.

[164] Krauskopf K, Federman DA, Mhango G and Klein RS, "Association of HIV status and non-AIDS comorbid diagnoses in a cohort of older HIV-infected and at-risk men.," *Journal of General Internal Medicine,* vol. 26, p. s58, 2011.

[165] Marder K, Liu XH, Stern Y, Malouf R, Dooneief G, Bell K, Todak G, Joseph M, Sorrell S, Sadr WE, Williams JBW, Ehrhardt A, Stein Z and Gorman J, "Risk of Human-Immunodeficiency-Virus Type 1-Related Neurologic Disease in a Cohort of Intravenous-Drug-Users," *Archives of Neurology,* vol. 52, no. 12, pp. 1174-1182, 1995.

[166] UK-CAB, "BHIVA guidelines writing groups and committees," UK-CAB, 2020. [Online]. Available: https://ukcab.net/community-representation/bhiva-guidelines-writing-groups-and-committees/. [Accessed June 2022].

[167] NHS Digital, "Terminology and classification - Read codes," [Online]. Available: https://digital.nhs.uk/services/terminology-and-classifications/read-codes. [Accessed 22 November 2021].

[168] Benson T, "The history of the Read codes: the inaugural James Read Memorial Lecture 2011," *Informatics in Primary Care,* vol. 19, p. 173–82, 2011.

[169] NHS Digital - UK Terminology Centre (UKTC) , "Read Codes," [Online]. Available: http://webarchive.nationalarchives.gov.uk/20160921145325/http://systems.digital.nhs.uk/data/uktc/readcodes. [Accessed 28 March 2016].

[170] NHS Digital, "Read Codes," [Online]. Available: https://digital.nhs.uk/services/terminology-and-classifications/read-codes. [Accessed 5 July 2018].

[171] Booth N, "What are Read codes?," *Health Libraries Review,* vol. 11, pp. 177-182, 1994.

[172] Scottish Information Management in Practice (SCIMP), “Read Terms for General Practice A Quick Introduction,” [Online]. Available: https://www.scimp.scot.nhs.uk/wp-content/uploads/SCIMP-Guide-to-Read-Codes-1-6.pdf.

[173] Harrison MI, Koppel R and Bar-Lev S, “Consequences of information technologies in healthcare—an interactive sociotechnical,” *J Am Med Inform Assoc,* vol. 14, p. 542–9, 2007.

[174] Maiorana A, Steward WT, Koester KA, Pearson C, Shade SB, Chakravarty D and Myers JJ, “Trust, confidentiality, and the acceptability of sharing HIV-related patient data: lessons learned from a mixed methods study about Health Information Exchanges,” *Implementation Science,* vol. 7, no. 34, 2012.

[175] Department of Health and Royal College of General Practitioners, Good practice guidelines for general practice electronic patient records (version 3.1), London: Department of Health & Royal College of General Practitioners, 2005.

[176] Tate AR, Beloff N, Al-Radwan B, Wickson J, Puri S, Williams T, Van Staa T and Bleach A, “Exploiting the potential of large databases of electronic health records for research using rapid search algorithms and an intuitive query interface,” *J Am Med Inform Assoc,* vol. 0, p. 1–7, 2013.

[177] World Medical Association, Medical Ethics Manual, 3rd edition ed., UK: World Health Communication Associates, 2015.

[178] General Medical Council, Confidentiality, General Medical Council, 2013.

[179] Simpson E, Brown R, Sillence E, Coventry L, Lloyd K, Gibbs J, Tariq S and Durrant AC, “Understanding the Barriers and Facilitators to Sharing Patient-Generated Health Data Using Digital Technology for People Living With Long-Term Health Conditions: A A Narrative Review Frontiers in Public Health,” *Public Health,* vol. 19, 2021.

[180] NHS England, “GP Patient Survey,” [Online]. Available: https://gp-patient.co.uk/. [Accessed 11 April 2016].

[181] Kall M, Kelly C, Auzenbergs M and Delpech V, “Positive Voices: The National Survey of People Living with HIV - findings from the 2017 survey.,” Public Health England., London, 2020.

[182] Teixeira PA, Gordon P, Camhi E and Bakken S, “HIV patients’ willingness to share personal health information electronically,” *Patient Education and Counseling,* vol. 84, p. e9–e12, 2011.

[183] Knapper C, Drayton R, Browning M and Lomax N, “Confidentiality: a continuing barrier to disclosure of HIV status to GPs? The experience and concerns of HIV patients in a Welsh HIV clinic.,” in *BHIVA Conference*, Birmingham, 2012.

[184] Campos-Castillo C and Anthony D, “The double-edged sword of electronic health records: implications for patient disclosure,” *J Am Med Inform Assoc ;:,* vol. 22, p. e130–e1, 2015.

[185] Williams CG and Miller A, “HIV testing: the danger of keeping secrets,” [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/23735384.

[186] Gulliford MC, Charlton J, Ashworth M, Rudd AG and Toschke AM, “Selection of Medical Diagnostic Codes for Analysis of Electronic Patient Records. Application to stroke in a Primary Care Database,” *PLoS ONE,* vol. 4, no. 9, p. e7168, 2009.

[187] Salomon JU, Blackford ST, Rosenbloom S, Seidel EW, Clayton DM, Dilts SG and Finder RM, “Openness of patients' reporting with use of electronic records: psychiatric clinicians' views,” *Journal of the American Medical Informatics Association,* vol. 17, no. 1, pp. 54-60, Jan-Feb 2010.

[188] Woodman J, Allister J, Rafi I, de Lusignan S, Belsey J, Petersen I and Gilbert R, “A simple approach to improve recording of concerns about child maltreatment in primary care records: developing a quality improvement intervention,” *British Journal of Generl Practices,* vol. 62, no. 600, pp. e478-86, 2012.

[189] Scott D, Tonmyr L, Fraser J, Walker SM and Vallmuur K, “The utility and challenges of using ICD codes in child maltreatment research: a review of existing literature.,” *Child Abuse Negl.,* vol. 33, no. 11, p. 791–808, 2009.

[190] González-Izquierdo A, Woodman J, Copley L, van der Meulen J, Brandon M, Hodes D, Lecky F and Gilbert R, “Variation in recording of child maltreatment in administrative records of hospital admissions for injury in England, 1997–2009.,” *Arch Dis Child.,* vol. 95, no. 11, p. 918–925, 2010.

[191] Evans HER, Mercer CH, Rait G, Hamill M, Delpech V, Hughes G, Brook MG, Williams T, Johnson AM, Singh S, Petersen I, Chadborn T and Cassell JA, “Trends in HIV testing and recording of HIV status in the UK primary care setting: a retrospective cohort study 1995–2005,” *Sex Transm Infect,* vol. 85, p. 520–526, 2009.

[192] Green J and Thoroughgood N, Qualitative methods for health research, 2nd ed., London: Sage Publications Ltd, 2009.

[193] Braun V and Clarke V, “Fundamentals of Qualitative Research,” in *Successful Qualitative Research*, London, Sage, 2013, pp. 19-14.

[194] Creswell JW, Research design: Qualitative and Quantitative approaches, London: Sage Publications Ltd, 1994.

[195] Britten N, “Qualitative interviews in medical research,” *BMJ,* vol. 311, pp. 251-253, 1995.

[196] Shaheen M, Pradhan S and Ranajee R, “Sampling in Qualitative Research,” in *Qualitative Techniques for Workplace Data Analysis*, 2019, pp. 25-51.

[197] Bryant A and Charmaz K, The SAGE Handbook of Grounded Theory, London : Sage, 2007.

[198] Glaser BG and Strauss AL, The Discovery of Grounded Theory: Strategies for Qualitative Research., Chicago: Aldine, 1967.

[199] Department of Health and Social Care (DH), “Attributing the costs of health and social care research,” June 2015. [Online]. Available: https://www.gov.uk/government/publications/guidance-on-attributing-the-costs-of-health-and-social-care-research. [Accessed 24 March 2018].

[200] Marshall C and Rossman GB, Designing qualitative research, London: Sage Publications Ltd, 2016.

[201] Creswell JW, Research design: Qualitative and Quantitative approaches, 4th ed., London: Sage Publications Ltd, 2014.

[202] Maguire M and Delahunt B, “Doing a Thematic Analysis: A Practical, Step-by-Step Guide for Learning and Teaching Scholars.,” *AISHE-J,* vol. 9, no. 3, pp. 3351-33513, 2017.

[203] Public Health England (PHE), “GUMCADv2 Genitourinary Medicine Clinic Activity Dataset (version 2): Technical guidance and extract specification for data extract,” London, 2015.

[204] Medical Foundation for Aids and Sexual Health, “Recommended standards for NHS HIV services,” Medical Foundation for Aids and Sexual Health, London, 2003.

[205] Jekel JF, Elmore JG and Katz DL, Epidemiology Biostatistics and Preventative Medicine, 3rd ed., Philadelphia: Saunders Company, 2007.

[206] Aughan JP, Victora C and Chowdhury AMR, “Epidemiological Principles,” in *Practical Epidemiology: Using Epidemiology to Support Primary Health Care*, Oxford University Press, 2021.

[207] Maguire A, Blak BT and Thompson M, “The importance of defining periods of complete mortality reporting for research using automated data from primary care.,” *Pharmacoepidemiology and Drug Safety,* vol. 18, no. 1, pp. 76-83, Jan 2009.

[208] Clinical Practice Research Datalink (CPRD), “CPRD linked data,” Clinical Practice Research Datalink, June 2022. [Online]. Available: https://cprd.com/cprd-linked-data. [Accessed September 2022].

[209] Lucas SB, Curtis H and Johnson MA, “National review of deaths among HIV infected adults,” *Clin Med,* vol. 8, no. 3, pp. 250-252, 2008.

[210] Department of Health (DH), “A Framework for Sexual Health Improvement in England,,” Department of Health, London, 2013.

[211] Public Health England (PHE), “PHOF,” 2015. [Online]. Available: https://www.gov.uk/government/collections/public-health-outcomes-framework.

[212] National Institute for Health and Care Excellence (NICE), “HIV testing: increasing uptake among people who may have undiagnosed HIV NICE guideline [NG60],” 01 December 2016. [Online]. Available: www.nice.org.uk/guidance/ng60. [Accessed 01 July 2020].

[213] Blak BT, Thompson M, Dattani H and Bourke A, “Generalisability of The Health Improvement Network (THIN) database: demographics, chronic disese prevalence and mortality rates,” *The Journal of Innovations in Health Informatics,* vol. 19, no. 4, pp. 251-5, 2011.

[214] Equator Network, “The Reporting of studies Conducted using Observational Routinely-collected health Data (RECORD) Statement,” [Online]. Available: http://www.equator-network.org/reporting-guidelines/record/. [Accessed 18 August 2016].

[215] Equator Network, “Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD),” [Online]. Available: https://www.equator-network.org/reporting-guidelines/tripod-statement/. [Accessed 24 March 2020].

[216] Collins GS, Reitsma JB, Altman DG and Moons KGM, “Research methods & reporting: Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement,” *BMJ,* vol. 350, p. g7594, 2014.

[217] Steyerberg E W, Moons KGM, van der Windt DA, Hayden JA, Perel P, Schroter S, Riley RD, Hemingway H and Altman DG, “Prognosis Research Strategy (PROGRESS) 3: Prognostic Model Research,” *PLOS Medicine,* February 2013.

[218] Heus P, Damen JAAG, Pajouheshnia R, Scholten RJPM, Reitsma JB, Collins GS, Altman DG, Moons KGM and Hooft L, “Poor reporting of multivariable

prediction model studies: towards a targeted implementation strategy of the TRIPOD statement," *BMC Medicine,* vol. 16, no. 120, 2018.

[219] Moons KG, Altman DG, Reitsma JB, Ioannidis JP, Macaskill P, Steyerberg EW, Vickers AJ, Ransohoff DF and Collins GS, "Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): explanation and elaboration," *Ann Intern Med,* vol. 162, no. 1, pp. W1-73, 2015.

[220] Steyerberg EW and Vergouwe Y, "Towards better clinical prediction models: seven steps for development and an ABCD for validation," *European Heart Journal,* vol. 35, p. 1925–1931, 2014.

[221] Seel RT, Steyerberg EW, Malec JF, Sherer M and Macciocchi SN, "Developing and Evaluating Prediction Models in Rehabilitation Populations," *Arch Phys Med Rehabil,* vol. 93 Suppl 2, 2012.

[222] Petersen I, Welch CA, Nazareth I, Walters K, Marston L, Morris RW, Carpenter JR, Morris TP and Pham TM, "Health indicator recording in UK primary care electronic health records: key implications for handling missing data," *Clinical Epidemiology ,* vol. 11, p. 157–167, 2019.

[223] The National Institute for Health and Care Excellence (NICE), "NICE Guidelines," [Online]. Available: https://www.nice.org.uk/guidance. [Accessed 16 December 2015].

[224] Forthofer RN, Lee ES and Hernandez M, Biostatistics: A guide to design, analysis, and discovery, 2nd ed., Oxford: Elsevier, 2007.

[225] Mathur R, Bhaskaran K, Chaturvedi N, Leon DA, vanStaa T, Grundy E and Smeeth L, "Completeness and usability of ethnicity data in UK-based primary care and hospital databases," *Journal of Public Health,* vol. 36, no. 4, p. 684–692, December 2014.

[226] Bennett DA, "How can I deal with missing data in my study?," *Australian and New Zealand Journal of Public Health,* vol. 25, p. 464–469, 2001.

[227] El-Masri SM and Fox-Wasylyshyn MM , "Best Practices in Research Methods; Missing Data: An Introductory Conceptual Overview for the Novice Researcher," *CJNR,* vol. 37, no. 4, p. 156–171, 2005.

[228] "Health Improvement Network – THIN," [Online]. Available: http://www.epic-uk.org/our-data/ethics.shtml.

[229] Groenwold RH, White IR, Donders AR, Carpenter JR, Altman DG and Moons KG, "Missing covariate data in clinical research: when and when not to use the missing-indicator method for analysis," *CMAJ,* vol. 184, no. 11, pp. 1265-9, 2012.

[230] van der Heijden GJMG, , Donders ART, Stijnen T and Moons KGM, “Imputation of missing values is superior to complete case analysis and the missing-indicator method in multivariable diagnostic research: A clinical example,” *Journal of Clinical Epidemiology,* vol. 59, no. 10, pp. 1102-1109, 2006.

[231] Carpenter J and Kenward M, “Brief comments on computational issues with multiple imputation,” 2018. [Online]. Available: https://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.394.9545. [Accessed August 2022].

[232] Riley RD, Snell KI, Ensor J, Burke DL, Harrell FE Jr, Moons KG, Collins GS, “Minimum sample size for developing a multivariable prediction model: PART II - binary and time-to-event outcomes,” *Stat Med,* vol. 38, no. 7, pp. 1276-1296, 2019 .

[233] Dunkler D, Plischke M, Leffondré K and Heinze G, “Augmented backward elimination: A pragmatic and purposeful way to develop statistical models,” *PloS one,* vol. 9, no. 11, pp. e113677-e113677, 2014.

[234] Austin PC, “Using the bootstrap to improve estimation and confidence intervals for regression coefficients selected using backwards variable elimination,” *Statistics in medicine,* vol. 27, no. 17, pp. 3286-3300, 2008.

[235] Sauerbrei W, “The Use of Resampling Methods to Simplify Regression Models in Medical Statistics,” *Journal of the Royal Statistical Society Series C - Applied Statis tics,* vol. 48, no. 3, pp. 313-329, 1999.

[236] Heinze G, Wallisch C and Dunkler D, “Variable selection – A review and recommendations for the practicing statistician,” *Biometrical Journa,* vol. 60, p. 431–449, 2018.

[237] Royston P, Moons KGM, Altman DG and Vergouwe Y, “Prognosis and prognostic research: Developing a prognostic model,” *BMJ,* vol. 338, p. b604, 2009.

[238] Lee YH, Bang H and Kim DJ, “How to Establish Clinical Prediction Models,” *Endocrinol Metab,* vol. 31, pp. 38-44, 2016.

[239] Steyerberg EW, Bleeker SE, Moll HA, Grobbee DE and Moons KGM, “Internal and external validation of predictive models: A simulation study of bias and precision in small samples,” *Journal of Clinical Epidemiology,* vol. 56, p. 441–447, 2003.

[240] Moons KG, Kengne AP, Grobbee DE, Royston P, Vergouwe Y, Altman DG and Woodward M, “Risk prediction models: II. External validation, model updating, and impact assessment,” *Heart,* vol. 98, pp. 691-8, 2012.

[241] Steyerberg EW, Harrell FE, Borsboom GJJM, Eijkemans MJC, Vergouwe Y and Habbema JDF, “Internal validation of predictive models: Efficiency of some procedures for logistic regression analysis.,” *Journal of Clinical Epidemiology,* vol. 54, no. 8, p. 774–781, 2001.

[242] Bleeker SE, Molla HA, Steyerberg EW, Donders ART, Derksen-Lubsen G, Grobbee DE and Moons KGM, “External validation is necessary in prediction research: A clinical example,” *Journal of Clinical Epidemiology,* vol. 56, no. 9, p. 826–832, September 2003.

[243] Riley RD, Ensor J, Snell KIE, Debray TPA, Altman DG, Moons KGM and Collins GS, “External validation of clinical prediction models using big datasets from e-health records or IPD meta-analysis: opportunities and challenges,” *BMJ ;,* vol. 353, no. i3140, 2016.

[244] Nash S, Desai S, Croxford S, Guerra L, Lowndes C, Connor N and Gill ON, “Progress towards ending the HIV epidemic in the United Kingdom 2018 report − summary, key messages and recommendations,” November 2018. [Online]. Available: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/813168/HIV_annual_report_2018_-_Summary_key_messages_and_recommendations.pdf. [Accessed 5 August 2020].

[245] Office of National Statistics, “Population estimates for the UK, England and Wales, Scotland and Northern Ireland,” Office of National Statistics, 28 June 2018. [Online]. Available: https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/bulletins/annualmidyearpopulationestimates/mid2017. [Accessed 28 August 2020].

[246] Fitzgerald M, Saville BR and Lewis RJ, “Decision Curve Analysis,” *JAMA,* vol. 313, no. 4, p. 409–410, 2015.

[247] Kerr KF, Brown MD, Zhu K and Janes H, “Assessing the Clinical Impact of Risk Prediction Models With Decision Curves: Guidance for Correct Interpretation and Appropriate Use.,” *J Clin Oncol,* vol. 34, no. 21, pp. 2534-40, 2016.

[248] Vickers AJ and Elkin EB, “Decision curve analysis: a novel method for evaluating prediction models,” *Med Decis Making,* vol. 26, no. 6, pp. 565-74, 2006.

[249] Vickers AJ, Van Calster B and Steyerberg EW, “Net benefit approaches to the evaluation of prediction models, molecular markers, and diagnostic tests,” *BMJ,* vol. 352 , 2016.

[250] Vickers AJ, Cronin AM, Elkin EB , “Extensions to decision curve analysis, a novel method for evaluating diagnostic tests, prediction models and molecular markers.,” *BMC Med Inform Decis Mak,* vol. 8, no. 53, 2008.

[251] Ensor J, Snell KIE and Martin EC, “PMCALPLOT: Stata module to produce calibration plot of prediction model performance,” Boston College Department of Economics, 2018.

[252] Hoots BE, MacDonald PD, Hightow-Weidman LB, Leone PA and Miller WC, “Developing a predictive model to prioritize human immunodeficiency virus partner notification in North Carolina,” *Sex Transm Dis,* vol. 39, no. 1, pp. 65-71, 2012 .

[253] Newman T and Kohn M, “Critical Appraisal of Studies of Diagnostic Test accuracy,” in *Evidence-Based Diagnosis: An Introduction to Clinical Epidemiology*, Cambridge, Cambridge University Press, 2020, pp. 75-109.

[254] Schedlbauer A, Prasad V, Mulvaney C, Phansalkar S, Stanton W, Bates DW and Avery AJ , “What Evidence Supports the Use of Computerized Alerts and Prompts to Improve Clinicians' Prescribing Behavior?,” *Journal of the American Medical Informatics Association,* vol. 16, no. 4, p. 531–538, 2009.

[255] Backman R, Bayliss S, Moore D and Litchfield I, “Clinical reminder alert fatigue in healthcare: a systematic literature review protocol using qualitative evidence Systematic Reviews,” vol. 6, no. 255, 2017.

[256] Murphy DR, Reis B, Sitting DF and Singh H, “Notifications Received by Primary Care Practitioners in Electronic Health Records: A Taxonomy and Time Analysis Clinical research study,” *The American Journal of Medicine,* vol. 125, no. 2, pp. e1-e7, 2012.

[257] Shojania KG, Jennings A, Mayhew A, Ramsay CR, Eccles MP and Grimshaw J, “The effects of on-screen, point of care computer reminders on processes and outcomes of care.,” *Cochrane Database of Systematic Reviews,* no. 3, 2009.

[258] Arditi C, Rège-Walther M, Durieux P and Burnand B, “Computer-generated reminders delivered on paper to healthcare professionals: effects on professional practice and healthcare outcomes.,” *Cochrane Database of Systematic Reviews,* no. 7, 2017.

[259] Cheung A, Weir M, Mayhew A, Kozloff N, Brown K and Grimshaw J, “Overview of systematic reviews of the effectiveness of reminders in improving healthcare professional behavior,” *Systematic Reviews,* vol. 1, no. 36, 2012.

[260] Kwan JL, Lo L, Ferguson J, Goldberg H, Diaz-Martinez JP, Tomlinson G, Grimshaw JM and Shojania KG, “Computerised clinical decision support

systems and absolute improvements in care: meta-analysis of controlled clinical trials," *BMJ,* vol. 370, 2020.

[261] Konerman MA, Thomson Mary, Gray K, Moore M, Choxi H, Seif E, and Lok ASF, "Impact of an Electronic Health Record Alert in Primary Care on Increasing Hepatitis C Screening and Curative Treatment for Baby Boomers," *Hepatology,* vol. 66, no. 6, pp. 1805-1813, 2017 .

[262] Guiriguet C, Muñoz-Ortiz L, Burón A, Rivero I, Grau J, Vela-Vallespín C, Vilarrubí M, Torres M, Hernández C, Méndez-Boo L, Toràn P, Caballeria L, Macià F and Castells A, "Alerts in electronic medical records to promote a colorectal cancer screening programme: a cluster randomised controlled trial in primary care," *British Journal of General Practice,* 2016.

[263] Redondo LC, Ruíz M, Kostov B, Sequeira E, Noguera P et al, "Indicator condition-guided HIV testing with an electronic prompt in primary healthcare: a before and after evaluation of an intervention.," *Sexually transmitted infections,* vol. 95, no. 4, pp. 238-243, 2019.

[264] Coma E, Medina M, Méndez L, Hermosilla E, Iglesias M, Olmos C and Calero S, "Effectiveness of electronic point-of-care reminders versus monthly feedback to improve adherence to 10 clinical recommendations in primary care: a cluster randomized clinical trial.," *BMC Med Inform Decis Mak,* vol. 19, p. 245, 2019.

[265] Agustí C, Cunillera O, Almeda J, Mascort J, Carrillo R, Olmos C, Montoliu A, Alberny M, Molina I, Cayuelas L and Casabona J, "Efficacy of an electronic reminder for HIV screening in primary healthcare based on indicator conditions in Catalonia (Spain)," *HIV Medicine,* vol. 00, pp. 1-12, 2022.

[266] Phytel, "Whitepaper: Predictive modeling and population health management," 2013. [Online]. Available: phytel.com. [Accessed 20 March 2018].

[267] Eichler K, Zoller M, Tschudi P and Steurer J, "Barriers to apply cardiovascular prediction rules in primary care: a postal survey," *BMC Family Practice,* vol. 8, no. 1, 2007.

[268] Baypinar F, Jan Kingma H, van der Hoeven RTM and Becker ML, "Physicians' Compliance with a Clinical Decision Support System Alerting during the Prescribing Process," *Journal of Medical Systems,* vol. 41, no. 96, 2017.

[269] Carroll AE, Anand V and Downs SM, "Understanding Why Clinicians Answer or Ignore Clinical Decision Support Prompts," *Applied Clinical Informatics,* vol. 3, p. 309–317, 2012.

[270] Bell H, Garfield S, Khosla S, Patel C and Franklin BD, "Mixed methods study of medication-related decision support alerts experienced during electronic

prescribing for inpatients at an English hospital.," *European Association of Hospital Pharmacists,* vol. 26, p. 318–322, 2019.

[271] Ancker JS, Edwards A, Nosal S, Hauser D, Mauer E and Kaushal R with the HITEC investigators, "Effects of workload, work complexity, and repeated alerts on alert fatigue in a clinical decision support system.," *BMC Medical Informatics and Decision Making,* vol. 17, no. 36, 2017.

[272] Moher D, Shamseer L, Clarke M, Ghersi D, Liberati A, Petticrew M, Shekelle P, Stewart LA and PRISMA-P Group, "Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement," *Systematic Reviews,* vol. 4, no. 1, 2015.

[273] Swinglehurst D, Roberts C, Li S, Weber O and Singy P, "Beyond the 'dyad': a qualitative re-evaluation of the changing clinical consultation," *BMJ Open,* vol. 4, p. e006017, 2014.

[274] Barnett-Page E and Thomas J, "Methods for the synthesis of qualitative research: a critical review," *BMC Med Res Methodol,* vol. 9, no. 59, 2009.

[275] France EF, Ring N , Thomas R, Noyes J, Maxwell M and Jepson R, "A methodological systematic review of what's wrong with meta-ethnography reporting," *BMC Medical Research Methodology,* vol. 14, no. 119, 2014.

[276] Thomas J and Harden A, "Methods for the thematic synthesis of qualitative research in systematic reviews," *BMC Medical Research Methodology,* vol. 8, no. 45, 2008.

[277] Kortteisto T, Komulainen J, Mäkelä M, Kunnamo I and Kaila M, "Clinical decision support must be useful, functional is not enough: a qualitative study of computer-based clinical decision support in primary care," *BMC Health Services Research,* vol. 12, no. 349, 2012.

[278] Harry ML, Truitt AR, Saman DM , Henzler-Buckingham HA, Allen CI, Walton KM, Ekstrom HL, O'Connor PJ, Sperl-Hillen2 JM, Bianco JM and Elliott TE, "Barriers and facilitators to implementing cancer prevention clinical decision support in primary care: a qualitative study," *BMC Health Services Research,* vol. 19, p. 534, 2019.

[279] Wan Q, Makeham M, Zwar NA and Petche S, "Qualitative evaluation of a diabetes electronic decision support tool: views of users," *BMC medical informatics and decision making,* vol. 12, 2012.

[280] Peiris DP, Joshi R, Webster RJ, Groenestein P, Usherwood TP, Heeley E, Turnbull FM, Lipman A and Patel AA, "An electronic clinical decision support tool to assist primary care providers in cardiovascular disease risk

management: development and mixed methods evaluation," *Journal of medical Internet research,* vol. 11, no. 4, p. e51, 2009.

[281] Dikomitis L, Green T and Macleod U, "Embedding electronic decision-support tools for suspected cancer in primary care: a qualitative study of GPs' experiences," *Primary Health Care Research & Development,* vol. 16, p. 548–555, 2015.

[282] Varonen H, Kortteisto T and Kaila M, "What may help or hinder the implementation of computerized decision support systems (CDSSs): a focus group study with physicians," *Family practice,* vol. 25, no. 3, pp. 163-7, 2008.

[283] Nair K, Malaeekeh R, Schabort I, Taenzer P, Radhakrishnan A and Guenter D, "A Clinical Decision Support System for Chronic Pain Management in Primary Care: Usability testing and its relevance," *Journal of innovation in health informatics,* vol. 22, no. 3, pp. 329-332, 2015.

[284] Dryden EM, Hardin J, McDonald J, Taveras EM and Hacker K, "Provider perspectives on electronic decision supports for obesity prevention," *Clinical pediatrics Clin Pediatr (Phila),* vol. 51, no. 5, pp. 490-497, 2012.

[285] Harle CA, Apathy NC, Cook RL, Danielson EC, DiIulio J; Downs SM, Hurley RW, Mamlin BW, Militello LG and Anders S, "Information Needs and Requirements for Decision Support in Primary Care: An Analysis of Chronic Pain Care," *AMIA Annual Symposium proceedings.,* pp. 527-534, 2018.

[286] Chase DA, Baron S and Asha JS, "Clinical Decision Support and Primary Care Acceptance of Genomic Medicine," *Precision Healthcare through Informatics,* pp. 700-703, 2017.

[287] Baron S, Filios MS, Marovich S, Chase D and Ash JS, "Recognition of the Relationship Between Patients' Work and Health: A Qualitative Evaluation of the Need for Clinical Decision Support (CDS) for Worker Health in Five Primary Care Practices," *Journal of occupational and environmental medicine,* vol. 59, no. 11, pp. e245-e250, 2017.

[288] Short D,Frischer M and Bashford J, "Barriers to the adoption of computerised decision support systems in general practice consultations: a qualitative study of GPs' perspectives," *International Journal of Medical Informatics,* vol. 73, p. 357—362, 2004.

[289] McParland CR, Cooper MA and Johnston B, "Differential Diagnosis Decision Support Systems in Primary and Out-of-Hours Care: A Qualitative Analysis of the Needs of Key Stakeholders in Scotland," *Journal of primary care & community health,* vol. 10, pp. 1-6, 2019.

[290] Polit DF and Beck CT, “Generalization in quantitative and qualitative research: Myths and strategies,” *International Journal of Nursing Studies ,* vol. 47 , p. 1451–1458, 2010.

[291] Publich Health in England (PHE), “HIV in England Data Tables 2019,” December 2020. [Online]. Available: https://www.gov.uk/government/statistics/hiv-annual-data-tables. [Accessed 1 November 2021].

[292] Baylis A, Buck D, Anderson J, Jabbal J and Ross S, “The future of HIV services in England Shaping the response to changing needs,” The King's Fund, London, 2017.

[293] Public Health England (PHE), “National HIV prevention programme,” 31 July 2014. [Online]. Available: https://www.gov.uk/government/collections/hiv-surveillance-data-and-management.

[294] Machin D, Campbell MJ, Tan S and Tan S, Sample Sizes for Clinical, Laboratory and Epidemiology Studies, 4th ed., John Wiley & Sons, Incorporated, 2018.

[295] Katki HA and Bebu I, “A simple framework to identify optimal cost-effective risk thresholds for a single screen: Comparison to Decision Curve Analysis,” *J R Stat Soc Ser A Stat Soc,* vol. 184, no. 3, pp. 887-903, 2021 Jul;.

[296] Billing C, Collinson S, Bramley G, Muccio ED, Rumbwere Dube B and Margreiter M, “Commercialisation: Bridging the University-Industry Gap City-REDI Research Brief,” 2018. [Online]. Available: https://www.birmingham.ac.uk/documents/college-social-sciences/business/research/city-redi/projects-docs/commercialisation-policy-briefing-oct-2019.pdf. [Accessed 24 October 2021].

[297] Kappen TH, van Klei WA, van Wolfswinkel L, Kalkman CJ, Vergouwe Y and Moons KGM, “Evaluating the impact of prediction models: lessons learned, challenges, and recommendations,” *Diagnostic and Prognostic Research,* vol. 2, no. 11, 2018.

[298] Vandewinckele L, Claessens M, Dinkla A, Brouwer C, Crijns W, Verellen D and van Elmpt W, “Overview of artificial intelligence-based applications in radiotherapy: Recommendations for implementation and quality assurance,” *Radiotherapy and Oncology,* vol. 153 , 2020.

[299] National Institute for Health Research (NIHR), “Self-testing kit for sexually transmitted infections increases diagnoses while reducing costs,” 2021.

[300] Kersh EN, Shukla M, Raphael BH, Habel M and Park I, “At-Home Specimen Self-Collection and Self-Testing for Sexually Transmitted Infection Screening Demand Accelerated by the COVID-19 Pandemic: a Review of Laboratory

Implementation Issues," *J Clin Microbiol,* vol. 59, no. 11, p. e0264620, 19 Oct 2021.

[301] Leenen J, Hoebe CJPA, Ackens RP, Posthouwer D, van Loo IHM, Wolffs PFG and Dukers-Muijrers NHTM, "Pilot implementation of a home-care programme with chlamydia, gonorrhoea, hepatitis B, and syphilis self-sampling in HIV-positive men who have sex with men," *BMC Infect Dis,* vol. 20, no. 1, p. 925, 4 Dec 2020.

[302] Chava L Ramspek, Kitty J Jager, Friedo W Dekker, Carmine Zoccali and Merel van Diepen, "External validation of prognostic models: what, why, how, when and where?," *Clinical Kidney Journal,* vol. 14, no. 1, p. 49–58, 2021.

[303] "Electronic Health Records in India: PTS – The Future of EHR in India – Patient Tracking System is a comprehensive, cost-effective & secure cloud based Patient Management Electronic Medical Records (EMR) App for Doctors and Hospitals in India.," [Online]. Available: https://ehr-india.com/. [Accessed 30 August 2022].

[304] Jacob N, Rice B, Kalk E, Heekes A, Morgan J, Brinkmann S, Hargreaves J, Orgill M and Boulle A, "Consolidating strategic information to monitor progress against the UNAIDS 90–90–90 targets: evaluating the operational feasibility of an electronic HIV testing register in Cape Town, South Africa. BMC Health Serv Res," *BMC Health Serv Res,* vol. 20, p. 720, 2020.

[305] Lucas RW, Card WI, Knill-Jones RP, Watkinson G and Crean GP, "Computer interrogation of patients," *BMJ,* vol. 2, pp. 623-5, 1976.

[306] Lilford RJ, Kelly M, Baines A, Cameron S, Cave M, Guthrie K and Thornton J, "Effect of using protocols on medical care: randomised trial of three methods of taking an antenatal history," *BMJ,* vol. 305, pp. 1181-4., 1992.

[307] Kirwan PD, Chau C, Brown AE, Gill ON, Delpech VC and contributors., "HIV in the UK - 2016 report.," Public Health England, London, 2016.

[308] Sanders GD, Bayoumi AM, Sundaram V, Bilir SP, Neukermans CP, Rydzak CE, Douglass LR, Lazzeroni LC, Holodniy M and Owens DK, "Cost-Effectiveness of Screening for HIV in the Era of Highly Active Antiretroviral Therapy," *The New England Journal of Medicine,* vol. 352, no. 6, pp. 570-85, 2015.

[309] Frew PM, Parker K, Vo L, Haley D, O'Leary A, Diallo DD, Golin CE, Kuo I, Soto-Torres L, Wang J, Adimora AA, Randall LA , del Rio C, Hodder S and the HIV Prevention Trials Network 064 (HTPN) Study Team, "Socioecological factors influencing women's HIV risk in the United States: qualitative findings from the women's HIV SeroIncidence study (HPTN 064),," *BMC Public Health,* vol. 16, p. 803, 2016.

[310] Fernandez M, Collazo JB, Bowen GS, Varga LM, Hernandez N and Perrino T, “Predictors of HIV testing and intention to test among Hispanic Farmworkers in South Florida,” *The Journal of Rural Health,* vol. 21, no. 1, pp. 56-64, 2006.

[311] Girardi E, Sabin CA and Monforte DD, “Late diagnosis of HIV infection: epidemiological features, consequences and strategies to encourage earlier testing,” *J Acquir Immune Defic Syndr,* vol. 46, p. S3–S8, 2007.

[312] Halve it, Early testing saves lives: HIV is a public health priority, 3rd ed., Gilead Sciences Ltd, n.d.

[313] Stevens RJ and Poppe KK, “Validation of clinical prediction models: what does the "calibration slope" really measure?,” *J Clin Epidemiol,* vol. 118, pp. 93-99, 2020 .

Appendices

Appendix 1: Screening criteria

Appendix 1-A: Screening stage 1 criteria

Screen using the title/abstracts and find out if studies fit the eligibility criteria

Eligibility criteria

1. The exposure of the study are risk factors or characteristics associated with HIV infection
    - demographic,
    - socio-economic or
    - clinical
2. The outcome of the study is
    - Human Immunodeficiency Virus or HIV or Acquired Immuno-deficiency Syndrome or AIDS
3. Published 1995 onwards
4. Reviews, cohort studies, case-control, RCTs

Exclusion criteria

1. Non-human studies.
2. Children (under 18) only
3. Non-HIV diagnosis
4. Setting is developing countries only

Appendix 1-B: Screening stage 2 criteria

Screen using the whole article and find out if studies fit the eligibility criteria

Eligibility criteria

1. The exposure of the study are risk factors or characteristics associated with HIV infection
    - demographic,
    - behavioural
    - socio-economic or
    - clinical
2. The outcome of the study is

    Human Immunodeficiency Virus or HIV or Acquired Immuno-deficiency Syndrome or AIDS

3. Published 1995 onwards
4. Reviews, cohort studies, case-control, RCTs

Exclusion criteria

5. Non-human studies.
6. Children (under 18) only
7. Non-HIV diagnosis or post HIV diagnosis
8. Setting is developing countries only
9. Studies on Health care or treatment

Appendix 2: Quality Assessment

	Article ID		
		A1	
		year	
Study design	1 - Cohort Study 2 – Case Control Study 3 –Guideline		
	1.1. The study addresses an appropriate and clearly focused question. [i]	Y/N/CS	
1.2 COHORT STUDY	1.2.1 The two groups being studied are selected from source populations that are comparable in all respects other than the factor under investigation. [ii]	Y/N/CS/DNA	
	1.2.2 The study indicates how many people participated, in each of the groups being studied. [iii]	Y/N/CS/DNA	
	1.2.3 The likelihood that some eligible subjects might have the outcome at the time of enrolment is assessed and taken into account in the analysis.[iv]	Y/N/CS/DNA	
	1.2.4 What percentage of individuals or clusters recruited into each arm of the study dropped out before the study was completed?[v]	Y/N/CS	
	1.2.5 Comparison is made between full participants and those lost to follow up, by exposure status.[vi]	Y/N/CS	
	1.2.6 The outcomes are clearly defined.[vii]	Y/N/CS	
	1.2.7 The assessment of outcome is made blind to exposure status. If the study is retrospective this may not be applicable.[viii]	Y/N/CS/DNA	
	1.2.8 Where blinding was not possible, there is some recognition that knowledge of exposure status could have influenced the assessment of outcome.[ix]	Y/N/CS	
	1.2.9 The method of assessment of exposure is reliable.[x]	Y/N/CS	
	1.2.10 Evidence from other sources is used to demonstrate that the method of outcome assessment is valid and reliable.[xi]	Y/N/CS/DNA	
	1.2.11 Exposure level or prognostic factor is assessed more than once.[xii]	Y/N/CS/DNA	
1.3 CASE CONTROL STUDY	1.3.1 The cases and controls are taken from comparable populations.[xiii]	Y/N/CS	
	1.3.2 The same exclusion criteria are used for both cases and controls.[xiv]	Y/N/CS	
	1.3.3 What percentage of each group (cases and controls) participated in the study?[xv]	Cases %: Controls%	
	1.3.4 Comparison is made between participants and non-participants to establish their similarities or differences.[xvi]	Y/N/CS	
	1.3.5 Cases are clearly defined and differentiated from controls.[xvii]	Y/N/CS	
	1.3.6 It is clearly established that controls are non-cases.[xviii]	Y/N/CS	
	1.3.7 Measures will have been taken to prevent knowledge of primary exposure influencing case ascertainment.[xix]	Y/N/CS/DNA	
	1.3.8 Exposure status is measured in a standard, valid and reliable way.[xx]	Y/N/CS	
Confounding	1.4 The main potential confounders are identified and taken into account in the design and analysis.[xxi]	Y/N/CS	
	1.5 Have confidence intervals been provided?[xxii]	Y/N	
Section 2: OVERALL ASSESSMENT OF THE STUDY	2.1 How well was the study done to minimise the risk of bias or confounding?[xxiii]	High quality (++) □ Acceptable (+) □ Unacceptable – reject 0	

	2.2 Taking into account clinical considerations, your evaluation of the methodology used, and the statistical power of the study, do you think there is clear evidence of an association between exposure and outcome?	Y/N/CS	
	2.3 Are the results of this study directly applicable to the patient group targeted in this guideline?	Y/N	
	Notes. Summarise the authors' conclusions. Add any comments on your own assessment of the study, and the extent to which it answers your question and mention any areas of uncertainty raised above.		

[1] Unless a clear and well-defined question is specified in the report of the review, it will be difficult to assess how well it has met its objectives or how relevant it is to the question you are trying to answer on the basis of the conclusions.

[1] This relates to selection bias.* It is important that the two groups selected for comparison are as similar as possible in all characteristics except for their exposure status, or the presence of specific prognostic factors or prognostic markers relevant to the study in question.

[1] This relates to selection bias.* The participation rate is defined as the number of study participants divided by the number of eligible subjects, and should be calculated separately for each branch of the study. A large difference in participation rate between the two arms of the study indicates that a significant degree of selection bias* may be present, and the study results should be treated with considerable caution.

[1] If some of the eligible subjects, particularly those in the unexposed group, already have the outcome at the start of the trial the final result will be subject to performance bias.* A well conducted study will attempt to estimate the likelihood of this occurring, and take it into account in the analysis through the use of sensitivity studies or other methods.

[1] This question relates to the risk of attrition bias.*The number of patients that drop out of a study should give concern if the number is very high. Conventionally, a 20% drop out rate is regarded as acceptable, but in observational studies conducted over a lengthy period of time a higher drop out rate is to be expected. A decision on whether to downgrade or reject a study because of a high drop out rate is a matter of judgement based on the reasons why people dropped out, and whether drop out rates were comparable in the exposed and unexposed groups. Reporting of efforts to follow up participants that dropped out may be regarded as an indicator of a well conducted study.

[1] For valid study results, it is essential that the study participants are truly representative of the source population. It is always possible that participants who dropped out of the study will differ in some significant way from those who remained part of the study throughout. A well conducted study will attempt to identify any such differences between full and partial participants in both the exposed and unexposed groups. This relates to the risk of attrition bias.* Any unexplained differences should lead to the study results being treated with caution.

[1] This relates to the risk of detection bias.* Once enrolled in the study, participants should be followed until specified end points or outcomes are reached. In a study of the effect of exercise on the death rates from heart disease in middle aged men, for example, participants might be followed up until death, or until reaching a predefined age. If outcomes and the criteria used for measuring them are not clearly defined, the study should be rejected.

[1] This relates to the risk of detection bias.* If the assessor is blinded to which participants received the exposure, and which did not, the prospects of unbiased results are significantly increased. Studies in which this is done should be rated more highly than those where it is not done, or not done adequately.

[1] This relates to the risk of detection bias.* Blinding is not possible in many cohort studies. In order to asses the extent of any bias that may be present, it may be helpful to compare process measures used on the participant groups - e.g. frequency of observations, who carried out the observations, the degree of detail and completeness of observations. If these process measures are comparable between the groups, the results may be regarded with more confidence.

[1] This relates to the risk of detection bias.* A well conducted study should indicate how the degree of exposure or presence of prognostic factors or markers was assessed. Whatever measures are used must be sufficient to establish clearly that participants have or have not received the exposure under investigation and the extent of such exposure, or that they do or do not possess a particular prognostic marker or factor. Clearly described, reliable measures should increase the confidence in the quality of the study

[1] This relates to the risk of detection bias.* The primary outcome measures used should be clearly stated in the study. If the outcome measures are not stated, or the study bases its main conclusions on secondary outcomes, the study should be rejected. Where outcome measures require any degree of subjectivity, some evidence should be provided that the measures used are reliable and have been validated prior to their use in the study.

[1] This relates to the risk of detection bias.* Confidence in data quality should be increased if exposure level is measured more than once in the course of the study. Independent assessment by more than one investigator is preferable.

[1] Study participants may be selected from the target population (all individuals to which the results of the study could be applied), the source population (a defined subset of the target population from which participants are selected), or from a pool of eligible subjects (a clearly defined and counted group selected from the source population. If the study does not include clear definitions of the source population it should be rejected.

[1] All selection and exclusion criteria should be applied equally to cases and controls. Failure to do so may introduce a significant degree of bias into the results of the study.

[1] Differences between the eligible population and the participants are important, as they may influence the validity of the study. A participation rate can be calculated by dividing the number of study participants by the number of eligible subjects. It is more useful if calculated separately for cases and controls. If the participation rate is low, or there is a large difference between the two groups, the study results may well be invalid due to differences between participants and non-participants. In these circumstances, the study should be downgraded, and rejected if the differences are very large.

[1] Even if participation rates are comparable and acceptable, it is still possible that the participants selected to act as cases or controls may differ from other members of the source population in some significant way. A well conducted case-control study will look at samples of the non-participants among the source population to ensure that the participants are a truly representative sample.

[1] The method of selection of cases is of critical importance to the validity of the study. Investigators have to be certain that cases are truly cases, but must balance this with the need to ensure that the cases admitted into the study are representative of the eligible population. The issues involved in case selection are complex, and should ideally be evaluated by someone with a good understanding of the design of case-control studies. If the study does not comment on how cases were selected, it is probably safest to reject it as a source of evidence.

[1] Just as it is important to be sure that cases are true cases, it is important to be sure that controls do not have the outcome under investigation. Control subjects should be chosen so that information on exposure status can be obtained or assessed in a similar way to that used for the selection of cases. If the methods of control selection are not described, the study should be rejected. If different methods of selection are used for cases and controls the study should be evaluated by someone with a good understanding of the design of case-control studies.

[1] If there is a possibility that case ascertainment can be influenced by knowledge of exposure status, assessment of any association is likely to be biased. A well conducted study should take this into account in the design of the study.

[1] The primary outcome measures used should be clearly stated in the study. If the outcome measures are not stated, or the study bases its main conclusions on secondary outcomes, the study should be rejected. Where outcome measures require any degree of subjectivity, some evidence should be provided that the measures used are reliable and have been validated prior to their use in the study.

[1] Confounding is the distortion of a link between exposure and outcome by another factor that is associated with both exposure and outcome. The possible presence of confounding factors is one of the principal reasons why observational studies are not more highly rated as a source of evidence. The report of the study should indicate which potential confounders have been considered, and how they have been assessed or allowed for in the analysis. Clinical judgement should be applied to consider whether all likely confounders have been considered. If the measures used to address confounding are considered inadequate, the study should be downgraded or rejected, depending on how serious the risk of confounding is considered to be. A study that does not address the possibility of confounding should be rejected.

[1] Confidence limits are the preferred method for indicating the precision of statistical results, and can be used to differentiate between an inconclusive study and a study that shows no effect. Studies that report a single value with no assessment of precision should be treated with extreme caution.

[1] Rate the overall methodological quality of the study, using the following as a guide: High quality (++): Majority of criteria met. Little or no risk of bias. Results unlikely to be changed by further research. Acceptable (+): Most criteria met. Some flaws in the study with an associated risk of bias, Conclusions may change in the light of further studies. Low quality (0): Either most criteria not met, or significant flaws relating to key aspects of study design. Conclusions likely to change in the light of further studies.

Appendix 3: BHIVA clinical indicator conditions

Table 1: Clinical indicator diseases for adult HIV infection

	AIDS-defining conditions	Other conditions where HIV testing should be offered
Respiratory	Tuberculosis	Bacterial pneumonia
	Pneumocystis	Aspergillosis
Neurology	Cerebral toxoplasmosis	Aseptic meningitis/encephalitis
	Primary cerebral lymphoma	Cerebral abscess
	Cryptococcal meningitis	Space occupying lesion of unknown cause
	Progressive multifocal leucoencephalopathy	Guillain–Barré syndrome
		Transverse myelitis
		Peripheral neuropathy
		Dementia
		Leucoencephalopathy
Dermatology	Kaposi's sarcoma	Severe or recalcitrant seborrhoeic dermatitis
		Severe or recalcitrant psoriasis
		Multidermatomal or recurrent herpes zoster
Gastroenterology	Persistent cryptosporidiosis	Oral candidiasis
		Oral hairy leukoplakia
		Chronic diarrhoea of unknown cause
		Weight loss of unknown cause
		Salmonella, shigella or campylobacter
		Hepatitis B infection
		Hepatitis C infection
Oncology	Non-Hodgkin's lymphoma	Anal cancer or anal intraepithelial dysplasia
		Lung cancer
		Seminoma
		Head and neck cancer
		Hodgkin's lymphoma
		Castleman's disease
Gynaecology	Cervical cancer	Vaginal intraepithelial neoplasia
		Cervical intraepithelial neoplasia Grade 2 or above
Haematology		Any unexplained blood dyscrasia including: • thrombocytopenia • neutropenia • lymphopenia
Ophthalmology	Cytomegalovirus retinitis	Infective retinal diseases including herpesviruses and toxoplasma
		Any unexplained retinopathy
ENT		Lymphadenopathy of unknown cause
		Chronic parotitis
		Lymphoepithelial parotid cysts
Other		Mononucleosis-like syndrome (primary HIV infection)
		Pyrexia of unknown origin
		Any lymphadenopathy of unknown cause
		Any sexually transmitted infection

Appendix 4: Predictor variables identified for code list development

		Predictor	Identified in systematic review	Code list developed	Reasons for inclusion
Demographic		Gender			Useful for analysis
		Age	Yes		Useful for analysis
		Ethnicity	Yes		
		Country of birth	Yes	Yes	
		Sexuality	Yes	Yes	
Socioeconomic		Deprivation		Yes	Proxy for socio-economic predictors
Geographic		Country		All codes included	For analysis
		Region			For analysis
		Urban/Rural			To cover rural/urban distinction shown with poverty index above.
		Urban/Rural type			
Lifestyle		Injected drugs users	Yes	Yes	Includes ever smoked crack cocaine
		Binge-drinking or alcohol misuse	Yes	Yes	
		Smoking	Yes	Yes	
		Unsafe sex	Yes	Yes	
		Contact abroad	Yes	Yes	
		Male anal sex in the last ≥ 6 months	Yes	Yes	
		Multiple life partners	Yes	Yes	
		Obesity	Yes	Yes	
		Partner Characteristics	Yes	Yes	
		Stressful events	Yes	Yes	
Clinical conditions and comorbidities	Respiratory Conditions	Aspergillosis		Yes	in BHIVA guidelines
		Cough	Yes	Yes	
		Pneumocystis carinii	Yes	Yes	
		Pneumonia	Yes	Yes	
		Tuberculosis		Yes	in BHIVA guidelines
	Neurology	Cerebral toxoplasmosis abscess		Yes	in BHIVA guidelines
		Cryptococcal meningitis		Yes	in BHIVA guidelines
		Dementia		Yes	in BHIVA guidelines
		Guillain–Barré syndrome		Yes	in BHIVA guidelines
		Progressive multifocal leukoencephalopathy		Yes	in BHIVA guidelines
		Aseptic meningitis/encephalitis		Yes	in BHIVA guidelines
		Transverse myelitis		Yes	in BHIVA guidelines
		Neurologic disability	Yes	Yes	
		Peripheral neuropathy	Yes	Yes	
	Ophthalmology	Cytomegalovirus retinitis	No	Yes	in BHIVA guidelines
		Any unexplained retinopathy	No	Yes	in BHIVA guidelines
	Dermatology	Herpes zoster	Yes	Yes	
		Psoriasis	Yes	Yes	
		Seborrhoeic dermatitis	No	Yes	in BHIVA guidelines
		Kaposi's sarcoma	Yes	Yes	
		Rash	Yes	Yes	
	Gastroenterology	Abdominal pain	Yes	Yes	
		Oral candidiasis	Yes	Yes	
		Cryptosporidiosis	No	Yes	in BHIVA guidelines
		Diarrhoea	Yes	Yes	
		Hepatitis B	Yes	Yes	
		Hepatitis C	No	Yes	in BHIVA guidelines
		Oral hairy leukoplakia	No	Yes	in BHIVA guidelines

		Predictor	Identified in systematic review	Code list developed	Reasons for inclusion
		Nausea/vomiting	Yes	Yes	
		Salmonella, shigella or campylobacter	No	Yes	in BHIVA guidelines
		Weight loss	Yes	Yes	
	Haematology	Blood dyscrasia	Yes	Yes	
	Oncology	Non-Hodgkin's lymphoma	Yes	Yes	
		Anal cancer or anal intraepithelial dysplasia	No	Yes	in BHIVA guidelines
		Castleman's disease	No	Yes	in BHIVA guidelines
		Cervical dysplasia	No	Yes	in BHIVA guidelines
		Head and neck cancer	No	Yes	in BHIVA guidelines
		Hodgkin lymphoma	No	Yes	in BHIVA guidelines
		Space-occupying lesion of unknown cause	No	Yes	in BHIVA guidelines
		Lung cancer	No	Yes	in BHIVA guidelines
		Cervical cancer	Yes	Yes	
	Gynaecology	Cervical intraepithelial neoplasia	Yes	Yes	
		Seminoma	No	Yes	in BHIVA guidelines
		Vaginal intraepithelial neoplasia	No	Yes	in BHIVA guidelines
	Ear, nose, and throat	Parotitis	Yes	Yes	
		Lymphoepithelial parotid cysts	No	Yes	in BHIVA guidelines
	Other Clinical conditions	Mononucleosis-like syndrome	Yes	Yes	
		Pyrexia of unknown origin	Yes	Yes	
		Hypertension	Yes	Yes	
		Chronic liver disease	Yes	Yes	
		Hyperlipidemia	Yes	Yes	
		Depression	Yes	Yes	
		Lymphogranuloma venereum	Yes	Yes	
		Lymphadenopathy	Yes	Yes	
		Trichomoniasis	Yes	Yes	
		Diabetes	Yes	Yes	
		Minor trauma	Yes	Yes	
		Sexually Transmitted Infections (STI)	Yes		
		Chlamydia		Yes	Expansion of STI
		Syphilis		Yes	Expansion of STI
		Gonorrhoea		Yes	Expansion of STI
		Genital herpes		Yes	Expansion of STI
		Other STIs (excludes above)		Yes	Expansion of STI
		STI		Yes	Expansion of STI
		Previous STI		Yes	Expansion of STI
		All STI (excluding Syphilis) and previous STIs		Yes	Expansion of STI

Predictors identified in systematic review but not available in primary care records

Socioeconomic	Housing problems	Yes	No	Used deprivation as a proxy
	Annual income <$10,000	Yes	No	
	Farmworker	Yes	No	
	Unemployed	Yes	No	
	Education beyond high school	Yes	No	
	Not a high school graduate	Yes	No	
	Poverty index in rural areas	Yes	No	Used urban/rural
	Poverty index in urban areas	Yes	No	
Lifestyle	Ever smoked crack cocaine	Yes	No	Included under drug misuse
	Substance use (combined)**	Yes	No	Disaggregated to alcohol and drug misuse
	HIV positive partner	Yes	No	Included under drug misuse
	Sex with drug user	Yes	No	Included under partner characteristics
	Ever exchanged money or drugs for sex	Yes	No	
	Multiple partners	Yes	No	
	Lifetime partners	Yes	No	Included under multiple life partners
Clinical conditions	Fever or chills	Yes	No	Included under pyrexia of unknown origin

Appendix 5: Read Code list for Predictor variables

Appendix 5-A: Demographic and socio-economic

Ethnicity

Read code	Description	Condition	Broader category
9S1..00	White	Ethnicity	White
9S10.00	White British	Ethnicity	White
9S11.00	White Irish	Ethnicity	White
9S12.00	Other white ethnic group	Ethnicity	White
9S13.00	White Scottish	Ethnicity	White
9S14.00	Other white British ethnic group	Ethnicity	White
9S2..00	Black Caribbean	Ethnicity	Black
9S3..00	Black African	Ethnicity	Black African
9S4..00	Black, other, non-mixed origin	Ethnicity	Black
9S41.00	Black British	Ethnicity	Black
9S42.00	Black Caribbean/W.I./Guyana	Ethnicity	Black
9S42.11	Black Caribbean	Ethnicity	Black
9S42.12	Black West Indian	Ethnicity	Black
9S42.13	Black Guyana	Ethnicity	Black
9S43.00	Black N African/Arab/Iranian	Ethnicity	Black
9S43.11	Black North African	Ethnicity	Black
9S43.12	Black Arab	Ethnicity	Black
9S43.13	Black Iranian	Ethnicity	Black
9S44.00	Black - other African country	Ethnicity	Black
9S45.00	Black E Afric Asia/Indo-Caribb	Ethnicity	Black
9S45.11	Black East African Asian	Ethnicity	Black
9S45.12	Black Indo-Caribbean	Ethnicity	Black
9S46.00	Black Indian sub-continent	Ethnicity	Black
9S47.00	Black - other Asian	Ethnicity	Black
9S48.00	Black Black - other	Ethnicity	Black
9S5..00	Black - other, mixed	Ethnicity	Mixed
9S51.00	Other Black - Black/White orig	Ethnicity	Mixed
9S52.00	Other Black - Black/Asian orig	Ethnicity	Mixed
9S6..00	Indian	Ethnicity	Asian
9S7..00	Pakistani	Ethnicity	Asian
9S8..00	Bangladeshi	Ethnicity	Asian
9S9..00	Chinese	Ethnicity	Chinese
9SA..00	Other ethnic non-mixed (NMO)	Ethnicity	Other
9SA1.00	Brit. ethnic minor. spec.(NMO)	Ethnicity	Other
9SA2.00	Brit. ethnic minor. unsp (NMO)	Ethnicity	Other
9SA3.00	Caribbean I./W.I./Guyana (NMO)	Ethnicity	Black
9SA3.11	Caribbean Island (NMO)	Ethnicity	Black
9SA3.12	West Indian (NMO)	Ethnicity	Black
9SA3.13	Guyana (NMO)	Ethnicity	Black
9SA4.00	N African Arab/Iranian (NMO)	Ethnicity	Other
9SA4.11	North African Arab (NMO)	Ethnicity	Other
9SA4.12	Iranian (NMO)	Ethnicity	Other
9SA5.00	Other African countries (NMO)	Ethnicity	Black
9SA6.00	E Afric Asian/Indo-Carib (NMO)	Ethnicity	Asian
9SA6.11	East African Asian (NMO)	Ethnicity	Asian
9SA6.12	Indo-Caribbean (NMO)	Ethnicity	Asian
9SA7.00	Indian sub-continent (NMO)	Ethnicity	Asian
9SA8.00	Other Asian (NMO)	Ethnicity	Asian
9SA9.00	Irish (NMO)	Ethnicity	White
9SAA.00	Greek/Greek Cypriot (NMO)	Ethnicity	White
9SAA.11	Greek (NMO)	Ethnicity	White
9SAA.12	Greek Cypriot (NMO)	Ethnicity	White
9SAB.00	Turkish/Turkish Cypriot (NMO)	Ethnicity	White
9SAB.11	Turkish (NMO)	Ethnicity	White
9SAB.12	Turkish Cypriot (NMO)	Ethnicity	White
9SAC.00	Other European (NMO)	Ethnicity	White
9SAD.00	Other ethnic NEC (NMO)	Ethnicity	White
9SB..00	Other ethnic, mixed origin	Ethnicity	Mixed
9SB1.00	Other ethnic, Black/White orig	Ethnicity	Mixed
9SB2.00	Other ethnic, Asian/White orig	Ethnicity	Mixed
9SB3.00	Other ethnic, mixed white orig	Ethnicity	Mixed
9SB4.00	Other ethnic, other mixed orig	Ethnicity	Mixed
9SB5.00	Black Caribbean and White	Ethnicity	Mixed
9SB6.00	Black African and White	Ethnicity	Mixed
9SC..00	Vietnamese	Ethnicity	Other
9SG..00	Other black ethnic group	Ethnicity	Black
9SH..00	Other Asian ethnic group	Ethnicity	Asian
9SI..00	Irish traveller	Ethnicity	White
9SJ..00	Other ethnic group	Ethnicity	Other
9T1..00	New Zealand ethnic groups	Ethnicity	Other
9T11.00	New Zealand European	Ethnicity	Other
9T11.11	Pakeha	Ethnicity	Other
9T12.00	Other European in New Zealand	Ethnicity	Other
9T13.00	New Zealand Maori	Ethnicity	Other
9T14.00	Samoan	Ethnicity	Other
9T15.00	Cook Island Maori	Ethnicity	Other
9T16.00	Tongan	Ethnicity	Other
9T17.00	Niuean	Ethnicity	Other
9T18.00	Tokelauan	Ethnicity	Other
9T19.00	Fijian	Ethnicity	Other
9T1A.00	Other Pacific ethnic group	Ethnicity	Other
9T1B.00	South East Asian	Ethnicity	Asian
9T1C.00	Chinese	Ethnicity	Chinese
9T1D.00	Indian	Ethnicity	Asian
9T1E.00	Other Asian	Ethnicity	Asian
9T1Y.00	Other New Zealand ethnic group	Ethnicity	Other
9T1Z.00	New Zealand ethnic group NOS	Ethnicity	Other
9T2..00	Traveller - gypsy	Ethnicity	White
9T3..00	Yemeni	Ethnicity	Other
9T4..00	Romanian	Ethnicity	Other
9T5..00	Bulgarian	Ethnicity	Other
9T6..00	Czech	Ethnicity	Other
9T7..00	Slovak	Ethnicity	Other
9T8..00	Portuguese	Ethnicity	Other
9T9..00	Nepali	Ethnicity	Other
9i0..00	British or mixed British - ethnic category 2001 census	Ethnicity	White
9i00.00	White British - ethnic category 2001 census	Ethnicity	White
9i1..00	Irish - ethnic category 2001 census	Ethnicity	White
9i10.00	White Irish - ethnic category 2001 census	Ethnicity	White
9i2..00	Other White background - ethnic category 2001 census	Ethnicity	White
9i20.00	English - ethnic category 2001 census	Ethnicity	White
9i21.00	Scottish - ethnic category 2001 census	Ethnicity	White
9i22.00	Welsh - ethnic category 2001 census	Ethnicity	White
9i23.00	Cornish - ethnic category 2001 census	Ethnicity	White
9i24.00	Northern Irish - ethnic category 2001 census	Ethnicity	White
9i25.00	Ulster Scots - ethnic category 2001 census	Ethnicity	White
9i26.00	Cypriot (part not stated) - ethnic category 2001 census	Ethnicity	White
9i27.00	Greek - ethnic category 2001 census	Ethnicity	White
9i28.00	Greek Cypriot - ethnic category 2001 census	Ethnicity	White
9i29.00	Turkish - ethnic category 2001 census	Ethnicity	White
9i2A.00	Turkish Cypriot - ethnic category 2001 census	Ethnicity	White
9i2B.00	Italian - ethnic category 2001 census	Ethnicity	White
9i2C.00	Irish Traveller - ethnic category 2001 census	Ethnicity	White
9i2D.00	Traveller - ethnic category 2001 census	Ethnicity	White
9i2E.00	Gypsy/Romany - ethnic category 2001 census	Ethnicity	White
9i2F.00	Polish - ethnic category 2001 census	Ethnicity	White
9i2G.00	Baltic Estonian/Latvian/Lithuanian - ethn categ 2001 census	Ethnicity	White
9i2H.00	Commonwealth (Russian) Indep States - ethn categ 2001 census	Ethnicity	White
9i2J.00	Kosovan - ethnic category 2001 census	Ethnicity	White
9i2K.00	Albanian - ethnic category 2001 census	Ethnicity	White
9i2L.00	Bosnian - ethnic category 2001 census	Ethnicity	White
9i2M.00	Croatian - ethnic category 2001 census	Ethnicity	White
9i2N.00	Serbian - ethnic category 2001 census	Ethnicity	White
9i2P.00	Other republics former Yugoslavia - ethnic categ 2001 census	Ethnicity	White
9i2Q.00	Mixed Irish and other White - ethnic category 2001 census	Ethnicity	White
9i2R.00	Oth White European/European unsp/Mixed European 2001 census	Ethnicity	White
9i2S.00	Other mixed White - ethnic category 2001 census	Ethnicity	White
9i2T.00	Other White or White unspecified ethnic category 2001 census	Ethnicity	White
9i3..00	White and Black Caribbean - ethnic category 2001 census	Ethnicity	Mixed
9i4..00	White and Black African - ethnic category 2001 census	Ethnicity	Mixed
9i5..00	White and Asian - ethnic category 2001 census	Ethnicity	Mixed
9i6..00	Other Mixed background - ethnic category 2001 census	Ethnicity	Mixed
9i60.00	Black and Asian - ethnic category 2001 census	Ethnicity	Mixed
9i61.00	Black and Chinese - ethnic category 2001 census	Ethnicity	Mixed
9i62.00	Black and White - ethnic category 2001 census	Ethnicity	Mixed
9i63.00	Chinese and White - ethnic category 2001 census	Ethnicity	Mixed
9i64.00	Asian and Chinese - ethnic category 2001 census	Ethnicity	Mixed
9i65.00	Other Mixed or Mixed unspecified ethnic category 2001 census	Ethnicity	Mixed
9i7..00	Indian or British Indian - ethnic category 2001 census	Ethnicity	Asian
9i8..00	Pakistani or British Pakistani - ethnic category 2001 census	Ethnicity	Asian
9i9..00	Bangladeshi or British Bangladeshi - ethn categ 2001 census	Ethnicity	Asian
9iA..00	Other Asian background - ethnic category 2001 census	Ethnicity	Asian
9iA1.00	Punjabi - ethnic category 2001 census	Ethnicity	Asian
9iA2.00	Kashmiri - ethnic category 2001 census	Ethnicity	Asian
9iA3.00	East African Asian - ethnic category 2001 census	Ethnicity	Asian
9iA4.00	Sri Lankan - ethnic category 2001 census	Ethnicity	Asian
9iA5.00	Tamil - ethnic category 2001 census	Ethnicity	Asian
9iA6.00	Sinhalese - ethnic category 2001 census	Ethnicity	Asian
9iA7.00	Caribbean Asian - ethnic category 2001 census	Ethnicity	Asian
9iA8.00	British Asian - ethnic category 2001 census	Ethnicity	Asian
9iA9.00	Mixed Asian - ethnic category 2001 census	Ethnicity	Asian
9iAA.00	Other Asian or Asian unspecified ethnic category 2001 census	Ethnicity	Asian
9iB..00	Caribbean - ethnic category 2001 census	Ethnicity	Black
9iC..00	African - ethnic category 2001 census	Ethnicity	Black African
9iD..00	Other Black background - ethnic category 2001 census	Ethnicity	Black
9iD0.00	Somali - ethnic category 2001 census	Ethnicity	Black African
9iD1.00	Nigerian - ethnic category 2001 census	Ethnicity	Black African
9iD2.00	Black British - ethnic category 2001 census	Ethnicity	Black
9iD3.00	Mixed Black - ethnic category 2001 census	Ethnicity	Black
9iD4.00	Other Black or Black unspecified ethnic category 2001 census	Ethnicity	Black
9iE..00	Chinese - ethnic category 2001 census	Ethnicity	Chinese
9iF..00	Other - ethnic category 2001 census	Ethnicity	Other
9iF0.00	Vietnamese - ethnic category 2001 census	Ethnicity	Other
9iF1.00	Japanese - ethnic category 2001 census	Ethnicity	Other
9iF2.00	Filipino - ethnic category 2001 census	Ethnicity	Other
9iF3.00	Malaysian - ethnic category 2001 census	Ethnicity	Other
9iF9.00	Arab - ethnic category 2001 census	Ethnicity	Other
9iFA.00	North African - ethnic category 2001 census	Ethnicity	Other
9iFB.00	Mid East (excl Israeli, Iranian & Arab) - eth cat 2001 cens	Ethnicity	Other
9iFC.00	Israeli - ethnic category 2001 census	Ethnicity	Other
9iFD.00	Iranian - ethnic category 2001 census	Ethnicity	Other
9iFE.00	Kurdish - ethnic category 2001 census	Ethnicity	Other
9iFF.00	Moroccan - ethnic category 2001 census	Ethnicity	Other
9iFG.00	Latin American - ethnic category 2001 census	Ethnicity	Other
9iFH.00	South and Central American - ethnic category 2001 census	Ethnicity	Other
9iFJ.00	Mauritian/Seychellois/Maldivian/St Helena eth cat 2001census	Ethnicity	Other
9iFK.00	Any other group - ethnic category 2001 census	Ethnicity	Other

Appendix 5-B: Lifestyle of behavioural

Drug misuse

Read code	Description
13c..00	Drug user
1P30.00	Compulsive uncontrollable drug taking
1P31.00	Compulsive drug taking
8B23.00	Drug addiction therapy
8B23.11	Drug addictn therap-methadone
8B23.12	FP10(MDA) issued
8B23.13	Drug dependence therapy
8B2N.00	Drug addiction detoxification therapy - methadone
8B2P.00	Drug addiction maintenance therapy - methadone
8B2Q.00	Drug addiction maintenance therapy - buprenorphine
8B2R.00	Drug addiction detoxification therapy - buprenorphine
8BA9.00	Detoxification dependence drug
8BAd.00	Opiate dependence detoxification
8FB0.00	Drug detoxification programme completed

Read code	Description
8FB..00	Drug rehabilitation
E02..00	Drug psychoses
E24..00	Drug dependence
E24..11	Drug addiction
E25..00	Nondependent abuse of drugs
SL50100	Heroin poisoning
SL50200	Methadone poisoning
SL50z00	Opiate or narcotic poisoning NOS
SL85000	Cocaine poisoning
SL96.00	Hallucinogen poisoning
SL97.00	Psychostimulant poisoning
SL97.11	Stimulant poisoning
SL9y.00	Other psychotropic agent poisoning
SL9z.00	Psychotropic agent poisoning NOS
1V0C.00	Drug addict

Smoking status

Read code	Description
137..00	Tobacco consumption
137..11	Smoker - amount smoked
1372.11	Occasional smoker
1373..0	Light smoker - 1-9 cigs/day
1374..0	Moderate smoker - 10-19 cigs/d
1375..0	Heavy smoker - 20-39 cigs/day
1376..0	Very heavy smoker - 40+cigs/d
1379..0	Ex-moderate smoker (10-19/day)
137A.00	Ex-heavy smoker (20-39/day)
137a.00	Pipe tobacco consumption
137B.00	Ex-very heavy smoker (40+/day)

Obesity

Read code	Description
22K2.00	Body Mass Index high K/M2
22K4.00	Body mass index index 25-29 - overweight
22K5.00	Body mass index 30+ - obesity
22K7.00	Body mass index 40+ - severely obese
22K9.00	Body mass index centile
22K9000	Baseline body mass index centile

Alcohol misuse

Read code	Description
136S.00	Hazardous alcohol use
136T.00	Harmful alcohol use
63C7.00	Maternal alcohol abuse
66e..00	Alcohol disorder monitoring
66e0.00	Alcohol abuse monitoring
8BA8.00	Alcohol detoxification
C150500	Alcohol-induced pseudo-Cushing's syndrome
E01..00	Alcoholic psychoses
E010.12	Delirium tremens
E010.11	DTs - delirium tremens
E010.00	Alcohol withdrawal delirium
E011.00	Alcohol amnestic syndrome
E011000	Korsakov's alcoholic psychosis
E011100	Korsakov's alcoholic psychosis with peripheral neuritis
E011200	Wernicke-Korsakov syndrome
E011z00	Alcohol amnestic syndrome NOS
E012.11	Alcoholic dementia NOS
E012.00	Other alcoholic dementia
E012000	Chronic alcoholic brain syndrome
E013.00	Alcohol withdrawal hallucinosis
E014.11	Drunkenness - pathological
E014.00	Pathological alcohol intoxication
E015.00	Alcoholic paranoia
E01y.00	Other alcoholic psychosis
E01y000	Alcohol withdrawal syndrome
E01yz00	Other alcoholic psychosis NOS
E01z.00	Alcoholic psychosis NOS
E23..11	Alcoholism
E23..12	Alcohol problem drinking
E23..00	Alcohol dependence syndrome
E230.00	Acute alcoholic intoxication in alcoholism
E230.11	Alcohol dependence with acute alcoholic intoxication
E230000	Acute alcoholic intoxication, unspecified, in alcoholism
E230100	Continuous acute alcoholic intoxication in alcoholism
E230200	Episodic acute alcoholic intoxication in alcoholism
E230300	Acute alcoholic intoxication in remission, in alcoholism
E230z00	Acute alcoholic intoxication in alcoholism NOS
E231.11	Dipsomania
E231.00	Chronic alcoholism
E231000	Unspecified chronic alcoholism
E231100	Continuous chronic alcoholism
E231200	Episodic chronic alcoholism
E231300	Chronic alcoholism in remission
E231z00	Chronic alcoholism NOS
E23z.00	Alcohol dependence syndrome NOS
E250.13	Inebriety NOS
E250.14	Intoxication - alcohol
E250.11	Drunkenness NOS
E250.00	Nondependent alcohol abuse
E250.12	Hangover (alcohol)
E250000	Nondependent alcohol abuse, unspecified
E250100	Nondependent alcohol abuse, continuous
E250200	Nondependent alcohol abuse, episodic
E250300	Nondependent alcohol abuse in remission

Read code	Description
E250z00	Nondependent alcohol abuse NOS
Eu10.00	[X]Mental and behavioural disorders due to use of alcohol
Eu10000	[X]Mental & behav dis due to use alcohol: acute intoxication
Eu10011	[X]Acute alcoholic drunkenness
Eu10100	[X]Mental and behav dis due to use of alcohol: harmful use
Eu10200	[X]Mental and behav dis due to use alcohol: dependence syndr
Eu10213	[X]Dipsomania
Eu10212	[X]Chronic alcoholism
Eu10211	[X]Alcohol addiction
Eu10300	[X]Mental and behav dis due to use alcohol: withdrawal state
Eu10400	[X]Men & behav dis due alcohl: withdrawl state with delirium
Eu10411	[X]Delirium tremens, alcohol induced
Eu10512	[X]Alcoholic jealousy
Eu10513	[X]Alcoholic paranoia
Eu10511	[X]Alcoholic hallucinosis
Eu10500	[X]Mental & behav dis due to use alcohol: psychotic disorder
Eu10514	[X]Alcoholic psychosis NOS
Eu10600	[X]Mental and behav dis due to use alcohol: amnesic syndrome
Eu10611	[X]Korsakov's psychosis, alcohol induced
Eu10700	[X]Men & behav dis due alcoh: resid & late-onset psychot dis
Eu10711	[X]Alcoholic dementia NOS
Eu10712	[X]Chronic alcoholic brain syndrome
Eu10800	[X]Alcohol withdrawal-induced seizure
Eu10y00	[X]Men & behav dis due to use alcohol: oth men & behav dis
Eu10z00	[X]Ment & behav dis due use alcohol: unsp ment & behav dis
F11x000	Cerebral degeneration due to alcoholism
F11x011	Alcoholic encephalopathy
F144000	Cerebellar ataxia due to alcoholism
F374700	Polyneuropathy in pellagra
F375.00	Alcoholic polyneuropathy
F394100	Alcoholic myopathy
G555.00	Alcoholic cardiomyopathy
G852300	Oesophageal varices in alcoholic cirrhosis of the liver
J153.00	Alcoholic gastritis
J610.00	Alcoholic fatty liver
J611.00	Acute alcoholic hepatitis
J612.11	Florid cirrhosis
J612.00	Alcoholic cirrhosis of liver
J612.12	Laennec's cirrhosis
J612000	Alcoholic fibrosis and sclerosis of liver
J613.00	Alcoholic liver damage unspecified
J613000	Alcoholic hepatic failure
J617.00	Alcoholic hepatitis
J617000	Chronic alcoholic hepatitis
J671000	Alcohol-induced chronic pancreatitis
136P.00	Heavy drinker
136Q.00	Very heavy drinker
136R.00	Binge drinker
136V.00	Alcohol units per week
136W.00	Alcohol misuse
136X.00	Alcohol units consumed on heaviest drinking day
136Y.00	Drinks in morning to get rid of hangover
136Z.00	Alcohol consumption NOS

Unsafe sex

Read code	Description
1565.11	Unprotected sex
1565.12	No protection used for sex
8CdA.00	Advice given about risks of unprotected sexual intercourse

Anal sex

Read code	Description
1ABA.00	Sexual activity - anal sex
1ABL.00	Homosexual activity
1b21.00	Male homosexual
1b21.11	Male homosexuality
E220000	Male homosexuality
E225200	Trans-sexuality with homosexual history

Lifetime partners

Read code	Description
0AL..00	Sex worker
13N5.12	Multiple sexual partners
14OP.00	At risk of sexually transmitted infection
14OP000	At high risk of sexually transmitted infection
1P70.00	Number of sexual partners in past year
1P71.00	New sexual partner
ZV4K400	[V]High-risk sexual behaviour
14Of.00	Former sex worker

Stressful events

Read code	Description
13M..00	Family bereavement
13M..11	Death of family member
13M1.00	Death of spouse
13M2.00	Death of infant
13M5.00	Death of son
13M6.00	Death of daughter
13M7.00	Death of father
13M8.00	Death of mother
13M9.00	Death of brother
13MA.00	Death of sister
13MB.00	Suicide of close relative
13MC.00	Death of sibling
13MD.00	Death of child
13ME.00	Relative killed
13MF.00	Death of partner
13MG.00	Death of wife
13MH.00	Husband died
13MI.00	Death of husband
13MZ.00	Family bereavement NOS
13MZ.11	Death of relative
ZV61000	[V]Family disruption
ZV61011	[V]Divorce
ZV61012	[V]Estrangement
ZV61100	[V]Marital problems

Read code	Description
13H2.00	Separation
13H3.00	Divorce
13H3.11	Divorce problems
13H3000	Divorce proceedings
13H3100	Divorce proceedings pending
13H4.00	Marital problems
13H4.11	Marital trouble
13H4.12	Marital stress
13H4100	Marital breakdown
13H4200	Marital conflict
13H4211	Marital discord
13H4212	Marital disharmony
13Hc.00	Bereavement
13HD.00	Violent spouse
13HE.00	Engaged
13He.00	Personal problems
13HF.00	Broken engagement
13Hf.00	Unable to cope
13HG.00	Broken with partner
13K1.00	Financial problem
13K1000	In debt
13K2.00	Financial circumstances change
13K3.00	Poverty
13K4.00	Bankruptcy

Appendix 5-C: Clinical and comorbid conditions

Respiratory

Aspergillosis

Read code	Description
AB63.00	Aspergillosis
AB63000	Invasive pulmonary aspergillosis
AB63100	Tonsillar aspergillosis
AB63200	Disseminated aspergillosis
AB63300	Allergic bronchopulmonary aspergillosis
AB63X00	Aspergillosis, unspecified
AyuEK00	[X]Other forms of aspergillosis
AyuEL00	[X]Aspergillosis, unspecified
AyuEU00	[X]Other pulmonary aspergillosis
F501B00	Chronic otitis externa due to aspergillosis
H246.00	Pneumonia with aspergillosis

Pneumocystis carinii

Read code	Description
43f8.00	Pneumocystis carinii IF
43fD.00	Pneumocystis carinii antigen level
H24y200	Pneumonia with pneumocystis carinii

Cough

Read code	Description
171..00	Cough
171..11	C/O - cough
1712	Dry cough
1713	Productive cough -clear sputum
1714	Productive cough -green sputum
1715	Productive cough-yellow sputum
1716	Productive cough NOS
1716.11	Coughing up phlegm
1717	Night cough present
1718	Night cough absent
1719	Chesty cough
1719.11	Bronchial cough
171A.00	Chronic cough
171B.00	Persistent cough
171C.00	Morning cough
171D.00	Evening cough
171E.00	Unexplained cough
171F.00	Cough with fever
171J.00	Reflux cough
171Z.00	Cough symptom NOS
E261100	Psychogenic cough
Eu45316	[X]Psychogenic cough
R062.00	[D]Cough
R062000	[D]Cough syncope
R062100	[D]Episodic dry cough
R063000	[D]Cough with haemorrhage

Tuberculosis

Read code	Description
A11..00	Pulmonary tuberculosis
A11..11	Lung tuberculosis
A110.00	Infiltrative lung tuberculosis
A111.00	Nodular lung tuberculosis
A112.00	Tuberculosis of lung with cavitation
A113.00	Tuberculosis of bronchus
A114.00	Tuberculous fibrosis of lung
A115.00	Tuberculous bronchiectasis
A116.00	Tuberculous pneumonia
A117.00	Tuberculous pneumothorax
A11y.00	Other specified pulmonary tuberculosis
A11z.00	Pulmonary tuberculosis NOS
A12..00	Other respiratory tuberculosis
A120.00	Tuberculous pleurisy
A120000	Tuberculosis of pleura
A120100	Tuberculous empyema
A120200	Tuberculous hydrothorax
A120z00	Tuberculous pleurisy NOS
A121.00	Tuberculosis of intrathoracic lymph nodes
A121000	Tuberculosis of hilar lymph nodes
A121100	Tuberculosis of mediastinal lymph nodes
A121200	Tuberculosis of tracheobronchial lymph nodes
A121z00	Tuberculosis of intrathoracic lymph nodes NOS
A122.00	Isolated tracheal or bronchial tuberculosis
A122000	Isolated tracheal tuberculosis
A122100	Isolated bronchial tuberculosis
A122z00	Isolated tracheal or bronchial tuberculosis NOS
A123.00	Tuberculous laryngitis
A124100	Tuberculosis of lung, confirmed by culture only
A124200	Tuberculosis of lung, confirmed histologically
A124300	Tuberculosis of lung, confirmed by unspecified means
A124500	Tuberculosis of larynx, trachea & bronchus conf bact/hist'y
A124600	Tuberculous pleurisy, conf bacteriologically/histologically
A125000	Tuberculosis of lung, bacteriologically & histolog'y neg
A125100	Tuberculosis lung bact and histological examin not done
A12y.00	Other specified respiratory tuberculosis
A12y000	Tuberculosis of mediastinum
A12y100	Tuberculosis of nasopharynx
A12y200	Tuberculosis of nasal septum
A12y300	Tuberculosis of nasal sinus
A12yz00	Other specified respiratory tuberculosis NOS
A13..00	Tuberculosis of meninges and central nervous system
A130.00	Tuberculous meningitis
A130000	Tuberculosis of cerebral meninges
A130100	Tuberculosis of spinal meninges
A130200	Tuberculous leptomeningitis
A130300	Tuberculous meningoencephalitis
A130z00	Tuberculous meningitis NOS
A131.00	Tuberculoma of meninges
A132.00	Tuberculoma of brain
A133.00	Tuberculous abscess of brain
A134.00	Tuberculoma of spinal cord
A135.00	Tuberculous abscess of spinal cord
A136.00	Tuberculous encephalitis or myelitis
A136000	Tuberculous encephalitis
A136100	Tuberculous myelitis
A136z00	Tuberculous encephalitis or myelitis NOS
A13y.00	Other specified tuberculosis of central nervous system
A13z.00	Tuberculosis of central nervous system NOS
A14..00	Tuberculosis of intestines, peritoneum and mesenteric glands
A140.00	Tuberculous peritonitis
A14y.00	Other gastrointestinal tract tuberculosis
A14y000	Tuberculosis of anus
A14y100	Tuberculosis of large intestine
A14y200	Tuberculosis of small intestine
A14y300	Tuberculosis of mesenteric lymph glands
A14y400	Tuberculosis of rectum
A14y500	Tuberculosis of retroperitoneal lymph nodes
A14yz00	Other gastrointestinal tract tuberculosis NOS
A14z.00	Tuberculosis of gastrointestinal tract NOS
A15..00	Tuberculosis of bones and joints
A15..11	Tuberculous osteomylelytis
A15..12	Tuberculous arthritis

Read code	Description
A15..13	Tuberculous synovitis
A150.00	Tuberculosis of vertebral column - Pott's
A151.00	Tuberculosis of hip
A152.00	Tuberculosis of knee
A153.00	Tuberculosis limb bones - Tuberculous dactylitis
A154.00	Tuberculous mastoiditis
A15x.00	Tuberculosis of other specified bones
A15y.00	Tuberculosis of other specified joint
A15z.00	Tuberculosis of bones or joints NOS
A16..00	Tuberculosis of genitourinary system
A160.00	Tuberculosis of kidney
A160.11	Renal tuberculosis
A160000	Tuberculous nephropathy
A160100	Tuberculous pyelitis
A160200	Tuberculous pyelonephritis
A160z00	Tuberculosis of kidney NOS
A161.00	Tuberculosis of bladder
A162.00	Tuberculosis of ureter
A163.00	Tuberculosis of other urinary organs
A164.00	Tuberculosis of epididymis
A165.00	Tuberculosis of other male genital organs
A165000	Tuberculosis of prostate
A165100	Tuberculosis seminal vesicle
A165200	Tuberculosis of testis
A165z00	Tuberculosis of other male genital organs NOS
A166.00	Tuberculous oophoritis or salpingitis
A166000	Tuberculous oophoritis
A166100	Tuberculous salpingitis
A166111	Fallopian tube tuberculosis
A166z00	Tuberculous oophoritis or salpingitis NOS
A167.00	Tuberculosis of other female genital organs
A167000	Tuberculous cervicitis
A167100	Tuberculous endometritis
A167z00	Tuberculosis of other female genital organs NOS
A168.00	Tuberculosis of urinary tract
A16z.00	Genitourinary tuberculosis NOS
A17..00	Tuberculosis of other organs
A170.00	Tuberculosis of skin and subcutaneous tissue
A170.11	Lupus - tuberculous
A170000	Tuberculosis - lupus exedens
A170100	Tuberculosis - lupus vulgaris
A170200	Tuberculosis - scrofuloderma
A170300	Tuberculosis - lupus NOS
A170400	Tuberculosis colliquativa
A170500	Tuberculosis cutis
A170600	Tuberculosis lichenoides
A170700	Tuberculosis papulonecrotica
A170800	Tuberculosis verrucosa cutis
A170z00	Tuberculosis of skin and subcutaneous tissue NOS
A171.00	Tuberculosis with erythema nodosum hypersensitivity reaction
A171100	Tuberculous erythema nodosum
A171z00	Erythema nodosum with tuberculosis NOS
A172.00	Tuberculosis of peripheral lymph nodes
A172000	Tuberculous - cervical lymphadenitis
A172011	Scrofula - tuberculous cervical lymph nodes
A172100	Scrofulous tuberculous abscess
A172200	Tuberculous adenitis
A172z00	Tuberculosis of peripheral lymph nodes NOS
A173.00	Tuberculosis of eye
A173000	Tuberculous chorioretinitis
A173100	Tuberculous episcleritis
A173200	Tuberculous interstitial keratitis
A173300	Tuberculous chronic iridocyclitis
A173400	Tuberculous keratoconjunctivitis
A173z00	Tuberculosis of eye NOS
A174.00	Tuberculosis of ear
A175.00	Tuberculosis of thyroid gland
A176.00	Tuberculosis of adrenal glands - Addison's disease
A177.00	Tuberculosis spleen
A178.00	Tuberculosis oesophagus
A17y.00	Tuberculosis of other specified organs
A17y000	Tuberculosis endocardium
A17y100	Tuberculosis myocardium
A17y200	Tuberculosis pericardium

Pneumonia

Read code	Description
H21..00	Lobar (pneumococcal) pneumonia
H21..11	Chest infection - pneumococcal pneumonia
H22..00	Other bacterial pneumonia
H22..11	Chest infection - other bacterial pneumonia
H220.00	Pneumonia due to klebsiella pneumoniae
H221.00	Pneumonia due to pseudomonas
H222.00	Pneumonia due to haemophilus influenzae
H222.11	Pneumonia due to haemophilus influenzae
H223.00	Pneumonia due to streptococcus
H223000	Pneumonia due to streptococcus, group B
H224.00	Pneumonia due to staphylococcus
H22y.00	Pneumonia due to other specified bacteria
H22y000	Pneumonia due to escherichia coli
H22y011	E.coli pneumonia
H22y100	Pneumonia due to proteus
H22y200	Pneumonia - Legionella
H22y300	Pneumonia due to Gram neg bact
H22yX00	Pneumonia due to other aerobic gram-negative bacteria
H22yz00	Pneumonia due to bacteria NOS
H22z.00	Bacterial pneumonia NOS
H23..00	Pneumonia due to other specified organisms
H23..11	Chest infection - pneumonia organism OS
H230.00	Pneumonia due to Eaton's agent
H231.00	Pneumonia due to mycoplasma pneumoniae
H232.00	Pneumonia due to pleuropneumonia like organisms
H233.00	Chlamydial pneumonia
H23z.00	Pneumonia due to specified organism NOS
H24..00	Pneumonia with infectious diseases EC
H24..11	Chest infection with infectious disease EC
H240.00	Pneumonia with measles
H241.00	Pneumonia with cytomegalic inclusion disease
H242.00	Pneumonia with ornithosis
H243.00	Pneumonia with whooping cough
H243.11	Pneumonia with pertussis
H244.00	Pneumonia with tularaemia
H245.00	Pneumonia with anthrax

Read code	Description
H246.00	Pneumonia with aspergillosis
H247.00	Pneumonia with other systemic mycoses
H247000	Pneumonia with candidiasis
H247100	Pneumonia with coccidioidomycosis
H247200	Pneumonia with histoplasmosis
H247z00	Pneumonia with systemic mycosis NOS
H24y.00	Pneumonia with other infectious diseases EC
H24y000	Pneumonia with actinomycosis
H24y100	Pneumonia with nocardiasis
H24y200	Pneumonia with pneumocystis carinii
H24y300	Pneumonia with Q-fever
H24y400	Pneumonia with salmonellosis
H24y500	Pneumonia with toxoplasmosis
H24y600	Pneumonia with typhoid fever
H24y700	Pneumonia with varicella
H24yz00	Pneumonia with other infectious diseases EC NOS
H24z.00	Pneumonia with infectious diseases EC NOS
H25..00	Bronchopneumonia due to unspecified organism
H25..11	Chest infection - unspecified bronchopneumonia
H26..00	Pneumonia due to unspecified organism
H26..11	Chest infection - pnemonia due to unspecified organism
H260.00	Lobar pneumonia due to unspecified organism
H260000	Lung consolidation
H261.00	Basal pneumonia due to unspecified organism
H262.00	Postoperative pneumonia
H263.00	Pneumonitis, unspecified
Hyu0900	[X]Pneumonia due to other aerobic gram-negative bacteria
Hyu0A00	[X]Other bacterial pneumonia
Hyu0B00	[X]Pneumonia due to other specified infectious organisms
Hyu0C00	[X]Pneumonia in bacterial diseases classified elsewhere
Hyu0D00	[X]Pneumonia in viral diseases classified elsewhere
Hyu0E00	[X]Pneumonia in mycoses classified elsewhere
Hyu0F00	[X]Pneumonia in parasitic diseases classified elsewhere
Hyu0G00	[X]Pneumonia in other diseases classified elsewhere
Hyu0H00	[X]Other pneumonia, organism unspecified

Neurology

Any unexplained retinopathy

Read code	Description
F421z00	Other background retinopathy NOS
F422.00	Other proliferative retinopathy
FyuF700	[X]Other proliferative retinopathy
F422z00	Proliferative retinopathy NOS
F421.00	Other background retinopathy
F421000	Unspecified background retinopathy

Transverse myelitis

Read code	Description
F03..13	Transverse myelitis
F037.00	Transverse myelitis
F037000	Varicella transverse myelitis

Aseptic meningitis/encephalitis

Read code	Description
A42z.11	Aseptic meningitis
F022.00	Chronic meningitis
F024.00	Benign recurrent meningitis
F02z.00	Unspecified meningitis
A130300	Tuberculous meningoencephalitis
A136.00	Tuberculous encephalitis or myelitis
A136000	Tuberculous encephalitis
A136z00	Tuberculous encephalitis or myelitis NOS
A361.00	Meningococcal encephalitis
A412.00	Subacute sclerosing panencephalitis
A412.11	Dawson's inclusion body encephalitis
A4y0.00	Enteroviral encephalitis
A4zy000	Acute inclusion body encephalitis
A4zy100	Acute necrotising encephalitis
A4zy200	Epidemic encephalitis
A4zy300	Encephalitis lethargica
A4zy400	Von Economo's encephalitis
A4zy500	Adenoviral encephalitis
A4zz.11	Viral encephalitis NOS
F03..00	Encephalitis, myelitis and encephalomyelitis
F030.00	Encephalitis in viral disease EC
F030000	Encephalitis due to kuru
F030011	Kuru encephalitis
F030100	Encephalitis due to subacute sclerosing panencephalitis
F030200	Encephalitis due to poliomyelitis
F030211	Poliomyelitis encephalitis
F030300	Encephalitis due to arthropod-borne virus
F030400	Encephalitis due to herpes simplex virus
F030411	Herpes simplex encephalitis
F030500	Encephalitis due to mumps virus
F030511	Mumps encephalitis
F030600	Encephalitis due to rubella virus
F030611	Rubella encephalitis
F030700	Encephalitis due to cytomegalovirus
F030711	Cytomegaloviral encephalitis
F030800	Encephalitis due to influenza-specific virus not identified

Read code	Description
F030900	Encephalitis due to herpes zoster
F030911	Herpes zoster encephalitis
F030A00	Encephalitis due to influenza-virus identified
F030z00	Encephalitis in viral disease NOS
F031.00	Encephalitis due to rickettsia EC
F032.00	Encephalitis due to protozoa EC
F032000	Encephalitis due to malaria
F032011	Malarial encephalitis
F032100	Encephalitis due to trypanosomiasis
F032111	Trypanosomiasis encephalitis
F032z00	Encephalitis due to protozoa EC NOS
F033.00	Encephalitis due to other infection EC
F033000	Encephalitis due to meningococcus
F033011	Meningococcal encephalitis
F033100	Encephalitis due to congenital syphilis
F033111	Syphilis encephalitis
F033200	Encephalitis due to syphilis unspecified
F033300	Encephalitis due to tuberculosis
F033311	Tuberculous encephalitis
F033400	Encephalitis due to toxoplasmosis
F033411	Toxoplasmosis encephalitis
F033z00	Unspecified encephalitis due to other infection EC
F035.00	Postinfectious encephalitis
F035000	Encephalitis following chickenpox
F035011	Encephalitis due to varicella
F035100	Encephalitis following measles
F03X.00	Bacterial meningoencephalitis+meningomyelitis,NEC
F03y.00	Other causes of encephalitis
F03z.00	Encephalitis NOS
F212.00	Acute and subacute haemorrhagic leukoencephalitis [Hurst]
Fyu0500	[X]Bacterial meningoencephalitis+meningomyelitis,NEC
Fyu0600	[X]Other encephalitis, myelitis and encephalomyelitis
Fyu0700	[X]Encephalitis,myelitis+encephalomyelitis/bactrl disease CE
Fyu0800	[X]Encephalitis,myelitis+encephalomyelitis/viral disease CE
Fyu0A00	[X]Encephalitis,myelitis+encephalomyelitis/other diseases CE
N000611	Systemic lupus erythematosus encephalitis
ZV05000	[V]Arthropod-borne viral encephalitis vaccination

Cerebral toxoplasmosis abscess

Read code	Description
AD00.00	Toxoplasma meningoencephalitis
F033400	Encephalitis due to toxoplasmosis
F033411	Toxoplasmosis encephalitis
F04..00	Intracranial and intraspinal abscesses
F040.00	Intracranial abscess
F040.11	Brain abscess
F040000	Cerebral intracranial abscess
F040011	Cerebral abscess
F040100	Cerebellar intracranial abscess
F040111	Cerebellar abscess
F040200	Otogenic intracranial abscess

Read code	Description
F040211	Otogenic intracranial abscess
F040300	Epidural intracranial abscess
F040311	Epidural intracranial abscess
F040400	Extradural intracranial abscess
F040500	Subdural intracranial abscess
F040511	Subdural intracranial abscess
F040600	Tuberculous intracranial abscess
F040z00	Intracranial abscess NOS
F04X.00	Extradural and subdural abscess, unspecified
F04z.00	Intracranial or intraspinal abscess NOS
F04z000	Epidural abscess

Cryptococcal meningitis

Read code	Description
AB65200	Cryptococcal meningitis

Neurologic disability

Read code	Description
1BZ2.00	Transient neurological symptoms
1BZ3.00	Progressive intellectual and neurological deterioration
1JA..00	Suspected neurological disease
1O3..00	Neurological disorder confirmed
7P21.00	Rehabilitation for neurological disorders
7P21y00	Other specified rehabilitation for neurological disorders
7P21z00	Rehabilitation for neurological disorders NOS
7Q04.00	High cost neurology drugs
7Q04y00	Other specified high cost neurology drugs
7Q04z00	High cost neurology drugs NOS
8A3..00	Neurological monitoring
8A3Z.00	Neurological monitoring NOS
8H13.00	Admit to neurological ITU
8H2E.00	Admit neurology emergency
8H3J.00	Non-urgent neurology admission
Z6I..13	Neurological physiotherapy treatments
Z6I..14	Neurological techniques
Z6I4.00	Neurological functional rehabilitation
ZS94.12	Soft neurological signs
13VCB00	Neurodisability
1BZ2.00	Transient neurological symptoms
1BZ3.00	Progressive intellectual and neurological deterioration
1J05.00	Suspected neuroblastoma
1JA..00	Suspected neurological disease
1M8..00	Diabetic peripheral neuropathic pain
1O3..00	Neurological disorder confirmed
285..00	Neurotic condition, insight present
B546.00	Neuroblastoma
B7F1000	Acoustic neuroma
B7Fy100	Morton neuroma
B927.00	Neurofibromatosis - Von Recklinghausen's disease
B927.12	Neurofibromatosis type 1
B929.00	Neurofibromatosis type 2
BB5R900	[M]Neuroendocrine carcinoma
BBa4.00	[M]Melanotic neuroectodermal tumour

Read code	Description
BBba.00	[M]Primitive neuroectodermal tumour
BBba000	[M]Peripheral neuroectodermal tumour
BBc..00	[M]Neuroepitheliomatous neoplasms
BBc0.00	[M]Ganglioneuromatous neoplasms
BBc0000	[M]Ganglioneuroma
BBc0100	[M]Ganglioneuroblastoma
BBc0200	[M]Ganglioneuromatosis
BBc0z00	[M]Ganglioneuromatous neoplasm NOS
BBc1.00	[M]Neuroblastoma NOS
BBc4.00	[M]Neuroepithelioma NOS
BBc5.00	[M]Spongioneuroblastoma
BBc6.11	[M]Glioneuroma
BBc7.00	[M]Neurocytoma
BBc7.11	[M]Neuroastrocytoma
BBcA.00	[M]Olfactory neurogenic tumour
BBcB.00	[M]Aesthesioneurocytoma
BBcC.00	[M]Aesthesioneuroblastoma
BBcC.11	[M]Olfactory neuroblastoma
BBcD.00	[M]Aesthesioneuroepithelioma
BBcD.11	[M]Olfactory neuroepithelioma
BBcz.00	[M]Neuroepitheliomatous neoplasm NOS
BBe..11	[M]Neurofibromas
BBe0.00	[M]Neurofibroma NOS
BBe1.00	[M]Neurofibromatosis NOS
BBe1.11	[M]Multiple neurofibromatosis
BBe2.00	[M]Neurofibrosarcoma
BBe3.00	[M]Melanotic neurofibroma
BBe4.00	[M]Plexiform neurofibroma
BBe5.11	[M]Acoustic neuroma
BBE7.00	[M]Neuronaevus
BBe8.00	[M]Neuroma NOS
BBeA.00	[M]Neurothekeoma
BBN5.00	[M]Clear cell sarcoma of tendons and aponeuroses
BBz0.00	[M]Neuroendocrine neoplasm

Dementia

Read code	Description
1461	H/O: dementia
E00..11	Senile dementia
E00..12	Senile/presenile dementia
E000.00	Uncomplicated senile dementia
E001.00	Presenile dementia
E001000	Uncomplicated presenile dementia
E001100	Presenile dementia with delirium
E001200	Presenile dementia with paranoia
E001300	Presenile dementia with depression
E001z00	Presenile dementia NOS
E002.00	Senile dementia with depressive or paranoid features
E002000	Senile dementia with paranoia
E002100	Senile dementia with depression
E002z00	Senile dementia with depressive or paranoid features NOS
E003.00	Senile dementia with delirium
E004.00	Arteriosclerotic dementia
E004.11	Multi infarct dementia
E004000	Uncomplicated arteriosclerotic dementia
E004100	Arteriosclerotic dementia with delirium
E004200	Arteriosclerotic dementia with paranoia
E004300	Arteriosclerotic dementia with depression
E004z00	Arteriosclerotic dementia NOS
E012.00	Other alcoholic dementia
E012.11	Alcoholic dementia NOS
E02y100	Drug-induced dementia
E041.00	Dementia in conditions EC
Eu00.00	[X]Dementia in Alzheimer's disease
Eu00000	[X]Dementia in Alzheimer's disease with early onset
Eu00011	[X]Presenile dementia,Alzheimer's type
Eu00012	[X]Primary degen dementia, Alzheimer's type, presenile onset
Eu00100	[X]Dementia in Alzheimer's disease with late onset
Eu00112	[X]Senile dementia, Alzheimer's type

Read code	Description
Eu00113	[X]Primary degen dementia of Alzheimer's type, senile onset
Eu00200	[X]Dementia in Alzheimer's dis, atypical or mixed type
Eu00z00	[X]Dementia in Alzheimer's disease, unspecified
Eu00z11	[X]Alzheimer's dementia unspec
Eu01.00	[X]Vascular dementia
Eu01.11	[X]Arteriosclerotic dementia
Eu01000	[X]Vascular dementia of acute onset
Eu01100	[X]Multi-infarct dementia
Eu01111	[X]Predominantly cortical dementia
Eu01200	[X]Subcortical vascular dementia
Eu01300	[X]Mixed cortical and subcortical vascular dementia
Eu01y00	[X]Other vascular dementia
Eu01z00	[X]Vascular dementia, unspecified
Eu02.00	[X]Dementia in other diseases classified elsewhere
Eu02000	[X]Dementia in Pick's disease
Eu02100	[X]Dementia in Creutzfeldt-Jakob disease
Eu02200	[X]Dementia in Huntington's disease
Eu02300	[X]Dementia in Parkinson's disease
Eu02400	[X]Dementia in human immunodef virus [HIV] disease
Eu02500	[X]Lewy body dementia
Eu02y00	[X]Dementia in other specified diseases classif elsewhere
Eu02z00	[X] Unspecified dementia
Eu02z11	[X] Presenile dementia NOS
Eu02z13	[X] Primary degenerative dementia NOS
Eu02z14	[X] Senile dementia NOS
Eu02z16	[X] Senile dementia, depressed or paranoid type
Eu04000	[X]Delirium not superimposed on dementia, so described
Eu04100	[X]Delirium superimposed on dementia
Eu10711	[X]Alcoholic dementia NOS
Eu84311	[X]Dementia infantalis
ZS7C500	Language disorder of dementia

Progressive multifocal leukoencephalopathy

Read code	Description
A41..00	Slow viral central nervous system infection
A413.00	Progressive multifocal leucoencephalopathy
A413.11	Progressive multifocal leukoencephalopathy
A41y.00	Other slow virus central nervous system infections
A41z.00	Slow virus central nervous system infection NOS
F11x.00	Cerebral degeneration in other disease EC
F11x800	Cerebral degeneration due to multifocal leucoencephalopathy
F11xz00	Cerebral degeneration other disease NOS

Peripheral neuropathy

Read code	Description
F35z.00	Mononeuritis of unspecified site NOS
F35z000	Diabetic mononeuritis NOS
F360000	Dejerine-Sottas disease
F361.00	Peroneal muscular atrophy
F361000	Charcot-Marie-Tooth disease
F361011	Charcot's atrophy
F361012	Charcot-Marie-Tooth syndrome
F361z00	Peroneal muscular atrophy NOS
F363.00	Refsum's disease
F364.00	Idiopathic progressive polyneuropathy
F366.00	Polyneuropathy
F367.00	Peripheral neuropathy
F36yz00	Other idiopathic peripheral neuropathy NOS
F37..00	Inflammatory and toxic neuropathy
F37..11	Toxic neuropathy
F170.00	Idiopathic peripheral autonomic neuropathy
F170000	Carotid sinus syndrome

Read code	Description
F170100	Cervical sympathetic paralysis
F170z00	Idiopathic peripheral autonomic neuropathy NOS
F171.00	Peripheral autonomic neuropathy disease EC
F171000	Autonomic neuropathy due to amyloid
F171100	Autonomic neuropathy due to diabetes
F171z00	Peripheral autonomic neuropathy due to disease NOS
F17z.00	Autonomic nervous system disorder NOS
F17z.11	Horner's syndrome
F17z.12	Autonomic failure
F36y.00	Other idiopathic peripheral neuropathy
F36y000	Supranuclear paralysis
1JA..00	Suspected neurological disease
F32..00	Other cranial nerve disorders
F320.00	Olfactory nerve disorders
F321.00	Glossopharyngeal neuralgia
F322.00	Other glossopharyngeal nerve disorder

Guillain–Barré syndrome

Read code	Description
F370000	Guillain-Barre syndrome

Ophthalmology

Cytomegalovirus retinitis

Read code	Description
F030711	Cytomegaloviral encephalitis
AyuD000	[X]Other cytomegaloviral diseases
AyuD100	[X]Cytomegaloviral disease, unspecified

Dermatology

Herpes zoster

Read code	Description
A53..00	Herpes zoster
A530.00	Herpes zoster with meningitis
A531.00	Herpes zoster with other central nervous system complication
A531000	Herpes zoster with other CNS complications
A531100	Geniculate herpes zoster
A531400	Zoster encephalitis
A531z00	Herpes zoster with other CNS complication NOS
A532.00	Herpes zoster with ophthalmic complication
A532000	Herpes zoster with dermatitis of eyelid
A532100	Herpes zoster with keratoconjunctivitis
A532200	Herpes zoster iridocyclitis

Read code	Description
A532300	Ophthalmic herpes zoster infection
A532400	Herpes zoster ophthalmicus
A532z00	Herpes zoster with other ophthalmic complication
A53x.00	Herpes zoster with other specified complication
A53x000	Herpes zoster otitis externa
A53x100	Disseminated zoster
A53xz00	Herpes zoster with other specified complication NOS
A53y.00	Herpes zoster with unspecified complication
A53z.00	Herpes zoster NOS
F374400	Polyneuropathy in herpes zoster
F501611	Herpes zoster - otitis externa

Psoriasis

Read code	Description
M16..00	Psoriasis and similar disorders
M160.00	Psoriatic arthropathy
M160.11	Psoriatic arthritis
M160000	Psoriasis spondylitica
M160100	Distal interphalangeal psoriatic arthropathy
M160200	Arthritis mutilans
M160z00	Psoriatic arthropathy NOS
M161.00	Other psoriasis
M161000	Psoriasis unspecified
M161100	Psoriasis annularis
M161200	Psoriasis circinata
M161300	Psoriasis diffusa
M161400	Psoriasis discoidea
M161500	Psoriasis geographica
M161600	Guttate psoriasis
M161700	Psoriasis gyrate

Read code	Description
M161800	Psoriasis inveterata
M161900	Psoriasis ostracea
M161A00	Psoriasis palmaris
M161B00	Psoriasis plantaris
M161C00	Psoriasis punctata
M161D00	Pustular psoriasis
M161E00	Psoriasis universalis
M161F00	Psoriasis vulgaris
M161F11	Chronic large plaque psoriasis
M161G00	Acrodermatitis continua
M161H00	Erythrodermic psoriasis
M161z00	Psoriasis NOS
M16y.00	Other psoriasis and similar disorders
M16y000	Scalp psoriasis
M16z.00	Psoriasis and similar disorders NOS
Myu3000	[X]Other psoriasis

Seborrhoeic dermatitis

Read code	Description
M101.00	Seborrhoeic dermatitis
M101.11	Seborrhoeic dermatitis capitis
Myu2000	[X]Other seborrhoeic dermatitis

Kaposi’s sarcoma

Read code	Description
A789500	HIV disease resulting in Kaposi's sarcoma
A789511	HIV disease resulting in Kaposi sarcoma
B05z000	Kaposi's sarcoma of palate
B31z000	Kaposi's sarcoma of soft tissue
B33z000	Kaposi's sarcoma of skin
B592X00	Kaposi's sarcoma of multiple organs
B59zX00	Kaposi's sarcoma, unspecified
B6z0.00	Kaposi's sarcoma of lymph nodes
BBTA.00	[M]Kaposi's sarcoma
Byu5200	[X]Kaposi's sarcoma of multiple organs
Byu5300	[X]Kaposi's sarcoma, unspecified
Byu5B00	[X]Kaposi's sarcoma of other sites

Oral candidiasis

Read code	Description
AB20.00	Candidiasis of mouth and oesophagus
AB20000	Candidiasis of mouth
AB20011	Oral thrush

Rash

Read code	description
15J..00	Vulval rash
1D14.00	C/O: a rash
2227	O/E - rash present
2227.12	O/E - itchy rash
2FR..00	Butterfly rash
2FU..00	O/E - erythematous rash
2I14.00	O/E - a rash
M2y4211	Vesicular rash
R021.00	[D]Rash and other nonspecific skin eruption
R021100	[D]Rash on genitals
R021z00	[D]Rash and other nonspecific skin eruption NOS
SP35200	Serum rash

Gastroenterology

Persistent cryptosporidiosis

Read code	Description
A064.00	Cryptosporidiosis
4JH4000	Stool culture cryptosporidium positive

Abdominal pain

Read code	Description
196..00	Type of GIT pain
196..11	Abdominal pain type
196..12	Type of GIT pain - symptom
1962..0	Colicky abdominal pain
1963..0	Non-colicky abdominal pain
1964..0	Shoulder pain from abdomen
1965..0	Biliary colic
1965.11	Biliary colic symptom
1966..0	Infantile colic
1966.11	Infantile colic - symptom
1967..0	Abdominal migraine - symptom
1968..0	Abdominal discomfort
1969..0	Abdominal pain
1969000	Abdominal wall pain
196A.00	Hunger pain
196B.00	Painful rectal bleeding
196Z.00	Type of GIT pain NOS
197..00	Site of GIT pain
197..11	Flank pain
197..12	Iliac fossa pain
197..13	Site of abdominal pain
197..14	Subcostal pain
1971..0	Central abdominal pain
1972..0	Epigastric pain
1973..0	Left subcostal pain
1974..0	Right subcostal pain
1975..0	Left flank pain
1976..0	Right flank pain
1977..0	Right iliac fossa pain
1978..0	Left iliac fossa pain
1979..0	Suprapubic pain
197A.00	Generalised abdominal pain
197A.11	General abdominal pain-symptom
197B.00	Upper abdominal pain
197C.00	Lower abdominal pain

Read code	Description
197D.00	Right upper quadrant pain
197Z.00	Site of GIT pain NOS
R090.00	[D]Abdominal pain
R090000	[D]Abdominal tenderness
R090100	[D]Abdominal colic
R090200	[D]Colic NOS
R090300	[D]Infantile colic
R090311	[D]Evening colic
R090312	[D]Three month colic
R090400	[D]Abdominal cramps
R090500	[D]Epigastric pain
R090600	[D]Umbilical pain
R090700	[D]Hypochondrial pain
R090800	[D]Suprapubic pain
R090900	[D]Pain in right iliac fossa
R090A00	[D]Pain in left iliac fossa
R090B00	[D]Groin pain
R090C00	[D]Loin pain
R090D00	[D]Abdominal migraine
R090E00	[D]Recurrent acute abdominal pain
R090F00	[D]Acute abdomen
R090G00	[D]Pelvic and perineal pain
R090G11	[D] Pelvic pain
R090G12	[D] Perineal pain
R090H00	[D]Upper abdominal pain
R090J00	[D]Right upper quadrant pain
R090K00	[D]Left upper quadrant pain
R090L00	[D]Left lower quadrant pain
R090M00	[D]Right lower quadrant pain
R090N00	[D]Nonspecific abdominal pain
R090P00	[D]Functional abdominal pain syndrome
R090y00	[D]Other specified abdominal pain
R090z00	[D]Abdominal pain NOS
R093200	[D]Abdominal lump

Oral hairy leukoplakia

Read code	Description
J086.00	Leukoplakia of oral mucosa
J086.11	Leucoplakia of oral mucosa
J086000	Leukoplakia of gingiva
J086100	Leukoplakia of lips
J086200	Leukoplakia of tongue
J086300	Leukoplakia of buccal mucosa
J086311	Buccal mucosa leukoplakia
J086400	Hairy leukoplakia
J086z00	Oral mucosa leukoplakia NOS

Nausea/vomiting

Read code	Description
198..00	Nausea
198..11	C/O - nausea
198..12	Nausea symptoms
1982	Nausea present
1983	Morning nausea
1984	Upset stomach
1984.11	Upset tummy
198Z.00	Nausea NOS
199..00	Vomiting
199..11	C/O - vomiting
199..12	Emesis
199..13	Rumination
199..14	Vomiting symptoms
1992	Vomiting
1992.11	Throwing up
1992.12	Bilious attack

Read code	Description
1993	Projectile vomiting
1994	Vomiting blood - fresh
1995	Vomiting blood - coffee ground
1996	Vomiting - bile stained
1997	Retching
1998	Posseting
199Z.00	Vomiting NOS
R070000	[D]Nausea
R070100	[D]Vomiting
R070111	[D] Sickness
R070200	[D]Emesis
R070300	[D]Drug induced vomiting
R070400	[D]Projectile vomiting
R070z00	[D]Nausea and vomiting NOS
R070z11	[D]Posseting
R070z12	[D]Retching

Chronic diarrhoea of unknown cause

Read code	Description
1040299	Antimotility drug
19F..00	Diarrhoea symptoms
19F..11	Diarrhoea
19F..12	Loose stools
19F2.00	Diarrhoea
19F3.00	Spurious (overflow) diarrhoea
19F4.00	Toddlers diarrhoea
19FZ.00	Diarrhoea symptom NOS
19FZ.11	Diarrhoea symptom & vomiting, symptom
J4z..00	Non-infective gastroenteritis NOS
J4z..11	Presumed non-infectious diarrhoea

Read code	Description
J4z0.00	Non-infective gastritis NOS
J4z1.00	Non-infective jejunitis NOS
J4z2.00	Non-infective ileitis NOS
J4z3.00	Non-infective colitis NOS
J4z4.00	Non-infective sigmoiditis NOS
J4z5.00	Exacerbation of non-infective colitis
J4zz.00	Non-infective gastroenteritis NOS
J4zz.11	Diarrhoea presumed non-infectious
A083.00	Diarrhoea of presumed infectious origin
A083.11	Diarrhoea & vomiting = ? infectious

Weight loss of unknown cause

Read code	Description
1625..0	Abnormal weight loss
1625.11	Abnormal weight loss - symptom
1627..0	Unintentional weight loss
1D1A.00	Complaining of weight loss
22A8.00	Weight loss from baseline weight
22A9.00	Percentage weight loss
R032.00	[D]Abnormal loss of weight

Salmonella, shigella or campylobacter

Read code	Description
65Q3.00	Salmonella carrier
A02..00	Other salmonella infections
A020.00	Salmonella gastroenteritis
A020.11	Salmonellosis
A020.12	Salmonella food poisoning
A021.00	Salmonella septicaemia
A022.00	Localised salmonella infection
A022000	Local salmonella infection unspecified
A022100	Salmonella meningitis
A022200	Salmonella pneumonia
A022300	Salmonella arthritis
A022400	Salmonella osteomyelitis
A022z00	Other local salmonella infection
A023.00	Salmonella sepsis
A02y.00	Other specified salmonella infection
A02z.00	Salmonella infection NOS
A03..00	Shigellosis
A030.00	Shigella dysenteriae (group A)

Read code	Description
A030.11	Bacillary dysentery
A031.00	Shigella flexneri (group B)
A032.00	Shigella boydii (group C)
A033.00	Shigella sonnei (group D)
A033.11	Bacillary dysentery Shigella sonnei
A03y.00	Other specified shigella infection
A03z.00	Shigellosis NOS
A074300	Campylobacter gastrointestinal tract infection
A074311	Diarrhoea due to Campylobacter jejuni
A074312	Campylobacter enteritis
4J27000	Salmonella not isolated
4J28.00	Shigella species not isolated
4J29.00	Campylobacter species not isolated
Ayu0200	[X]Other specified salmonella infections
Ayu0300	[X]Salmonella infection, unspecified
Ayu0400	[X]Other shigellosis
Ayu0500	[X]Shigellosis, unspecified
F007300	Meningitis due to salmonella

Hepatitis B infection

Read code	Description
141E.00	History of hepatitis B
9kZ..00	Hepatitis B screening positive - enhanced services admin
9kZ..11	Hepatitis B screening positive
A702.00	Viral hepatitis B with coma
A702000	Acute hep B with delta-agent (coinfection) with hep coma
A703.00	Viral (serum) hepatitis B
A703000	Acute hep B with delta-agent (coinfectn) without hep coma
A704.00	Other specified viral hepatitis with coma
A705100	Acute delta-(super)infection of hepatitis B carrier
A707000	Chronic viral hepatitis B with delta-agent

Read code	Description
A707100	Chronic viral hepatitis B without delta-agent
A707300	Chronic viral hepatitis B
A707X00	Chronic viral hepatitis, unspecified
A708.00	Viral hepatitis with hepatic coma
A709.00	Viral hepatitis without hepatic coma
ZV02B00	[V]Hepatitis B carrier
141E.00	History of hepatitis B
4J3D.00	Hepatitis B viral load
7Q05200	Hepatitis B treatment drugs band 1

Hepatitis C infection

Read code	Description
2126700	Hepatitis C resolved
14i..00	H/O hepatitis C antiviral drug therapy
2J1..00	Hepatitis C status
2J11.00	Hepatitis C immune
2J12.00	Hepatitis C non immune
43dD.00	Hepatitis C recombinant immunoblot assay
4JQD.00	Hepatitis C viral ribonucleic acid PCR positive
4JQD.11	Hepatitis C PCR positive
4JQF.00	Hepatitis C antigen positive
9kV..00	Hepatitis C screening positive - enhanced services admin
9kV..11	Hepatitis C screening positive
A704000	Viral hepatitis C with coma

Read code	Description
A705000	Viral hepatitis C without mention of hepatic coma
A707200	Chronic viral hepatitis C
A70A.00	Hepatitis C genotype 1
A70B.00	Hepatitis C genotype 2
A70C.00	Hepatitis C genotype 3
A70D.00	Hepatitis C genotype 4
A70E.00	Hepatitis C genotype 5
A70F.00	Hepatitis C genotype 6
A70G.00	Acute hepatitis C
A70z.00	Unspecified viral hepatitis
A70z000	Hepatitis C

Oncology

Non-Hodgkin's lymphoma

Read code	Description
B627.00	Non - Hodgkin's lymphoma
B627000	Follicular non-Hodgkin's small cleaved cell lymphoma
B627100	Follicular non-Hodg mixed sml cleavd & lge cell lymphoma
B627200	Follicular non-Hodgkin's large cell lymphoma
B627300	Diffuse non-Hodgkin's small cell (diffuse) lymphoma
B627400	Diffuse non-Hodgkin's small cleaved cell (diffuse) lymphoma
B627500	Diffuse non-Hodgkin mixed sml & lge cell (diffuse) lymphoma
B627600	Diffuse non-Hodgkin's immunoblastic (diffuse) lymphoma
B627700	Diffuse non-Hodgkin's lymphoblastic (diffuse) lymphoma
B627800	Diffuse non-Hodgkin's lymphoma undifferentiated (diffuse)
B627A00	Diffuse non-Hodgkin's large cell lymphoma
B627B00	Other types of follicular non-Hodgkin's lymphoma
B627C00	Follicular non-Hodgkin's lymphoma

Read code	Description
B627D00	Diffuse non-Hodgkin's centroblastic lymphoma
B627W00	Unspecified B-cell non-Hodgkin's lymphoma
B627X00	Diffuse non-Hodgkin's lymphoma, unspecified
BBg2.00	[M]Malignant lymphoma, non Hodgkin's type
BBg2.11	[M]Non Hodgkins lymphoma
ByuD100	[X]Other types of follicular non-Hodgkin's lymphoma
ByuD200	[X]Other types of diffuse non-Hodgkin's lymphoma
ByuD300	[X]Other specified types of non-Hodgkin's lymphoma
ByuDC00	[X]Diffuse non-Hodgkin's lymphoma, unspecified
ByuDD00	[X]Oth and unspecif peripheral & cutaneous T-cell lymphomas
ByuDE00	[X]Unspecified B-cell non-Hodgkin's lymphoma
ByuDF00	[X]Non-Hodgkin's lymphoma, unspecified type
ByuDF11	[X]Non-Hodgkin's lymphoma NOS

Anal cancer or anal intraepithelial dysplasia

Read code	Description
B142.00	Malignant neoplasm of anal canal
B142.11	Anal carcinoma
B142000	Malignant neoplasm of cloacogenic zone
B143.00	Malignant neoplasm of anus unspecified
B14y.00	Malig neop other site rectum, rectosigmoid junction and anus
B14z.00	Malignant neoplasm rectum,rectosigmoid junction and anus NOS
B805000	Anal intraepithelial neoplasia grade III
J574K00	Anal intraepithelial neoplasia grade I
J574L00	Anal intraepithelial neoplasia grade II

Castleman's disease

Read code	Description
R056200	[D]Castleman's disease

Primary cerebral lymphoma

Read code	Description
B51..00	Malignant neoplasm of brain
B51..11	Cerebral tumour - malignant
B510.00	Malignant neoplasm cerebrum (excluding lobes and ventricles)
B510000	Malignant neoplasm of basal ganglia
B510100	Malignant neoplasm of cerebral cortex
B510200	Malignant neoplasm of corpus striatum
B510300	Malignant neoplasm of globus pallidus
B510400	Malignant neoplasm of hypothalamus
B510500	Malignant neoplasm of thalamus
B510z00	Malignant neoplasm of cerebrum NOS
B511.00	Malignant neoplasm of frontal lobe
B512.00	Malignant neoplasm of temporal lobe
B512000	Malignant neoplasm of hippocampus
B512100	Malignant neoplasm of uncus
B512z00	Malignant neoplasm of temporal lobe NOS
B513.00	Malignant neoplasm of parietal lobe
B514.00	Malignant neoplasm of occipital lobe

Read code	Description
B515.00	Malignant neoplasm of cerebral ventricles
B515000	Malignant neoplasm of choroid plexus
B515100	Malignant neoplasm of floor of cerebral ventricle
B515z00	Malignant neoplasm of cerebral ventricle NOS
B516.00	Malignant neoplasm of cerebellum
B517.00	Malignant neoplasm of brain stem
B517000	Malignant neoplasm of cerebral peduncle
B517100	Malignant neoplasm of medulla oblongata
B517200	Malignant neoplasm of midbrain
B517300	Malignant neoplasm of pons
B517z00	Malignant neoplasm of brain stem NOS
B51y.00	Malignant neoplasm of other parts of brain
B51y000	Malignant neoplasm of corpus callosum
B51y100	Malignant neoplasm of tapetum
B51y200	Malignant neoplasm, overlapping lesion of brain
B51yz00	Malignant neoplasm of other part of brain NOS
B51z.00	Malignant neoplasm of brain NOS

Cervical dysplasia

Read code	Description
K551.00	Dysplasia of cervix uteri
K551.11	Atypism - cervical
K551.12	CIN I - II, cervical dysplasia
K551100	Epidermidization of cervix
K551300	Mild cervical dysplasia
K551400	Moderate cervical dysplasia
K551411	Cervical intraepithelial neoplasia grade II
K551X00	Severe cervical dysplasia, not elsewhere classified
K551z00	Dysplasia of cervix NOS
Kyu9K00	[X]Severe cervical dysplasia, not elsewhere classified

Lung cancer

Read code	Description
B22..00	Malignant neoplasm of trachea, bronchus and lung
B221100	Malignant neoplasm of hilus of lung
B222.00	Malignant neoplasm of upper lobe, bronchus or lung
B222100	Malignant neoplasm of upper lobe of lung
B222z00	Malignant neoplasm of upper lobe, bronchus or lung NOS
B223.00	Malignant neoplasm of middle lobe, bronchus or lung
B223100	Malignant neoplasm of middle lobe of lung
B223z00	Malignant neoplasm of middle lobe, bronchus or lung NOS
B224.00	Malignant neoplasm of lower lobe, bronchus or lung
B224100	Malignant neoplasm of lower lobe of lung

Read code	Description
B224z00	Malignant neoplasm of lower lobe, bronchus or lung NOS
B225.00	Malignant neoplasm of overlapping lesion of bronchus & lung
B22y.00	Malignant neoplasm of other sites of bronchus or lung
B22z.00	Malignant neoplasm of bronchus or lung NOS
B22z.11	Lung cancer
B570.00	Secondary malignant neoplasm of lung
B907.00	Neoplasm of uncertain behaviour trachea, bronchus and lung
B907200	Neoplasm of uncertain behaviour of lung
B907z00	Neop of uncertain behaviour of trachea, bronchus or lung NOS

Hodgkin lymphoma

Read code	Description
B61..11	Hodgkin lymphoma
B610.00	Hodgkin's paragranuloma
B610000	Hodgkin's paragranuloma of unspecified site
B610100	Hodgkin's paragranuloma of lymph nodes of head, face, neck
B610200	Hodgkin's paragranuloma of intrathoracic lymph nodes
B610300	Hodgkin's paragranuloma of intra-abdominal lymph nodes
B610400	Hodgkin's paragranuloma of lymph nodes of axilla and arm
B610500	Hodgkin's paragranuloma lymph nodes inguinal region and leg
B610600	Hodgkin's paragranuloma of intrapelvic lymph nodes
B610700	Hodgkin's paragranuloma of spleen
B610800	Hodgkin's paragranuloma of lymph nodes of multiple sites
B610z00	Hodgkin's paragranuloma NOS
B611.00	Hodgkin's granuloma
B611000	Hodgkin's granuloma of unspecified site
B611100	Hodgkin's granuloma of lymph nodes of head, face and neck
B611200	Hodgkin's granuloma of intrathoracic lymph nodes
B611300	Hodgkin's granuloma of intra-abdominal lymph nodes
B611400	Hodgkin's granuloma of lymph nodes of axilla and upper limb
B611500	Hodgkin's granuloma lymph nodes of inguinal region and leg
B611600	Hodgkin's granuloma of intrapelvic lymph nodes
B611700	Hodgkin's granuloma of spleen
B611800	Hodgkin's granuloma of lymph nodes of multiple sites
B611z00	Hodgkin's granuloma NOS
B612000	Hodgkin's sarcoma of unspecified site
B612100	Hodgkin's sarcoma of lymph nodes of head, face and neck
B612200	Hodgkin's sarcoma of intrathoracic lymph nodes
B612300	Hodgkin's sarcoma of intra-abdominal lymph nodes
B612400	Hodgkin's sarcoma of lymph nodes of axilla and upper limb
B612500	Hodgkin's sarcoma of lymph nodes of inguinal region and leg
B612600	Hodgkin's sarcoma of intrapelvic lymph nodes
B612700	Hodgkin's sarcoma of spleen
B612800	Hodgkin's sarcoma of lymph nodes of multiple sites
B612z00	Hodgkin's sarcoma NOS
B613.00	Hodgkin's disease, lymphocytic-histiocytic predominance
B613.11	Hodgkin lymphma, lymphcyte-rch
B613000	Hodgkin's, lymphocytic-histiocytic predominance unspec site
B613100	Hodgkin's, lymphocytic-histiocytic pred of head, face, neck
B613200	Hodgkin's, lymphocytic-histiocytic pred intrathoracic nodes
B613300	Hodgkin's, lymphocytic-histiocytic pred intra-abdominal node
B613400	Hodgkin's, lymphocytic-histiocytic pred axilla and arm
B613500	Hodgkin's, lymphocytic-histiocytic pred inguinal and leg
B613600	Hodgkin's, lymphocytic-histiocytic pred intrapelvic nodes
B613700	Hodgkin's, lymphocytic-histiocytic predominance of spleen
B613800	Hodgkin's, lymphocytic-histiocytic pred of multiple sites
B613z00	Hodgkin's, lymphocytic-histiocytic predominance NOS
B614.00	Hodgkin's disease, nodular sclerosis
B614000	Hodgkin's disease, nodular sclerosis of unspecified site
B614100	Hodgkin's nodular sclerosis of head, face and neck
B614200	Hodgkin's nodular sclerosis of intrathoracic lymph nodes
B614300	Hodgkin's nodular sclerosis of intra-abdominal lymph nodes
B614400	Hodgkin's nodular sclerosis of lymph nodes of axilla and arm
B614500	Hodgkin's nodular sclerosis of inguinal region and leg
B614600	Hodgkin's nodular sclerosis of intrapelvic lymph nodes
B614700	Hodgkin's disease, nodular sclerosis of spleen
B614800	Hodgkin's nodular sclerosis of lymph nodes of multiple sites
B614z00	Hodgkin's disease, nodular sclerosis NOS
B615.00	Hodgkin's disease, mixed cellularity
B615000	Hodgkin's disease, mixed cellularity of unspecified site

Read code	Description
B615100	Hodgkin's mixed cellularity of lymph nodes head, face, neck
B615200	Hodgkin's mixed cellularity of intrathoracic lymph nodes
B615300	Hodgkin's mixed cellularity of intra-abdominal lymph nodes
B615400	Hodgkin's mixed cellularity of lymph nodes of axilla and arm
B615500	Hodgkin's mixed cellularity of lymph nodes inguinal and leg
B615600	Hodgkin's mixed cellularity of intrapelvic lymph nodes
B615700	Hodgkin's disease, mixed cellularity of spleen
B615800	Hodgkin's mixed cellularity of lymph nodes of multiple sites
B615z00	Hodgkin's disease, mixed cellularity NOS
B616.00	Hodgkin's disease, lymphocytic depletion
B616000	Hodgkin's lymphocytic depletion of unspecified site
B616100	Hodgkin's lymphocytic depletion of head, face and neck
B616200	Hodgkin's lymphocytic depletion of intrathoracic lymph nodes
B616300	Hodgkin's lymphocytic depletion intra-abdominal lymph nodes
B616400	Hodgkin's lymphocytic depletion lymph nodes axilla and arm
B616500	Hodgkin's lymphocytic depletion lymph nodes inguinal and leg
B616600	Hodgkin's lymphocytic depletion of intrapelvic lymph nodes
B616700	Hodgkin's disease, lymphocytic depletion of spleen
B616800	Hodgkin's lymphocytic depletion lymph nodes multiple sites
B616z00	Hodgkin's disease, lymphocytic depletion NOS
B617.00	Nodular lymphocyte predominant Hodgkin lymphoma
B618.00	Nodular sclerosis classical Hodgkin lymphoma
B619.00	Mixed cellularity classical Hodgkin lymphoma
B61A.00	Lymphocyte depleted classical Hodgkin lymphoma
B61B.00	Lymphocyte-rich classical Hodgkin lymphoma
B61C.00	Other classical Hodgkin lymphoma
B61z.00	Hodgkin's disease NOS
B61z.11	Hodgkin lymphoma NOS
B61z000	Hodgkin's disease NOS, unspecified site
B61z100	Hodgkin's disease NOS of lymph nodes of head, face and neck
B61z200	Hodgkin's disease NOS of intrathoracic lymph nodes
B61z300	Hodgkin's disease NOS of intra-abdominal lymph nodes
B61z400	Hodgkin's disease NOS of lymph nodes of axilla and arm
B61z500	Hodgkin's disease NOS of lymph nodes inguinal region and leg
B61z600	Hodgkin's disease NOS of intrapelvic lymph nodes
B61z700	Hodgkin's disease NOS of spleen
B61z800	Hodgkin's disease NOS of lymph nodes of multiple sites
B61zz00	Hodgkin's disease NOS
BBj..00	[M]Hodgkin's disease
BBj0.00	[M]Hodgkin's disease NOS
BBj1.00	[M]Hodgkin's disease, lymphocytic predominance
BBj1000	[M]Hodgkin,s disease, lymphocytic predominance, diffuse
BBj1100	[M]Hodgkin,s disease, lymphocytic predominance, nodular
BBj2.00	[M]Hodgkin's disease, mixed cellularity
BBj3.00	[M]Hodgkin's disease, lymphocytic depletion NOS
BBj4.00	[M]Hodgkin's disease,lymphocytic depletion,diffuse fibrosis
BBj5.00	[M]Hodgkin's disease, lymphocytic depletion, reticular type
BBj6.00	[M]Hodgkin's disease, nodular sclerosis NOS
BBj6000	[M]Hodgkin,s disease, nodular sclerosis, lymphocytic predom
BBj6100	[M]Hodgkin,s disease, nodular sclerosis, mixed cellularity
BBj6200	[M]Hodgkin,s disease, nodular sclerosis, lymphocytic deplet
BBj7.00	[M]Hodgkin's disease, nodular sclerosis, cellular phase
BBj8.00	[M]Hodgkin's paragranuloma
BBj9.00	[M]Hodgkin's granuloma
BBjA.00	[M]Hodgkin's sarcoma
BBjz.00	[M]Hodgkin's disease NOS
ByuD000	[X]Other Hodgkin's disease
ZV10711	[V]Personal history of Hodgkin's disease

Head and neck cancer

Read code	Description
B310000	Malignant neoplasm of soft tissue of head
B310100	Malignant neoplasm of soft tissue of face
B310200	Malignant neoplasm of soft tissue of neck
B310300	Malignant neoplasm of cartilage of ear
B310400	Malignant neoplasm of tarsus of eyelid
B330.00	Malignant neoplasm of skin of lip
B331.00	Malignant neoplasm of eyelid including canthus
B331000	Malignant neoplasm of canthus
B331100	Malignant neoplasm of upper eyelid
B331200	Malignant neoplasm of lower eyelid
B332.00	Malignant neoplasm skin of ear and external auricular canal
B332000	Malignant neoplasm of skin of auricle (ear)
B332100	Malignant neoplasm of skin of external auditory meatus
B332200	Malignant neoplasm of pinna NEC
B333.00	Malignant neoplasm skin of other and unspecified parts face
B333000	Malignant neoplasm of skin of cheek, external
B333100	Malignant neoplasm of skin of chin
B333200	Malignant neoplasm of skin of eyebrow
B333300	Malignant neoplasm of skin of forehead
B333400	Malignant neoplasm of skin of nose (external)
B333500	Malignant neoplasm of skin of temple
B333z00	Malignant neoplasm skin other and unspec part of face NOS
B334.00	Malignant neoplasm of scalp and skin of neck
B334000	Malignant neoplasm of scalp
B334100	Malignant neoplasm of skin of neck
B334z00	Malignant neoplasm of scalp or skin of neck NOS
B50..00	Malignant neoplasm of eye
B500000	Malignant neoplasm of ciliary body
B500100	Malignant neoplasm of iris
B500200	Malignant neoplasm of crystalline lens
B500300	Malignant neoplasm of sclera
B500z00	Malignant neoplasm of eyeball NOS
B501.00	Malignant neoplasm of orbit
B501000	Malignant neoplasm of connective tissue of orbit
B501100	Malignant neoplasm of extraocular muscle of orbit
B501z00	Malignant neoplasm of orbit NOS
B502.00	Malignant neoplasm of lacrimal gland
B503.00	Malignant neoplasm of conjunctiva
B504.00	Malignant neoplasm of cornea
B505.00	Malignant neoplasm of retina
B506.00	Malignant neoplasm of choroid
B507.00	Malignant neoplasm of lacrimal duct
B507000	Malignant neoplasm of lacrimal sac
B507100	Malignant neoplasm of nasolacrimal duct
B507z00	Malignant neoplasm of lacrimal duct NOS
B508.00	Malignant neoplasm, overlapping lesion of eye and adnexa
B50y.00	Malignant neoplasm of other specified site of eye
B50z.00	Malignant neoplasm of eye NOS
B51..00	Malignant neoplasm of brain
B510.00	Malignant neoplasm cerebrum (excluding lobes and ventricles)
B510000	Malignant neoplasm of basal ganglia
B510100	Malignant neoplasm of cerebral cortex
B510200	Malignant neoplasm of corpus striatum
B510300	Malignant neoplasm of globus pallidus
B510400	Malignant neoplasm of hypothalamus
B510500	Malignant neoplasm of thalamus
B510z00	Malignant neoplasm of cerebrum NOS
B511.00	Malignant neoplasm of frontal lobe
B512.00	Malignant neoplasm of temporal lobe
B512000	Malignant neoplasm of hippocampus
B512100	Malignant neoplasm of uncus
B512z00	Malignant neoplasm of temporal lobe NOS
B513.00	Malignant neoplasm of parietal lobe

Read code	Description
B514.00	Malignant neoplasm of occipital lobe
B515.00	Malignant neoplasm of cerebral ventricles
B515000	Malignant neoplasm of choroid plexus
B515100	Malignant neoplasm of floor of cerebral ventricle
B515z00	Malignant neoplasm of cerebral ventricle NOS
B516.00	Malignant neoplasm of cerebellum
B517.00	Malignant neoplasm of brain stem
B517000	Malignant neoplasm of cerebral peduncle
B517100	Malignant neoplasm of medulla oblongata
B517200	Malignant neoplasm of midbrain
B517300	Malignant neoplasm of pons
B517z00	Malignant neoplasm of brain stem NOS
B51y.00	Malignant neoplasm of other parts of brain
B51y000	Malignant neoplasm of corpus callosum
B51y100	Malignant neoplasm of tapetum
B51y200	Malignant neoplasm, overlapping lesion of brain
B51yz00	Malignant neoplasm of other part of brain NOS
B51z.00	Malignant neoplasm of brain NOS
B520.00	Malignant neoplasm of cranial nerves
B520000	Malignant neoplasm of olfactory bulb
B520100	Malignant neoplasm of optic nerve
B520200	Malignant neoplasm of acoustic nerve
B520z00	Malignant neoplasm of cranial nerves NOS
B521.00	Malignant neoplasm of cerebral meninges
B521000	Malignant neoplasm of cerebral dura mater
B521100	Malignant neoplasm of cerebral arachnoid mater
B521200	Malignant neoplasm of cerebral pia mater
B521z00	Malignant neoplasm of cerebral meninges NOS
B525.00	Malignant neoplasm of cauda equina
B52X.00	Malignant neoplasm of meninges, unspecified
B52y.00	Malignant neoplasm of other specified part of nervous system
B52z.00	Malignant neoplasm of nervous system NOS
B53..00	Malignant neoplasm of thyroid gland
B540.00	Malignant neoplasm of adrenal gland
B540000	Malignant neoplasm of adrenal cortex
B540100	Malignant neoplasm of adrenal medulla
B540z00	Malignant neoplasm of adrenal gland NOS
B541.00	Malignant neoplasm of parathyroid gland
B542.00	Malignant neoplasm pituitary gland and craniopharyngeal duct
B542000	Malignant neoplasm of pituitary gland
B542100	Malignant neoplasm of craniopharyngeal duct
B543.00	Malignant neoplasm of pineal gland
B544.00	Malignant neoplasm of carotid body
B545.00	Malignant neoplasm of aortic body and other paraganglia
B545000	Malignant neoplasm of glomus jugulare
B545100	Malignant neoplasm of aortic body
B545200	Malignant neoplasm of coccygeal body
B545z00	Malignant neoplasm of aortic body or paraganglia NOS
B54X.00	Malignant neoplasm-pluriglandular involvement,unspecified
B54y.00	Malignant neoplasm of other specified endocrine gland
B55..00	Malignant neoplasm of other and ill-defined sites
B550.00	Malignant neoplasm of head, neck and face
B550000	Malignant neoplasm of head NOS
B550100	Malignant neoplasm of cheek NOS
B550200	Malignant neoplasm of nose NOS
B550300	Malignant neoplasm of jaw NOS
B550400	Malignant neoplasm of neck NOS
B550500	Malignant neoplasm of supraclavicular fossa NOS
B550z00	Malignant neoplasm of head, neck and face NOS
B0...11	Carcinoma of lip, oral cavity and pharynx
B00..11	Carcinoma of lip

Space-occupying lesion of unknown cause

Read code	Description
R042300	[D]Space-occupying intracranial lesion NOS

Gynaecology

Cervical cancer

Read code	Description
B41..00	Malignant neoplasm of cervix uteri
B41..11	Cervical carcinoma (uterus)
B410.00	Malignant neoplasm of endocervix
B410000	Malignant neoplasm of endocervical canal
B410100	Malignant neoplasm of endocervical gland
B410z00	Malignant neoplasm of endocervix NOS
B411.00	Malignant neoplasm of exocervix

Read code	Description
B412.00	Malignant neoplasm, overlapping lesion of cervix uteri
B41y.00	Malignant neoplasm of other site of cervix
B41y000	Malignant neoplasm of cervical stump
B41y100	Malignant neoplasm of squamocolumnar junction of cervix
B41yz00	Malignant neoplasm of other site of cervix NOS
B41z.00	Malignant neoplasm of cervix uteri NOS

Cervical intraepithelial neoplasia

Read code	Description
B831.13	Cervical intraepithelial neoplasia grade III
K551411	Cervical intraepithelial neoplasia grade II
ZV13C11	[V]PH of cervical intraepithelial neoplasia grade II
ZV13E00	[V]PH of cervical intraepithelial neoplasia, grade III

Vaginal intraepithelial neoplasia

Read code	Description
B833311	Vulval intraepithelial neoplasia
B833700	Vaginal intraepithelial neoplasia grade 1
B833800	Vaginal intraepithelial neoplasia grade 2
B833900	Vaginal intraepithelial neoplasia grade 3

Ear, nose, and throat

Parotitis

Read code	Description
A72y.11	Parotitis - epidemic
A72z.00	Mumps with no complication
A72z.11	Epidemic parotitis
J072.00	Sialoadenitis
J072.11	Parotitis
J072000	Allergic parotitis
J072100	Toxic parotitis
J072200	Parotitis NOS
J072300	Sialoadenitis of the submandibular gland

Read code	Description
J072400	Sialoadenitis of the sublingual gland
J072500	Sialoangiitis
J072600	Sialodochitis
J072700	Infective sialoadenitis
J072z00	Sialoadenitis NOS
A72..00	Mumps
A72x.00	Mumps with other specified complications
A72xz00	Mumps with other specified complications NOS
A72y.00	Mumps with unspecified complication

Lymphoepithelial parotid cysts

Read code	Description
J076400	Parotid cyst
J076z00	Salivary gland mucocele NOS
J077.00	Salivary secretion disturbance
J077000	Salivary hyposecretion
J07y.00	Other salivary gland diseases
J07y.11	Sialectasia
J07y.12	Stenosis of salivary duct
J07y000	Benign lymphoepithelial salivary gland lesion
J07y100	Parotid sialectasia
J07y200	Submandibular sialectasia
J07y300	Sublingual sialectasia
J07y400	Sialectasia NOS

Haematology

Blood dyscrasia

Read code	Description
D410.00	Secondary polycythaemia
D410000	Stress polycythaemia
D410011	Spurious polycythaemia
D410100	High altitude polycythaemia
D410200	Polycythaemia due to cyanotic heart disease
D410300	Polycythaemia due to cyanotic respiratory disease
D410400	Renal polycythaemia
D410z00	Secondary polycythaemia NOS
42P2.00	Thrombocytopenia
42P2.11	Auto-immune thrombocytopenia
D313.00	Primary thrombocytopenia
D313.11	Evan's syndrome
D313.12	Idiopathic thrombocytopenic purpura
D313.13	Idiopathic purpura
D313.14	Megakaryocytic hypoplasia
D313.15	Thrombocytopenic purpura
D313000	Idiopathic thrombocytopenic purpura
D313011	Idiopathic purpura
D313012	ITP - idiopathic thrombocytopenic purpura
D313100	Congenital thrombocytopenic purpura
D313111	Hereditary thrombocytopenia NEC
D313200	Thrombocytopenic purpura with absent radius
D313211	TAR syndrome
D313300	[X]Essential thrombocytopenia NOS
D313y00	Other specified primary thrombocytopenia
D313z00	Primary thrombocytopenia NOS
D313z11	Essential thrombocytopenia NOS
D314.00	Secondary thrombocytopenia
D314000	Post-transfusion purpura
D314011	Thrombocytopenia due to massive blood transfusion
D314100	Thrombocytopenia due to drugs
D314200	Thrombocytopenia due to extracorporeal circulation of blood
D314y00	Other specified secondary thrombocytopenia
D314z00	Secondary thrombocytopenia NOS

Read code	Description
D314z11	Dilutional thrombocytopenia
D315.00	Thrombocytopenia NOS
Dyu3200	[X]Other primary thrombocytopenia
D41y300	Bone marrow depression
42H2.00	Leucopenia - low white count
42H2.11	Leucopenia
42J2.00	Neutropenia
42J3.00	Neutropenia
D400.00	Agranulocytosis
D400.11	Kostmann's syndrome
D400.12	Neutropenia
D400000	Idiopathic agranulocytosis
D400011	Idiopathic neutropenia
D400100	Primary splenic neutropenia
D400200	Agranulocytosis - drug induced
D400211	Neutropenia - drug induced
D400300	Agranulocytosis due to irradiation
D400311	Radiation agranulocytosis
D400312	Neutropenia due to irradiation
D400400	Agranulocytosis due to infection
D400411	Neutropenia due to infection
D400500	Congenital neutropenia
D400511	Congenital agranulocytosis NEC
D400600	Drug-induced neutropenia
D400700	Acquired neutropenia in newborn
D400800	Acquired neutropenia NEC
D400811	Acquired agranulocytosis NEC
D400900	Cyclical neutropenia
D400A00	Leucopenia
D400A00	Leucopenia
D400y00	Other specified agranulocytosis
D400z00	Agranulocytosis NOS
D40y300	Lymphopenia

Other

Mononucleosis-like syndrome

Read code	Description
AyuD.00	[X]Other viral diseases
AyuD000	[X]Other cytomegaloviral diseases
AyuD100	[X]Cytomegaloviral disease, unspecified
AyuD200	[X]Mumps with other complications
AyuD300	[X]Mumps without complication
AyuD400	[X]Other infectious mononucleosis
AyuD500	[X]Infectious mononucleosis, unspecified
AyuD600	[X]Other viral conjunctivitis
AyuD700	[X]Viral conjunctivitis, unspecified
AyuD800	[X]Retrovirus infections, not elsewhere classified

Read code	Description
AyuD900	[X]Other specified viral diseases
AyuDA00	[X]Adenovirus infection, unspecified
AyuDB00	[X]Enterovirus infection, unspecified
AyuDC00	[X]Coronavirus infection, unspecified
AyuDD00	[X]Parvovirus infection, unspecified
AyuDE00	[X]Papovavirus infection, unspecified
AyuDF00	[X]Other viral infections of unspecified site
AyuDG00	[X]Viral infection, unspecified
J631200	Hepatitis in infectious mononucleosis

Hypertension

Read code	Description
G2...00	Hypertensive disease
G2...11	BP - hypertensive disease
G20..00	Essential hypertension
G20..11	High blood pressure
G20..12	Primary hypertension
G200.00	Malignant essential hypertension
G201.00	Benign essential hypertension
G202.00	Systolic hypertension
G203.00	Diastolic hypertension
G20z.00	Essential hypertension NOS
G20z.11	Hypertension NOS

Minor trauma

Read code	Description
ZRLs.00	Hospital trauma index
ZRqU.00	Trauma index
ZRqV.00	Trauma score
ZRqV.11	TS - Trauma score
ZRqV100	Revised trauma score
ZRqV111	RTS - Revised trauma score
ZRqW.00	Trauma and injury severity score
ZRqW.11	TRISS - Trauma and injury severity score
ZRrX.00	Westmeade post-traumatic amnesia test
ZS7C700	Post-traumatic mutism
ZV15400	[V]Personal history of psychological trauma
ZVu6U00	[X]Personal history of other physical trauma

Depression

Read code	Description
1B17.00	Depressed
1B17.11	C/O - feeling depressed
1B1U.00	Symptoms of depression
1B1U.11	Depressive symptoms
1BT..00	Depressed mood
2257000	O/E - depressed
8BK0.00	Depression management programme
8CAa.00	Patient given advice about management of depression
8HHq.00	Referral for guided self-help for depression
9k4..00	Depression - enhanced services administration
9k40.00	Depression - enhanced service completed
9kQ..00	On full dose long term treatment depression - enh serv admin
9kQ..11	On full dose long term treatment for depression
E001300	Presenile dementia with depression
E002100	Senile dementia with depression
E004300	Arteriosclerotic dementia with depression
E02y300	Drug-induced depressive state
E11..00	Affective psychoses
E11..12	Depressive psychoses
E112.00	Single major depressive episode
E112.11	Agitated depression
E112.12	Endogenous depression first episode
E112.13	Endogenous depression first episode
E112.14	Endogenous depression
E112000	Single major depressive episode, unspecified
E112100	Single major depressive episode, mild
E112200	Single major depressive episode, moderate
E112300	Single major depressive episode, severe, without psychosis
E112400	Single major depressive episode, severe, with psychosis
E112z00	Single major depressive episode NOS
E118.00	Seasonal affective disorder
E11y200	Atypical depressive disorder
E11z.00	Other and unspecified affective psychoses
E11z000	Unspecified affective psychoses NOS
E11z200	Masked depression
E11zz00	Other affective psychosis NOS
E130.00	Reactive depressive psychosis
E130.11	Psychotic reactive depression
E135.00	Agitated depression
E200300	Anxiety with depression
E204.00	Neurotic depression reactive type
E291.00	Prolonged depressive reaction
E2B..00	Depressive disorder NEC
E2B0.00	Postviral depression

Read code	Description
E2B1.00	Chronic depression
Eu02z16	[X] Senile dementia, depressed or paranoid type
Eu32.00	[X]Depressive episode
Eu32.11	[X]Single episode of depressive reaction
Eu32.12	[X]Single episode of psychogenic depression
Eu32.13	[X]Single episode of reactive depression
Eu32000	[X]Mild depressive episode
Eu32100	[X]Moderate depressive episode
Eu32200	[X]Severe depressive episode without psychotic symptoms
Eu32211	[X]Single episode agitated depressn w'out psychotic symptoms
Eu32212	[X]Single episode major depression w'out psychotic symptoms
Eu32213	[X]Single episode vital depression w'out psychotic symptoms
Eu32300	[X]Severe depressive episode with psychotic symptoms
Eu32311	[X]Single episode of major depression and psychotic symptoms
Eu32312	[X]Single episode of psychogenic depressive psychosis
Eu32313	[X]Single episode of psychotic depression
Eu32314	[X]Single episode of reactive depressive psychosis
Eu32400	[X]Mild depression
Eu32500	[X]Major depression, mild
Eu32600	[X]Major depression, moderately severe
Eu32700	[X]Major depression, severe without psychotic symptoms
Eu32800	[X]Major depression, severe with psychotic symptoms
Eu32y00	[X]Other depressive episodes
Eu32y11	[X]Atypical depression
Eu32y12	[X]Single episode of masked depression NOS
Eu32z00	[X]Depressive episode, unspecified
Eu32z11	[X]Depression NOS
Eu32z12	[X]Depressive disorder NOS
Eu32z13	[X]Prolonged single episode of reactive depression
Eu32z14	[X] Reactive depression NOS
Eu33.14	[X]Seasonal depressive disorder
Eu33.15	[X]SAD - Seasonal affective disorder
Eu33211	[X]Endogenous depression without psychotic symptoms
Eu33311	[X]Endogenous depression with psychotic symptoms
Eu33z11	[X]Monopolar depression NOS
Eu34100	[X]Dysthymia
Eu34111	[X]Depressive neurosis
Eu34112	[X]Depressive personality disorder
Eu34113	[X]Neurotic depression
Eu34114	[X]Persistant anxiety depression
Eu41200	[X]Mixed anxiety and depressive disorder
Eu41211	[X]Mild anxiety depression
ZV11100	[V]Personal history of affective disorder

Lymphogranuloma venereum

Read code	Description
A991.00	Lymphogranuloma venereum
A991000	Pharyngeal lymphogranuloma venereum
A991100	Rectal lymphogranuloma venereum

Pyrexia of unknown origin

Read code	Description
165..00	Temperature symptoms
165..11	Fever symptoms
165..12	Pyrexia symptoms
1652..0	Feels hot/feverish
1653..0	Fever with sweating
1654..0	Having rigors
1654.11	Rigor - symptom
1655..0	C/O shivering
1656..0	Feverish cold
1657..0	Hot flushes
1658..0	Different temperature in opposite limbs
165Z.00	Temperature symptom NOS
2E...00	Examination of fever
2E...11	O/E - fever
2E1..00	O/E - fever - general
2E12.00	O/E - fever examination - NAD
2E13.00	O/E -pyrexia of unknown origin
2E13.11	O/E - pyrexia - ? cause
2E1Z.00	O/E - fever - general NOS
2E34.00	O/E - temperature elevated
2E35.00	O/E - hyperpyrexia-> 40.5 oCEL
2E3Z.00	O/E - level of fever NOS

Read code	Description
2E4..00	O/E - character of fever
2E4..11	O/E - temperature character
2E41.00	O/E - fever - acute rise
2E42.00	O/E - fever - gradual rise
2E43.00	O/E - fever - continuous
2E44.00	O/E - fever - remittent
2E45.00	O/E - fever - intermittent
2E46.00	O/E - staircase fever
2E47.00	O/E - fever - irregular
2E48.00	O/E - fever - fast fall-crisis
2E49.00	O/E - fever-gradual fall-lysis
2E4Z.00	O/E - fever character NOS
R003000	[D]Convulsions, febrile
R003011	[D]Pyrexial convulsion
R006.00	[D]Pyrexia of unknown origin
R006.11	[D]Fever of unknown origin
R006000	[D]Chills with fever
R006100	[D]Hyperpyrexia NOS
R006200	[D]Fever NOS
R006300	[D]Persistent fever
R006z00	[D]Pyrexia of unknown origin NOS

Lymphadenopathy

Read code	Description
2C3..00	O/E - lymphadenopathy
2C3..11	O/E - adenopathy
2C3..12	O/E - enlarged lymph nodes
2C32.00	O/E -cervical lymphadenopathy
2C33.00	O/E-sub-mental lymphadenopathy
2C34.00	O/E-supraclav.lymphadenopathy
2C35.00	O/E -axillary lymphadenopathy
2C36.00	O/E - inguinal lymphadenopathy
2C37.00	O/E -popliteal lymphadenopathy
2C38.00	O/E - post-auricular lymphadenopathy
2C3Z.00	O/E - lymphadenopathy NOS

Read code	Description
BBm8.00	[M] Angioimmunoblastic lymphadenopathy
R056.00	[D]Lymph node enlargement
R056000	[D]Lymphadenopathy
R056100	[D]Swollen glands
R056300	[D]Generalized enlarged lymph nodes
R056400	[D]Localized enlarged lymph nodes
R056500	[D]Reactive lymphadenopathy
R056z00	[D]Lymph node enlargement NOS
R056z11	[D]Adenitis NOS
R056z12	[D]Enlarged submandibular lymph gland

Hyperlipidemia

Read code	description
8CR3.00	Hyperlipidaemia clinical management plan
C320200	Hyperlipidaemia, group A
C322.00	Mixed hyperlipidaemia
C324.00	Hyperlipidaemia NOS
Cyu8D00	[X]Other hyperlipidaemia

Chronic liver disease

Read code	Description
J61..00	Cirrhosis and chronic liver disease
J610.00	Alcoholic fatty liver
J611.00	Acute alcoholic hepatitis
J612.00	Alcoholic cirrhosis of liver
J612.11	Florid cirrhosis
J612.12	Laennec's cirrhosis
J612000	Alcoholic fibrosis and sclerosis of liver
J613.00	Alcoholic liver damage unspecified
J613000	Alcoholic hepatic failure
J614.00	Chronic hepatitis
J614000	Chronic persistent hepatitis
J614100	Chronic active hepatitis
J614111	Autoimmune chronic active hepatitis
J614200	Chronic aggressive hepatitis
J614300	Recurrent hepatitis
J614400	Chronic lobular hepatitis
J614y00	Chronic hepatitis unspecified
J614z00	Chronic hepatitis NOS
J615.00	Cirrhosis - non alcoholic
J615.11	Portal cirrhosis
J615000	Unilobular portal cirrhosis
J615100	Multilobular portal cirrhosis
J615111	Postnecrotic cirrhosis of liver
J615200	Mixed portal cirrhosis
J615300	Diffuse nodular cirrhosis
J615400	Fatty portal cirrhosis
J615500	Hypertrophic portal cirrhosis
J615600	Capsular portal cirrhosis
J615700	Cardiac portal cirrhosis
J615711	Congestive cirrhosis
J615800	Juvenile portal cirrhosis
J615811	Childhood function cirrhosis
J615812	Indian childhood cirrhosis
J615900	Pigmentary portal cirrhosis
J615A00	Pipe-stem portal cirrhosis
J615B00	Toxic portal cirrhosis

Read code	Description
J615C00	Xanthomatous portal cirrhosis
J615D00	Bacterial portal cirrhosis
J615E00	Cardituberculous cirrhosis
J615F00	Syphilitic portal cirrhosis
J615G00	Zooparasitic portal cirrhosis
J615H00	Infectious cirrhosis NOS
J615y00	Portal cirrhosis unspecified
J615z00	Non-alcoholic cirrhosis NOS
J615z11	Macronodular cirrhosis of liver
J615z12	Cryptogenic cirrhosis of liver
J615z13	Cirrhosis of liver NOS
J615z14	Laennec's cirrhosis, non-alcoholic
J615z15	Hepatic fibrosis
J616.00	Biliary cirrhosis
J616000	Primary biliary cirrhosis
J616100	Secondary biliary cirrhosis
J616200	Biliary cirrhosis of children
J616z00	Biliary cirrhosis NOS
J617.00	Alcoholic hepatitis
J617000	Chronic alcoholic hepatitis
J61y.00	Other non-alcoholic chronic liver disease
J61y000	Chronic yellow liver atrophy
J61y100	Non-alcoholic fatty liver
J61y200	Hepatosplenomegaly
J61y300	Portal fibrosis without cirrhosis
J61y400	Hepatic fibrosis
J61y500	Hepatic sclerosis
J61y600	Hepatic fibrosis with hepatic sclerosis
J61y700	Steatosis of liver
J61y800	Nonalcoholic steatohepatitis
J61y900	Fatty change of liver
J61y911	Fatty liver
J61yz00	Other non-alcoholic chronic liver disease NOS
J61z.00	Chronic liver disease NOS
14C5.00	H/O: liver disease

Diabetes

Read code	Description
C10..00	Diabetes mellitus
C100.00	Diabetes mellitus with no mention of complication
C100100	Diabetes mellitus, adult onset, no mention of complication
C100111	Maturity onset diabetes
C100112	Non-insulin dependent diabetes mellitus
C100z00	Diabetes mellitus NOS with no mention of complication
C101.00	Diabetes mellitus with ketoacidosis
C101100	Diabetes mellitus, adult onset, with ketoacidosis
C101y00	Other specified diabetes mellitus with ketoacidosis
C101z00	Diabetes mellitus NOS with ketoacidosis
C102.00	Diabetes mellitus with hyperosmolar coma
C102100	Diabetes mellitus, adult onset, with hyperosmolar coma
C102z00	Diabetes mellitus NOS with hyperosmolar coma
C103.00	Diabetes mellitus with ketoacidotic coma
C103100	Diabetes mellitus, adult onset, with ketoacidotic coma
C103y00	Other specified diabetes mellitus with coma
C103z00	Diabetes mellitus NOS with ketoacidotic coma
C104.00	Diabetes mellitus with renal manifestation
C104.11	Diabetic nephropathy
C104100	Diabetes mellitus, adult onset, with renal manifestation
C104y00	Other specified diabetes mellitus with renal complications
C104z00	Diabetes mellitus with nephropathy NOS
C105.00	Diabetes mellitus with ophthalmic manifestation
C105100	Diabetes mellitus, adult onset, + ophthalmic manifestation
C105y00	Other specified diabetes mellitus with ophthalmic complicatn
C105z00	Diabetes mellitus NOS with ophthalmic manifestation
C106.00	Diabetes mellitus with neurological manifestation
C106.11	Diabetic amyotrophy
C106.12	Diabetes mellitus with neuropathy
C106.13	Diabetes mellitus with polyneuropathy
C106100	Diabetes mellitus, adult onset, + neurological manifestation
C106y00	Other specified diabetes mellitus with neurological comps
C106z00	Diabetes mellitus NOS with neurological manifestation
C107.00	Diabetes mellitus with peripheral circulatory disorder
C107.11	Diabetes mellitus with gangrene
C107.12	Diabetes with gangrene
C107100	Diabetes mellitus, adult, + peripheral circulatory disorder
C107200	Diabetes mellitus, adult with gangrene
C107300	IDDM with peripheral circulatory disorder
C107400	NIDDM with peripheral circulatory disorder
C107y00	Other specified diabetes mellitus with periph circ comps
C107z00	Diabetes mellitus NOS with peripheral circulatory disorder
C108y00	Other specified diabetes mellitus with multiple comps
C108z00	Unspecified diabetes mellitus with multiple complications
C109.00	Non-insulin dependent diabetes mellitus
C109.11	NIDDM - Non-insulin dependent diabetes mellitus
C109.12	Type 2 diabetes mellitus
C109.13	Type II diabetes mellitus
C109000	Non-insulin-dependent diabetes mellitus with renal comps
C109011	Type II diabetes mellitus with renal complications
C109012	Type 2 diabetes mellitus with renal complications
C109100	Non-insulin-dependent diabetes mellitus with ophthalm comps
C109111	Type II diabetes mellitus with ophthalmic complications
C109112	Type 2 diabetes mellitus with ophthalmic complications
C109200	Non-insulin-dependent diabetes mellitus with neuro comps
C109211	Type II diabetes mellitus with neurological complications
C109212	Type 2 diabetes mellitus with neurological complications
C109300	Non-insulin-dependent diabetes mellitus with multiple comps
C109311	Type II diabetes mellitus with multiple complications
C109312	Type 2 diabetes mellitus with multiple complications
C109400	Non-insulin dependent diabetes mellitus with ulcer
C109411	Type II diabetes mellitus with ulcer
C109412	Type 2 diabetes mellitus with ulcer
C109500	Non-insulin dependent diabetes mellitus with gangrene

Read code	Description
C109511	Type II diabetes mellitus with gangrene
C109512	Type 2 diabetes mellitus with gangrene
C109600	Non-insulin-dependent diabetes mellitus with retinopathy
C109611	Type II diabetes mellitus with retinopathy
C109612	Type 2 diabetes mellitus with retinopathy
C109700	Non-insulin dependent diabetes mellitus - poor control
C109711	Type II diabetes mellitus - poor control
C109712	Type 2 diabetes mellitus - poor control
C109900	Non-insulin-dependent diabetes mellitus without complication
C109911	Type II diabetes mellitus without complication
C109912	Type 2 diabetes mellitus without complication
C109A00	Non-insulin dependent diabetes mellitus with mononeuropathy
C109A11	Type II diabetes mellitus with mononeuropathy
C109A12	Type 2 diabetes mellitus with mononeuropathy
C109B00	Non-insulin dependent diabetes mellitus with polyneuropathy
C109B11	Type II diabetes mellitus with polyneuropathy
C109B12	Type 2 diabetes mellitus with polyneuropathy
C109C00	Non-insulin dependent diabetes mellitus with nephropathy
C109C11	Type II diabetes mellitus with nephropathy
C109C12	Type 2 diabetes mellitus with nephropathy
C109D00	Non-insulin dependent diabetes mellitus with hypoglyca coma
C109D11	Type II diabetes mellitus with hypoglycaemic coma
C109D12	Type 2 diabetes mellitus with hypoglycaemic coma
C109E00	Non-insulin depend diabetes mellitus with diabetic cataract
C109E11	Type II diabetes mellitus with diabetic cataract
C109E12	Type 2 diabetes mellitus with diabetic cataract
C109F00	Non-insulin-dependent d m with peripheral angiopath
C109F11	Type II diabetes mellitus with peripheral angiopathy
C109F12	Type 2 diabetes mellitus with peripheral angiopathy
C109G00	Non-insulin dependent diabetes mellitus with arthropathy
C109G11	Type II diabetes mellitus with arthropathy
C109G12	Type 2 diabetes mellitus with arthropathy
C109H00	Non-insulin dependent d m with neuropathic arthropathy
C109H11	Type II diabetes mellitus with neuropathic arthropathy
C109H12	Type 2 diabetes mellitus with neuropathic arthropathy
C109J00	Insulin treated Type 2 diabetes mellitus
C109J11	Insulin treated non-insulin dependent diabetes mellitus
C109J12	Insulin treated Type II diabetes mellitus
C109K00	Hyperosmolar non-ketotic state in type 2 diabetes mellitus
C10A.00	Malnutrition-related diabetes mellitus
C10A.11	Jamaica type diabetes
C10A000	Malnutrition-related diabetes mellitus with coma
C10A100	Malnutrition-related diabetes mellitus with ketoacidosis
C10A200	Malnutrition-related diabetes mellitus with renal complicatn
C10A300	Malnutrit-related diabetes mellitus wth ophthalmic complicat
C10A400	Malnutrition-related diabetes mellitus wth neuro complicatns
C10A500	Malnutritn-relat diabetes melitus wth periph circul complctn
C10A600	Malnutrition-related diabetes mellitus with multiple comps
C10A700	Malnutrition-related diabetes mellitus without complications
C10AW00	Malnutrit-related diabetes mellitus with unspec complics
C10AX00	Malnutrit-relat diabetes mellitus with other spec comps
C10ER00	Latent autoimmune diabetes mellitus in adult
C10F.00	Type 2 diabetes mellitus
C10F.11	Type II diabetes mellitus
C10F000	Type 2 diabetes mellitus with renal complications
C10F011	Type II diabetes mellitus with renal complications
C10F100	Type 2 diabetes mellitus with ophthalmic complications
C10F111	Type II diabetes mellitus with ophthalmic complications
C10F200	Type 2 diabetes mellitus with neurological complications
C10F211	Type II diabetes mellitus with neurological complications
C10F300	Type 2 diabetes mellitus with multiple complications
C10F311	Type II diabetes mellitus with multiple complications
C10F400	Type 2 diabetes mellitus with ulcer
C10F411	Type II diabetes mellitus with ulcer

Read code	Description
C10F500	Type 2 diabetes mellitus with gangrene
C10F511	Type II diabetes mellitus with gangrene
C10F600	Type 2 diabetes mellitus with retinopathy
C10F611	Type II diabetes mellitus with retinopathy
C10F700	Type 2 diabetes mellitus - poor control
C10F711	Type II diabetes mellitus - poor control
C10F800	Reaven's syndrome
C10F811	Metabolic syndrome X
C10F900	Type 2 diabetes mellitus without complication
C10F911	Type II diabetes mellitus without complication
C10FA00	Type 2 diabetes mellitus with mononeuropathy
C10FA11	Type II diabetes mellitus with mononeuropathy
C10FB00	Type 2 diabetes mellitus with polyneuropathy
C10FB11	Type II diabetes mellitus with polyneuropathy
C10FC00	Type 2 diabetes mellitus with nephropathy
C10FC11	Type II diabetes mellitus with nephropathy
C10FD00	Type 2 diabetes mellitus with hypoglycaemic coma
C10FD11	Type II diabetes mellitus with hypoglycaemic coma
C10FE00	Type 2 diabetes mellitus with diabetic cataract
C10FE11	Type II diabetes mellitus with diabetic cataract
C10FF00	Type 2 diabetes mellitus with peripheral angiopathy
C10FF11	Type II diabetes mellitus with peripheral angiopathy
C10FG00	Type 2 diabetes mellitus with arthropathy
C10FG11	Type II diabetes mellitus with arthropathy
C10FH00	Type 2 diabetes mellitus with neuropathic arthropathy
C10FH11	Type II diabetes mellitus with neuropathic arthropathy
C10FJ00	Insulin treated Type 2 diabetes mellitus
C10FJ11	Insulin treated Type II diabetes mellitus
C10FK00	Hyperosmolar non-ketotic state in type 2 diabetes mellitus
C10FL00	Type 2 diabetes mellitus with persistent proteinuria

Read code	Description
C10FL11	Type II diabetes mellitus with persistent proteinuria
C10FM00	Type 2 diabetes mellitus with persistent microalbuminuria
C10FM11	Type II diabetes mellitus with persistent microalbuminuria
C10FN00	Type 2 diabetes mellitus with ketoacidosis
C10FN11	Type II diabetes mellitus with ketoacidosis
C10FP00	Type 2 diabetes mellitus with ketoacidotic coma
C10FP11	Type II diabetes mellitus with ketoacidotic coma
C10FQ00	Type 2 diabetes mellitus with exudative maculopathy
C10FQ11	Type II diabetes mellitus with exudative maculopathy
C10FR00	Type 2 diabetes mellitus with gastroparesis
C10FS00	Maternally inherited diabetes mellitus
C10G.00	Secondary pancreatic diabetes mellitus
C10G000	Secondary pancreatic diabetes mellitus without complication
C10H.00	Diabetes mellitus induced by non-steroid drugs
C10H000	DM induced by non-steroid drugs without complication
C10L.00	Fibrocalculous pancreatopathy
C10L000	Fibrocalculous pancreatopathy without complication
C10M.00	Lipoatrophic diabetes mellitus
C10M000	Lipoatrophic diabetes mellitus without complication
C10N.00	Secondary diabetes mellitus
C10N000	Secondary diabetes mellitus without complication
C10N100	Cystic fibrosis related diabetes mellitus
C10y.00	Diabetes mellitus with other specified manifestation
C10y100	Diabetes mellitus, adult, + other specified manifestation
C10yy00	Other specified diabetes mellitus with other spec comps
C10yz00	Diabetes mellitus NOS with other specified manifestation
C10z.00	Diabetes mellitus with unspecified complication
C10z100	Diabetes mellitus, adult onset, + unspecified complication
C10zy00	Other specified diabetes mellitus with unspecified comps
C10zz00	Diabetes mellitus NOS with unspecified complication

Syphilis

Read code	Description
A9...00	Syphilis and other venereal diseases
A91..00	Early symptomatic syphilis
A910.00	Primary genital syphilis
A910.11	Genital chancre
A911.00	Primary anal syphilis
A912.00	Other primary syphilis
A912000	Primary breast syphilis
A912100	Primary finger syphilis
A912200	Primary lip syphilis
A912300	Primary tonsil syphilis
A912z00	Other primary syphilis NOS
A913.00	Secondary syphilis of skin or mucus membranes
A913000	Secondary syphilis of anus
A913100	Secondary syphilis of mouth
A913200	Secondary syphilis of pharynx
A913300	Secondary syphilis of skin
A913400	Secondary syphilis of tonsils
A913500	Secondary syphilis of vulva
A913z00	Secondary syphilis of skin or mucus membranes NOS
A914.00	Adenopathy due to secondary syphilis
A915.00	Uveitis due to secondary syphilis
A915000	Syphilitic uveitis unspecified
A915100	Secondary syphilitic chorioretinitis
A915200	Secondary syphilitic iridocyclitis
A915z00	Secondary syphilitic uveitis NOS
A916.00	Secondary syphilis of viscera or bone
A916.11	Secondary syphilis of bone
A916.12	Secondary syphilis of viscera
A916000	Secondary syphilitic periostitis
A916100	Secondary syphilitic hepatitis
A916z00	Secondary syphilis of viscera and bone NOS
A917.00	Secondary syphilis relapse
A918.00	Other forms of secondary syphilis
A918000	Acute secondary syphilitic meningitis
A918100	Syphilitic alopecia
A918z00	Other secondary syphilis NOS
A919.00	Unspecified secondary syphilis
A91X.00	Early syphilis, unspecified
A92..00	Latent early syphilis
A920.00	Serological relapse after treatment of latent early syphilis
A92z.00	Latent early syphilis NOS
A93..00	Cardiovascular syphilis
A930.00	Syphilitic aortic aneurysm
A931.00	Syphilitic aortitis
A932.00	Syphilitic endocarditis
A932.11	Syphilitic valve disease
A932000	Syphilitic endocarditis of unspecified valve
A932100	Syphilitic endocarditis of mitral valve
A932200	Syphilitic endocarditis of aortic valve
A932300	Syphilitic endocarditis of tricuspid valve
A932400	Syphilitic endocarditis of pulmonary valve
A932z00	Syphilitic endocarditis of heart valve NOS
A93y.00	Other specified cardiovascular system syphilis
A93y000	Syphilitic pericarditis
A93y100	Syphilitic myocarditis
A93yz00	Other specified cardiovascular system syphilis NOS

Read code	Description
A93z.00	Cardiovascular syphilis NOS
A94..00	Neurosyphilis
A940.00	Tabes dorsalis - neurosyphilis
A940.11	Locomotor ataxia
A940.12	Syphilitic posterior spinal sclerosis
A941.00	General paresis - neurosyphilis
A941.11	General paralysis of insane
A941.12	Taboparesis
A942.00	Syphilitic meningitis
A943.00	Asymptomatic neurosyphilis
A94y.00	Other specified neurosyphilis
A94y000	Syphilitic encephalitis
A94y100	Syphilitic parkinsonism
A94y200	Syphilitic disseminated retinochoroiditis
A94y300	Syphilitic optic atrophy
A94y400	Syphilitic retrobulbar neuritis
A94y500	Syphilitic acoustic neuritis
A94y600	Rupture of syphilitic cerebral aneurysm
A94yz00	Other specified neurosyphilis NOS
A94z.00	Neurosyphilis NOS
A95..00	Other forms of late syphilis with symptoms
A950.00	Syphilitic episcleritis
A951.00	Syphilis of lung
A952.00	Syphilitic peritonitis
A953.00	Syphilis of liver
A954.00	Syphilis of kidney
A954.11	Renal syphilis
A955.00	Syphilis of bone
A956.00	Syphilis of muscle
A957.00	Syphilis of synovium, tendon or bursa
A957000	Syphilis of synovium
A957100	Syphilis of tendon
A957200	Syphilis of bursa
A957z00	Syphilis of synovium, tendon or bursa NOS
A95y.00	Other specified late syphilis
A95z.00	Late symptomatic syphilis NOS
A96..00	Late latent syphilis
A97..00	Other and unspecified syphilis
A97..11	Syphilis
A97..12	Treponemal infection
A970.00	Late syphilis unspecified
A971.00	Latent syphilis unspecified
A97z.00	Syphilis NOS
A990.00	Chancroid
A990.11	Chancroidal bubo
A990.12	Ducrey's chancre
A9y..00	Other specified syphilis or other venereal diseases
A9z..00	Syphilis or venereal disease NOS
Ayu4400	[X]Primary syphilis of other sites
Ayu4500	[X]Other secondary syphilis
Ayu4600	[X]Early syphilis, unspecified
Ayu4700	[X]Neurosyphilis, unspecified
Ayu4800	[X]Other symptomatic late syphilis
Ayu4900	[X]Late syphilis, unspecified
Ayu4A00	[X]Syphilis, unspecified

Any sexually transmitted infection (includes other STI, chlamydia, gonorrhoea, genital herpes)

Read code	Description
A9...11	Sexually transmitted diseases
A9...12	Venereal diseases
A9...13	Sexually transmitted infectious diseases
A541.00	Genital herpes simplex
A541000	Genital herpes unspecified
A541100	Herpetic vulvovaginitis
A541200	Herpetic ulceration of vulva
A541300	Herpetic infection of penis
A541400	Herpesviral infection of perianal skin and rectum
A541500	Anogenital herpesviral infection
A541600	Genital herpes simplex type 1
A541700	Genital herpes simplex type 2
A541800	Recurrent genital herpes simplex type 1
A541900	Recurrent genital herpes simplex type 2
A541z00	Genital herpes simplex NOS
Ayu4G00	[X]Anogenital herpes viral infection, unspecified
A781200	Genital warts
A781211	Condylomata acuminatum
A781212	Penile warts
A781213	Venereal warts
A98..00	Gonococcal infections
A980.00	Acute gonorrhoea of lower genitourinary tract
A980000	Acute gonococcal Bartholinitis
A980100	Acute gonococcal urethritis
A980200	Acute gonococcal vulvovaginitis
A980z00	Acute gonorrhoea of lower genitourinary tract NOS
A981.00	Acute gonorrhoea of upper genitourinary tract
A981000	Acute unspecified gonorrhoea of upper genitourinary tract
A981100	Acute gonococcal cystitis
A981111	Bladder gonorrhoea - acute
A981200	Acute gonococcal prostatitis
A981300	Acute gonococcal epididymo-orchitis
A981311	Acute gonococcal orchitis
A981400	Acute gonococcal seminal vesiculitis
A981500	Acute gonococcal cervicitis
A981600	Acute gonococcal endometritis
A981611	Uterus - acute gonorrhoea
A981700	Acute gonococcal salpingitis
A981z00	Acute gonorrhoea upper genitourinary tract NOS
A982.00	Chronic gonorrhoea lower genitourinary tract
A982000	Chronic gonococcal bartholinitis
A982100	Chronic gonococcal urethritis
A982200	Chronic gonococcal vulvovaginitis
A982z00	Chronic gonorrhoea of lower genitourinary tract NOS
A983.00	Chronic gonorrhoea of upper genitourinary tract
A983000	Chronic unspecified gonorrhoea of upper genitourinary tract
A983100	Chronic gonococcal cystitis
A983200	Chronic gonococcal prostatitis
A983300	Chronic gonococcal epididymo-orchitis
A983400	Chronic gonococcal seminal vesiculitis
A983500	Chronic gonococcal cervicitis
A983600	Chronic gonococcal endometritis
A983611	Uterus - chronic gonorrhoea

Read code	Description
A983700	Chronic gonococcal salpingitis
A983z00	Chronic gonorrhoea of upper genitourinary tract NOS
A984.00	Gonococcal eye infection
A984000	Neonatal gonococcal conjunctivits
A984011	Ophthalmia neonatorum - gonococcal
A984100	Gonococcal iridocyclitis
A984200	Gonococcal endophthalmia
A984300	Gonococcal keratitis
A984z00	Gonococcal eye infection NOS
A985.00	Gonococcal joint infection
A985000	Gonococcal arthritis
A985100	Gonococcal synovitis or tenosynovitis
A985111	Gonococcal synovitis
A985112	Gonococcal tenosynovitis
A985200	Gonococcal bursitis
A985300	Gonococcal spondylitis
A985z00	Gonococcal joint infection NOS
A985z11	Rheumatism - gonococcal
A986.00	Gonococcal pharynx infection
A987.00	Gonococcal proctitis
A987000	Gonococcal anal infection
A987100	Gonococcal rectal infection
A987z00	Gonococcal proctitis NOS
A98y.00	Gonococcal infection of other specified sites
A98y000	Gonococcal keratosis
A98y100	Gonococcal meningitis
A98y200	Gonococcal pericarditis
A98y300	Gonococcal endocarditis
A98y400	Other gonococcal heart disease
A98y500	Gonococcal peritonitis
A98y600	Fitzhugh Curtis syndrome
A98yy00	Other gonococcal infection of other specified site
A98yy11	Gonococcal hepatitis
A98yy12	Abscess gonococcal
A98yy13	Gonococcal perihepatitis
A98yy14	Gonococcal cellulitis
A98yz00	Gonococcal infection of other site NOS
A98yz11	Gonococcaemia NOS
A98yz12	Gonococcal septicaemia
A98z.00	Gonococcal infections NOS
A98z.11	Gonorrhoea
Ayu4B00	[X]Other gonococcal infections
Ayu4C00	[X]Gonococcal infection, unspecified
A99..00	Other venereal diseases
A991.00	Lymphogranuloma venereum
A992.00	Granuloma inguinale
A992.11	Donovanosis
A992.12	Pudendal ulcer
A993.00	Reiter's disease / syndrome
A993.11	Reiter's syndrome
A994.00	Nonspecific urethritis
A99y.00	Other specified venereal diseases
A99z.00	Venereal disease NOS

Appendix 6: Recommended and permitted Read codes for HIV from GUMCAD STI Surveillance System Clinical guidelines

CTV3 Read Code Option	Read Code Description
Recommended Read Codes	
6827.	HIV screening
XaIOI	Antenatal HIV screening
XaLI7	HIV screening declined
43C2.	HIV negative
43C3.	HIV positive
New request	New HIV diagnosis
New request	New HIV diagnosis (Acute)
New request	New HIV diagnosis (Late)
XaXU7	Human immunodeficiency virus test not appropriate
65VE.	Notification of AIDS
All Permitted Read Codes	
XaILa	HIV 1 nucleic acid detection
XaFuK	HIV 1 PCR
XaFuN	HIV antibody/antigen (Duo)
A7896	HIV disease resulting in Burkitt's lymphoma
A7892	HIV disease resulting in candidiasis
A7891	HIV disease resulting in cytomegaloviral disease
A7895	HIV disease resulting in Kaposi's sarcoma
A7899	HIV disease resulting in lymphoid interstitial pneumonitis
A7894	HIV disease resulting in multiple infections
A7898	HIV disease resulting in multiple malignant neoplasms
A7890	HIV disease resulting in mycobacterial infection
A7897	HIV disease resulting in other types of non-Hodgkin's lymphoma
A7893	HIV disease resulting in Pneumocystis carinii pneumonia
43C2.	HIV negative
XaFuO	HIV p24 antigen level
43C3.	HIV positive
6827.	HIV screening
XaLI7	HIV screening declined
XaIon	HIV screening test
XaKEv	HIV serology
Xa0v3	HIV status
XaFuL	HIV viral load
XaEQb	HIV1 antibody level
XaEME	HIV2 antibody level
X00cZ	HIV-associated retinitis
Xa1k1	HIV-related sclerosing cholangitis
XaIV7	HPV - Human papillomavirus test consent given
XaIV8	HPV - Human papillomavirus test declined
XaIVB	HPV - Human papillomavirus test negative
XaIVA	HPV - Human papillomavirus test positive
XaILb	HTLV 1 nucleic acid detection
XaIMB	HTLV 2 nucleic acid detection
43CZ.	HTLV-3 antibody NOS
XaJGg	Human herpes virus serology
A789.	Human immunodef virus resulting in other disease
X70O5	Human immunodeficiency viral myelitis
XaFuM	Human immunodeficiency virus antibody level
X76za	Human immunodeficiency virus antibody titre
XaLu9	Human immunodeficiency virus blood test
XaLHr	Human immunodeficiency virus contact
XaLHr	Human immunodeficiency virus contact
X303J	Human immunodeficiency virus enteropathy
X70M6	Human immunodeficiency virus infection
A7883	Human immunodeficiency virus infection constitutional disease
X70O1	Human immunodeficiency virus infection wasting syndrome
A7882	Human immunodeficiency virus infection with persistent generalised lymphadenopathy

A7885	Human immunodeficiency virus infection with secondary clinical infectious disease
X001h	Human immunodeficiency virus myelitis
X709a	Human immunodeficiency virus myopathy
X70O6	Human immunodeficiency virus neuropathy
XaMBF	Human immunodeficiency virus proviral deoxyribonucleic acid polymerase chain reaction
43C..	Human immunodeficiency virus test
XaLot	Human immunodeficiency virus test equivocal
XaXU7	Human immunodeficiency virus test not appropriate
XaO9K	Human immunodeficiency virus viral load by log rank
A7884	Human immunodeficiency virus with neurological disease
A788y	Human immunodeficiency virus with other clinical findings
A7886	Human immunodeficiency virus with secondary cancers
X20Q9	Human immunodeficiency virus-associated periodontitis
X001B	Human immunodefiency virus encephalitis
X001C	Human immunodefiency virus leukoencephalopathy

Appendix 7: Participant Information sheet and consent forms

Appendix 7-A: Participant Information sheet

Research into coding of information on Human Immunodeficiency Virus (HIV) in primary care records

What this study is about

In this study we will interview primary care clinicians to explore their perspectives on coding practice in relation to sensitive health conditions, particularly HIV, in primary care. Understanding what clinicians do in practice, and their reasons for doing this, will assist us to develop tools to identify patients with HIV in primary care records where the coding is not systematically searchable as clinicians were using free text or alternative clinical codes. The next step will be to investigate risk factors and patient characteristics that are associated with a diagnosis of HIV. This will allow us to develop a prediction score to identify patients who might have HIV infection from information recorded in their primary care records and help primary care clinicians to decide which patients should be offered HIV testing.

Why you have been asked to take part

We are inviting you to take part in this study because you expressed an interest in being interviewed on coding practice in relation to sensitive health conditions, particularly HIV, in primary care records.

You are under no obligation to take part in this interview and should you decide to withdraw, you may do so at any time during the interview and data can be withdrawn up to two weeks after the interview.

What the study involves

You will be interviewed about what you do and your perspective on what other clinicians may do when coding information on HIV diagnosis in primary care records. The interview will take

about 20 minutes and will be recorded and transcribed. Afterwards the student conducting the study will review the transcripts and draw some conclusions about coding practice in relation to HIV diagnosis in primary care.

Confidentiality

The recordings will be identified only by a number. The transcriptions will have any names or potential identifiers removed. The transcriptions will therefore be anonymous. No individual will be identified in any report. The recordings and data collected in the interviews will be stored at the University of Birmingham and it will only be used for research and audit purposes. The data will be destroyed after 10 years, in accordance with the University of Birmingham guidelines.

Will I receive compensation for my participation?

Interviewees will be paid for their time using the standard rate for service support costs (£80 per hour paid to your practice).

How will the findings of the study be used?

Findings from this research will be published in an academic journal and presented at conferences. If requested, you can be sent a summary of the findings.

Contact details

For any queries about this study please contact:

Benhildah Rumbwere Dube	**Prof Tom Marshall**
Email: [redacted]	**Email:** [redacted]
Phone: [redacted]	**Phone:** [redacted]

Appendix 7-B: Interview Consent Form

This information is collected as part of a research project by the Institute of Applied Health Research at the University of Birmingham. The research aims to explore GPs' perspectives on coding practice in relation to sensitive health conditions, particularly HIV, in primary care. The aim is to develop prediction rules to help General Practitioners identify patients who should be offered HIV testing. For this we need to find out how General Practitioners might record information about a diagnosis of HIV infection in medical records. We therefore would like interview you about how General Practitioners might record information about a diagnosis of HIV infection in medical records.

The interview will be recorded and transcribed. No information allowing the participant to be identified such as name, address or date of birth will be included in the transcriptions. The recordings and data collected in the interviews will be stored in the University of Birmingham computer network: access will be restricted to the study investigators and they will only be used for research and audit purposes. The anonymised transcriptions will be stored on an encrypted laptop. The data will be destroyed after 10 years, in accordance with the University of Birmingham guidelines.

- I confirm that I have read and understood the participant information sheet for this study. I have had the opportunity to ask questions if necessary and have had these answered satisfactorily. ☐
- I understand that my participation is voluntary and that I am free to withdraw at any stage of the interview or data can be withdrawn up to two weeks after the interview by contacting the researcher without giving any reason. ☐
- I understand that this interview will be recorded. ☐

- I understand that information used in publications and reports will be anonymised. ☐

- I understand that my personal data will be processed for the purposes detailed above, in accordance with the Data Protection Act 1998. ☐

Based upon the above, I agree to take part in this study.

Name of participant	**Date**	**Signature**
Name of researcher	Date	Signature

Appendix 8: Read Code list for confirmed and" probable" HIV

Confirmed HIV

Read code	Description
R109.0	[D]Laboratory evidence of human immunodeficiency virus [HIV]
AyuCB	[X]Hiv disease resulting in haematological and immunological abnormalities not elsewhere classified
AyuCA	[X]HIV disease resulting in multiple diseases classified elsewhere
AyuC0	[X]HIV disease resulting in other bacterial infections
AyuC4	[X]Hiv disease resulting in other infectious and parasitic diseases
AyuC8	[X]HIV disease resulting in other malignant neoplasms
AyuC7	[X]Hiv disease resulting in other malignant neoplasms of lymphoid haematopoietic and related tissue
AyuC2	[X]HIV disease resulting in other mycoses
AyuC6	[X]HIV disease resulting in other non-Hodgkin's lymphoma
AyuCC	[X]HIV disease resulting in other specified conditions
AyuC1	[X]HIV disease resulting in other viral infections
AyuC5	[X]Hiv disease resulting in unspecified infectious and parasitic disease
AyuC9	[X]HIV disease resulting in unspecified malignant neoplasm
AyuCD	[X]Unspecified human immunodeficiency virus [HIV] disease
XaFuN	HIV antibody/antigen (Duo)
A7896	HIV disease resulting in Burkitt's lymphoma
A7892	HIV disease resulting in candidiasis
A7891	HIV disease resulting in cytomegaloviral disease
A7895	HIV disease resulting in Kaposi's sarcoma
A7899	HIV disease resulting in lymphoid interstitial pneumonitis
A7894	HIV disease resulting in multiple infections
A7898	HIV disease resulting in multiple malignant neoplasms
A7890	HIV disease resulting in mycobacterial infection
A7897	HIV disease resulting in other types of non-Hodgkin's lymphoma
A7893	HIV disease resulting in Pneumocystis carinii pneumonia
43C3.	HIV positive
Xa0v3	HIV status
XaFuL	HIV viral load
XaEQb	HIV1 antibody level
XaEME	HIV2 antibody level
X00cZ	HIV-associated retinitis
Xa1k1	HIV-related sclerosing cholangitis

“Probable HIV”

Immunodeficiency

medcode	description
2J30.00	Patient immunocompromised
2J31.00	Patient immunosuppressed
C39X.00	Immunodeficiency associated+major defect, unspecified
Cyu0100	[X]Other combined immunodeficiency disorders
Cyu0200	[X]Immunodeficiency associatd+other specified major defects
Cyu0300	[X]Immunodeficiency associated+major defect, unspecified
Cyu0500	[X]Other specified immunodeficiency disorders

Viral load

medcode	description
4J39.00	Viral load

Immunity problems

medcode	description
C393.00	Unspecified immunity deficiency

Appendix 9: The RECORD statement – checklist of items, extended from the STROBE statement that should be reported in observational studies using routinely collected health data.

	Item No.	STROBE items	Location in manuscript where items are reported	RECORD items	Location in manuscript where items are reported
Title and abstract					
	1	(a) Indicate the study's design with a commonly used term in the title or the abstract (b) Provide in the abstract an informative and balanced summary of what was done and what was found		RECORD 1.1: The type of data used should be specified in the title or abstract. When possible, the name of the databases used should be included. RECORD 1.2: If applicable, the geographic region and timeframe within which the study took place	

				should be reported in the title or abstract. RECORD 1.3: If linkage between databases was conducted for the study, this should be clearly stated in the title or abstract.	
Introduction					
Background rationale	2	Explain the scientific background and rationale for the investigation being reported			
Objectives	3	State specific objectives, including any prespecified hypotheses			
Methods					
Study Design	4	Present key elements of study			

		design early in the paper			
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection			
Participants	6	*(a) Cohort study* - Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up *Case-control study* - Give the eligibility criteria, and the sources and methods of case ascertainment and control selection.		RECORD 6.1: The methods of study population selection (such as codes or algorithms used to identify subjects) should be listed in detail. If this is not possible, an explanation should be provided. RECORD 6.2: Any validation studies of the codes or algorithms used to select the	

		Give the rationale for the choice of cases and controls *Cross-sectional study* - Give the eligibility criteria, and the sources and methods of selection of participants *(b) Cohort study* - For matched studies, give matching criteria and number of exposed and unexposed *Case-control study* - For matched studies, give matching criteria and the number of controls per case		population should be referenced. If validation was conducted for this study and not published elsewhere, detailed methods and results should be provided. RECORD 6.3: If the study involved linkage of databases, consider use of a flow diagram or other graphical display to demonstrate the data linkage process, including the number of individuals with linked data at each stage.	

Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable.		RECORD 7.1: A complete list of codes and algorithms used to classify exposures, outcomes, confounders, and effect modifiers should be provided. If these cannot be reported, an explanation should be provided.	
Data sources/ measure ment	8	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group			
Bias	9	Describe any efforts to address potential sources of bias			

Study size	10	Explain how the study size was arrived at			
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen, and why			
Statistical methods	12	(a) Describe all statistical methods, including those used to control for confounding (b) Describe any methods used to examine subgroups and interactions (c) Explain how missing data were addressed			

		(d) *Cohort study* - If applicable, explain how loss to follow-up was addressed *Case-control study* - If applicable, explain how matching of cases and controls was addressed *Cross-sectional study* - If applicable, describe analytical methods taking account of sampling strategy (e) Describe any sensitivity analyses			
Data access and cleaning methods		..		RECORD 12.1: Authors should describe the extent to which the investigators had access to the database population used to	

				create the study population. RECORD 12.2: Authors should provide information on the data cleaning methods used in the study.	
Linkage		..		RECORD 12.3: State whether the study included person-level, institutional-level, or other data linkage across two or more databases. The methods of linkage and methods of linkage quality evaluation should be provided.	
Results					
Participants	13	(a) Report the numbers of individuals at each stage of the study (*e.g.*, numbers		RECORD 13.1: Describe in detail the selection of the persons included in the study (*i.e.*, study population	

		potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed) (b) Give reasons for non-participation at each stage. (c) Consider use of a flow diagram		selection) including filtering based on data quality, data availability and linkage. The selection of included persons can be described in the text and/or by means of the study flow diagram.	
Descriptive data	14	(a) Give characteristics of study participants (*e.g.*, demographic, clinical, social) and information on exposures and potential confounders (b) Indicate the number of participants with missing data for			

		each variable of interest (c) *Cohort study* - summarise follow-up time (*e.g.*, average and total amount)			
Outcome data	15	*Cohort study* - Report numbers of outcome events or summary measures over time *Case-control study* - Report numbers in each exposure category, or summary measures of exposure *Cross-sectional study* - Report numbers of outcome events or summary measures			
Main results	16	(a) Give unadjusted estimates and, if			

		applicable, confounder-adjusted estimates and their precision (e.g., 95% confidence interval). Make clear which confounders were adjusted for and why they were included (b) Report category boundaries when continuous variables were categorized (c) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period			
Other analyses	17	Report other analyses done—			

		e.g., analyses of subgroups and interactions, and sensitivity analyses			
Discussion					
Key results	18	Summarise key results with reference to study objectives			
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias		RECORD 19.1: Discuss the implications of using data that were not created or collected to answer the specific research question(s). Include discussion of misclassification bias, unmeasured confounding, missing data, and changing eligibility over time, as they pertain to the study being reported.	
Interpretation	20	Give a cautious overall			

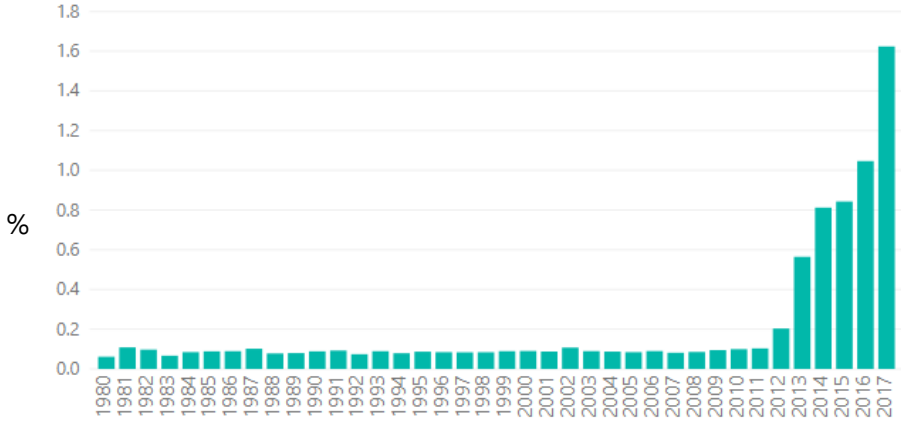
		interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence			
Generalisability	21	Discuss the generalisability (external validity) of the study results			
Other Information					
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based			
Accessibility of		..		RECORD 22.1: Authors should provide	

protocol, raw data, and programming code				information on how to access any supplemental information such as the study protocol, raw data, or programming code.	

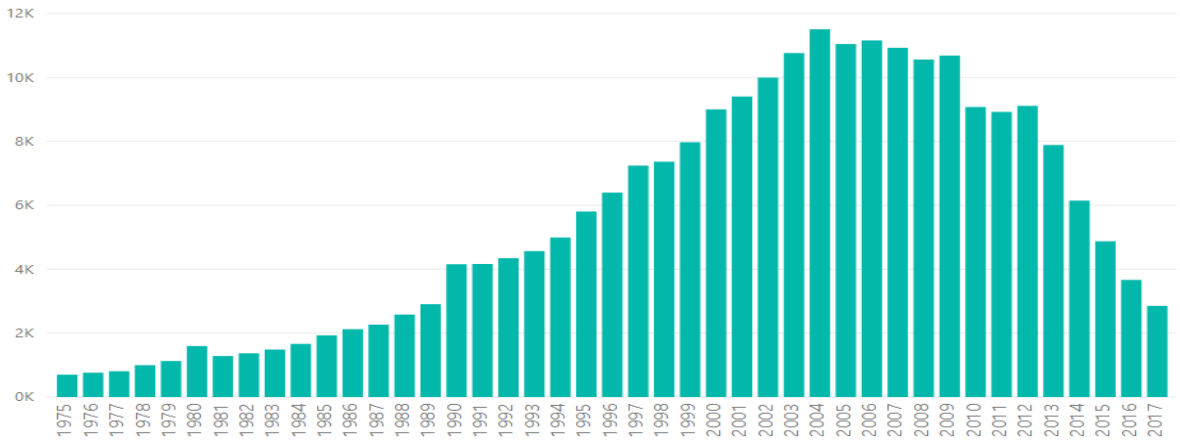
*Reference: Benchimol EI, Smeeth L, Guttmann A, Harron K, Moher D, Petersen I, Sørensen HT, von Elm E, Langan SM, the RECORD Working Committee. The REporting of studies Conducted using Observational Routinely collected health Data (RECORD) Statement. *PLoS Medicine* 2015; in press.

Appendix 10: Consistency of prevalence of predictor variables over time

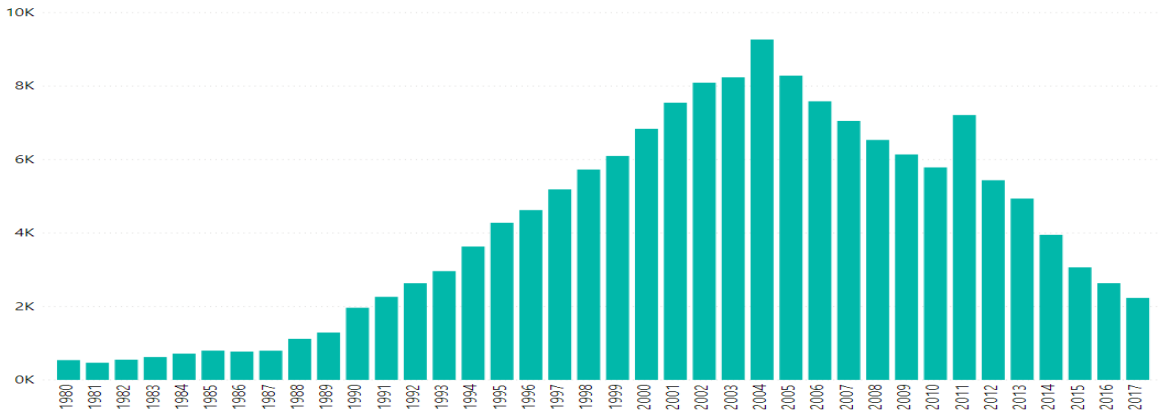
A. Proportion of GP registered patients with sexuality recorded by year of registration



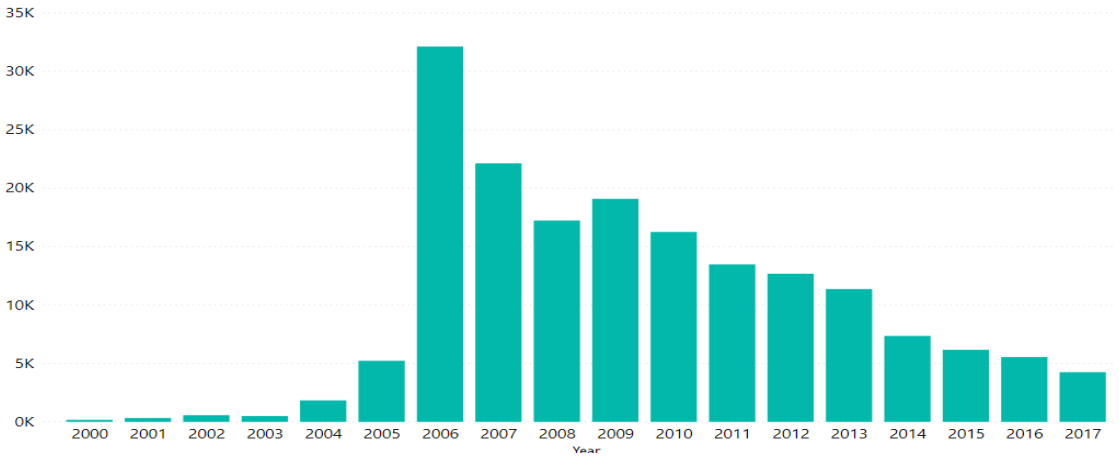
B. Number of GP registered patients with alcohol misuse by earliest year of first recording



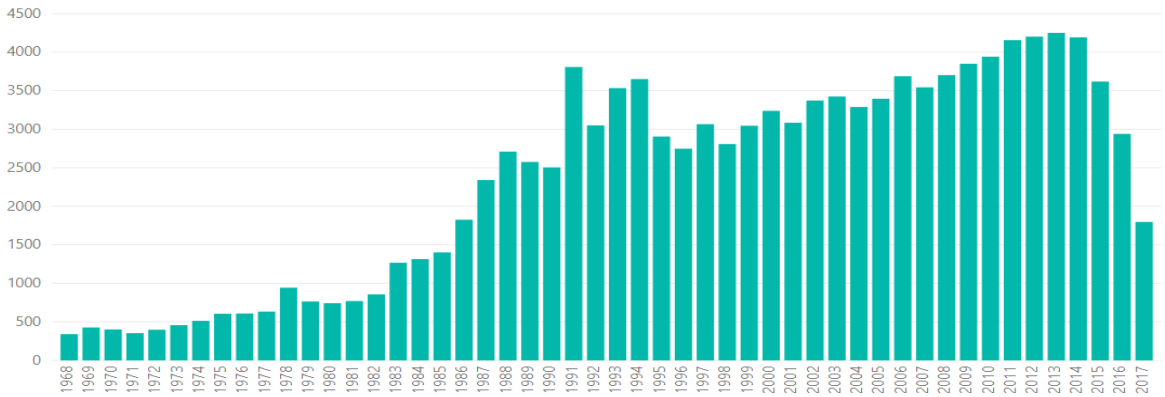
C. Number of GP registered patients with drug misuse by earliest year of first recording



D. Number of GP registered patients with obesity by earliest year of first recording

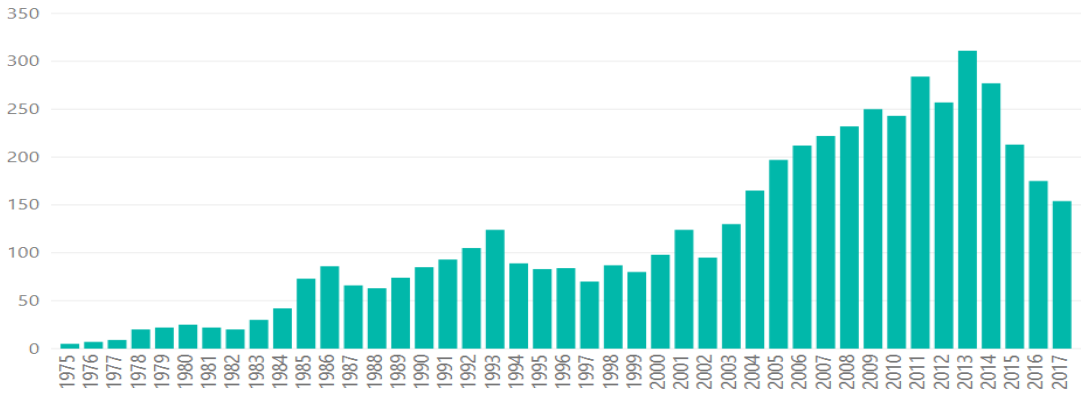


E. Number of GP registered patients with contact abroad recorded by year of registration

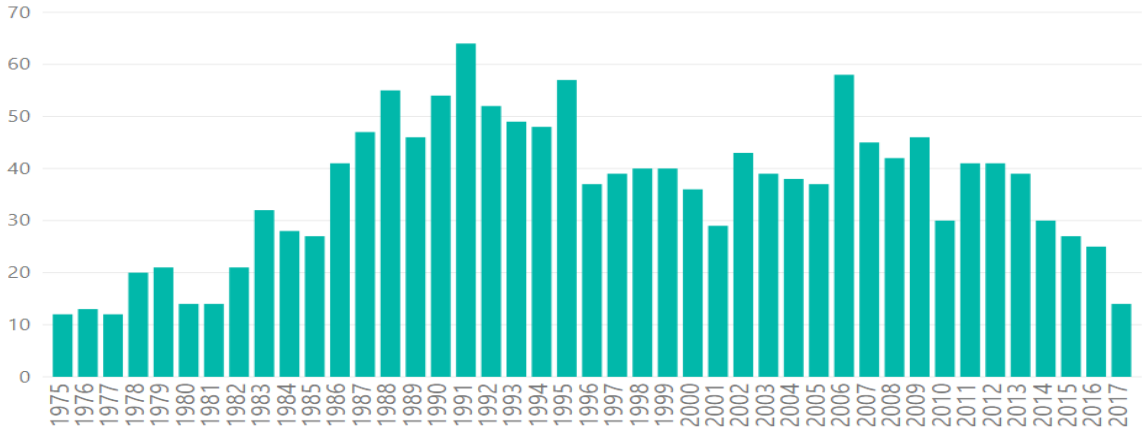


F. Number of GP registered patients recorded by year of registration

i) Multiple number of life partners



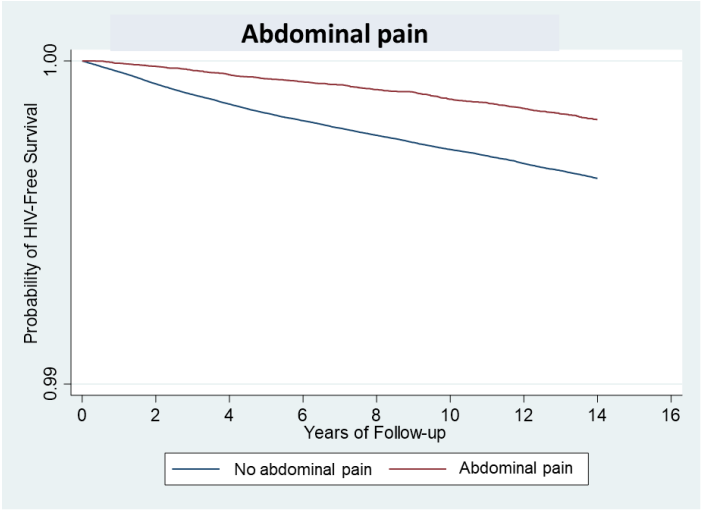
ii) Unsafe partner characteristics

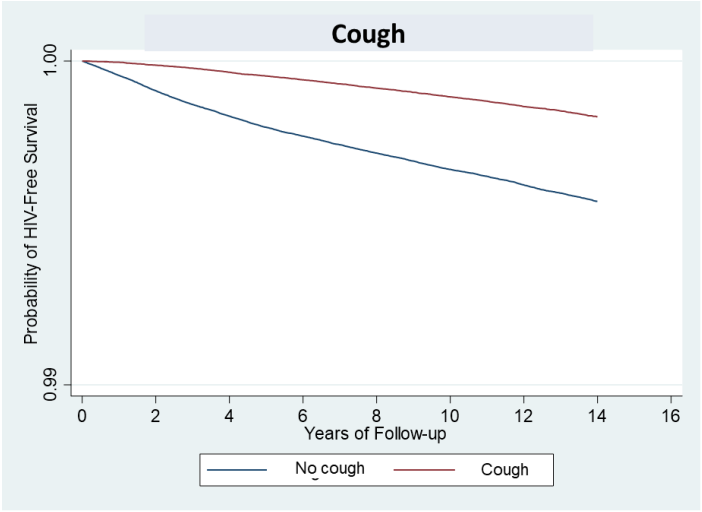


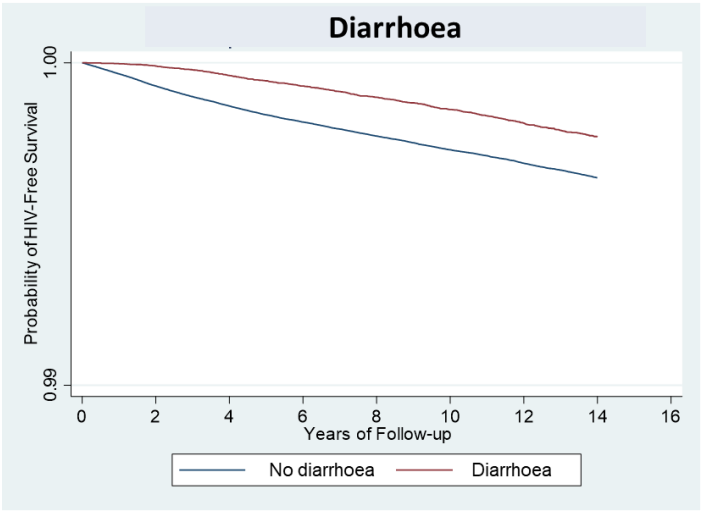
Appendix 11: Predictor variables excluded from subsequent analysis

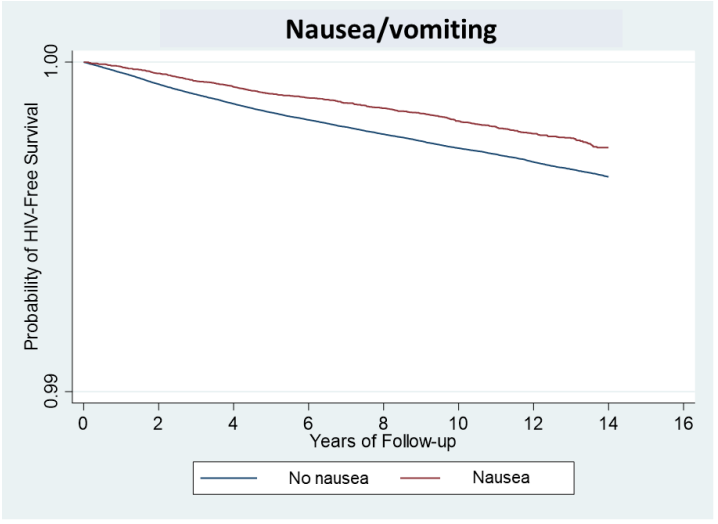
	Category	Predictor	Frequency in population	Prevalence of HIV	Decision
1.	Demographic	Sexuality	<0.2%		Exclude
2.	Demographic	Area of origin	Missing>98%		Exclude
3.	Demographic	Region		Too varied	Exclude
4.	Demographic	Urban/rural type	<0.1% by categories		Exclude
5.	Lifestyle	Unsafe sex	<0.2%		Exclude
6.	Lifestyle	Anal sex	<0.2%		Exclude
7.	Lifestyle	Number of lifetime partners	<0.2%		Exclude
8.	Lifestyle	Partner characteristics	<0.2%		Exclude
9.	Lifestyle	Obesity	<0.2%		Exclude
10.	Clinical conditions and co-morbid	Cryptococcal meningitis		HIV +ve = 0	Exclude
11.	Clinical conditions and co-morbid	Transverse myelitis		HIV +ve = 0	Exclude
12.	Clinical conditions and co-morbid	Any unexplained retinopathy		HIV +ve = 0	Exclude
13.	Clinical conditions and co-morbid	Space-occupying lesion of unknown cause		HIV +ve = 0	Exclude
14.	Clinical conditions and co-morbid	Vaginal intraepithelial neoplasia		HIV +ve = 0	Exclude
15.	Clinical conditions and co-morbid	Mononucleosis-like syndrome		HIV +ve = 0	Exclude
16.	Clinical conditions and co-morbid	Lymphogranuloma venereum		HIV +ve = 0	Exclude
17.	Clinical conditions and co-morbid	Lymphoepithelial parotid cysts		HIV +ve = 0	Exclude
18.	Clinical conditions and co-morbid	Minor trauma		HIV +ve = 0	Exclude
19.	Clinical conditions and co-morbid	Aspergillosis	Exclude <0.2%		Exclude
20.	Clinical conditions and co-morbid	Dementia	Exclude <0.2%		Exclude
21.	Clinical conditions and co-morbid	Guillain–Barré syndrome	Exclude <0.2%		Exclude
22.	Clinical conditions and co-morbid	Cytomegalovirus retinitis	Exclude <0.2%		Exclude
23.	Clinical conditions and co-morbid	1. Cryptosporidiosis	Exclude <0.2%		Exclude
24.	Clinical conditions and co-morbid	Oral hairy leukoplakia	Exclude <0.2%		Exclude
25.	Clinical conditions and co-morbid	Castleman's disease	Exclude <0.2%		Exclude
26.	Clinical conditions and co-morbid	Head and neck cancer	Exclude <0.2%		Exclude
27.	Clinical conditions and co-morbid	Hodgkin lymphoma	Exclude <0.2%		Exclude
28.	Clinical conditions and co-morbid	Seminoma	Exclude <0.2%		Exclude
29.	Clinical conditions and co-morbid	Hepatitis B	Combine hepatitis		Exclude
30.	Clinical conditions and co-morbid	Hepatitis C	Combine hepatitis		Exclude
31.	Clinical conditions and co-morbid	Cervical dysplasia	Combine cervical		Exclude
32.	Clinical conditions and co-morbid	Cervical cancer	Combine cervical		Exclude
33.	Clinical conditions and co-morbid	Cervical intraepithelial neoplasia	Combine cervical		Exclude
34.	Clinical conditions and co-morbid	Chlamydia	Combine STIs		Exclude
35.	Clinical conditions and co-morbid	Gonorrhoea	Combine STIs		Exclude
36.	Clinical conditions and co-morbid	Genital herpes	Combine STIs		Exclude
37.	Clinical conditions and co-morbid	Other STIs (excludes above)	Combine STIs		Exclude
38.	Clinical conditions and co-morbid	Previous STI	Combine STIs		Exclude

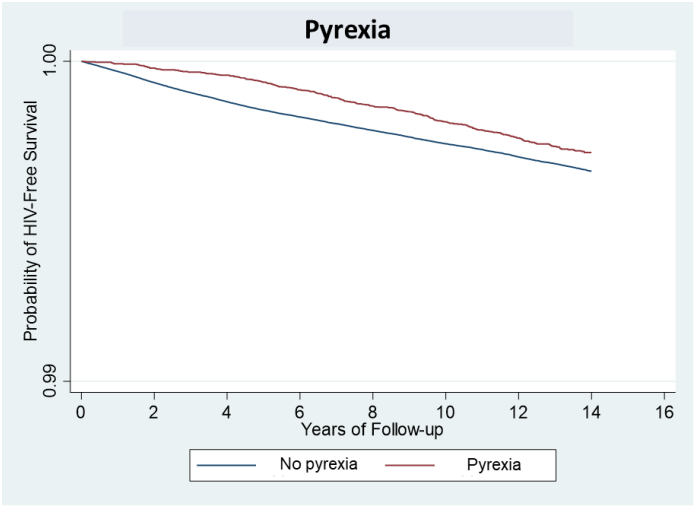
Appendix 12: Kaplan-Meier curves for abdominal pain, cough, diarrhoea, nausea/vomiting, pyrexia of unknown origin, rash and weight loss.

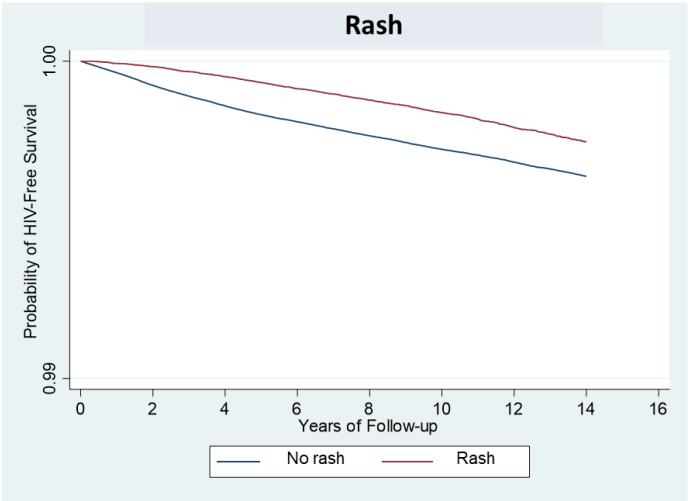


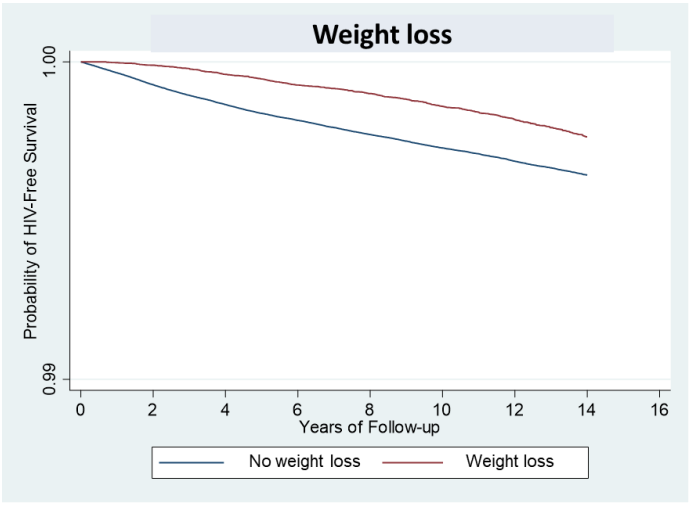







Pyrexia
Probability of HIV-Free Survival
1.00
0.99
0
2
4
6
8
10
12
14
16
Years of Follow-up
No pyrexia
Pyrexia


Rash
Probability of HIV-Free Survival
1.00
0.99
0
2
4
6
8
10
12
14
16
Years of Follow-up
No rash
Rash


Weight loss
Probability of HIV-Free Survival
1.00
0.99
0
2
4
6
8
10
12
14
16
Years of Follow-up
No weight loss
Weight loss

Appendix 13: Multivariable analysis of relationship between predictors and confirmed HIV, excluding outlier practices.

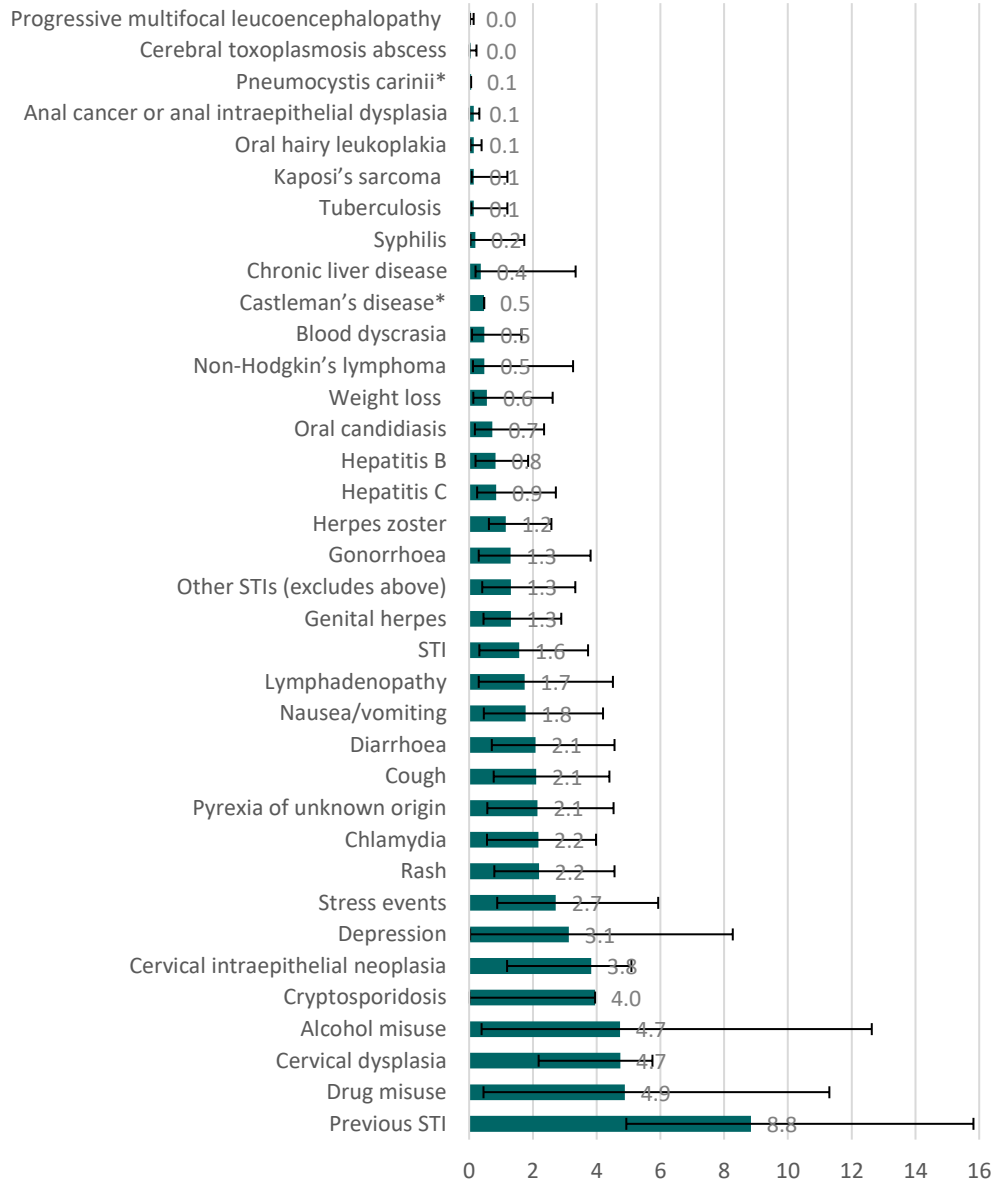
	Predictor	Hazard Ratio	[95% Confidence Interval]	p-value
Socio-Demographic and geographic	Gender			
	Male	1 (reference)		
	Female	0.7	(0.66, 0.74)	<0.001
	Age group			
	18-24	1 (reference)		
	25-34	1.62	(1.48, 1.77)	<0.001
	35-49	1.46	(1.33, 1.59)	<0.001
	50-59	0.85	(0.75, 0.96)	0.01
	60+	0.33	(0.28, 0.38)	<0.001
	Ethnicity			
	White/Asian/missing	1 (reference)		
	Black	11.39	(10.46, 12.4)	<0.001
	Mixed and other	2.76	(2.31, 3.28)	<0.001
	Deprivation			
	1	1 (reference)		
	3	1.24	(1.13, 1.36)	<0.001
	4	1.49	(1.36, 1.63)	<0.001
	5	2	(1.83, 2.19)	<0.001
	Not stated	1.64	(1.46, 1.83)	<0.001
	Urban/rural			
	Rural	1 (reference)		
	Urban	1.69	(1.48, 1.93)	<0.001
	Missing	1.74	(1.5, 2.02)	<0.001
Lifestyle	Drug misuse	2.91	(2.58, 3.27)	<0.001
Clinical and comorbid conditions	Kaposi's sarcoma	141.13	(72.75, 273.78)	<0.001
	Pneumocystis carinii	73.23	(10.16, 527.66)	<0.001
	Progressive multifocal leukoencephalopathy	65.95	(16.47, 264.16)	<0.001
	Syphilis	11.68	(7.3, 18.69)	<0.001
	Non-Hodgkin's lymphoma	9.53	(7.17, 12.67)	<0.001
	Cerebral toxoplasmosis abscess	7.6	(2.84, 20.31)	<0.001
	Anal cancer or anal intraepithelial dysplasia	6.47	(2.08, 20.07)	0.01
	Hepatitis	4.55	(3.8, 5.45)	<0.001
	Chronic liver disease	3.9	(3.39, 4.49)	<0.001
	Aseptic meningitis/encephalitis	3.78	(1.57, 9.1)	<0.001
	Oral candidiasis	2.42	(1.86, 3.13)	<0.001
	Tuberculosis	2.36	(1.18, 4.73)	0.02
	Blood dyscrasia	2.04	(1.57, 2.66)	<0.001
	STI (excludes syphilis) or previous STI	1.78	(1.58, 2.02)	<0.001
	Herpes zoster	1.63	(1.13, 2.33)	0.01
	Lymphadenopathy	1.55	(1.3, 1.85)	<0.001
	Pneumonia	1.5	(1.09, 2.06)	0.01
	Depression	1.44	(1.35, 1.54)	<0.001
	Rash	0.92	(0.85, 1)	0.04
	Hypertension	0.91	(0.83, 1.01)	0.06
	Weightloss	0.82	(0.74, 0.91)	<0.001
	Stressful events	0.82	(0.73, 0.93)	0.001
	Nausea/vomiting	0.78	(0.69, 0.89)	<0.001
	Peripheral neuropathy	0.54	(0.28, 1.04)	0.07
	Abdominal pain	0.53	(0.47, 0.6)	<0.001
	Cough	0.52	(0.48, 0.56)	<0.001
	Lung cancer	0.32	(0.1, 1.01)	0.05
	Neurologic disability	0.28	(0.09, 0.87)	0.03
	Hyperlipidemia	0.25	(0.17, 0.39)	<0.001

Appendix 14: Time between presentation of predictor to GP and confirmed HIV diagnosis

Predictor	Time prior to HIV diagnosis (years)
Pneumocystis carinii	0.06
Progressive multifocal leukoencephalopathy	0.08
Cerebral toxoplasmosis abscess	0.23
Anal cancer or anal intraepithelial dysplasia	0.44
Aseptic meningitis/encephalitis	0.62
Blood dyscrasia	0.95
Kaposi's sarcoma	1.08
Oral candidiasis	1.32
Syphilis	1.46
Tuberculosis	1.61
Herpes zoster	1.68
Non-Hodgkin's lymphoma	2.26
Lymphadenopathy	2.47
Chronic liver disease	4.33
Depression	7.42
Drug misuse	9.47

Survival analysis between the time when predictors occur and recording of confirmed HIV shows that most predictors were recorded in THIN database within a year before HIV status was recorded.

Duration (in years) between recording of predictor and HIV in THIN database: 2000 to 2017.



*Small numbers for HIV cases

Appendix 15: Sensitivity, specificity and predictive values at selected Cut-off points for different geographies

A. Sensitivity, specificity and predictive values at selected Cut-off points for England (excluding outlier practice).

	Cut-off points				
	0.075%	0.1%	0.25%	0.5%	1%
Sensitivity (%)	59%	51%	27%	18%	10%
Specificity (%)	56%	67%	93%	98%	99%
PPV (%)	0.08%	0.09%	0.22%	0.44%	0.48%
NPV (%)	99.96%	99.96%	99.95%	99.95%	99.95%
Total	6169640	6169640	6169640	6169640	6169640
With HIV	3536	3536	3536	3536	3536
Likelihood Ratio +	1.3	1.5	3.8	7.7	8.4
Likelihood Ratio -	0.7	0.7	0.8	0.8	0.9

B. Sensitivity, specificity and predictive values at selected Cut-off points for Northern Ireland (excluding outlier practice).

	Cut-off points				
	0.075%	0.1%	0.25%	0.5%	1%
Sensitivity (%)	65%	56%	25%	10%	5%
Specificity (%)	43%	52%	84%	96%	99%
PPV (%)	0.10%	0.10%	0.13%	0.21%	0.70%
NPV (%)	99.93%	99.93%	99.93%	99.92%	99.92%
Total	304564	304564	304564	304564	304564
With HIV	256	256	256	256	256
Likelihood Ratio +	1.1	1.2	1.6	2.5	8.4
Likelihood Ratio -	0.8	0.9	0.9	0.9	1.0

C. Sensitivity, specificity and predictive values at selected Cut-off points for Scotland (excluding outlier practices).

	Cut-off points				
	0.075%	0.1%	0.25%	0.5%	1%
Sensitivity (%)	70%	62%	36%	27%	12%
Specificity (%)	53%	66%	94%	97%	99%
PPV (%)	0.08%	0.10%	0.31%	0.50%	0.48%
NPV (%)	99.97%	99.97%	99.96%	99.96%	99.95%
Total	1235093	1235093	1235093	1235093	1235093
With HIV	650	650	650	650	650
Likelihood Ratio +	1.5	1.8	6.0	9.6	9.2
Likelihood Ratio -	0.6	0.6	0.7	0.8	0.9

D. Sensitivity, specificity and predictive values at selected Cut-off points for Wales.

	Cut-off points				
	0.075%	0.1%	0.25%	0.5%	1%
Sensitivity (%)	49%	39%	16%	8%	3%
Specificity (%)	72%	80%	97%	99%	100%
PPV (%)	0.06%	0.07%	0.18%	0.22%	0.23%
NPV (%)	99.98%	99.97%	99.97%	99.97%	99.97%
Total	984358	984358	984358	984358	984358
With HIV	336	336	336	336	336
Likelihood Ratio +	1.7	1.9	5.2	6.3	6.7
Likelihood Ratio -	0.7	0.8	0.9	0.9	1.0

E. Sensitivity, specificity and predictive values at selected Cut-off points for the 4 Outlier practices.

	Cut-off points				
	0.075%	0.1%	0.25%	0.5%	1%
Sensitivity (%)	100%	99%	98%	96%	95%
Specificity (%)	13%	46%	55%	61%	70%
PPV (%)	3.02%	4.72%	5.60%	6.26%	7.81%
NPV (%)	99.92%	99.92%	99.89%	99.83%	99.79%
Total	50963	50963	50963	50963	50963
With HIV	1354	1354	1354	1354	1354
Likelihood Ratio +	1.1	1.8	2.2	2.4	3.1
Likelihood Ratio -	0	0	0	0.1	0.1

Appendix 16: Screening criteria

Appendix 16-A: Screening stage 1 criteria

Screen using the title/abstracts and find out if studies fit the eligibility criteria

Eligibility criteria

1. The population of the study is healthcare providers in primary care/ General Practice (GP)

2. The phenomenon of interest for the study should be:

point-of-care alert or clinical prompt or clinician alert or pop-up alert or electronic decision-support tools

3. The setting should be:

- UK
- Europe (European Union and European Free Trade Association nations)
- North America (USA and Canada)
- Australia
- New Zealand

4. Published: 2000 onwards

5. Studies: Qualitative studies, mixed method, systematic review of qualitative research, surveys with analysis of free text.

Exclusion criteria

1. Non-human studies.

2. Setting is developing countries only

3. Quantitative studies (trials, cohort, case control and sample surveys) only.

Appendix 16-B: Screening stage 2 criteria

Screen using the whole article and find out if studies fit the eligibility criteria

Eligibility criteria

1. The population of the study is healthcare providers in primary care/ General Practice (GP)

2. The phenomenon of interest for the study should be:

point-of-care alert or clinical prompt or clinician alert or pop-up alert or electronic decision-support tools (as alert)

3. The setting should be:

- UK
- Europe (European Union and European Free Trade Association nations)
- North America (USA and Canada)
- Australia
- New Zealand

4. Published: 2000 onwards

5. Studies: Qualitative studies, mixed method, systematic review of qualitative research, surveys with analysis of free text.

Exclusion criteria

1. Non-human studies.
2. Alerts or tools that aid in selection of medication.
3. Setting is developing countries only.
4. Quantitative studies (trials, cohort, case control and sample surveys) only.

Appendix 17: Quality Assessment on barriers and facilitators of use of pop-up alerts.

(Adopted from **CASP Checklist: 10 questions to help you make sense of a Qualitative research)**

	Paper for appraisal and reference:		
Section A: Are the results valid?	1. Was there a clear statement of the aims of the research?	HINT: Consider • what was the goal of the research • why it was thought important • its relevance	Yes/Can't Tell/No Comments:
	2. Is a qualitative methodology appropriate?	HINT: Consider • If the research seeks to interpret or illuminate the actions and/or subjective experiences of research participants • Is qualitative research the right methodology for addressing the research goal	Yes/Can't Tell/No Comments: Is it worth continuing?
	3. Was the research design appropriate to address the aims of the research?	HINT: Consider • if the researcher has justified the research design (e.g. have they discussed how they decided which method to use)	Yes/Can't Tell/No Comments:
	4. Was the recruitment strategy appropriate to the aims of the research?	HINT: Consider • If the researcher has explained how the participants were selected • If they explained why the participants they selected were the most appropriate to provide access to the type of knowledge sought by the study • If there are any discussions around recruitment (e.g. why some people chose not to take part)	Yes/Can't Tell/No Comments:
	5. Was the data collected in a way that addressed the research issue?	HINT: Consider • If the setting for the data collection was justified • If it is clear how data were collected (e.g. focus group, semi-structured interview etc.) • If the researcher has justified the methods chosen • If the researcher has made the methods explicit (e.g. for interview method, is there an indication of how interviews are conducted, or did they use a topic guide) • If methods were modified during the study. If so, has the researcher explained how and why • If the form of data is clear (e.g. tape recordings, video material, notes etc.) • If the researcher has discussed saturation of data	Yes/Can't Tell/No Comments:
	6. Has the relationship between researcher and participants been adequately considered?	HINT: Consider • If the researcher critically examined their own role, potential bias and influence during (a) formulation of the research questions (b) data collection, including sample recruitment and choice of location	Yes/Can't Tell/No Comments:

		• How the researcher responded to events during the study and whether they considered the implications of any changes in the research design	
Section B: What are the results?	7. Have ethical issues been taken into consideration?	HINT: Consider • If there are sufficient details of how the research was explained to participants for the reader to assess whether ethical standards were maintained • If the researcher has discussed issues raised by the study (e.g. issues around informed consent or confidentiality or how they have handled the effects of the study on the participants during and after the study) • If approval has been sought from the ethics committee	Yes/Can't Tell/No Comments:
	8. Was the data analysis sufficiently rigorous?	HINT: Consider • If there is an in-depth description of the analysis process • If thematic analysis is used. If so, is it clear how the categories/themes were derived from the data • Whether the researcher explains how the data presented were selected from the original sample to demonstrate the analysis process • If sufficient data are presented to support the findings • To what extent contradictory data are taken into account • Whether the researcher critically examined their own role, potential bias and influence during analysis and selection of data for presentation	Yes/Can't Tell/No Comments:
	9. Is there a clear statement of findings?	HINT: Consider whether • If the findings are explicit • If there is adequate discussion of the evidence both for and against the researcher's arguments • If the researcher has discussed the credibility of their findings (e.g. triangulation, respondent validation, more than one analyst) • If the findings are discussed in relation to the original research question	Yes/Can't Tell/No Comments:
Section C: Will the results help locally?	10. How valuable is the research?	HINT: Consider • If the researcher discusses the contribution the study makes to existing knowledge or understanding (e.g. do they consider the findings in relation to current practice or policy, or relevant research based literature • If they identify new areas where research is necessary • If the researchers have discussed whether or how the findings can be transferred to other populations or considered other ways the research may be used	Comments: