EXAMINING THE EPIDEMIOLOGY OF TUBERCULOSIS IN MIGRANTS TO THE UK TO INFORM EVIDENCE-BASED SCREENING POLICIES.

DR ROBERT WILLIAM ALDRIDGE

UNIVERSITY COLLEGE LONDON

RESEARCH DEPARTMENT OF INFECTION AND POPULATION HEALTH

PHD THESIS

I, Robert Aldridge, confirm that the work presented in this thesis is my own. Where information has been derived from other sources I confirm that this has been indicated in the thesis.

Signed _____

Abstract

Background: In high-income countries an increasing proportion of all tuberculosis cases are detected in migrants. Understanding the epidemiology of tuberculosis in migrants to inform evidence-based screening policies is a priority.

Methods: A systematic review and meta-analysis of pre-entry screening for tuberculosis was undertaken (chapter 2). Data from a pilot pre-entry programme in migrants to the UK was described, and risk factors for prevalent cases examined (chapter 3). The accuracy of a novel method for identifying individuals between two datasets was studied (chapter 4). This linkage method was used to combine data from migrants screened pre-entry to the UK tuberculosis register including molecular strain typing data. The linked datasets enabled estimates of the incidence of tuberculosis to be calculated, and risk factors were identified (chapters 5 and 6).

Results: The systematic review identified 15 studies and found that culture confirmation increased with WHO prevalence in the country of origin. The crude prevalence of bacteriologically confirmed tuberculosis identified by UK pre-entry screening was 92 per 100,000 population screened. Migrants reporting a history of contact with a case of tuberculosis, and those from higher prevalence countries were at greatest risk. Compared to a gold standard of NHS number, probabilistic linkage identified individuals in two datasets with high sensitivity and specificity. The estimated incidence of tuberculosis notified in the UK in migrants screened pre-entry was 194 per 100,000 person years at risk. Migrants with a chest radiograph classified as suspected tuberculosis and those from higher prevalence countries had a higher risk post-migration. Compared to other non-UK born individuals, migrants screened pre-entry were less likely to be the first case in a cluster of tuberculosis.

Conclusions: This thesis generated new knowledge that improves our understanding of the epidemiology of tuberculosis in migrants to the UK. Based on these findings, evidence-based screening recommendations were made.

List of Contents

List of Boxes

List of Images

List of Figures

List of Tables

Abbreviations

AFB	Acid-fast Bacilli
CDC	Centers for Disease Control and Prevention, USA
CI	Confidence Interval
CXR	Chest Radiograph
EMBASE	Excerpta Medica dataBASE
EMS	Enhanced Matching System
ETS	Enhanced Tuberculosis Surveillance
FCO	Foreign and Commonwealth Office
HIV	Human Immunodeficiency Virus
IOM	International Organization for Migration
NHS	National Health Service
LILACS	Latin American and Caribbean Health Sciences Literature
PACS	Picture Archiving and Communications System
PHE	Public Health England
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses
SQL	Structured Query Language
TB	Tuberculosis
TST	Tuberculin Skin Test
UKBA	UK Border Agency
WHO	World Health Organization

Acknowledgements

I would firstly like to thank my supervisors Professor Andrew Hayward, Professor Ibrahim Abubakar and Dr Peter White for their patience, time, support, encouragement and sense of humour. I would also like to thank Dr Dominik Zenner and Maeve Lalor at Public Health England, as well as the rest of the tuberculosis section, for their expert insights into the surveillance data. None of this work would have been possible without the migrants taking part in the pre-entry screening programme and the high quality data provided by the epidemiology unit at International Organization for Migration. In particular I would like to thank Dr Poonan Dhavan at IOM for her support and help in accessing and interpreting their data, as well as being such an excellent host during my visit to Manila in 2014. Elizabeth Williamson provided invaluable support and advice for the imputation work carried out in Chapters 5 & 6 and was on hand when the code crashed or just didn't run – thank you very much! Dr Tom Yates has been a supportive friend throughout the PhD, and along with Dominik Zenner, performed the rather tedious task of double screening abstracts for the systematic review - thank you both. Finally, and most importantly, I am always grateful to Helen and Hazel for providing a sense of perspective on things as well as their love, support and humour.

CHAPTER 1

Tuberculosis, migration and UK immigration screening policy: an introduction

1.1 Abstract

Tuberculosis is a major global cause of morbidity and mortality. In 2013, tuberculosis was estimated to have resulted in 1.3 million deaths, and there were 9 million cases of tuberculosis across the globe in 2014. Tuberculosis is therefore a major population health challenge requiring a coordinated international effort, particularly given the increasing movement of people across countries and regions. For over a century, immigration medical screening has targeted people moving from high to low tuberculosis burden countries, and this screening has taken several forms. Investigating the epidemiology of tuberculosis in screened migrants would enable a better understanding of those at greatest risk, and would facilitate the development of evidence-based medical screening interventions to improve the health of this population.

1.2 Natural history, detection and treatment of tuberculosis

Mycobacterium tuberculosis was first described as the cause of tuberculosis by the German physician and scientist Robert Koch in 1882.(1) At the time of this discovery, tuberculosis was a major cause of morbidity and mortality, with one in seven people dying from the disease - a statistic Koch highlighted when he reported his discovery for the first time.(2) When Koch identified the aetiological agent of tuberculosis (TB), little was known about its natural history.(3) Understanding the natural history of tuberculosis is important, particularly when aiming to improve evidence-based screening in migrant populations, as it helps illustrate when and where screening

interventions could be targeted. This chapter begins with a brief overview of the natural history of tuberculosis and focuses on the distinction between active disease and latent infection, as understanding these two entities clarifies the potential opportunities for public health interventions aimed at ultimately reducing the burden of tuberculosis in migrants.

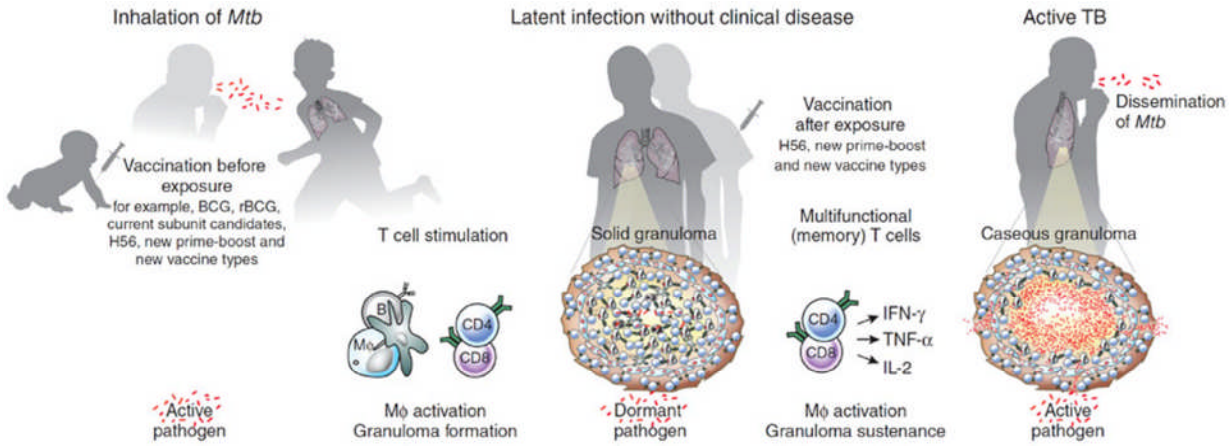

Image 1. Natural history of M.tuberculosis: from latent infection to active disease.

Source: Tuberculosis vaccines - a new kid on the block. Kaufmann.(4) Reproduced with permission Copyright © 2014 Nature Publishing Group.

Latent infection and active disease are both caused by *M.tuberculosis,* which is spread primarily through air droplets, although transmission can occur through ingestion, for example, by drinking unpasteurised milk. The immune system response to infection includes the involvement of T lymphocytes which cause the formation of granulomas, where this initial infection is kept under control by macrophages (Image 1).(4) An individual that has been infected with *M.tuberculosis* has no clinical symptoms during this period of latent infection, and they are not infectious to others as they do not produce the air droplets containing *M.tuberculosis*.

1.2.1 Latent tuberculosis infection

The WHO defines latent tuberculosis infection as: "a state of persistent immune response to stimulation by *M.tuberculosis* antigens without evidence of clinically manifested active tuberculosis."(5) The US Centers for Disease Control has outlined a series of criteria that help define the characteristics of a person with latent infection (Box 1).

Box 1. CDC defined characteristics of a person with latent tuberculosis infection.(6)

A person with latent tuberculosis infection:

- Usually has a skin test or blood test result indicating TB infection
- Has a normal chest x-ray and a negative sputum test
- Has TB bacteria in his/her body that are alive, but inactive
- Does not feel sick
- Cannot spread TB bacteria to others
- Needs treatment for latent TB infection to prevent TB disease; however, if exposed and infected by a person with multidrug-resistant TB (MDR TB) or extensively drug-resistant TB (XDR TB), preventive treatment may not be an option

Whilst the CDC criteria for latent tuberculosis set out in Box 1 indicate that an individual with latent tuberculosis "has a normal chest x-ray and negative sputum", the UK tuberculosis technical instructions for classifying chest radiographs, which is based on the Canadian immigration system, has a category for minor findings on chest radiographs that are occasionally associated with tuberculosis infection.(7,8)

There are several tests for detecting latent infection including the Tuberculin Skin Test (TST), and Interferon Gamma Release Assays (IGRAs) blood tests.(9) IGRA tests require only a single contact with the patient compared to TST, which needs two visits to a health worker - one to carry out the test and another between 48 and 72 hours later to read it.(10) IGRA tests have comparable sensitivity and specificity to TST, are unaffected by BCG vaccination (unlike TST), may have less false positives in people exposed to environmental mycobacteria, but are more costly and the predictive validity for the development of active tuberculosis (see Box 2) in migrants is currently unknown.(11,12)

Based on a systematic review of the scientific literature published before the availability of anti-tuberculosis therapy (and therefore also HIV), the mean time between onset of tuberculosis disease and cure or death was three years. It is estimated that individuals

with latent infection who do not undergo treatment have a life time risk of progressing to active tuberculosis of approximately 10%, and whilst this often quoted historical figure has a high degree of uncertainty, those co-infected with HIV have a 30% lifetime risk which is almost certainly higher.(13,14) The progression from latent infection to active tuberculosis disease occurs when the immune system is no longer able to control the infection, therefore explaining the increased rates of progression in those co-infected with HIV, which impairs the ability of the immune system.(13,15,16) Tuberculosis disease most commonly affects the lung, but it can cause disease in any organ and can be exclusively extra-pulmonary i.e. have no lung involvement at the time of clinical presentation.(17)

Effective treatments for latent tuberculosis infection exist for both HIV positive and negative individuals. Systematic reviews estimate that treatment for latent infection can reduce the risk of progression to active disease in HIV positive individuals with a positive tuberculin skin test (relative risk in treated vs. untreated patients 0.38; 95% CIs: 0.25, 0.57), and in HIV negative individuals (relative risk in treated vs. untreated patients 0.40; 95% CIs: 0.31, 0.52).(14,18)

Treatment for latent infection has side effects in some individuals, including serious events such as liver failure.(19) Therefore not everyone identified with latent infection will take up the offer of treatment based on personal preferences and values around risk versus benefit of treatment. There are other reasons why not everyone with latent infection is treated, including the fact that as latent infection is an asymptomatic state, it is only through screening that eligible patients are identified and offered treatment, and such screening programmes are not universally available in most countries. Despite these issues, treatment for tuberculosis infection is widely available in many high-income countries, and if cases of infection are correctly identified and appropriately treated, there is a potential to reduce the burden of morbidity and mortality from tuberculosis.(20)

In 1985 Karel Styblo, a medical advisor to the Royal Netherlands Tuberculosis Association, described the relationship between tuberculosis infection and the risk of developing infectious tuberculosis (i.e. disease that can spread to others).(21) Styblo's relationship can be summarised as follows: two people will develop active tuberculosis disease for every 20 cases of latent infection, and one of these cases of active disease

will become an infectious case capable of spreading the disease to others. This "rule of thumb" was based upon tuberculosis surveillance data from the Netherlands and USA, as well as prevalence surveys from 12 low-income countries. This rule does not necessarily apply in all contexts(22), and can be modified with additional data, but it illustrates the relationship between infection, disease and transmission, and the potential opportunities at each stage and associated benefits with regard to prevention of future cases.

1.2.2 Active tuberculosis disease

An individual infected with *M.tuberculosis* who develops active disease with symptoms such as fever, cough and weight loss has the ability to transmit tuberculosis to others – the onset of these symptoms are characteristic of the development of active disease (Box 2).

Box 2. CDC defined characteristics of a person with active tuberculosis disease.(6)

A person with TB disease:

- Usually has a skin test or blood test result indicating TB infection
- May have an abnormal chest x-ray, or positive sputum smear or culture
- Has active TB bacteria in his/her body
- Usually feels sick and may have symptoms such as coughing, fever, and weight loss
- May spread TB bacteria to others
- Needs treatment to treat TB disease

Tuberculosis disease can be detected in a variety of ways. Guidelines by the National Institute for Health and Care Excellence (NICE) suggest that individuals with suspected tuberculosis of the lungs should first have a chest radiograph, and if this has any of the characteristic appearances of tuberculosis lesions then multiple sputum samples should be taken.(23) These multiple sputum samples can be tested in a variety of ways, but historically two forms of testing have been performed: sputum smear microscopy; and mycobacterial culture testing.(17) Smear microscopy formed the basis of tuberculosis

disease detection for many years because, compared to culture methods, it required less complex equipment, training of technicians was simpler, it correlated with the infectiousness of a case, and results were available more rapidly.(24) Despite these advantages, smear microscopy has several limitations including the rates of false positives, false negatives, inter-observer reliability and problems with reproducibility in the preparation of sputum samples.(24,25)

The probability of a positive smear microscopy result is proportional to how many bacilli are present in a sputum sample produced by an individual with disease. When concentrations of the organism in a sample from an individual with *M.tuberculosis* disease are below 1,000 per ml, the probability of a positive test becomes less than 10%.(24) This is in contrast to a sputum tested by culture, that can detect smaller numbers of bacilli, with a limit of around 100 organisms per ml.(24) Smear microscopy also has higher false negative rates for detecting tuberculosis disease in HIV positive individuals.(25,26) Individuals with culture positive, but smear negative disease are less likely to be infectious, as well as more likely to have a positive outcome without treatment compared to smear positive cases.(27,28) For these reasons, NICE recommends that microscopy and culture are both used: "multiple sputum samples (at least three, with one early morning sample) should be sent for tuberculosis microscopy and culture for suspected respiratory tuberculosis before starting treatment if possible or, failing that, within 7 days of starting".(23) Once a case of active tuberculosis is detected, the World Health Organisation (WHO) and NICE both recommend a standard regimen of six months of treatment of active respiratory disease using four initial drugs (6 months of isoniazid and rifampicin supplemented in the first 2 months with pyrazinamide and ethambutol).(23,29) More complicated treatment recommendations, such as those for drug resistant cases of tuberculosis, or disease that affects the bones or central nervous system, are not discussed in this chapter as these are rarely detected at pre-entry screening.(23)

1.3 Global tuberculosis epidemiology

Globally, there were an estimated 9 million cases of tuberculosis in 2014(30) and in 2010 it ranked as the 13th largest cause of disability-adjusted life years.(31) The largest burden of disease occurs in the WHO regions of South-East Asia and Western Pacific,

which account for over half of all cases. Although the largest number of cases occurs in these regions, prevalence rates are highest in the African region. The WHO African region has approximately a quarter of all cases globally, and has seen a slower decline in prevalence rates compared to most other regions since 1990 (Figure 1). Mortality rates in the African region are also higher than others, and it has the highest death rates (Figure 2). With the exception of Europe (which saw an increase in rates through the 1990s followed by a decline), there has been a gradual decline since 1990 in the prevalence of tuberculosis in all WHO regions. There is disagreement about the actual rates of change, particularly when comparing different methodologies and rates for adults and children, but general agreement in the overall direction of change towards a decline. (30,32–34)

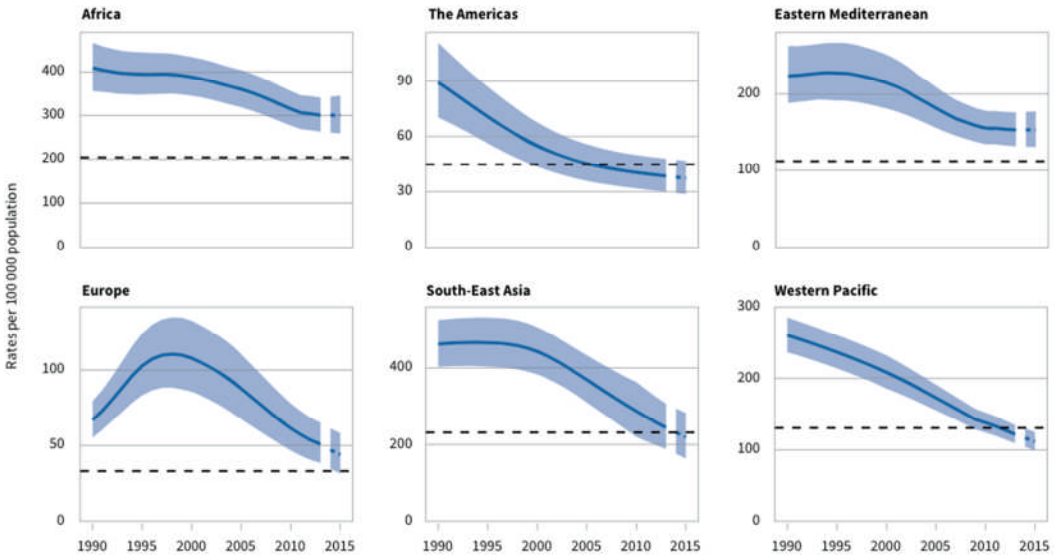

Figure 1. Trends in estimated TB prevalence rates 1990–2013 and forecast TB prevalence rates 2014–2015, by WHO region.

Source: World Health Organisation. Global tuberculosis report 2014.**(30)** Reproduced in accordance with WHO guidelines.

Figure 2. Trends in estimated TB mortality rates 1990–2013 and forecast TB mortality rates 2014–2015, by WHO region.

Source: World Health Organisation. Global tuberculosis report 2014.**(30)** Reproduced in accordance with WHO guidelines.

The total number of cases is higher in men, who account for approximately 60% of all reported disease internationally.(30) Incidence rates in HIV-negative men and women peak in the 25-29 year age groups and gradually decline after this point, with men having a higher incidence compared to women in all age groups except those aged 5-19 years (Figure 3).(32) The total number of deaths is higher in men than women in all age groups except in those aged 5-19. Total number of deaths, as presented in Figure 3, should be considered in the context of the population age structure, as age-specific rates increase with age up to 70 years in men.(32) Despite increasing death rates with age, 83% of cases and 59% of all global deaths were in individuals younger than 60 years in 2013 as a result of the young age-structure in countries with substantial burden of tuberculosis in individuals who are HIV-negative.

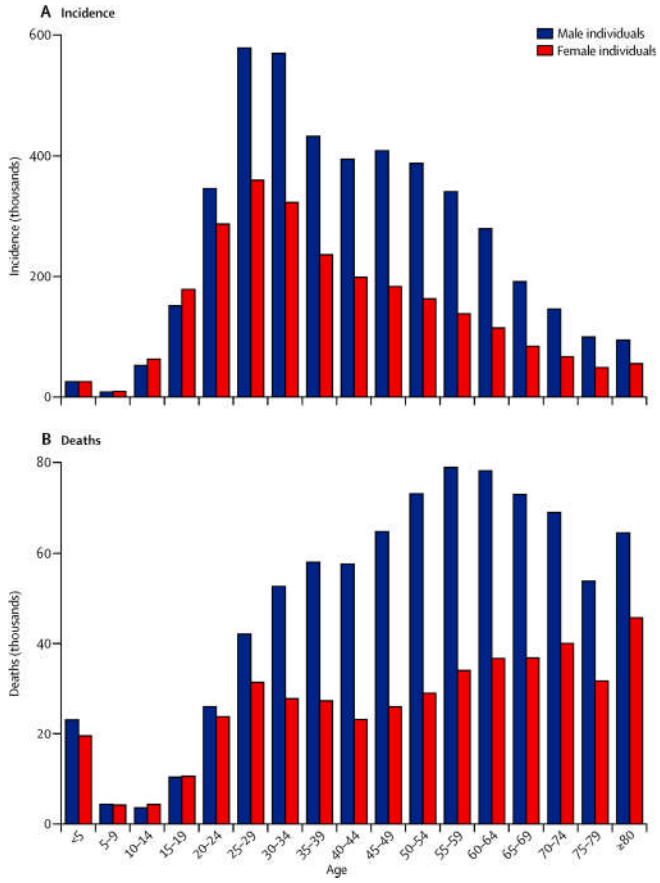

Figure 3. Global age-sex distribution of tuberculosis incidence (A) and deaths (B) in HIV-negative individuals in 2013.

Source: Murray CJL, Ortblad KF, Guinovart C, Lim SS, Wolock TM, Roberts DA, et al. Global, regional, and national incidence and mortality for HIV, tuberculosis, and malaria during 1990–2013: a systematic analysis for the Global Burden of Disease Study 2013. The Lancet. 2014 Sep;384(9947):1005–70.**(32)**.(4) Reproduced with permission Copyright © 2014 Elsevier Ltd.

1.4 Tuberculosis epidemiology in the UK

In the UK, a total of 7,892 tuberculosis cases were notified during 2013, representing a slight decline of the annual number of cases compared to 2012, and a reversal of the increasing number of notifications that had generally been going up since 1987 and peaked in 2011.(35) During this period of time, the rate of tuberculosis in UK born individuals has remained constant at around 4.1 per 100,000 population (Figure 4). The total number of cases in non-UK born individuals was 5,529 (70%) in 2013, which was

more than double the 2,103 (27%) cases in the UK born population.[a] The rates of tuberculosis were nearly 18 times higher in the non-UK born population at 70 per 100,000 population and approximately 80% of all cases in individuals born outside the UK had lived in the country for two or more years. The non-UK born rate has been declining annually from a peak of 98 per 100,000 in 2005.(35)

Men accounted for over half of all cases in 2013 (58%, 4,560/7,892). In the UK born, the highest rates of tuberculosis are found in those aged over 80, which is likely to reflect historical exposure to tuberculosis and waning immunity in the elderly.(36–39) In the non-UK born population, the highest number of cases is found in those aged 30-34, but the highest rates are in those aged 20-29.(35)

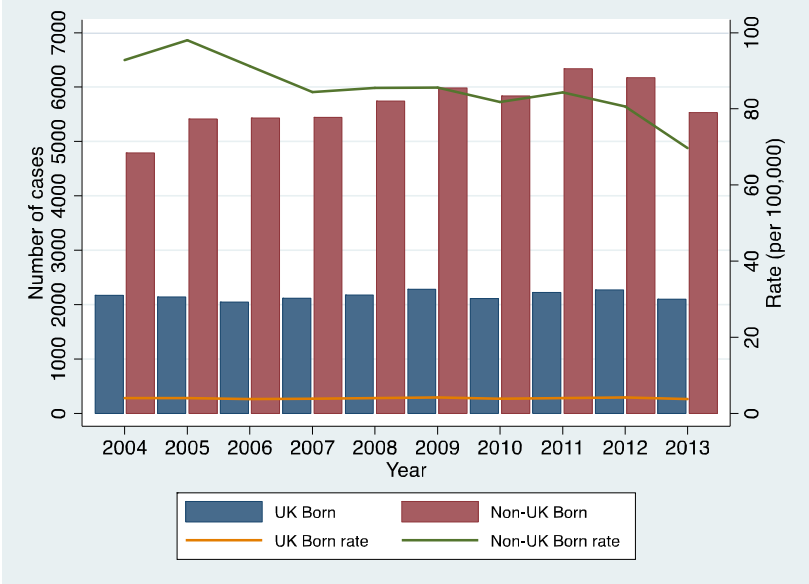

Figure 4. Tuberculosis case reports and rates by place of birth, UK, 2000-2013.

Data source: Public Health England. Tuberculosis in the UK: annual report - 2014.**(35)**

The burden of tuberculosis in individuals not born in the UK is therefore substantial and despite encouraging declines in both the total number of cases and rates of disease, further work is required to better understand who is at highest risk and the potential opportunities for health improvement interventions if we are to fully tackle and eliminate tuberculosis as a public health problem in the UK. This is particularly important as England has some of the highest rates of disease in Europe, particularly in

[a] percentages do not add up to 100 because of missing data on country of birth.

urban areas like London and Birmingham which have a larger number of cases and much higher rates than all other big European cities.(40,41)

Tuberculosis cases in England, Wales and Northern Ireland are notified to the Enhanced Tuberculosis Surveillance system (ETS). In Scotland, cases are notified to the Enhanced Surveillance of Mycobacterial Infections (ESMI), and data from ETS and ESMI are combined for the purpose of UK reporting. Using case notification data in their current configuration has several important limitations for the understanding of the burden of disease in migrant populations. Firstly, it has only been possible to estimate the incidence of active tuberculosis in non-UK born populations based on country of birth and not country of migration, which is more likely to represent risk of disease, particularly where these two countries may have very different population levels of tuberculosis. Secondly, there is a lack of reliable population denominator data for the UK based on country of birth, and therefore these estimates are generally presented by ethnicity, which has limitations for informing public health practice, as these groups are extremely heterogeneous. Thirdly, incidence estimates in non-UK born individuals includes prevalent cases in migrants who arrive in the UK with active disease – these are therefore not truly incident cases. Fourthly, limited risk factor data on non-UK born individuals are available, and very little of these data relate to migrant history prior to arrival in the UK, which could be extremely informative for improving our understanding of the disease in this group. Finally, the follow-up of migrants for epidemiological and public health purposes after entering the UK has not previously been possible as no dataset existed to facilitate this process. Therefore there is an urgent need for an improved dataset that would overcome these limitations in existing surveillance data.

1.5 Migration and tuberculosis

In recent decades, migration patterns have led to a change in the epidemiological profile of tuberculosis in many low-incidence countries.(42) In the United States, the total number of tuberculosis cases has been decreasing, but notifications in foreign born individuals are 11.5 times higher than those born in the country.(43) In Europe, the overall proportion of tuberculosis cases in individuals of foreign origin is 25.8%, but

many countries have much higher proportions, such as Sweden (89.4%), Norway (87.8%), and United Kingdom (70%).(44,45)

Internationally, the number of people residing outside their country of birth is considerable. The United Nations Population Division estimated that globally this population consisted of a total of 232 million people in 2013.(46,47) Between 1990 and 2013, North America accepted the largest gross inflow of migrants at 25 million, and Europe had the second largest at 23 million. Every year a substantial number of migrants move from high-incidence countries to those with a low-incidence. Reasons for international migration include economics (to work in the receiving country or move away from financial crises in the country of origin), education, political instability or war, natural disasters, and reunion (joining family members in the receiving country).(48,49)

In the UK, migration is a topic of great political and public interest, and there is a disconnect between public perception of migration numbers and the actual number of individuals arriving and leaving each year.(50–52) The Office for National Statistics (ONS) publishes regular statistics on net migration, which is defined as "the difference between the number of long-term immigrants coming to the UK and the number of long-term emigrants leaving the UK".(50) Between 1980 and 1997, net migration was fairly stable, varying between a reduction of 79,000 and an increase of 76,000 per annum (Figure 5). After 1997 net migration increased annually to a peak of 273,000 in 2007 and has seen a gradual decline since, although it remains substantially higher than it was between 1980 and 1997.

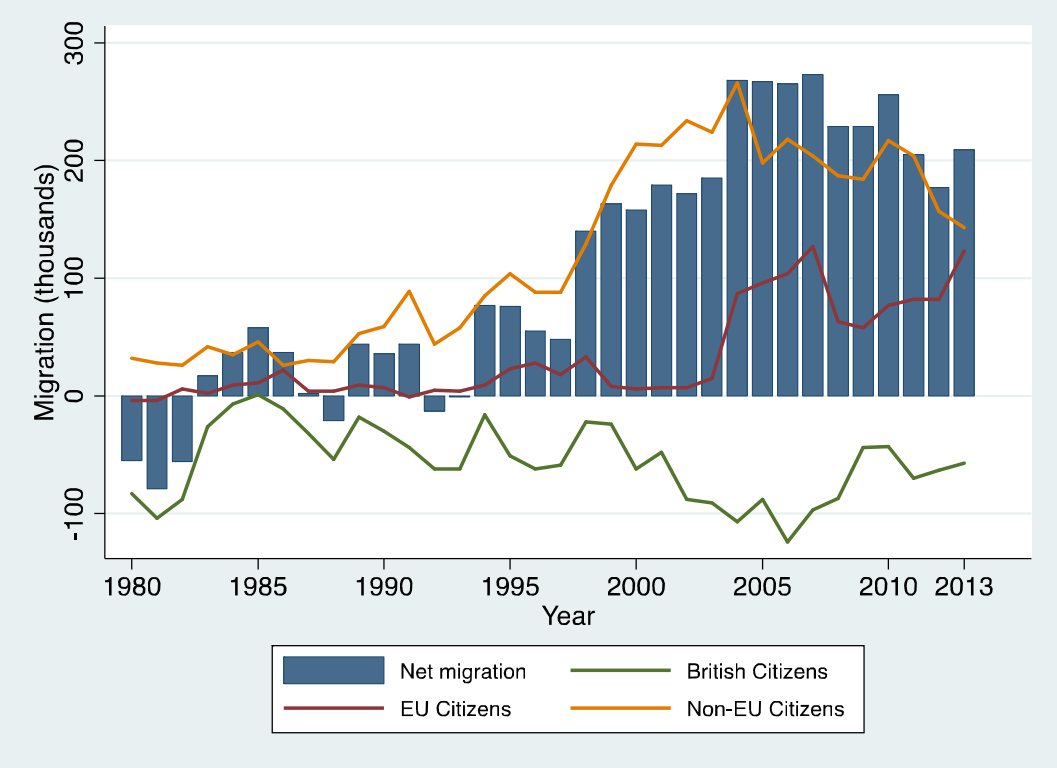

Figure 5. Long-term international net migration the UK and by citizenship, 1980 to 2013.

Data source: Office for National Statistics. Migration Statistics Quarterly Report.**(50)**

Net migration of Non-EU citizens has also been higher than EU citizens since 1980, and net migration of British citizens has been negative during this whole period (i.e. more British citizens have been leaving than returning; Figure 5). Consistent with international data on the reasons for migration, the two main reasons for immigration to the UK were for work or formal study, with a smaller number of individuals accompanying or joining family members.(50,53)

1.6 Medical screening of international migrants

As described in previous sections, globally and in the UK there are high levels of migration, and in countries with low tuberculosis incidence, there has been a changing epidemiological situation with an increasing proportion of cases detected in migrants. Medical screening of migrants for tuberculosis has been a focus of immigration health for over a century(54), but this recent change in the epidemiological profile of cases has

meant that there has been increased political and public scrutiny of such programmes. As a result, many low incidence countries have been re-examining their policies and interventions to tackle the burden of tuberculosis in migrant populations.

Tuberculosis screening in migrants can occur at three points in time: 1) prior to entering the country (pre-entry screening); 2) at the point of entry (upon entry or port of entry); 3) or post-arrival. Many European countries have implemented post-arrival screening programmes.(55,56) Whilst there are differences in the screening approach implemented, the characteristics of such post-arrival programmes are well documented.(55) Canada, USA, Australia, New Zealand and Israel all have pre-entry screening programmes for tuberculosis. Before the work undertaken in this thesis, no systematic review of the published literature on existing pre-entry programmes had been performed. The lack of such a review limited the ability to draw conclusions about the international effectiveness of pre-entry screening, and the potential strengths and weaknesses of the different approaches taken by each country.

The UK historically used a combination of upon- and post-entry screening.(57,58) In 2005 the UK Border Agency (UKBA) and Foreign and Commonwealth Office (FCO) funded the set up costs for a trial of pre-entry screening for active tuberculosis disease in migrants to the UK.(59) The pilot programme was run in partnership with the International Organization for Migration (IOM), an inter-governmental organisation that works with migrants, governments and the international community.(60) The pilot was launched in 2005 by IOM in Kenya and was rolled out to seven other locations screening migrants from a total of 15 countries. Local IOM clinics ran the programme with oversight from an epidemiological unit based in Manila, Philippines. The UK technical instructions describe the pre-entry screening programme in detail and involve a combination of symptom screening, clinical examination, chest radiography and sputum sampling (reviewed in detail in Chapter 3).(7) In May 2012 it was announced that the UK would close its existing upon entry screening programme for migrants and move fully to a pre-entry system, expanding it from the 15 pilot locations to 101 countries with a WHO estimated incidence of greater than 40 per 100,000 population. This transition from upon to pre-entry screening was conducted in four phases and was completed on 31st March 2014 (Figure 6).

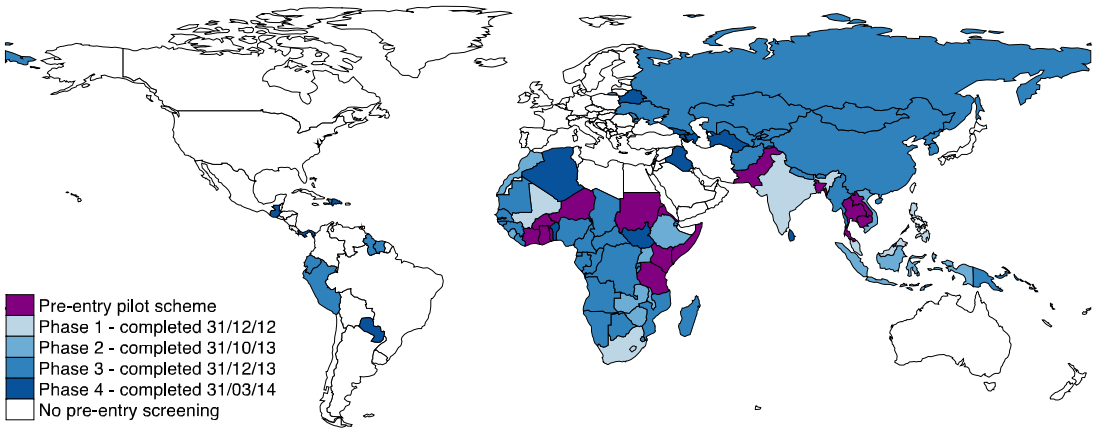

Figure 6. Location of pre-entry screening sites globally (includes IOM and non-IOM sites).

Note: Pre-entry pilot scheme countries were Bangladesh, Burkina Faso, Cambodia, Cote D'Ivoire, Eritrea, Ghana, Kenya, Laos, Niger, Pakistan, Somalia, Sudan, Tanzania, Thailand, and Togo.

In addition to pre-entry screening for active disease, NICE guidelines recommend post-migration screening for latent tuberculosis infection in individuals aged 16–35 years who enter the UK from high-incidence countries.(23) At present, the implementation levels of these post-entry screening recommendations for latent tuberculosis vary across the country and are dependent on local policies and funding, and there is evidence that levels of screening do not reflect the local burden of disease.(61–63) There is also a great deal of variation internationally in the application of such screening programmes, including a lack of evidence regarding the cost effectiveness of the different approaches.(56) A national collaborative tuberculosis strategy formulated by Public Health England in 2015 called for new entrant latent tuberculosis screening to be systematically implemented across England.(64)

1.6.1 Informing evidence-based screening programme for migrants to the UK.

As described in previous sections, the benefits of screening for latent infection and active disease relate to the natural history of tuberculosis. The current pre-entry tuberculosis screening programme detects active pulmonary tuberculosis using a combination of chest radiographs and sputum smear and culture tests. Several novel diagnostic tools have been developed that may offer substantial improvements to the existing migrant screening pathway and are outlined in Figure 7.(65,66) Active

tuberculosis can be confirmed with the Xpert® MTB/RIF assay, which uses molecular techniques and has increased the sensitivity of non-culture based tests and allows rapid detection of resistance (67,68). Despite these advantages, Xpert® is expensive and limited by its throughput capacity, which given the large number of migrants attending pre-entry screening clinics could be problematic operationally, as well as its lower accuracy for detecting extra-pulmonary disease compared to traditional culture methods.(69) Screening for latent tuberculosis can be performed using new IGRA tests, which require only a single contact with the patient and have comparable sensitivity and specificity to TST testing(11,12). However, IGRAs are costly and the ability of this test to predict development of active tuberculosis in migrants is uncertain at present (9,70–72).

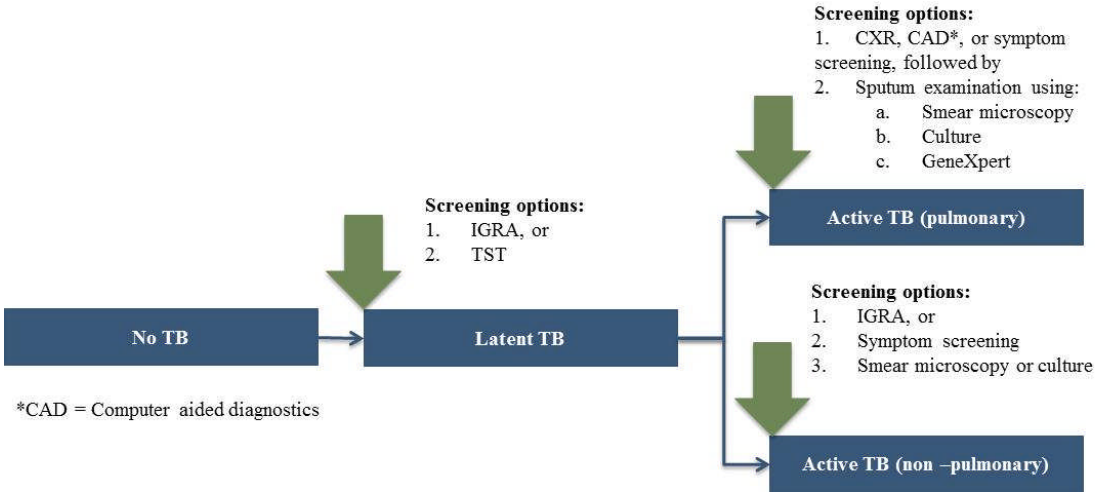

Figure 7. Screening options for TB in UK migrants.

The exact combination of tests for latent and active disease that would be most effective could be informed by a better understanding of the epidemiology of tuberculosis in migrants to the UK. As described earlier in this chapter, approximately 80% of tuberculosis disease notified in non-UK born individuals have lived in the country for two or more years.(35) This suggests that many cases do not enter with active disease, and despite various cross-sectional studies in UK migrants, it is uncertain what proportion of all individuals enter the country with latent tuberculosis, and who would therefore be amenable to intervention through a pre-entry screening programme for latent infection.(73) Similarly, few data exist to estimate the number of migrants who become infected on trips back to their country of origin, or acquire tuberculosis in the

UK due to other risk factors that are prevalent in new entrants such as poor housing, behavioural and lifestyle factors or immunosuppression due to HIV.(74–76,76) Therefore, despite the high number of cases in individuals not born in the UK, it has not been feasible to calculate the incidence rates of active tuberculosis in migrants, or the number of these cases that were due to reactivation of disease acquired prior to entry and therefore potentially amenable to intervention if detected after arrival in the UK.

Probabilistic matching algorithms developed by Public Health England are able to identify individuals across datasets without unique identifiers such as an NHS number, whilst also allowing for missing data and errors such as misspelling of names, and switching of day and month for dates of birth in record entries. Probabilistic matching therefore has the potential to identify incident cases of disease notified in the UK among individuals screened by the pre-entry programme by linking data from the UK tuberculosis register with that from migrants screened pre-entry. Such a dataset would enable estimates of the incidence of tuberculosis in migrants to be calculated, as well as the risk factors associated with these cases. Analysis of strain typing data, also contained within the UK tuberculosis register, would also facilitate estimates of the proportion of incident cases among migrants screened via the pre-entry programme that were due to reactivation. By combining these new epidemiological data about prevalent and incident tuberculosis in migrants to the UK, and using existing mathematical models, it would be possible to establish the best combination of tests to reduce the burden of infection and disease in this vulnerable group. This information could also be used to examine the cost effectiveness of new diagnostic tests for latent and active tuberculosis in migrants undergoing pre-entry screening.

1.7 Aims and objectives of the PhD thesis

This PhD aimed to inform the development of evidence-based migrant screening by examining the epidemiology of tuberculosis in migrants screened by a pre-entry programme.

Specific research objectives were:

1. To systematically review the published literature on pre-entry screening for tuberculosis in migrants to low incidence countries (Chapter 2).
2. To estimate the prevalence and risk factors for tuberculosis in migrants from high incidence countries screened prior to entering the UK (Chapter 3).
3. To establish the accuracy of probabilistic linkage for identifying individuals across datasets where no standard unique identifier or address data exist (Chapter 4).
4. To estimate the incidence and risk factors for tuberculosis cases notified in the UK among migrants screened by a pre-entry programme (Chapter 5).
5. To use molecular strain typing data to infer whether migrants screened pre-entry are less likely than non-UK born individuals to transmit tuberculosis (using the proxy marker of being the first case in a cluster of tuberculosis cases) and describe the incidence and risk factors of first in cluster cases in pre-entry screened migrants Chapter 6).
6. To use molecular strain typing data to infer the incidence of disease in pre-entry screened migrants that is potentially preventable through additional screening for latent tuberculosis infection by measuring the post-migration incidence of tuberculosis cases with unique molecular strain typing fingerprints (suggestive of disease reactivation rather than local transmission Chapter 6)

CHAPTER 2

Pre-entry screening for tuberculosis in migrants to low-incidence countries: a systematic review and meta-analysis

2.1 Abstract

Background: Several high income countries have pre-entry screening programmes for tuberculosis. This chapter aims to describe the existing literature on pre-entry screening programmes in order to inform evidence-based policy for migrant health screening.

Methods: Six bibliographic databases were searched for experimental or observational studies and systematic reviews, which reported data on migrant screening for active or latent tuberculosis by any method before migration to a low-incidence country. Primary outcomes were principal reported screening prevalence of active tuberculosis; prevalence of culture confirmed cases; and prevalence of sputum smear for acid-fast bacilli cases. Where appropriate, fixed effects models were used to summarise the prevalence of pre-entry screening across included studies.

Results: A total of 15 unique studies with data on 3,739,266 migrants screened pre-entry for tuberculosis between 1982 and 2010 were identified. Heterogeneity was high for all primary outcomes. After stratifying by prevalence in country of origin heterogeneity was reduced for culture and smear confirmed cases. Culture confirmed prevalence increased with prevalence in the country of origin, and summary estimates ranged from 20 (95%CIs: 10, 32) to 336 per 100,000 individuals screened (95%CIs: 283, 393) in countries with a prevalence of 50-149 and greater than 350 per 100,000 population respectively.

Conclusion: Targeting high-prevalence countries is likely to result in the highest levels of prevalent active disease to be identified.

2.2 Introduction

Several high income countries (Australia, Austria, Canada, France, Israel, Jordan, New Zealand and USA) have pre-entry screening programmes for tuberculosis.(56) The UK has used a combination of upon- and post-entry screening for several decades, but fully transitioned to pre-entry screening on 1st April 2014.(7)

Because of the high burden of tuberculosis in migrants, many governments in low-incidence settings have implemented screening programmes. Tuberculosis screening programmes for migrants can occur at three points in time: 1) pre-entry (prior to entering the country); 2) upon-entry; 3) or post-entry. Many European countries have implemented post-entry screening and, whilst there are differences in the screening approach, the characteristics of such programmes are well documented.(55)

The prevalence of pre-entry screening programmes for tuberculosis may differ from upon- and post-entry programmes. With some exceptions, upon- and post-entry screening tends not to be a compulsory part of visa applications; therefore, individuals undergoing screening may not be representative of the wider migrant population. Attendance for post-entry screening may be determined by patient health seeking behaviour or the opinion of immigration staff in upon-entry settings. Conversely, pre-entry screening programmes are typically a compulsory part of the visa application process and as a result coverage is higher, if not complete, and such studies should be fully representative of the populations screened and intending to migrate.

The characteristics of post- and upon-entry screening programmes have been well documented previously, but pre-entry screening programmes have not been systematically reviewed.(55,77,78) The aim of this systematic review was therefore to establish the ability of pre-entry screening programmes to detect prevalent cases of active disease and latent infection in order to inform evidence-based policy for migrant health screening.

2.3 Methods

2.3.1 Inclusion criteria

The following study types were eligible for inclusion in this review: experimental studies (randomized controlled trials as well as quasi-randomized controlled trials, including before and after studies); observational studies (including retrospective and prospective cohort studies, case-control studies, cross-sectional and case series); and systematic reviews. Additional inclusion criteria were that a study needed to be published with an abstract in English; to report the total number of individuals screened who plan to migrate as well as the number of cases of tuberculosis infection or disease identified; and screening was required to have taken place prior to the migrant entering a low-incidence country. Eligible studies could screen for tuberculosis by any method including radiographic, microbiological and a clinician's recommendation to treat an individual on the basis of clinical and/or radiological signs and/or symptoms compatible with tuberculosis.

2.3.2 Important definitions used in this systematic review and meta-analysis

We used the definition of migrants developed by Rieder et al. and used in a recent systematic review of screening in the EU.(55,79) This review classifies migrants into the following groups: Migrant (a foreigner legally admitted and expected to settle in a host country); Asylum seeker (a person wishing to be admitted to a country as a refugee and awaiting decision on their application for refugee status under relevant international instruments); Foreign-born citizen (a person who is a national of the state in which they are present but who was born in another country); Undocumented foreigner/migrant (formerly classified as 'illegal', describing an individual who enters, stays or works in a host country without an appropriate residence permit or visa).

There is no universally accepted definition of a low-incidence tuberculosis country. For the purpose of this analysis we used the European Centre for Disease Prevention and Control (ECDC) definition of a low-incidence country as one with a notification rate below 20 cases per 100,000 in the general population.(80)

2.3.3 Primary and secondary outcomes

Three primary outcomes were considered: 1) the principal prevalence of pre-entry screening for active tuberculosis reported for each study (detected by any method); 2) prevalence of active tuberculosis cases confirmed by culture; 3) prevalence of active tuberculosis cases confirmed by smear for acid-fast bacilli (AFB). Secondary outcomes included: 1) prevalence of active cases detected by radiography; 2) prevalence of drug resistant active disease; 3) prevalence of latent tuberculosis (diagnosed by any method); 4) costs associated with screening individual migrants, and; 5) costs of treatment for individuals screened and found to have tuberculosis. PRISMA reporting guidelines were followed.(81)

2.3.4 Search strategy

The following sources were searched to identify published literature: Medline, EMBASE, LILACS; Cochrane Infectious Diseases Group Specialized Register; Cochrane Library; Conference Proceedings Citation Index- Science; Conference Proceedings Citation Index- Social Science & Humanities. Reference lists of included studies were hand-searched in order to identify further relevant work. Only studies published after 1980 were included.

Detailed search terms for the bibliographic databases are presented in Appendix 1. In summary, terms covered the populations of interest (migrants, refugees, asylum seekers, new entrants, and undocumented migrants), the intervention (pre-entry screening) and standard terms for tuberculosis.

Initial search results were imported into EPPI-Reviewer 4 where duplicates were identified and removed. An updated search carried out on 1st April 2014 was performed in Zotero.(82,83) Three researchers (Robert Aldridge, Tom Yates, and Dominik Zenner) screened titles, abstracts and full text publications. Disagreements were resolved by discussion and remaining issues were assessed in conjunction with a fourth reviewer (Andrew Hayward – primary supervisor of this PhD thesis). Data from included studies were extracted in duplicate to an Excel spread sheet.

2.3.5 Methods for the meta-analysis

Fixed effects models with Freeman-Tukey transformation of data were used to estimate the summary prevalence of pre-entry screening across studies and subgroups where appropriate.(84,85) The I-squared transformation was used to describe the proportion of total variation in study estimates that is due to heterogeneity.(86) Where overlapping data on an individual screening programme were identified, the publication with the largest amount of data (by time period or number of individuals screened) was included in this review. Economic components of the studies identified were presented in a narrative format where identified.

Subgroup analysis was carried out for the primary outcomes to examine the impact of: prevalence in the country of origin; the screening method used (e.g. radiographic, microbiological, and clinical); receiving country; and type of migrants screened. As there are no universally accepted categories to classify prevalence of tuberculosis at the country level, we chose to use the following groups: 20-49; 50–149; 150–249; 250–349, and ≥350/100 000 cases per 100,000 population. We used WHO prevalence estimates for the middle year in which screening was performed.(42) Where possible, data for primary outcomes were extracted for each of the subgroups (e.g. different countries of origin) and then included in the subgroup analysis.

2.3.6 Assessment of the risk of bias of included studies

The risk of bias for included studies was assessed using the GRADE approach, and was carried out independently by two reviewers (Robert Aldridge and Tom Yates).(87) Any disagreements were discussed and resolved with the help of a third reviewer (Andrew Hayward) where necessary.

2.4 Results

A search of all bibliographic databases was performed on 5th April 2013 and updated on 1st April 2014. A total of 1,887 studies were found (Figure 8). A further 15 publications were identified through other sources including reviewing references of included studies. A total of 157 full text articles were retrieved and assessed for eligibility and 19 manuscripts met the inclusion criteria after double screening and review.(88–106) After

further review and extraction of data, four studies were excluded from the final analysis as they contained overlapping data for the primary outcomes.(103–106)

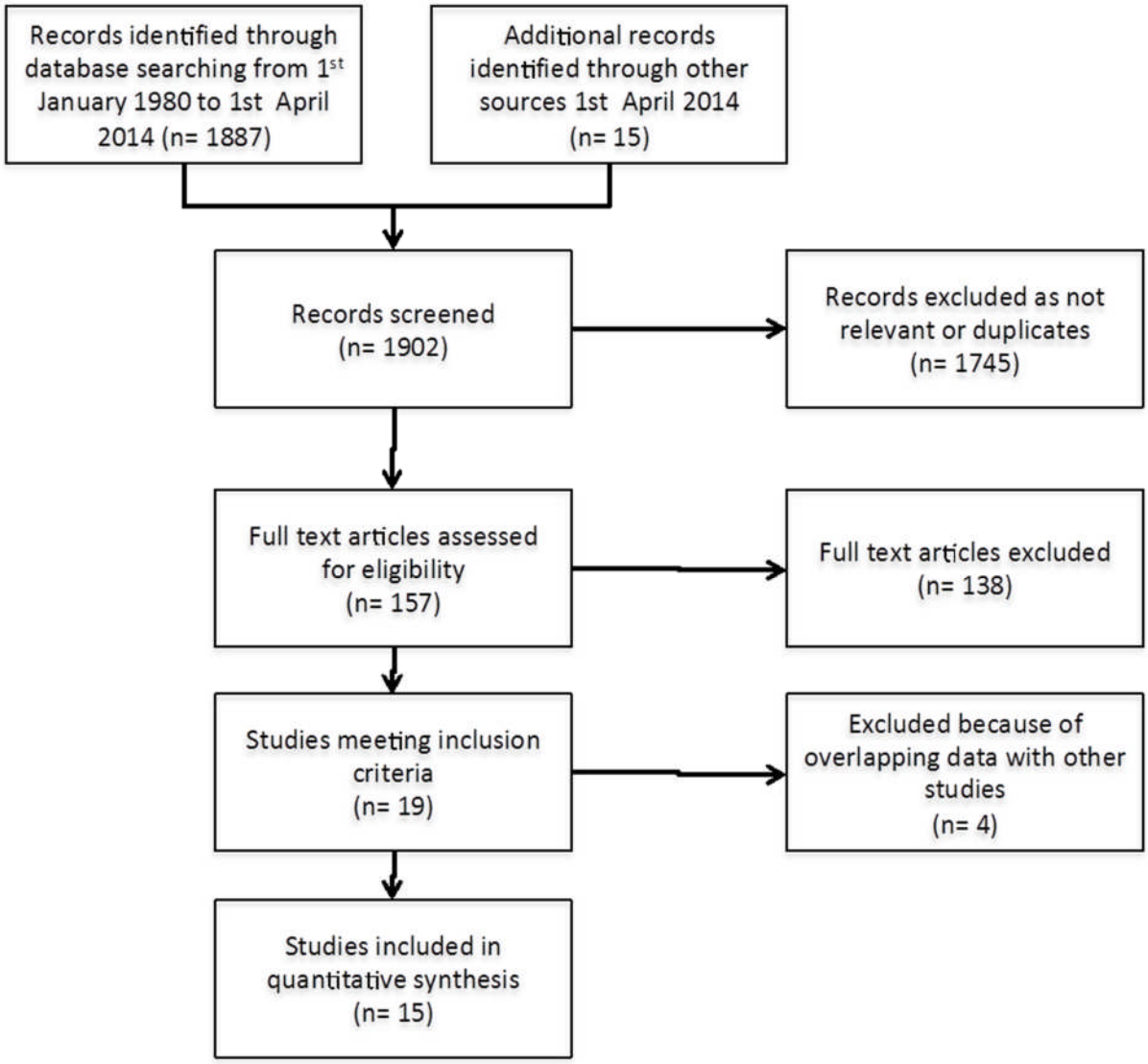

Figure 8. Study flow diagram.

The 15 studies included in the final analysis reported data on 3,739,266 individual migrants screened between 1982 and 2010 (Table 1).(88–102) Data were published on migrants to four low-incidence countries (Figure 9). The smallest study reported data on 873 migrants and the largest 3,092,729 migrants. Screening protocols varied between studies, but many involved an initial chest radiograph, clinical examination and testing of sputum smear and culture in selected individuals. The principal outcome for 10 studies reporting data on active tuberculosis included a combination of smear, culture or intention to treat on the basis of clinical findings as part of their case definition (Figure

10). Meta-analyses of prevalence for all three primary outcomes demonstrated high levels of heterogeneity (I-squared greater than 90%) and therefore summary effect estimates across studies were not calculated.

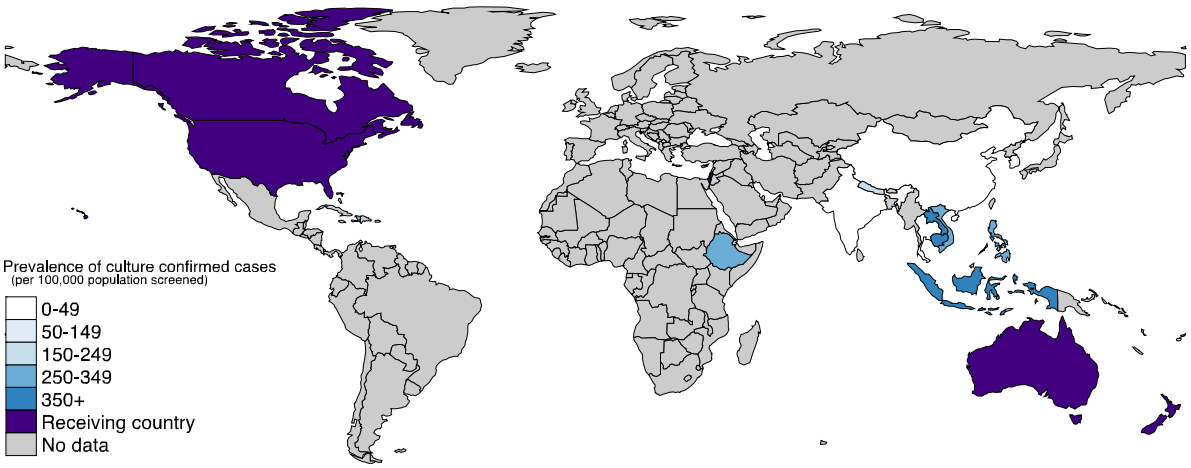

Figure 9. Map of countries conducting pre-entry screening of migrants (receiving countries) and the prevalence of culture confirmed cases for migrants by country of origin.

Note: For countries with yield for culture confirmed cases in more than one study, data from the most recent study is presented.

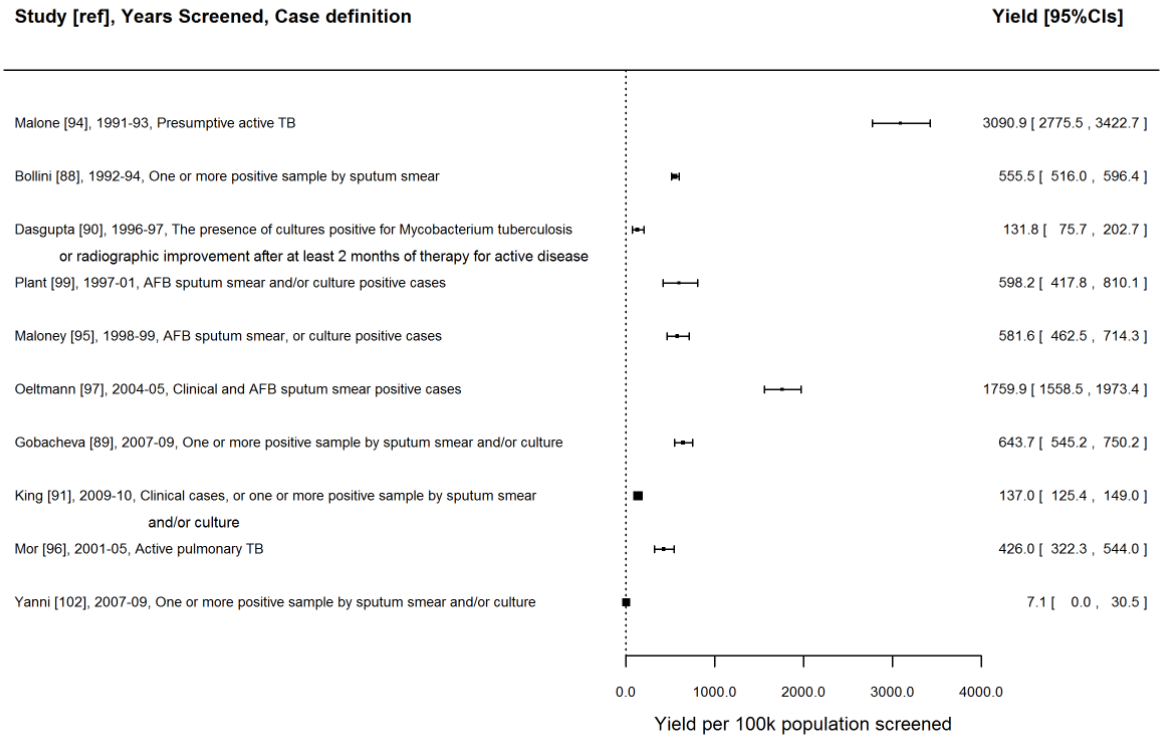

Figure 10. Forrest plot of pre-entry screening programme prevalence for principal outcome of active tuberculosis cases found by each study (case definition varies between studies, sorted by year of publication).

No studies reported the number of individuals tested by sputum culture or smear and therefore it was only possible to calculate prevalence based on the total number of individuals screened, and not by total number of microbiological tests performed (Table 2). Six studies presented data on a total of 755 cases that were culture confirmed among 452,971 individuals initially screened. (90,91,94–98) Six studies presented data on smear positive cases of tuberculosis, with a total of 987 cases found in the 569,210 individuals initially screened.(88,89,91,95,97,99) The majority of studies performed sputum smears on three samples for those individuals with a chest radiograph or clinical symptoms suggestive of tuberculosis (Full details in Appendix 2). There was some variation in the number of positive samples required to classify individuals with smear positive disease.

2.4.1 Stratifying by prevalence in the country of origin, population, screening method and receiving country.

After stratifying results by prevalence of tuberculosis in the country of origin, heterogeneity was reduced for culture and smear confirmed cases, but not principal outcome - active tuberculosis cases (Figure 11 and Appendix 3, Figures 39-41). Increasing prevalence of culture and smear positive cases were seen with increasing prevalence in the country of origin. Summary estimates of prevalence of culture positive cases ranged from 20 (95%CIs 10, 32) to 336 per 100,000 individuals screened (95%CIs: 283, 393) in countries with an incidence of 50-149 or greater than 350 per 100,000 population respectively (Figure 11). The results of the meta-analyses are dominated by one large study, which acknowledged limitations with data for smear and culture testing as this was not uniformly performed across all sites and for all cases.(91) Across all included studies, prevalence of culture confirmed cases was highest in migrants to USA from Vietnam with 1298 cases per 100,000 individuals screened (95%CIs: 1118, 1492; Figure 40 in Appendix 3).

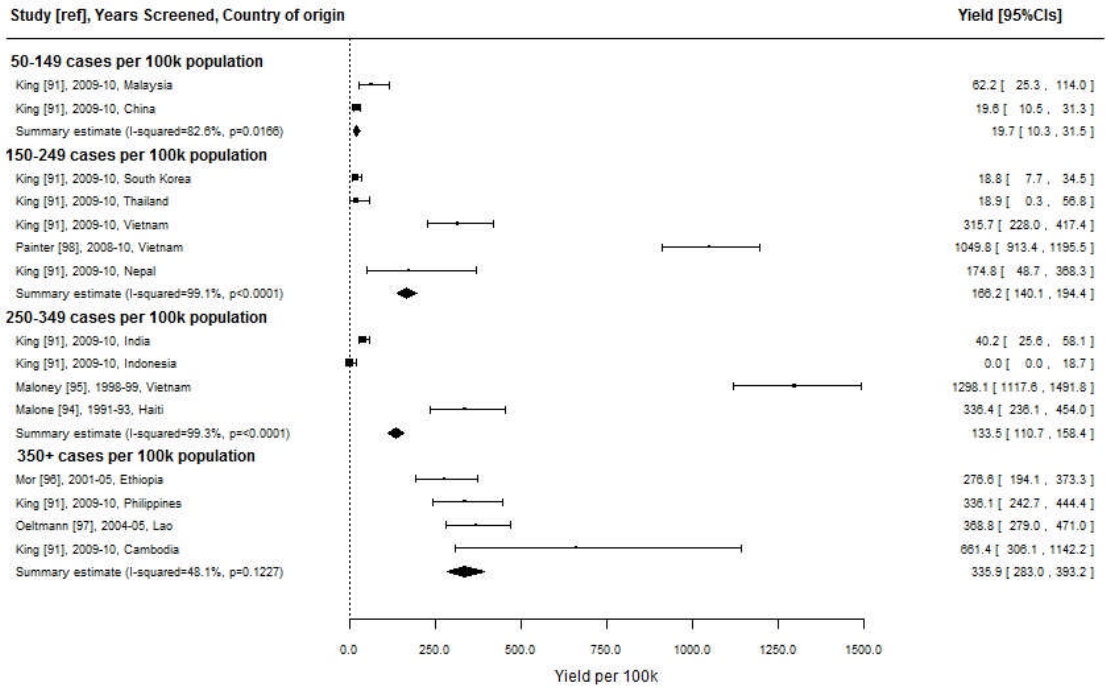

Figure 11. Forrest plot of pre-entry screening programme prevalence of culture positive cases of tuberculosis, stratified by WHO prevalence of tuberculosis in country of origin (sorted by prevalence in country of origin).

With the exception of culture confirmed cases in refugees (I-squared 0%, p= 0.85), heterogeneity remained high for all three primary outcomes (I-squared greater than 90%) after stratifying by population, screening method and receiving country (full results presented in the Appendix 3, Figures 39-53).

2.4.2 Secondary outcomes

Among studies reporting data on culture confirmed cases, three described prevalence of Multi-drug resistant (MDR) tuberculosis. A total of 33 cases in 183 individuals with culture confirmed disease were found in these three studies.(91,95,97) Whilst the majority of studies performed radiographic screening as a first line test, numerator and denominator data for this specific outcome were only presented in five studies.(90,93,95,96,101) A total 34,495 positive cases by chest radiograph were reported among the 3,154,873 individuals screened. This is likely to include both active and signs of previous tuberculosis infection or disease that has been treated or cleared spontaneously. Not all studies provided details as to how chest radiographs were analysed and classified which is likely to explain some of the variation in findings between studies.

Three studies reported data on latent tuberculosis infection, with a total of 1,884 latent tuberculosis infections found among 20,587 (9.2%) individuals screened (varying tests and cut offs were used – see Appendix 2 for full details).(92,97,102) One study tested for latent tuberculosis on a sample of migrants on the basis of chest radiograph results (testing 1,000 applicants with radiographic findings consistent with active tuberculosis and 500 applicants with a normal radiograph). Therefore the prevalence of latent tuberculosis from this study does not represent population prevalence of LTBI and the results were not included in this secondary analysis.(98)

Cost effectiveness was examined by one study using data from the Canadian pre-entry migrant screening programme from June 1996 to June 1997.(90) Compared to passive detection of cases after arrival in Canada, this study estimated the incremental cost (savings) to treat each case of prevalent active tuberculosis detected pre-entry as $39,409.(90) A further study, using data presented in this systematic review(96), estimated the cost of running a health station for an active tuberculosis screening programme in Ethiopia at $60,100 for approximately 3,500 individuals screened per

annum.(107) No data were found on costs of treatment for individuals screened and found to have tuberculosis.

2.4.3 Risk of bias of included studies

GRADE criteria were used to assess the risk of bias of included studies (Table 3). All included studies were observational in nature and therefore the evidence for each outcome was initially determined as low as per the GRADE methodology. This systematic review focused on describing prevalence of existing screening programmes in operational settings and therefore observational studies are an appropriate study design. The majority of studies were at risk of bias as a result of the eligibility criteria applied, and the reporting and measurement of exposure and outcome data. Substantial heterogeneity existed for primary outcomes, with confidence intervals across studies showing minimal or no overlap with the exception of culture and smear confirmed disease when stratified by prevalence in country of origin. As a result of these limitations, the quality of evidence for all outcomes was downgraded to very low as per the GRADE methodology.

Table 1. Summary descriptive information for each study included in review.

Study	Year of publication	Method of screening	Principal case definition	Number screened	Cases found	Prevalence per 100K population screened	Population screened	Country of Origin	Receiving Country	Country where screening took place	Years screened
Bollini	1998	X-ray. If compatible with TB, sputum smear samples were taken on three consecutive days	One or more positive sample by sputum smear	131241	729	555	Migrants	Vietnam	USA, Australia, Canada	Vietnam	1992-94
Dasgupta	2000	X-ray, If compatible with TB, sputum smear samples and tuberculin tests when judged as appropriate	Culture positive or radiographic improvement after at least 2 months therapy	12898	17	132	Migrants	Multiple	Canada	Multiple	1996-97
Gorbacheva	2010	X-ray, clinical examination, history and TST. Three sputum specimens in those with findings suggestive of TB	One or more positive sample by sputum smear and/or culture	23459	151	644	Refugees	Bhutan	USA, Canada, Australia, New Zealand, Denmark and Norway‡	Nepal	2007-09
King	2011	X-ray. If compatible with tuberculosis, sputum smear and culture testing*	Clinical cases, or one or more positive sample by sputum smear and/or culture	378939	519	137	Migrants	Multiple	Australia	Multiple	2009-10
Lange	1989	5 tuberculin units of purified protein derivative	10mm induration after PPD	873	9	1031	Adoptees	South Korea	USA	South Korea	1985-88
Liu	2009	X-ray. If compatible with TB, sputum smear samples were taken on three consecutive days	Inactive TB: X-ray positive, AFB sputum smear-negative tuberculosis	3092729	29,998	970	Mixed	Multiple	USA	Multiple	1999-05
Malone	1994	X-ray and physical examination. If compatible with TB, sputum smear and culture testing on three consecutive samples	Presumptive active TB†	11000	340	3091	Migrants	Haiti	USA	U.S. Naval Base in Guantanamo Bay, Cuba	1991-93
Maloney	2006	X-ray. If compatible with TB, sputum smear and culture testing on three	AFB sputum smear, or culture positive cases	14098	183	582	Migrants	Vietnam	USA	Vietnam	1998-99

		consecutive samples									
Mor	2012	X-ray, clinical examination, history and TST. Three sputum specimens in those with findings suggestive of TB	Active pulmonary TB‡	13379	57	426	Migrants	Ethiopia	Israel	Ethiopia	2001-05
Oeltman	2008	X-ray, clinical examination, history. Three sputum specimens in those with findings suggestive of TB	Clinical and AFB sputum smear positive cases	15455	272	1760	Refugees	Lao People's Democratic Republic	USA	Thailand	2004-05
Painter	2013	X-ray, clinical examination, history and sputum testing for M. tuberculosis as per CDC 2009 technical instructions.§	QuantiFERON ®-TB Gold In-Tube Assay and TST positive	1475	859	-**	Migrants	Vietnam	USA	Vietnam	2008-10
Plant	2004	X-ray, clinical examination, history. Three sputum specimens in those with findings suggestive of TB	AFB sputum smear and/or culture positive cases	6018	36	598	Migrants	Vietnam	Australia	Vietnam	1997-01
Wang	1991	X-ray followed by three sputum cultures in those with findings suggestive of TB	Inactive tuberculosis***	21956	1173	5343	Migrants	Multiple	Canada	Multiple	1982-85
Watkins	2005	X-ray	X-ray positive cases	1669	170	10186	Migrants	Vietnam	Australia	Vietnam	Not stated
Yanni	2013	X-ray, clinical examination, history and sputum testing for M. tuberculosis as per CDC 2009 technical instructions	One or more positive sample by sputum smear and/or culture	14077	1	7	Refugees	Iraq	USA	Jordan	2007-09

* Limitations in sputum smear and culture methods reported by study authors.
† Full definition of presumptive active TB not provided and unable to contact corresponding author to confirm what this encompasses.
‡ Active pulmonary TB defined as a symptomatic patient with pulmonary disease and confirmed *M.tuberculosis* complex culture.
§ Following the results of chest X-ray, applicants were invited to participate in a study of TST and QFT for which they would be provided the results, but the result of which would not affect their visa application. Varying size of TST was used as cut off
**Prevalence for latent tuberculosis for this study is not presented as the primary aim of the study was to compare the sensitivity of QuantiFERON ®-TB Gold In-Tube Assay (QFT) and TST for culture- positive pulmonary. It was therefore performed on a sample of migrants with and without abnormal X-ray results, and therefore prevalence of latent tuberculosis will not be representative.
*** Inactive tuberculosis defined by authors as: “radiograph shows evidence of tuberculosis, it is repeated at a minimum interval of 3 months to confirm stability of the lesion. In addition, 3 sputum cultures, incubated for 7-8 weeks, taken at least 24h apart, are required to be negative.”
‡ Whilst the abstract mentions these countries, it was unclear whether refugees went to anywhere other than USA and Canada and authors were uncontactable to confirm.

Table 2. Summary descriptive information for each study included within the quantitative review

Study	Number screened	Total cases of active disease found (%)		Smear positive cases (%)		Culture positive cases (%)		MDR cases (%)		X-ray positive cases (%)		Latent TB cases (%)		Population screened	Country of origin	Receiving country
Bollini	131241	729	(0.6)	729	(0.6)									Migrants	Vietnam	US, Australia, Canada
Dasgupta	12898	17	(0.1)							722	(5.6)	353	(2.7)	Migrants	Multiple	Canada
Gobacheva	23459	151	(0.6)	54	(0.2)									Refugees	Bhutan	USA, Canada, Australia, New Zealand, Denmark and Norway
King†	12795	113	(0.9)	4	(0.0)	43	(0.3)							Migrants	Philippines	Australia
King†	59666	87	(0.1)	2	(0.0)	24	(0.0)							Migrants	India	Australia
King†	13621	84	(0.6)	6	(0.0)	43	(0.3)							Migrants	Vietnam	Australia
King†	71600	43	(0.1)	1	(0.0)	14	(0.0)							Migrants	China	Australia
King†	42503	24	(0.1)	2	(0.0)	8	(0.0)							Migrants	South Korea	Australia
King†	12859	20	(0.2)	0	(0.0)	8	(0.1)							Migrants	Malaysia	Australia
King†	9192	15	(0.2)	1	(0.0)	0	(0.0)							Migrants	Indonesia	Australia
King†	1512	14	(0.9)	1	(0.1)	10	(0.7)							Migrants	Cambodia	Australia
King†	10608	13	(0.1)	0	(0.0)	2	(0.0)							Migrants	Thailand	Australia
King†	2861	12	(0.4	0	(0.0)	5	(0.2)							Migrants	Nepal	Australia
Lange	873	9	(1.0)									9	(1.0)	Adoptees	South Korea	USA
Lui	2714223	26075	(1.0)							26075	(1.0)			Migrants	Multiple	USA
Lui	378506	3923	(1.0)							3923	(1.0)			Refugees	Multiple	USA
Malone	11000	340	(3.1)			37	(0.3)							Migrants	Haiti	USA
Maloney	14098	82	(0.6)	82	(0.6)	183	(1.3)	5	(0.0)	1331	(9.4)			Migrants	Vietnam	USA
Mor	13379	57	(0.4)			37	(0.3)			150	(1.1)			Migrants	Ethiopia	Israel
Oeltman	15455	272	(1.8)	34	(0.2)	57	(0.4)	24	(0.2)					Refugees	Lao People's Democratic Republic	USA

Oeltman	5637	1624	(28.8)							1624	(28.8)	Refugees	Lao People's Democratic Republic	USA
Painter	20100	211	(1.0)			211	(1.0)		2087	(10.4)			Migrants	Vietnam	USA
Plant	5108	25	(0.5)	15	(0.3)								Migrants	Vietnam	Australia
Plant	910	11	(1.2)	6	(0.7)								Migrants	Cambodia	Australia
Wang	21956	1173	(5.3)										Migrants	Multiple	Canada
Watkins	1669	170	(10.2)						170	(10.2)			Migrants	Vietnam	Australia
Yanni	14077	1	(0.0)								251	(1.8)	Refugees	Iraq	USA

† Study reports that overall 230 cases were culture confirmed and 67 were smear positive, but not all of these data are included as the data on number of migrants screened was not presented for all countries.

Table 3. GRADE summary of findings and quality of evidence for the primary and secondary outcomes

	Quality assessment							Quality	Importance
	No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations		
Principal outcome	15	Observational studies*	Serious†‡	Very serious§	Serious¶‖	No serious imprecision	Reporting bias** Dose response gradient††	Very Low	Important
Sputum Culture	6	Observational studies*	Serious†‡‡	Serious§§	Serious¶¶	No serious imprecision	Reporting bias** Dose response gradient††	Very Low	Important
Sputum Smear	6	Observational studies*	Serious†‡‡	Serious§§	Serious¶¶	No serious imprecision	Reporting bias**	Very Low	Important
Chest radiograph	5	Observational studies*	Serious†	Very serious§	Very serious¶‖	No serious imprecision	Reporting bias**	Very Low	Important
Latent tuberculosis	3	Observational studies*	Serious†	Very serious§	Very serious¶	No serious imprecision	Reporting bias**	Very Low	Not Important
Multi-drug resistant tuberculosis	3	Observational studies*	Serious†‡‡	Serious§§	Serious¶¶	No serious imprecision	Reporting bias**	Very Low	Important

*Case series.

†Most studies were at some risk of bias for failure to develop and apply appropriate eligibility criteria and measurements of outcome that had limitations.

‡Largest study for analysis by subgroup(91) stated that "smear and culture testing may be offered but of variable quality". Several studies only look back at results of individuals who arrived in the low-incidence country—a potential bias if there was a difference in the proportion who travelled by test result, which is likely to be the case.

§Substantial heterogeneity existed among studies with CIs that minimally overlapped. The proportion of the variation in point estimates due to among-study differences was large and exploration of a-priori subgroup analyses did not substantially explain this.

¶Populations across studies varied; however, the evidence summaries are highly relevant to policy makers and those interpreting the studies, and outcomes (such as active tuberculosis) are likely to be of interest and important to migrants.

‖Interventions and outcomes varied greatly, particularly as smear and culture testing was offered but of variable quality in the largest included study for analysis by subgroup(91) and as many studies included radiographic diagnoses with substantial variation in the radiographic case definition used. Additionally, studies with high detection rates(101) seem likely to have included inactive and old tuberculosis scars in addition to active disease.

**Data for all years from countries conducting pre-entry screening were not available in the published literature.

††Some evidence to suggest that higher tuberculosis prevalence in country conducting pre-entry screening was associated with a higher yield of cases.

‡‡There was the potential for outbreak bias in one study(97) because it was initiated as a result of an unusually high number of cases.

§§Substantial heterogeneity among studies with CIs that showed minimal or no overlap. The proportion of the variation in point estimates due to between-study differences was large. Exploration of a-priori subgroups reduced heterogeneity.

¶¶Interventions and outcomes for multidrug-resistant cases are likely to be less variable due to procedures involved in laboratory testing being somewhat uniform across sites, although the consistency with which these were applied across studies might cause some issues in relation to indirectness.

2.5 Discussion

This systematic review found 15 unique studies with data on 3,739,266 individuals screened by pre-entry tuberculosis programmes. Heterogeneity was high for all primary outcomes examined, but reduced after stratification by prevalence of tuberculosis in the country of origin for culture and smear confirmed cases, and culture confirmed cases in refugees, but not principal outcome (of active tuberculosis) reported by each study. Culture and smear confirmed prevalence increased with prevalence in the country of origin. Summary estimates of culture confirmed cases ranged from 20 (95%CIs 10, 32) to 336 per 100,000 individuals screened (95%CIs: 283, 393) in countries with an incidence of 50-149 or greater than 350 per 100,000 population respectively.

2.5.1 Strengths and weaknesses of the systematic review

As far as it has been possible to identify, this is the first systematic review and meta-analysis of pre-entry screening programme data for tuberculosis. Established systematic review procedures were used including double screening review, and PRISMA reporting guidelines for systematic reviews and meta-analyses.(81) Attempts were made to reduce bias in the review process by following empirically based systematic review and meta-analysis guidelines.(87)

There was substantial heterogeneity between studies, limiting the ability to synthesise results across settings and outcomes. With the exception of prevalence of culture and smear confirmed cases, when stratified by prevalence of tuberculosis in country of origin, and culture confirmed cases in refugees, heterogeneity in the primary and secondary outcomes remained high after exploring potential *a priori* explanatory variables. Data on the age of those screened was not provided consistently, which might be particularly important for latent tuberculosis and studies that included old tuberculosis detected by chest radiograph. The top five countries of origin for migrants from developing to developed countries in 2010 were Mexico, India, China, Philippines, and Turkey.(49) Although data were presented for India, China, and the Philippines, the migrants were not entirely representative of migrant flow between developing and developed countries. It was not clear from most studies whether there was uniform drug susceptibility testing or whether only retreatment cases were tested.

There was a risk of misclassification in the principal outcomes reported by many studies, particularly for those that included clinically identified cases (with an intention to treat) as part of the case definition.

3.5.2 Strengths and weaknesses in relation to other studies

A previously published systematic review focused on all types of migrant screening programmes for tuberculosis in the EU/EEA, independent of where the screening took place.(55) This review found no studies of pre-entry screening in Europe, but data were separately reported from three pre-entry screening programmes conducted outside EU/EEA, all of which were included in this systematic review. A total of 14 studies reported data from upon-entry screening programmes in EU/EEA with a median active tuberculosis prevalence of 360 cases per 100,000 people screened (interquartile range 100–520). Five studies reported data on community post-entry screening with a summary active tuberculosis prevalence of 220 cases per 100,000 people invited to screening (interquartile range 100–380). Direct comparisons with upon- and post- entry screening programmes are difficult to make due to lack of comparability between study designs, secular trends and populations considered. Pre-entry screening, when conducted in countries with a prevalence of tuberculosis greater than 350 per 100,000 population appears to be within a similar range as these upon- and post-entry programmes.

2.6 Conclusion

Pre-entry screening programmes aim to identify cases of active tuberculosis before arrival of the migrant in the host country. This review provides evidence that pre-entry screening programmes have varying prevalence that increases with prevalence in the country of origin. Screening in countries with prevalence <150 per 100,000 is likely to result in a low prevalence of culture and smear confirmed cases detected. The full public health implications of this analysis, recommendations, and directions of future research are discussed in the final chapter of the thesis.

2.7 Publications and presentations arising from this chapter

1. Aldridge RW, Yates TA, Zenner D, White PJ, Abubakar I, Hayward AC. Pre-entry screening programmes for tuberculosis in migrants to low-incidence countries: a systematic review and meta-analysis. Lancet Infect Dis. 2014 Dec;14(12):1240–9.
2. Aldridge RW, Yates TA, Zenner D, White PJ, Abubakar I, Hayward AC. Pre-entry tuberculosis screening of migrants to low incidence countries: a systematic review. 44th Union World Conference in Paris. 3 November 2013.
3. Aldridge RW. Investigating the epidemiology of tuberculosis in migrants to the UK and the cost effectiveness of pre-entry screening. World TB Day 2014: Reaching the 3 Million. 24 March 2014. LSHTM

CHAPTER 3

Prevalence and risk factors for active tuberculosis in migrants to the UK screened pre-entry.

3.1 Abstract

Background: A pilot pre-entry screening programme was set up for migrants to the UK from 15 countries in 2005. This study aimed to investigate the prevalence and risk factors for tuberculosis in migrants screened prior to entering the UK from the 15 pilot high-burden countries.

Methods: A cross-sectional study was performed on pre-entry screening data collected between 1st October 2005 and 31st December 2013. The primary outcome was the prevalence of bacteriologically confirmed tuberculosis. Poisson regression was used to estimate crude prevalence and a multivariable logistic regression model was created to identify risk factors for the primary outcome.

Results: A total of 476,455 visa applications were screened for tuberculosis at sites where sputum samples were tested by culture. Crude prevalence was 92 per 100,000 population screened (95%CIs: 84, 101). After adjusting for age, sex and clustering by individual, there was evidence that having contact with a case of tuberculosis (OR 11.6; 95%CIs: 7.0, 19.3; p-value <0.001) was associated with an increased risk of bacteriologically confirmed tuberculosis at pre-entry screening.

Conclusion: The results of this study provide a comprehensive analysis of the historical data from a pilot pre-entry screening programme that has been running since 2005. The study identified issues that could inform evidence-based migrant screening policies, and groups of migrants that would benefit from increased health improvement interventions.

3.2 Introduction

In 2005 the UK Border Agency (UKBA) and Foreign and Commonwealth Office (FCO) funded the set up costs for a trial of pre-entry screening of migrants to the UK in conjunction with the International Organization for Migration (IOM).(59,60) The pilot was launched in November 2005 in Kenya, Sudan and Thailand, and was subsequently rolled out to eight locations screening migrants from a total of 15 countries (Table 4). Local IOM clinics ran the programme with oversight from an epidemiological unit based in Manila, Philippines.

Table 4. Countries from the IOM pre-entry screening programme.

Country where IOM clinic is located	Migrants from other countries screened at clinic	Start of screening	UKTB Global Software Start
Thailand	Laos	November 2005	July 2008
Bangladesh		January 2006	October 2008
Cambodia		February 2006	October 2008
Pakistan		March 2007	March 2009
Ghana	Cote D'Ivoire	February 2007	December 2009
	Burkina Faso		
	Niger		
	Togo		
Tanzania		October 2005	December 2009
Sudan	Eritrea	November 2005	December 2009
Kenya	Somalia	November 2005	August 2010

From 2005 to 2012 the pre-entry pilot scheme required migrants from 15 countries intending to stay in the UK for longer than six months to be certified free of pulmonary tuberculosis as part of the visa application process. IOM clinics carrying out pre-entry screening for the UK government used a set of technical instructions that specified how screening should be conducted, and included:(7)

- A symptom screen (asking for details of cough, haemoptysis, weight loss, night sweats, history of previous tuberculosis).
- A clinical history including details of any recent contact with a case of active pulmonary tuberculosis.

- A physical examination if considered necessary by the certifying physician.
- A chest radiograph (CXR) in all applicants except pregnant women and children under the age of 11 years.
- All applicants with radiological findings classified as "major findings sometimes seen in active tuberculosis" (see Table 5 for further details) are required to undergo sputum testing for tuberculosis with three sputum samples tested by microscopy for acid fast bacilli (AFB).

In May 2012 it was announced that the pre-entry system would be expanded from the 15 pilot locations to 101 countries with a WHO incidence of greater than 40 per 100,000 population. As a result many more migrants to the UK were required to be certified free from active tuberculosis before a visa could be issued. Analysis of the historical data from the 15 pilot countries can be used to inform and improve the expanded pre-entry screening programme, identifying its strengths, weaknesses and quality assurance opportunities, as well as suggesting potential opportunities to intervene and improve the health of migrants. Using data from the 15 pilot countries, this study aimed to investigate the prevalence and risk factors for tuberculosis in migrants from high incidence countries screened prior to entering the UK, and put these data in the context of previously published literature.

3.2.1 Research questions:

The analysis presented in this chapter aims to answer the following research questions:

1. What is the prevalence of bacteriologically confirmed tuberculosis in migrants at the time of pre-entry screening?
2. What are the risk factors for bacteriologically confirmed tuberculosis in migrants at the time of pre-entry screening?

3.3 Methods

3.3.1 Study design and setting

This was a cross-sectional study of migrants applying for visas to stay in the UK for more than six months, screened for tuberculosis in 15 countries taking part in a pre-

entry screening pilot programme. The data included in this analysis were collected between 1st October 2005 and 31st December 2013.

3.3.2 Chest radiography

The UK technical instructions set out how tuberculosis screening was performed by IOM from 2005.(7) All applicants completed an informed consent form in a language they were able to understand, or steps were taken to ensure they understood the form if they were unable to read it (See Appendix 4 for a copy of the consent form). The applicant signed the consent form before the screening process started. Applicants of 11 years of age and above received a standard postero-anterior view chest radiograph. All visa applicants with radiological findings classified in group four "major findings sometimes seen in active tuberculosis" were required to undergo sputum testing for tuberculosis (Table 5). Individuals with chest radiographs classified as group three "minor findings occasionally associated with tuberculosis infection" were not mandatorily required to undergo sputum testing, but the panel physician responsible for the migrant screening (a medical doctor in charge of screening at an IOM clinic) was requested to consider sputum testing on a case-by-case basis. A radiologist and the panel physician both interpreted chest radiographs. Applicants unwilling or unable to undergo radiographic screening were required to provide three consecutive daily sputum specimens that were tested in a designated laboratory for smear and culture.

Table 5. Recording of radiographic findings

Minor Findings	
1.1	Single fibrous streak/band/scar
1.2	Bony islets
2.1	Pleural capping with a smooth inferior border (<1cm thick at all points)
2.2	Unilateral or bilateral costophrenic angle blunting (below the horizontal)
2.3	Calcified nodule(s) in the hilum / mediastinum with no pulmonary granulomas
Minor findings occasionally associated with tuberculosis infection	
3.1	Solitary Granuloma (< 1 cm and of any lobe) with an unremarkable hilum
3.2	Solitary Granuloma (< 1 cm and of any lobe) with calcified / enlarged hilar lymph nodes

3.3	Single / Multiple calcified pulmonary nodules / micronodules with distinct borders
3.4	Calcified pleural lesions
3.5	Costophrenic Angle blunting (either side above the horizontal)
Major findings sometimes seen in active tuberculosis (or other conditions)	
4.0	Notable apical pleural capping (rough or ragged inferior border and/or ≥ 1cm thick at any point)
4.1	Apical fibronodular / fibrocalcific lesions or apical microcalcifications
4.2	Multiple / single pulmonary nodules / micronodules (noncalcified or poorly defined) 4.3 Isolated hilar or mediastinal mass/lymphadenopathy (non-calcified)
4.4	Single / multiple pulmonary nodules / masses ≥ 1 cm.
4.5	Non-calcified pleural fibrosis and / or effusion.
4.6	Interstitial fibrosis/ parenchymal lung disease/ acute pulmonary disease
4.7	Any cavitating lesion OR “fluffy” or “Soft” lesions felt likely to represent active TB

3.3.3 Laboratory testing

Laboratory examination for *M.tuberculosis* consisted of the provision of at least three sputum specimens, taken on 3 separate occasions, not less than 24 hours apart and ideally in the early morning. All specimens underwent microscopy for acid-fast bacilli (AFB) by an auramine stain (or, if necessary, by Ziehl-Neelsen stain). At the start of the pilot programme, culture testing was not universally available at all IOM screening clinics. Where culture testing was available, this was undertaken as a culture on liquid or solid media for mycobacteria and confirmation of the Mycobacterium species at least to the *M.tuberculosis* complex level. Specimens were cultured for a minimum of six weeks in liquid media and eight weeks in solid media, unless a positive result was obtained earlier than this. If there was no growth after these time periods, specimens were reported as negative. Where available, positive *M.tuberculosis* cultures underwent drug susceptibility testing (DST) in a designated laboratory in accordance with World Health Organisation guidelines.(108)

3.3.4 Outcomes of pre-entry screening

The panel physician could issue applicants with a clearance certificate if they had a chest radiograph classified as free of any radiological changes, or findings in groups one or two. In addition to chest radiographs and sputum testing results, physicians were able to use their clinical judgment in the evaluation of an applicant, and if active pulmonary tuberculosis was suspected then they were able to refuse to issue a medical clearance certificate. Panel physicians were under no obligation to treat applicants diagnosed with tuberculosis, but were required to provide clear and unambiguous advice about the need to seek treatment immediately and provide a treatment referral letter.

Applicants diagnosed with active tuberculosis were able to restart the screening process having successfully completed a full course of approved treatment, but not within six months of the original examination. A written treatment summary from the treatment provider was required at the time of repeat screening. At rescreening, panel physicians were required to compare the chest radiograph taken at the time of the original application with an updated image. Where the panel physician was satisfied that the applicant no longer had active tuberculosis, a medical clearance certificate could be issued.

3.3.5 Primary outcome

The primary outcome for this study was the prevalence of bacteriologically confirmed tuberculosis (culture testing on liquid or solid media, or microscopy for acid fast bacilli). Cases of bacteriologically confirmed tuberculosis in this study were specified according to the WHO revised definition as: "one from whom a biological specimen is positive by smear microscopy, culture or WHO-approved rapid diagnostics (such as Xpert MTB/RIF)".(109) No WHO-approved rapid diagnostics were used in this study.

3.3.6 Secondary outcomes

1. Prevalence of tuberculosis
    a. Confirmed by culture testing on liquid or solid media.
    b. Confirmed by microscopy for acid fast bacilli.
    c. Confirmed by culture testing on liquid or solid media and resistant to one or more anti-tuberculosis drugs.

2. Prevalence of clinically confirmed tuberculosis.

The WHO definition for a case of clinically diagnosed tuberculosis implies that the diagnosing clinician or other medical practitioner intends to give the patient a full course of treatment. As panel physicians were not obliged to treat cases diagnosed as part of the screening process, the definition of a clinical case for this study was adapted from the WHO guidelines, and defined as: "A clinically diagnosed case is one that does not fulfil the criteria for bacteriological confirmation but has been diagnosed with active tuberculosis by a clinician or other medical practitioner who has decided to give, **or refer** the patient for a full course of tuberculosis treatment." The change to the WHO definition used in this chapter is highlighted in bold and includes cases diagnosed on the basis of chest radiograph abnormalities or suggestive histology and extra pulmonary cases without laboratory confirmation. Clinically diagnosed cases subsequently found to be bacteriologically positive (before or after starting treatment) were reclassified as bacteriologically confirmed.(109)

3.3.7 Data sources

The analysis presented in this chapter used data collected by IOM as part of the screening process, including demographic and clinical data for all individuals screened. At the start of the pilot study, data were collected by IOM clinics on Microsoft excel spreadsheets and sent on a monthly basis to the IOM epidemiology unit in Manila, Philippines. From July 2008, an internet based system called "Global Software" was rolled out by IOM to replace the previous use of spreadsheets for the collection of data on migrants (see Table 4 for roll out dates of Global Software at each IOM clinic). Global Software was a single database with a web interface and integrated all requirements of the UK technical screening instructions. Global Software included an automated facility to detect individuals attending multiple sites for screening, and therefore minimised duplicate entries for individuals. Global Software also contained digital image capture for personal identification purposes and digital chest radiographic imaging using Picture Archiving and Communications System (PACS) technology. The IOM epidemiology unit in Manila checked and reconciled data (from Global Software or excel sheets) on a monthly basis to ensure they were consistent and that outcomes of

screening were fully completed. Data were cleaned by the IOM epidemiology unit in coordination with clinics to ensure that records included all laboratory results on individuals screened and that any duplicate entries resulting from administrative error were removed or consolidated into one record. Monthly reports containing aggregate numbers on the screening process from all 15 clinics were sent to UKBA and Public Health England's tuberculosis screening surveillance unit. The dataset used in this analysis contained a variable to indicate which individuals were screened at clinics where culture testing was being performed.

3.3.8 Sample size

Before access to the IOM dataset was granted, it was estimated that the database would contain records on 350,000 migrants. Confidence intervals for the prevalence of bacteriologically confirmed tuberculosis were estimated for a country with a prevalence of 100 per 100,000 population, and for three different scenarios based on the number of migrants screened: 1000, 10,000 and 50,000. It was assumed that the prevalence in migrants was representative of the general population. Under these assumptions, a prevalence of 100 per 100,000 population would be estimated with the following 95% CIs:

- 3 to 556 if 1,000 individuals were screened
- 48 to 184 if 10,000 individuals were screened
- 74 to 132 if 50,000 individuals were screened

3.3.9 Duplicates

It was possible for individuals screened pre-entry to have multiple entries in the IOM database for several reasons, including: 1) clearance certificates once issued only last for six months, and if a UK visa application was not processed during this time period, the applicant was required to undergo repeat screening with a duplicate record being created; 2) administrative errors leading to two records being created for one individual during the same screening application. This scenario was more likely to occur prior to the roll out of the Global Software which automatically checked for these errors; 3) migrants undergoing a clearance screen after an initial positive screen for tuberculosis.

Duplicates were therefore analysed on the basis of whether they occurred less than 12 months apart or not. A period of 12 months was chosen a-priori as this was long enough to capture individuals found to have tuberculosis on their first screen, who were then undergoing repeat screening for visa clearance, but not too long a time period such that tuberculosis exposure and risk factors may have changed significantly for a majority of individuals. On the basis of these considerations, rules used to determine whether duplicate entries for an individual should be included in the cohort analysis are provided in Table 6.

Table 6. Rules for dealing with duplicate screens.

		First screen TB result	
		Positive	Negative
Last screen TB result	Positive	*Less than or equal to 12 months apart:* 1st screen - Include 2nd screen - Exclude *Greater than 12 months apart:* Include all duplicates	Include all
	Negative	*Less than or equal to 12 months apart:* 1st screen - Include 2nd screen - Exclude *Greater than 12 months apart:* Include all duplicates	Include all

For duplicates occurring less than 12 months apart where the first screen was positive, the second screen has been undertaken for clearance purposes. This second screen is therefore treated as a continuation of the initial screening process taken for the purposes of clearance and was excluded as it is not a newly prevalent case, the primary outcome for this analysis. All duplicates occurring greater than 12 months apart were included as this is sufficient time for the exposure to tuberculosis in an individual to have changed.

3.3.10 Exclusion criteria

During the start of the pilot phase, not all sites conducted culture testing for *M.tuberculosis*. To ensure comparability of estimates for the primary and secondary outcomes across countries and locations, this analysis was restricted to sites were culture and smear testing was performed on all sputum samples collected. This analysis would also be the most useful for informing current screening policy due to the fact that is now a requirement that sites conducting pre-entry screening perform culture testing for *M.tuberculosis* on all sputum samples.

3.3.11 Statistical analysis

Poisson regression (suitable for modelling rare event data) was used to estimate crude prevalence for the primary and secondary outcomes and was calculated per 100,000 population screened. Adjusted estimates for primary and secondary outcomes in the population screened were then calculated using the results of multivariable Poisson regression that included terms for age and sex and using the Margins command in Stata. Adjusted estimates for each country were compared to WHO estimates of prevalence for tuberculosis in 2010.(110) To account for duplicate screens included in this analysis, all crude and adjusted estimates accounted for clustering by individual. A multivariable logistic regression model was created to identify risk factors for the primary outcome. Final results were presented as odds ratios, 95% confidence intervals with p-values. Stata v.13 (Statacorp LP, College Station, TX, USA) was used for all statistical analyses.

To compare the results with the published literature, the meta-analyses presented in chapter two (prevalence of all cases, and prevalence of culture confirmed cases) were updated using the crude estimates from this analysis. Countries with less than 1,000 migrants screened were excluded from the updated meta-analysis of culture confirmed cases stratified by country of origin due to small numbers of cases detected.

3.3.12 Sensitivity analysis

The main analysis presented in this study included only migrants screened at sites where culture testing was performed on sputum samples. To examine the impact of the introduction of this sputum testing on the rates of bacteriologically confirmed cases of

tuberculosis, a sensitivity analysis was conducted that included all migrants screened pre-entry by IOM, regardless of whether they were seen at sites undertaking culture testing or not.

3.3.13 Ethics approval

Ethical approval was received for this analysis from UCL research ethics committee (3294/002; See Appendix 4 for copy of approval letter). The work was conducted with Public Health England which has Health Research Authority approval to hold and analyse national surveillance data for public health purposes under Section 251 of the NHS Act 2006.

3.4 Results

Between 1st October 2005 and 31st December 2013 a total of 692,362 visa applications were screened for tuberculosis (Figure 12). A total of 106 duplicate screens were excluded, all less than 12 months apart, one with a positive first and second screen, and 105 with a positive first screen and negative second screen. After excluding duplicates screens, and applicants screened at sites not performing culture testing on sputum, there were a total of 476,455 screening records included in this analysis. A total of 470,223 chest radiographs were performed and 21,772 sputum samples were collected. Chest radiographs were not carried out on 3,911 children and 2,319 pregnant women.

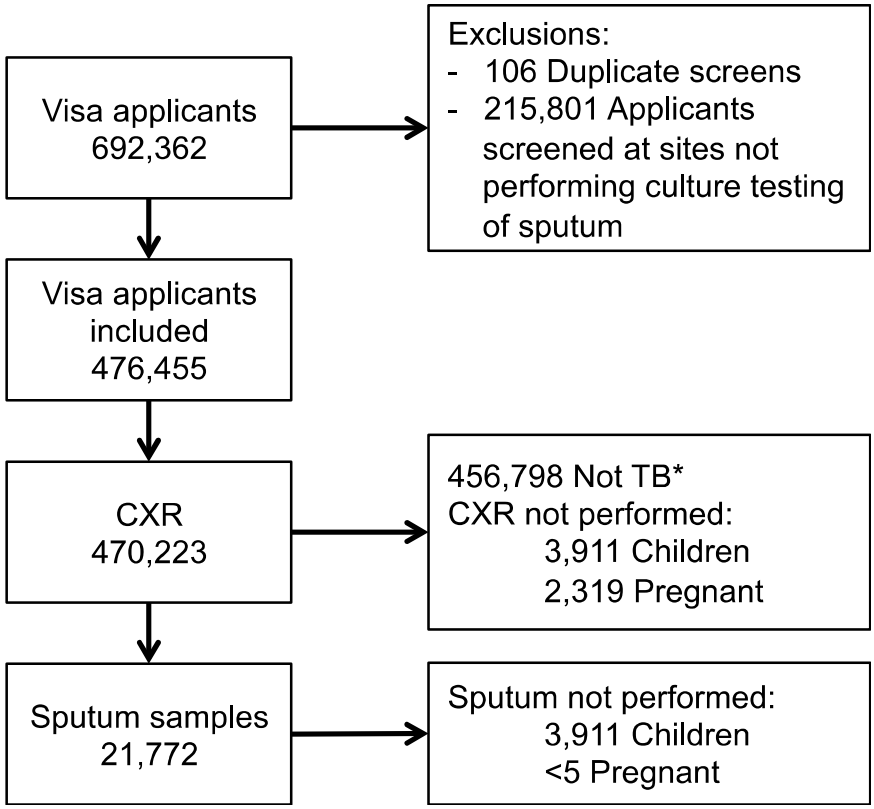

Figure 12. Results of the screening process between 17th October 2005 and 31st December 2013 in countries with IOM screening.

*No abnormality, or abnormality in classified in groups 1, 2, 3 and not requiring sputum testing as determined by the panel physician

The majority of individuals screened were aged between 16 and 44 (444,579; 93.3%) and male (309,062, 64.9%; Table 7). More migrants were screened in 2009 than any other year (116,899; 24.5%) and the country conducting the most screening was Pakistan (243,243; 51.1%), which accounted for just over half of all migrants screened. The majority of those screened were students (281,703; 59.1%). A total of 1,239 (0.3%) individuals being screened reported close contact with a case of tuberculosis. There were no missing data on any of the variables included in the final dataset.

After excluding duplicate records, 439 cases of bacteriologically confirmed tuberculosis were diagnosed providing a crude prevalence of 92 per 100,000 population screened (95%CIs: 84, 101). Crude prevalence of smear positive tuberculosis (55 per 100,000; 95%CIs: 49, 62) was lower than culture confirmed tuberculosis (83 per 100,000; 95%CIs: 75, 92). The crude prevalence of clinically diagnosed cases (3 per 100,000;

95%CIs: 2, 4) and overall prevalence of culture confirmed samples with resistance to one or more tuberculosis drugs was 3 per 100,000 (95%CIs: 2, 5).

The crude prevalence of bacteriologically confirmed tuberculosis was highest in those with a history of contact with an infectious case of tuberculosis (1,372 per 100,000; 95%CIs: 855, 2,201) and those over the age of 65 (329 per 100,000; 95%CIs: 172, 631), but due to the low number of cases in both of these categories, estimates had wide confidence intervals. Women had higher crude prevalence of bacteriologically confirmed tuberculosis (116 per 100,000; 95%CIs: 101, 134) than men (79 per 100,000; 95%CIs: 70, 90).

Migrants applying under the visa category of Settlement and Dependents had the highest crude prevalence of bacteriologically confirmed tuberculosis (108 per 100,000; 95%CIs: 93, 125). Crude prevalence of bacteriologically confirmed tuberculosis did not increase with WHO prevalence in country of origin, and was highest in migrants from countries with an estimated WHO prevalence of between 150 and 349 per 100,000 population (225 per 100,000; 95%CIs: 192, 264).

Table 7. Baseline characteristics of applicants screened for tuberculosis and prevalence of primary and secondary outcomes per 100,000 individuals screened.

	N (%)	Bacteriologically confirmed (95%CIs)	Culture positive (95%CIs)	Smear positive (95%CIs)
All	476455 (100%)	92 (84, 101)	83 (75, 92)	55 (49, 62)
Age group				
0-15	18729 (3.9%)	37 (18, 78)	37 (18, 78)	11 (3, 43)
16-44	444579 (93.3%)	92 (83, 101)	83 (75, 92)	53 (47, 60)
45-64	10413 (2.2%)	134 (80, 227)	115 (65, 203)	163 (102, 262)
>65	2734 (0.6%)	329 (172, 631)	293 (147, 584)	256 (122, 536)
Sex				
Female	167393 (35.1%)	116 (101, 134)	108 (93, 125)	83 (70, 98)
Male	309062 (64.9%)	79 (70, 90)	70 (61, 80)	40 (33, 47)
Family contact with infectious case of TB				
No	475216 (99.7%)	89 (81, 98)	80 (72, 88)	53 (47, 60)
Yes	1239 (0.3%)	1372 (855, 2201)	1211 (732, 2003)	726 (379, 1393)
Visa Type				
Student	281703 (59.1%)	85 (75, 96)	76 (66, 86)	52 (44, 61)
Settlement and Dependent	160436 (33.7%)	108 (93, 125)	99 (85, 116)	60 (49, 73)
Work	14748 (3.1%)	88 (51, 152)	68 (36, 126)	102 (61, 169)
Working Holiday Maker	7380 (1.6%)	0 (0, 0)	81 (37, 181)	14 (2, 96)
Family Reunion	3389 (0.7%)	59 (15, 236)	59 (15, 236)	0 (0, 0)
Other	8799 (1.9%)	68 (31, 152)	57 (24, 136)	45 (17, 121)
CXR				
No abnormality	449401 (94.3%)	-	-	-
TB suspected	19654 (4.1%)	2234 (2036, 2450)	2010 (1822, 2216)	1308 (1158, 1476)
Abnormality not TB	7400 (1.6%)	-	-	-
WHO prevalence of TB in country of migration				
40-149	18910 (4.0%)	32 (14, 71)	11 (3, 42)	26 (11, 64)
150-349	67574 (14.2%)	225 (192, 264)	223 (190, 263)	200 (169, 236)
350+	389971 (81.9%)	72 (64, 81)	62 (55, 70)	31 (26, 37)
Year of examination				
2007	5489 (1.2%)	146 (73, 291)	128 (61, 267)	109 (49, 243)
2008	34343 (7.2%)	166 (128, 215)	154 (118, 202)	122 (90, 165)
2009	116899 (24.5%)	87 (72, 106)	71 (57, 88)	67 (53, 83)
2010	109356 (23.0%)	68 (54, 85)	56 (43, 72)	47 (35, 61)
2011	97455 (20.5%)	87 (71, 108)	82 (66, 102)	33 (23, 46)
2012	62338 (13.1%)	106 (83, 135)	103 (80, 131)	59 (43, 82)
2013	50575 (10.6%)	93 (70, 124)	93 (70, 124)	32 (19, 52)

Country of screening				
Burkina Faso	73 (0.0%)	-	-	-
Bangladesh	143154 (30.1%)	85 (71, 101)	77 (64, 93)	39 (30, 51)
Cambodia	621 (0.1%)	161 (23, 1144)	161 (23, 1144)	-
Cote D'Ivoire	1026 (0.2%)	-	-	-
Eritrea	152 (0.0%)	658 (92, 4684)	-	658 (92, 4684)
Ghana	18649 (3.9%)	32 (14, 72)	11 (3, 43)	27 (11, 64)
Kenya	12867 (2.7%)	101 (59, 174)	101 (59, 174)	39 (16, 93)
Laos	193 (0.0%)	-	-	-
Niger	36 (0.0%)	-	-	-
Pakistan	243243 (51.1%)	63 (54, 74)	52 (44, 62)	26 (20, 33)
Sudan	4025 (0.8%)	25 (4, 176)	25 (4, 176)	-
Somalia	2760 (0.6%)	181 (76, 435)	145 (54, 386)	109 (35, 337)
Togo	188 (0%)	-	-	-
Tanzania	4166 (0.9%)	120 (50, 288)	120 (50, 288)	24 (3, 170)
Thailand	45302 (9.5%)	291 (245, 346)	291 (245, 346)	283 (238, 336)

The crude prevalence of bacteriologically confirmed tuberculosis varied greatly between countries with the highest rates found in Eritrea (658 per 100,000; 95%CIs: 92, 4,684) and Thailand (291 per 100,000; 95%CIs: 245, 346; Figure 13). Adjusting for age and sex had a minimal impact on the estimates. Adjusted prevalence was highest in Eritrea at 556 per 100,000 individuals (95%CIs: 0, 1,659), but less than five cases were detected and therefore the confidence intervals were large for this estimate (Figure 14). Pakistan had the highest number of bacteriologically confirmed cases (156) and an adjusted prevalence of 63 per 100,000 individuals screened (95%CIs: 53, 73). Eritrea and Tanzania had age and sex adjusted estimates for bacteriologically confirmed tuberculosis that were consistent with WHO estimates of prevalence due to overlapping 95% confidence intervals with the 2010 country estimate. Thailand had an age and sex adjusted prevalence of bacteriologically confirmed tuberculosis detected at pre-entry screening greater than WHO prevalence country estimates in 2010. All other country estimates were lower than WHO prevalence country estimates in 2010.

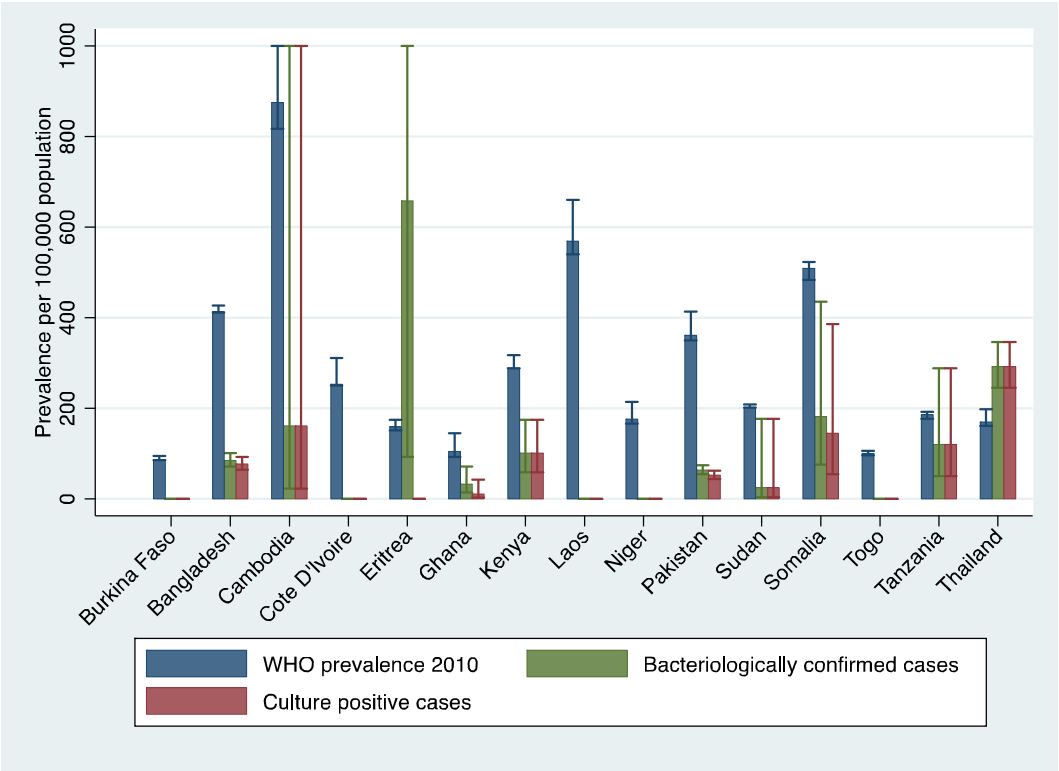

Figure 13. Crude prevalence of bacteriological and culture confirmed TB diagnosed at pre-entry screening compared to 2010 WHO country prevalence estimates.

Note: Error bars on bacteriological and culture confirmed tuberculosis estimates are 95%confidence intervals. Error bars on WHO 2010 prevalence country estimates are highest and lowest prevalence estimate for each country between 2007 and 2013. Confidence intervals limited to a maximum of 1,000 per 100,000 population.

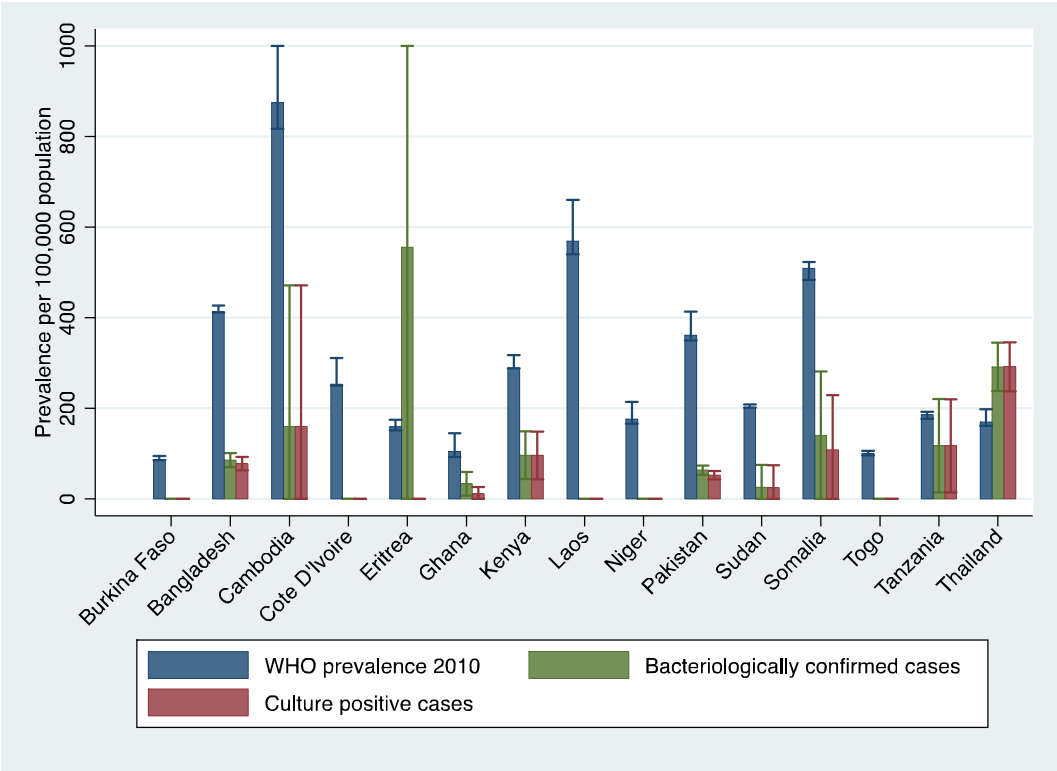

Figure 14. Age and Sex adjusted prevalence of bacteriological and culture confirmed TB diagnosed at pre-entry screening compared to 2010 WHO country prevalence estimates.

Note: Error bars on bacteriological and culture confirmed tuberculosis estimates are 95%confidence intervals. Error bars on WHO 2010 prevalence country estimates are highest and lowest prevalence estimate for each country between 2007 and 2013. Confidence intervals limited to a maximum of 1,000 per 100,000 population.

In order to examine the results of this study in the context of previously published data on pre-entry screening, the meta-analyses presented in chapter two was updated. The primary outcome of crude prevalence of bacteriologically confirmed cases across all countries was used in this updated meta-analysis (92 per 100,000 population screened; 95% CIs 84, 101). This crude estimate was found to be the second lowest of all published studies (Figure 15) and overlapped with the estimate from one other study.(90)

The meta-analysis of culture confirmed cases by country of origin was also updated, including all countries where more than 1,000 migrants had been screened (Figure 16). Compared to the meta-analysis in chapter two, the level of heterogeneity increased

when including estimates from this study, with prevalence of culture positive cases no longer increasing with prevalence of tuberculosis in the country of origin. The summary estimate of culture confirmed cases was highest in countries with a WHO estimate of prevalence of tuberculosis between 150-249 per 100,000 population (192 per 100,000 individuals screened; 95%CIs 170, 216). With the exception of Bangladesh and Pakistan, 95% confidence intervals for the estimates of culture confirmed disease overlapped with published estimates within each stratified subgroup.

To determine the risk factors for bacteriologically confirmed tuberculosis in migrants at the time of pre-entry screening, a multivariable logistic regression was conducted (Table 8). After adjusting for age, sex and clustering by individual, there was strong evidence that having contact with a case of tuberculosis (OR 11.6; 95%CIs: 7.0, 19.3; p-value <0.001) was associated with an increased risk of bacteriologically confirmed tuberculosis at pre-entry screening. Migrants screened in countries with a WHO prevalence of 40-149 per 100,000 population were at reduced risk of bacteriologically confirmed tuberculosis at pre-entry screening (OR 0.1; 95%CIs 0.1, 0.3; p-value <0.001), as were those from countries with a prevalence greater than 350 (OR 0.3; 95%CIs 0.3, 0.4; p-value <0.001) compared to migrants from countries with a prevalence of 150-349 after adjusting for age and sex. Migrants on a settlement and dependant visa also had a higher risk after adjusting for age and sex (OR 1.3; 95%CIs 1.0, 1.6; p-value 0.02).

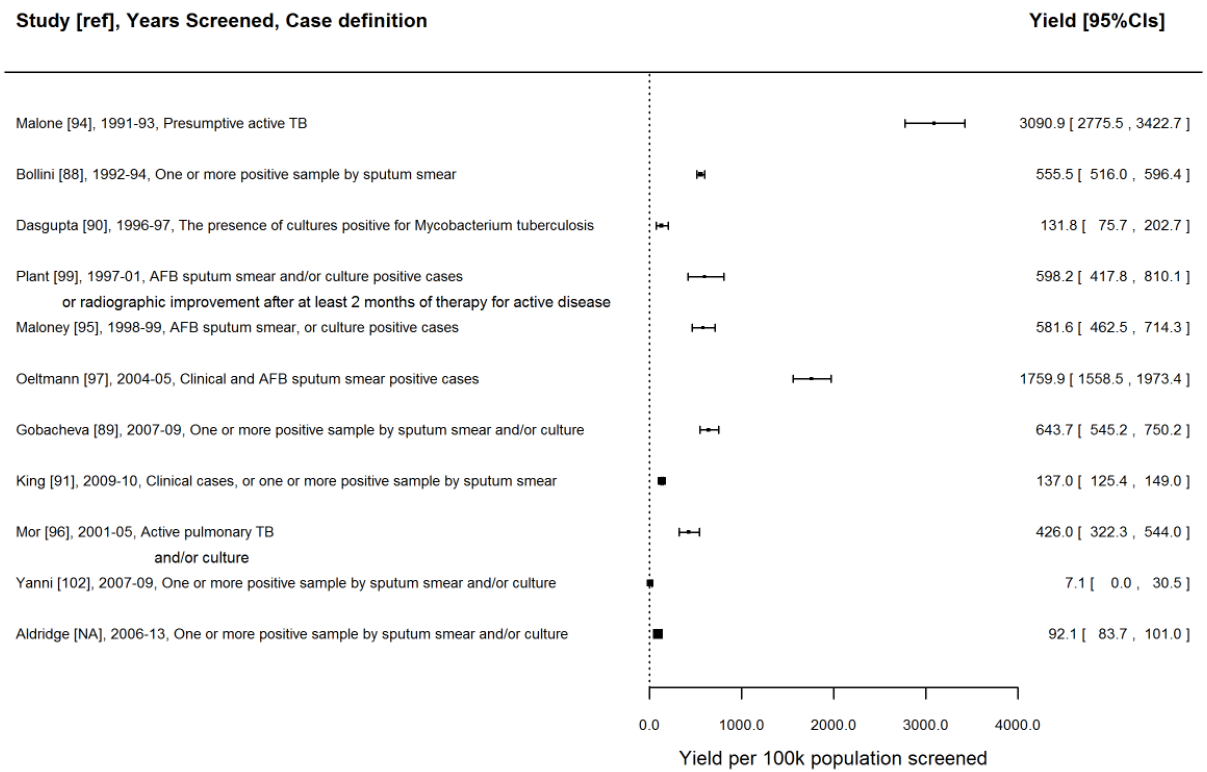

Figure 15. Forrest plot of pre-entry screening programme prevalence for principal outcome of active tuberculosis cases found by each study (case definition varies between studies, sorted by year of publication).

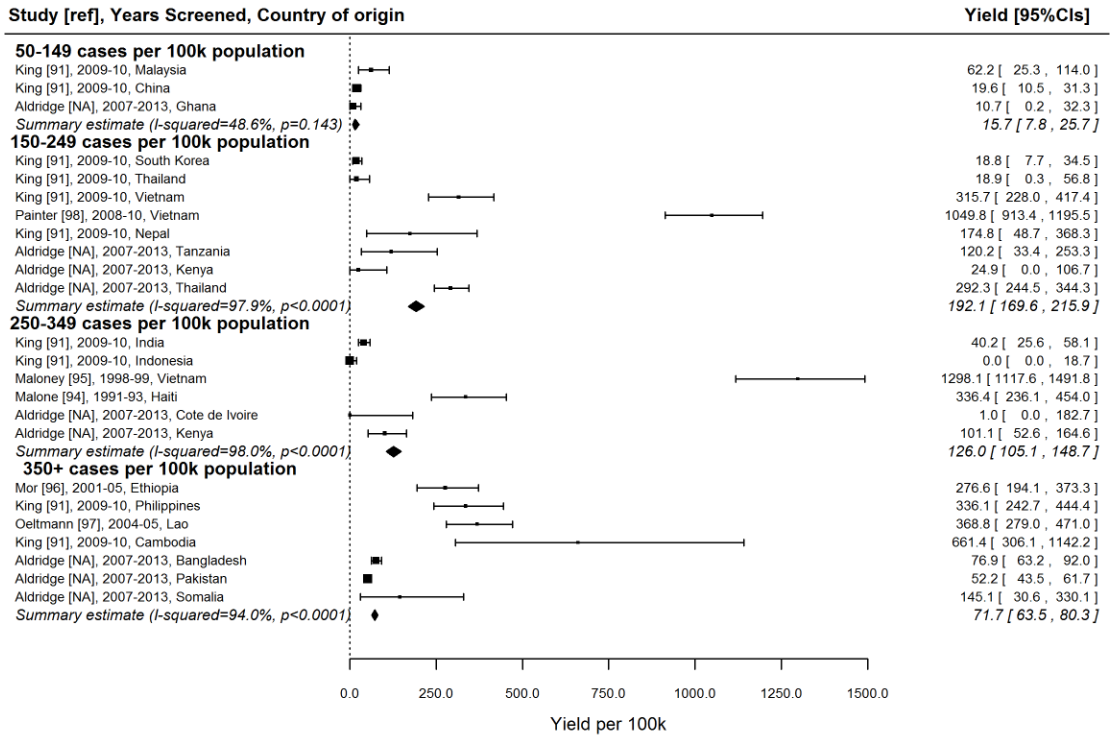

Figure 16. Forrest plot of pre-entry screening programme prevalence of culture positive cases of tuberculosis, stratified by WHO prevalence of tuberculosis in country of origin (sorted by prevalence in country of origin).

Table 8. Multivariable analysis examining risk factors for bacteriologically confirmed tuberculosis

Risk Factor	Univariable IRR (95% CIs)	Multivariable IRR (95%CIs)	p-value
Age			
0-15	0.4 (0.2, 0.9)	0.3 (0.2, 0.7)	0.01
16-44	1.0	1.0	
45-64	1.5 (0.9, 2.5)	1.2 (0.7, 2.0)	0.56
>65	3.6 (1.9, 6.9)	3.2 (1.6, 6.3)	<0.001
Sex			
Female	1.0	1.0	
Male	0.7 (0.6, 0.8)	1.0 (0.8, 1.3)	0.73
Contact with case TB			
No	1.0	1.0	
Yes	15.7 (9.6, 25.5)	11.6 (7.0, 19.3)	<0.001
Visa			
Students	1.0	1.0	
Settlement and dependents	1.3 (1.0, 1.5)	1.3 (1.0, 1.6)	0.02
Work	1 (0.6, 1.8)	0.9 (0.5, 1.6)	0.73
Working holiday maker	1 (0.4, 2.2)	1.2 (0.5, 2.8)	0.63
Family reunion	0.7 (0.2, 2.8)	0.4 (0.1, 1.7)	0.21
Other	0.8 (0.4, 1.8)	0.9 (0.4, 2.1)	0.84
WHO category			
40-149	0.1 (0.1, 0.3)	0.1 (0.1, 0.3)	<0.001
150-349	1.0	1.0	
350+	0.3 (0.3, 0.4)	0.3 (0.3, 0.4)	<0.001

A sensitivity analysis was conducted to examine the prevalence rates of the primary and secondary outcomes when including all migrants screened pre-entry, not just those seen at clinics where culture and smear testing was performed on all sputum samples collected (See Appendix 5 for all results). A total of 692,232 migrants were screened under this protocol and the overall rate of bacteriologically confirmed tuberculosis was lower at 75 (95%CIs: 69, 82; Table 31). In a multivariable analysis adjusted for age and sex, being screened at a site where sputum culture testing was performed on all samples was associated with increased odds of being bacteriologically confirmed as a case of tuberculosis (OR 2.4; 95%CIs: 1.9, 3.0; p-value <0.001; Table 32). Eritrea still had the highest crude rate of bacteriologically confirmed tuberculosis (Figure 54), and after

adjusting for age and sex, only Thailand had bacteriologically confirmed tuberculosis rates higher than WHO country prevalence estimates in 2010, Eretria had results that were consistent with WHO levels, and all other countries had rates lower than WHO estimates (Figure 55).

3.5 Discussion

Between 1st October 2005 and 31st December 2013 nearly seven hundred thousand pre-entry screening episodes for tuberculosis were conducted. Almost five hundred thousand of these migrants were screened at locations using culture testing of sputum samples. The overall crude prevalence of bacteriologically confirmed tuberculosis was 92 per 100,000 population screened and migrants over the age of 65, those who had been in contact with an infectious case of tuberculosis, and those with a chest radiograph classified as suspected tuberculosis had the highest crude prevalence of bacteriologically confirmed tuberculosis. After adjusting for age and sex, all countries except Thailand had a prevalence of tuberculosis detected at pre-entry screening that was consistent with or lower than WHO population estimates in 2010. After adjusting for age and sex, migrants that had a history of a close contact with a case of tuberculosis and applicants screened in countries with a WHO prevalence of between 151-349 per 100,000 population, and those on settlement and dependant visas were all associated with an increased risk of being detected with bacteriologically confirmed tuberculosis at pre-entry screening. In a sensitivity analysis that included all migrants screened pre-entry, those individuals screened at a clinics where culture and smear testing was performed on all sputum samples collected were associated with an increased risk of being detected with bacteriologically confirmed tuberculosis.

3.5.1 Strengths and weaknesses of the study

There were several strengths to this study including the large sample size and the fact that the data were highly representative of long-term migrants from the 15 countries taking part in the pre-entry screening pilot, as a result of the compulsory nature of the process for all migrants applying to stay for six months or more. The use of WHO definitions for primary and secondary outcomes means that data were internationally comparable. After excluding those sites were culture testing was not routinely being

performed, screening across different sites should be consistent as a result of the quality assured UK technical instructions that all clinics must conform to.(7) These guidelines specify in detail how screening should be undertaken and include standardised reporting guidelines for chest radiographs based on the Canadian immigration system for the classification of tuberculosis.(8) The technical instructions should therefore reduce measurement error and misclassification bias for exposures and outcomes, including in the interpretation and reading of chest radiographs due to the established classification system used. Prevalence estimates included only the first screen for individuals with duplicate screens (less than 12 months apart) that were repeatedly positive as these were not newly prevalent cases. These records represent migrants with tuberculosis detected at the initial screen who were undergoing repeat screening but found still to have tuberculosis. Country estimates for the prevalence of bacteriological and culture confirmed tuberculosis in migrants were adjusted by age and sex, and therefore comparable across countries in this study for these two confounding factors.

A major limitation of this study is that whilst it is highly representative of migrants staying for longer than six months in the UK, it does not include data on undocumented migrants, refugees and those on short visas. Undocumented migrants and refugees in particular will be at a higher risk of tuberculosis compared to the individuals included in this dataset for complex reasons including malnutrition, history of living in overcrowded situations such as refugee camps, higher rates of HIV, and a disruption in access to health services.(111–116) A large proportion of migrants to the UK on visas greater than 6 months were also likely to come from higher socio-economic groups in their country of origin. These biases (compared to a representative country random sample) were likely to account for some of differences between prevalence estimates in this analysis and WHO country prevalence estimates.

An alternative explanation for the difference between WHO prevalence estimates and country prevalence estimates in this study is the potential for misclassification bias in the WHO country data. WHO estimates were modelled using a beta distribution from country level case notification data (where available), prevalence surveys, expert opinion, and UNAIDS estimates of HIV. Recent papers have examined in detail why WHO estimates were different to those estimated by the Global Burden of Disease study, which in general found lower estimates of incidence and prevalence, and some of

these issues about parameterisation and modelling strategies are proposed as potential explanations.(32,110,117) In particular, the country level case notification data used by WHO is extremely variable, and this may explain why prevalence did not increase with WHO prevalence and why those from countries with prevalence between 151-349 per 100,000 population had the highest risk of being detected as a bacteriologically confirmed case at pre-entry screening.

The number of variables included in the risk factor analysis was limited as the data used was primarily collected for operational and not epidemiological purposes. Information on risk factors for tuberculosis, including potentially important exposures such as social deprivation, HIV, immunosuppressive drugs and social risk factors such as history of drug use or imprisonment were not available to include in these analyses.(118–124) Unmeasured confounding may therefore also explain discrepancies in the rates of tuberculosis found by the pre-entry screening process and WHO estimates of disease. Drug sensitivity testing on culture positive samples was recorded in very few cases, and there were a low number of clinically confirmed cases, limiting the conclusions that can be drawn on the basis of these data. Over half of the migrants in this analysis came from Pakistan, and as a result the risk factor analysis will be influenced highly by factors found to be important in this population, but not measured in this study such as overcrowding, poverty, and language spoken.(125–127) No data were available on the site of diagnosed disease, and therefore it was not possible to provide prevalence estimates for pulmonary or extra-pulmonary disease.

3.5.2 Strengths and weaknesses in relation to other studies

The updated meta-analyses comparing overall cases of active tuberculosis identified by this pre-entry programme demonstrates that compared to 10 other published studies, the estimates of prevalence of screening (detected by any method) presented here were lower than all other studies except one.(102) There are several potential explanations for this finding including the fact that descriptions of how cases were identified and classified as having tuberculosis in these other published studies was not always clear, therefore limiting the ability to make strong comparisons.(128) An updated meta-analysis including only culture confirmed cases of tuberculosis found that there was only evidence that two countries (Bangladesh and Pakistan) had a lower prevalence of

culture confirmed tuberculosis in migrants screened pre-entry when compared to other published studies that had presented data that allowed stratification into subgroups by WHO population prevalence estimates of tuberculosis. As Bangladesh and Pakistan accounted jointly for over 80% of all migrants in the study, it is therefore also not surprising that the combined rate across all countries was highly influenced by these two countries when combined in this simplistic manner.

Results from this study may differ to previous literature on pre-entry screening for several additional reasons other than those discussed above. Firstly, the data presented are for migrants intending to stay in the UK for a minimum of six months, either as a student, to work, or as a family reunion. An extremely large proportion of migrants screened were students or young working age adults. No data were available from the other published studies that would allow adjusted estimates to be compared to take of differences such as age and socio-economic status. If greater proportion of older migrants were included in these previously published studies then it is not surprising that they found a higher prevalence of tuberculosis detected at pre-entry screening compared to the large number of students in this analysis, who will be younger, healthier, from a higher socio-economic status (by virtue of the fact they can afford to study in the UK), all of which reduce their risk of tuberculosis. Secondly, not all other studies provided exact details of how culture confirmation was performed, and one large study highlighted the fact that this may not have been uniform across screening sites.(91) Due to the UK technical instructions, such variability should not be an issue as once culture testing was introduced it would have been carried out consistently and continuously by screening centres, and subject to quality assurance inspections by Public Health England. Finally, the rules used to exclude duplicates in this analysis were also likely to vary compared to those in other studies. It was not clear from these other published studies whether repeat screens in an individual were excluded or not and therefore differences in the way these duplicate screens were handled could also have accounted for some of the differences found.

3.6 Conclusion

The results of this study provide a comprehensive analysis of the historical data from a pilot pre-entry screening programme that has been running since 2005. Restricting the

data to sites where culture testing was routinely performed provides the strongest basis for which to make recommendations for the improvement of pre-entry screening, as it is comparable to current processes in the UK technical instructions. There was a great deal of variation in the prevalence of bacteriologically confirmed tuberculosis by country, with crude prevalence highest in Cambodia, Eritrea, Somalia and Thailand. Except for Thailand, prevalence estimates in migrants were lower than WHO country prevalence estimates, which is likely explained by the fact that migrants to the UK represent a higher socio-economic group compared to the whole population in the country of origin. The high number of students is likely to have had a significant influence on the results presented, as this group had the second lowest of bacteriologically confirmed tuberculosis. A lack of socio-economic and clinical risk factors within this dataset limited the ability to explore this issue further. The implications for tuberculosis control, recommendations, and directions of future research are discussed in the final chapter of the thesis.

CHAPTER 4

Accuracy of probabilistic linkage using the Enhanced Matching System for public health and epidemiological studies.

4.1 Abstract

Background: The Enhanced Matching System (EMS) is a probabilistic record linkage program developed by the tuberculosis section at Public Health England to match data for individuals across datasets. This chapter outlines how EMS works and investigates its accuracy for linkage across public health datasets.

Methods: To examine the accuracy of EMS, two public health databases were matched exactly using NHS number as a gold standard unique identifier. Probabilistic linkage was then performed on the same two datasets without inclusion of NHS number.

Results: Exact matching using NHS number between two datasets (containing 5931 and 1759 records) identified 1071 record pairs. EMS probabilistic linkage identified 1068 record pairs. The sensitivity of probabilistic linkage was calculated as 99.5% (95%CI: 98.9, 99.8), specificity 100.0% (95%CI: 99.9, 100.0), positive predictive value 99.8% (95%CI: 99.3, 100.0), and negative predictive value 99.9% (95%CI: 99.8, 100.0). Probabilistic matching was most accurate when including address variables with manual review, but performed well without manual review, without address information and in a dataset only containing non-UK born individuals.

Conclusion: The Enhanced Matching System examined in this chapter has been found to have high accuracy for the linkage of public health datasets that do not contain a unique identifying variable and in non-UK born individuals.

4.2 Introduction

The routine collection of electronic health records provides unique opportunities to investigate important research questions in an efficient and powerful way by linking individuals across datasets collected by different organisations. Record linkage has been performed for a number of years in various epidemiological study designs including case control, cohort studies, capture recapture studies and economic evaluations.(129–132)

In a majority of studies, three methods have been used to match records between datasets: Exact matching, deterministic matching, and probabilistic linkage. Exact matching requires records within the two data sets to contain a universally available and unique identifying variable. Many databases across health and social care do not contain such a unique and universally available variable, or accurate and fully available personal identifiable information, limiting the ability to perform exact matching. Deterministic matching is defined as: "Record linkage of two (or more) files based on exact agreement of matching variables"(133) and has many variants, but typically contains a series of rules to identify record pairs using combinations of personal identifiers such as first name, surname, date of birth and sex. Probabilistic linkage is defined as: "Record linkage of two (or more) files that utilizes the probabilities of agreement and disagreement between a range of matching variables".(133)

The Enhanced Matching System (EMS) is a probabilistic record linkage program developed to combine data for individuals across datasets or within a single dataset for the purposes of de-duplication. EMS was developed over several years and can be configured with ease for different matching projects.

EMS was designed and developed by the tuberculosis section at Public Health England and builds upon the classic methods described by Newcombe.(134,135) EMS is used operationally by the tuberculosis section in Public Health England for many types of analysis including measuring the levels of drug resistance in tuberculosis cases notified in the UK, and establishing the amount transmission among these cases.(136) Historically, probabilistic linkage has been necessary for this work due to the low recording rates of a unique identifier between the two datasets (Case notifications of tuberculosis to Public Health England and culture positive isolates from tuberculosis

reference laboratories across UK) used to establish these estimates. These datasets are probabilistically linked and de-duplicated to form the Enhanced Tuberculosis Surveillance (ETS) database.

A potential new application of EMS is to use it for linking data on individuals entering the UK that have been screened by the pre-entry tuberculosis screening programme. At present the pre-entry screening database does not contain a unique identifying variable such as NHS number, and therefore any linkage must be performed using variables such as first name, surname, date of birth and sex. The accuracy of EMS linkage has not been previously investigated. This chapter outlines the main features of EMS and presents an analysis that examines its accuracy at matching two public health tuberculosis datasets.

4.3 Methods

4.3.1 Enhanced Matching System

EMS is a configurable Microsoft SQL Server database program, currently implemented on Windows 7 and SQL Management studio 2012 and written using the Transact-SQL programming language. The first step in using EMS is data mediation, whereby variables are converted into a pre-specified EMS table structure. To increase the accuracy of the matching, data are then standardised by splitting and parsing of forenames, postcodes, date of birth and addresses, cleaning of country and hospital names when available (for example, through the removal of erroneous spaces at the beginning or end of these country or hospital names), and generating Soundex codes (an algorithm generated index based on the way a name sounds(137)) for name variables. Address information, where available, is split into house or flat number, street name, town or city, postcode and country. EMS is capable of undertaking de-duplication, but this is not discussed in this chapter.

Pairing and blocking is used by EMS to improve the efficiency of the matching process by reducing the number of comparisons required. Depending on available fields within the databases to be matched, blocking involves EMS breaking the records into smaller blocks in three ways: records having the same surname soundex, year of birth, or postcode area. Data are then grouped (or blocked) and links within these groups are examined. Blocks are combined at the end of the linkage using the SQL command

'UNION' that removes any duplicate pairs and merges data into a final single output table. Blocking reduces the number of record comparisons between two datasets. Three way blocking is used because, for example, if year of birth is missing for one record pair, this record pair will still be linked in one of the other two blocks (surname soundex or postcode district) as long as data are also not missing on these variables in both datasets.

Probabilistic linkage relies on the generation of weights to identify agreements and disagreements between the identifying set of fields in two datasets.(138) EMS generates a weight for matching fields based on the m probability (the probability that the matching variable agrees given that the comparison pair being examined is a match(133)) and the u probability (the probability that a matching variable agrees given that the comparison pair being examined is a non-match(133)). For example, the probability of random agreement for month of birth in two records that are not a true match is approximately 0.08 (or 1/12), which corresponds to the u probability. The m probability depends mainly on the data quality for a matching field. A typical m probability, for an outcome of agreement, is around 0.9 indicating that 90% of matches (for two records are in fact a true match) will have the same value for this matching field. The m probabilities at the start of a matching project are estimates, and several matching runs can be performed within EMS in order to refine values.

Weights are then calculated for each matching field in a pair of records, depending on whether they agree or disagree (Equation 1). The weight for each possible matching pair field is the logarithm to the base 2 ($\log_2$) of the likelihood ratio. Logarithms to the base 2 are used in probabilistic linkage as per information theory convention.(139)

$$W = \log_2\left(\frac{m\ probability}{u\ probability}\right)$$

Equation 1. Formula for calculating individual weights for agreed matches. The m probability and u probability are replaced by (1 – m probability) and (1- u probability) for disagreement matches.

A weight of zero is used when one or both of the fields have missing or unknown values. Where the m probability is greater than the u probability, the weight is positive, and where the 1 – m probability is less than the 1 – u probability then the weight is

negative. Therefore agreements are generally positive and disagreements generally negative. EMS carries out matching across multiple fields under the assumption that they are independent. A total weight is obtained for all matching fields in each record pair, as per Equation 2. The logs for individual weights are summed, a process equivalent to multiplying the likelihood ratios.

$$W = \sum W_i$$

Equation 2. Formula for calculating total weight for a record pair

The prior probability of a random pair matching is then calculated (Equation 3(134)) and a threshold is calculated (Equation 4 adapted from the work by Newcombe(134)), above which a record pair is considered matched, using a pre-specified positive predictive value. This threshold is calculated independently of the blocking and is therefore based on all possible pairs, and not the actual number of pairs compared during the matching process.

$$P(M) = \frac{m}{n}$$

where:

m = estimated number of matches

n = possible matching pairs = $(n_1 \cdot n_2) / z$

n_1 and n_2 = dataset sizes (obtained directly from the data)

y = number of years

z = correction factor for restrictions on matching, and where z =

$1 - \ all\ matches\ allowed\ (i.e.no\ effect);$

or

$$\frac{y^2}{(2y-1)} \approx \frac{y}{2} - \ matches\ restricted\ to\ within\ 1\ year;$$

or

$$y - \ matches\ further\ restricted\ to\ single\ core\ year.$$

Equation 3. The prior probability of a random pair matching

$$Threshold = \log_2\left(\frac{P(U)}{P(M)}\right) + \log_2\left(\frac{ppv}{1-ppv}\right)$$

$$P(U) = prior\ probability\ of\ not\ matching\ =\ 1 - P(M)$$

ppv = positive predictive value, the probability that record pairs with a total weight above the threshold are truly matches and is specified by the user performing the matching.

Equation 4. Threshold formula

Records below the threshold are considered unmatched and human (manual) review can then be carried out to examine matches close to this threshold (above and below) in order to identify false positives and false negative.

4.3.2 Accuracy of EMS

Two datasets were used to examine the accuracy of EMS: case notifications of tuberculosis to Public Health England and a laboratory database of all bacteriologically confirmed isolates from tuberculosis reference laboratories from England, Wales and Northern Ireland. Case notifications are made to Public Health England by healthcare workers looking after patients with tuberculosis and include demographic and clinical details of cases. The laboratory database contains basic demographic, address information as well as mycobacterial species and drug susceptibility testing results for positive tuberculosis specimens. It should be noted that the case notifications dataset referred to in this chapter is not the same as what is referred to as ETS throughout the rest of the thesis. Instead, the linkage between the two datasets described in this chapter (case notifications and laboratory data) is part of what ultimately contributes to forming ETS. Therefore the case notifications data is not referred to as ETS throughout this chapter.

All case notifications with a NHS number for the calendar year 2012 were used for this analysis, along with all laboratory database records with an NHS number for the period 1st October 2011 – 31st March 2013. Laboratory database records for the three months before and after the calendar year of 2012 were included as tuberculosis cases may be

notified before or after the microbiological result. This strategy therefore aims to increase the number of possible matches across the two datasets. The laboratory database was de-duplicated (by linking to itself using EMS) before it was linked to the case notifications database.

4.3.3 Gold standard comparator

Exact matching using NHS number was used as a gold standard to identify linked records. NHS number is a ten digit unique identifier for a patient and is used throughout the health service. It is often formatted 3-4-4, with separating space or hyphen (e.g. 123-4567-8901). NHS numbers included in the final analysis as the gold standard were checked for validity. Simple descriptive analysis was performed to examine differences between tuberculosis notification and laboratory isolate records with and without NHS numbers.

4.3.4 Probabilistic matching

After exact matching, case notifications and the laboratory database were probabilistically linked by EMS using first name, surname, date of birth, sex, address details (including postcode) data, first name soundex and surname soundex. NHS number was not included as a matching variable in the probabilistic linkage in order that the matching was independent of this gold standard identifier. Blocking was performed on surname, year of birth and postcode. A descriptive analysis stratified by NHS number availability and validity was performed to examine missing data on variables used for the linkage from the laboratory and case notifications datasets. All records except for those with a missing or invalid NHS number were included in the accuracy analysis.

The matching threshold was calculated using a value of y=1 and ppv =0.99. A decision was taken a-priori to manually review records with a matching score of 10 above and below the matching threshold. As probabilistic linkage was performed without including NHS number as matching variable, manual review included all records regardless of whether or not they were exact matched by NHS number. Manual review was also performed without data on NHS number for each record pair.

4.3.5 Primary outcomes and statistical analysis

Outcomes used to assess accuracy of the matching process were sensitivity, specificity, positive and negative predictive values. A full description of how these were calculated is provided in Table 33 in Appendix 6. Exact confidence intervals (to the 95% level) were calculated in Stata version 13 using a binomial distribution.

4.3.6 Sensitivity analyses

Four sensitivity analyses were carried out. Firstly, probabilistic linkage was performed using first name, surname, date of birth, sex, address details (including postcode), first name soundex, and surname soundex but without manual review. Secondly, probabilistic linkage was performed without address variables and without manual review. Thirdly, to examine the effect that a larger proportion of non-English names would have on the accuracy, matching was performed without address variables and manual review, but in a case notifications dataset that only included non-UK born individuals. This analysis most closely replicates the linkage that would be performed between ETS and IOM pre-entry screening data. Fourthly, a sensitivity analysis was carried out to determine the impact of varying the automatically calculated weight threshold on outcome measures without manual review, but including all matching variables except NHS number.

4.3.7 Ethics approval

Ethical approval was not required for this accuracy analysis, as Public Health England has Health Research Authority approval to hold and analyse national surveillance data for public health purposes under Section 251 of the NHS Act 2006.

4.4 Results

A total of 8,751 records were extracted from the case notifications dataset and 7,538 unique records from the laboratory database. 67.8% of case notifications and 23.3% of the laboratory database records contained valid NHS numbers. Comparing the characteristics of records with and without valid NHS number in the laboratory dataset showed differences in age and higher rates of Isoniazid resistance in those records without an NHS number (Table 9). No other differences were found.

A greater number of differences were seen between those records with and without a valid NHS number in the case notifications dataset. All variables apart from those for drug sensitivity testing and site of tuberculosis disease showed a difference between those with and without an NHS number (Table 10). These data show that women, ethnic minority groups, individuals not born in the UK, and individuals with at least one social risk factor for tuberculosis (including drug use, homelessness, alcohol misuse/ abuse, prison) were more likely not to have an NHS number.

Table 9. Descriptive analysis of laboratory dataset for records with and without an NHS number.

		NHS Number				
		Available and valid		Not available or invalid		
	All	N	%	N	%	p-value*
All	7538	1759	23.3	5779	76.7	
Age group in years						
0 to 14	122	40	32.8	82	67.2	
15 to 44	4724	990	21.0	3734	79.0	
45 to 64	1576	409	26.0	1167	74.0	
65 and over	1061	320	30.2	741	69.8	<0.001
Missing**	55	0	0	55	100.0	
Sex of case						
Female	2941	726	24.7	2215	75.3	
Male	4355	1012	23.2	3343	76.8	
Missing	242	21	8.7	221	91.3	0.15
Isoniazid sensitivity result						
Resistant	508	95	18.7	413	81.3	
Sensitive	6801	1629	24.0	5172	76.0	
Missing	229	35	15.3	194	84.7	0.007
Ethambutol sensitivity result						
Resistant	84	19	22.6	65	77.4	
Sensitive	7217	1699	23.5	5518	76.5	
Missing	237	41	17.3	196	82.7	0.84
Rifampicin sensitivity result						
Resistant	142	29	20.4	113	79.6	
Sensitive	7181	1697	23.6	5484	76.4	
Missing	215	33	15.3	182	84.7	0.37
Pyrazinamide sensitivity result						
Resistant	108	26	24.1	82	75.9	
Sensitive	7161	1690	23.6	5471	76.4	
Missing	269	43	16.0	226	84.0	0.91

*Chi squared test, not including missing data for each variable other than NHS number

**It was not possible to calculate the exact age for these records as the date of their laboratory result was not recorded, but date of birth was available for all records.

Table 10. Descriptive analysis of case notifications dataset for records with and without an NHS number.

		NHS Number				
		Available and valid		Not available or invalid		
	All	N	%	N	%	p-value*
Total	8751	5931	67.8	2820	32.2	
Age						
<14	414	277	66.9	137	33.1	
15-44	5291	3495	66.1	1796	33.9	
45-64	1830	1273	69.6	557	30.4	
65+	1216	886	72.9	330	27.1	<0.001
Sex						
Female	3706	2619	70.7	1087	29.3	
Male	5045	3312	65.6	1733	34.4	<0.001
Ethnic group						
White	1814	1316	72.5	498	27.5	
Black-Caribbean	175	121	69.1	54	30.9	
Black-African	1358	859	63.3	499	36.7	
Black-other	71	41	57.7	30	42.3	
Indian	2295	1471	64.1	824	35.9	
Pakistani	1418	1098	77.4	320	22.6	
Bangladeshi	320	194	60.6	126	39.4	
Chinese	95	67	70.5	28	29.5	
Mixed/other	979	625	63.8	354	36.2	
Missing	226	139	61.5	87	38.5	<0.001
UK Born						
No	6125	4049	66.1	2076	33.9	
Yes	2256	1652	73.2	604	26.8	
Missing	370	230	62.2	140	37.8	<0.001
Site of disease						
Extra-pulmonary disease only	4095	2754	67.3	1341	32.7	
Pulmonary, with or without extra-pulmonary disease	4563	3128	68.6	1435	31.4	
Missing	93	49	52.7	44	47.3	0.20
Social risk factor **						
No	7683	5210	67.8	2473	32.2	
Yes	637	390	61.2	247	38.8	
Missing	431	331	76.7	100	23.3	<0.001

Isoniazid sensitivity result						
Sensitive	4801	3206	66.8	1595	33.2	
Resistant	351	235	67.0	116	33.0	
Missing	3599	2490	69.2	1109	30.8	0.95
Ethambutol sensitivity result						
Sensitive	5087	3396	66.8	1691	33.2	
Resistant	51	35	68.6	16	31.4	
Missing	3613	2500	69.2	1113	30.8	0.78
Rifampicin sensitivity result						
Sensitive	5060	3377	66.7	1683	33.3	
Resistant	91	63	69.2	28	30.8	
Missing	3600	2491	69.2	1109	30.8	0.62
Pyrazinamide sensitivity result						
Sensitive	5043	3364	66.7	1679	33.3	
Resistant	45	33	73.3	12	26.7	
Missing	3663	2534	69.2	1129	30.8	0.35

*Chi squared test, not including missing data for each variable other than NHS number

**At least one social risk factor including drug use, homelessness, alcohol misuse/ abuse, prison

Table 11. Description of missing data on variables used for the linkage from the laboratory, case notifications and an example pre-entry screening dataset, by NHS number availability and validity.

	Missing data for linkage variables			
	NHS number available and valid		NHS number not available or invalid	
	N	%	N	%
Laboratory dataset				
All	1759	100%	5779	100%
First name	1	0%	13	0%
Surname	1	0%	0	0%
Date of birth	0	0%	0	0%
Sex	21	1%	173	3%
Address line 1*	232	13%	4513	78%
Address line 2**	1023	58%	5387	93%
Postcode	126	7%	3939	68%
Case notifications dataset				
All	5931	100%	2820	100%
First name	0	0%	0	0%
Surname	1	0%	0	0%
Date of birth	1	0%	1	0%
Sex	0	0%	0	0%
Address line 1*	13	0%	8	0%
Address line 2**	2918	49%	1302	46%
Postcode	6	0%	9	0%

*E.g. house number and street name
**E.g. city.

For assessment of accuracy, only records with an available and valid NHS number were included in the probabilistic linkage. The final probabilistic linkage dataset therefore consisted of 5,931 records from the case notifications database and 1,759 records from the laboratory database (Figure 17). In this final dataset used for the assessment of accuracy there was one record in the case notification dataset missing surname and date of birth, but not first name (Table 11). One record in the laboratory dataset was missing first name and surname information, and no records were missing date of birth. 13 (0.2%) records were missing the first line of their address in the case notifications database and 232 (15.1%) in the laboratory database. 6 (0.1%) records were missing

postcode information in the case notifications database, and 126 (7.2%) were missing in the laboratory dataset. 21 (1.1%) records were missing information on sex in the laboratory database, and none in the case notifications. With the exception of the address line 2 (typically city, town, or local area) in the case notifications dataset, records with a missing or invalid NHS number had slightly higher levels of missing data than those with a valid and available NHS number.

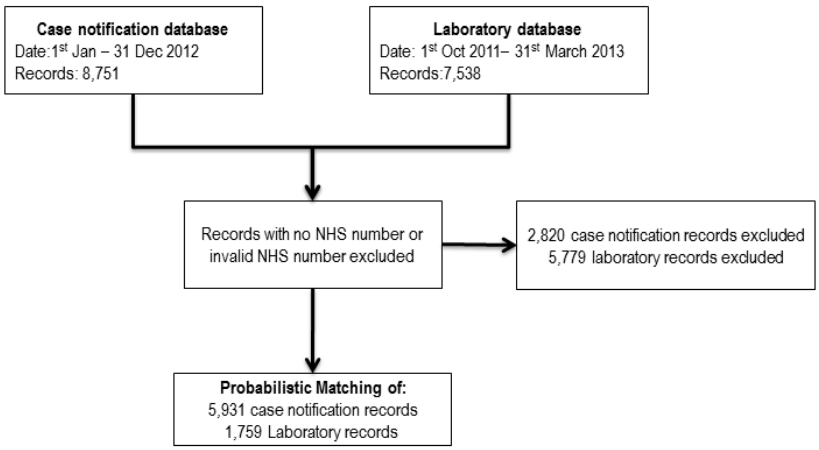

Figure 17. Flow chart of datasets used for study.

Exact matching between the two datasets using NHS number identified 1,071 matched pairs. In the case notifications database 4,860 (81.9%) records had no matching pair in the laboratory database, and 688 (39.1%) records from the laboratory database had no matching pair in the case notifications database.

Probabilistic linkage of the case notifications database to the laboratory database identified 1088 linked pairs using the EMS generated threshold of 19.98. Manual review of 67 pairs with a weight between 10 and 30 resulted in two below the threshold being changed to matches, and three results above the threshold being marked as not matching. A total of 19 records that represented multiple matches were removed (without manual review) – the pair with the highest weight was chosen as the final match to include in the analysis. A total of 1068 matches were therefore identified by the EMS process after manual review and de-duplication.

4.4.1 Accuracy of probabilistic linkage

Using the threshold of 19.98, and after manual review of records above and below the threshold and de-duplication of matches, 1,066 records were identified as true positives and 5,546 as true negatives (Table 12). At the threshold of 19.98, and with manual review, there were 2 false positives and 5 false negatives. The false negatives all had the same date of birth, several had first names and surnames that were switched, sex was unknown for one case, and all had different addresses. Both false positives had the same first name, surname, date of birth, sex and address, but different NHS numbers suggesting an error in the recording of NHS number. The sensitivity of the probabilistic linkage was 99.5% (95%CI: 98.9, 99.8) and specificity 100.0% (95%CI: 99.9, 100.0). The corresponding positive predictive value was 99.8% (95%CI: 99.3, 100.0) and negative predictive value 99.9% (95%CI: 99.8, 100.0).

Table 12. Comparison of matches identified by exact linkage using NHS number, and the probabilistic linkage process (without NHS number) and with de-duplication and manual review.

		Exact matching (NHS Number)		
		+ve	-ve	Total
Probabilistic (EMS)	+ve	1066	2	1068
	-ve	5	5546	5551
	Total	1071	5548	6619

Note: the total denominator is calculated using the number of exactly linked pairs (1071), plus the 4,860 records in the case notifications database had no matching pair in the laboratory database, and 688 records from the laboratory database had no matching pair in the case notifications database (i.e. 1071 + 4860 + 688 = 6619).

A series of sensitivity analyses were carried out to examine the performance of the linkage compared to a gold standard using different assumptions (Table 13). Without manual review, but using all linkage variables except NHS number, sensitivity was 99.3% (95%CI: 98.7, 99.7) and specificity 99.9% (95%CI: 99.8, 100.0). Matching without NHS number, address variables and manual review resulted in a sensitivity of 97.1% (95%CI: 95.9, 98.0) and specificity 100.0% (95%CI: 99.9, 100.0). To examine the effect of having a larger proportion of non-English names has on accuracy, matching was performed without address variables and manual review in a case

notifications dataset with only non-UK born individuals and found a sensitivity of 96.5% (95%CI: 94.9, 97.8) and specificity 100.0% (95%CI: 99.8, 100.0).

Varying the threshold between a credible range of 10 and 50 resulted in sensitivity changing from 86.1% (95%CI: 83.9, 88.1) to 99.6% (95%CI: 99.0, 99.9) and specificity ranging from 99.5% (95%CI: 99.3, 99.7) to 100.0% (95%CI: 99.8, 100.0; Table 14). The distribution of the weights for the matching process is shown in figure 18, which presents pairs with a total weight score greater than zero and therefore excludes the very large number of non-matches. The number of matches increased rapidly after a weight of around 50 and decreasing rapidly after the mode of 77.

Table 13. Calculation of sensitivity and specificity for probabilistic matching, without manual review, not including address variables and using a case notification dataset that only including non-UK born individuals.

Variables used for matching	Sensitivity	Specificity
All with manual review	99.5% (95%CI: 98.9, 99.8)	100.0% (95%CI: 99.9, 100.0)
All without manual review	99.3% (95%CI: 98.7, 99.7)	99.9% (95%CI: 99.8, 100.0)
No address variables (without manual review)	97.1% (95%CI: 95.9, 98.0)	100.0% (95%CI: 99.9, 100.0)
No address variables, only individuals born outside UK (without manual review)	96.5% (95%CI: 94.9, 97.8)	100.0% (95%CI: 99.8, 100.0)

Table 14. Sensitivity, specificity, positive predictive value and negative predictive value when varying the thresholds used to determine matched pairs without manual review.

Threshold weight score	True positives	Probabilistic matches	True negatives	Sensitivity	Specificity	Positive predictive value	Negative predictive value
10	1067	1094	5521	99.6%	99.5%	97.5%	99.9%
15	1065	1086	5527	99.4%	99.6%	98.1%	99.9%
20	1064	1069	5543	99.3%	99.9%	99.5%	99.9%
25	1060	1062	5546	99.0%	100.0%	99.8%	99.8%
30	1047	1049	5546	97.8%	100.0%	99.8%	99.6%
35	1022	1024	5546	95.4%	100.0%	99.8%	99.1%
40	987	989	5546	92.2%	100.0%	99.8%	98.5%
45	946	948	5546	88.3%	100.0%	99.8%	97.8%
50	922	924	5546	86.1%	100.0%	99.8%	97.4%

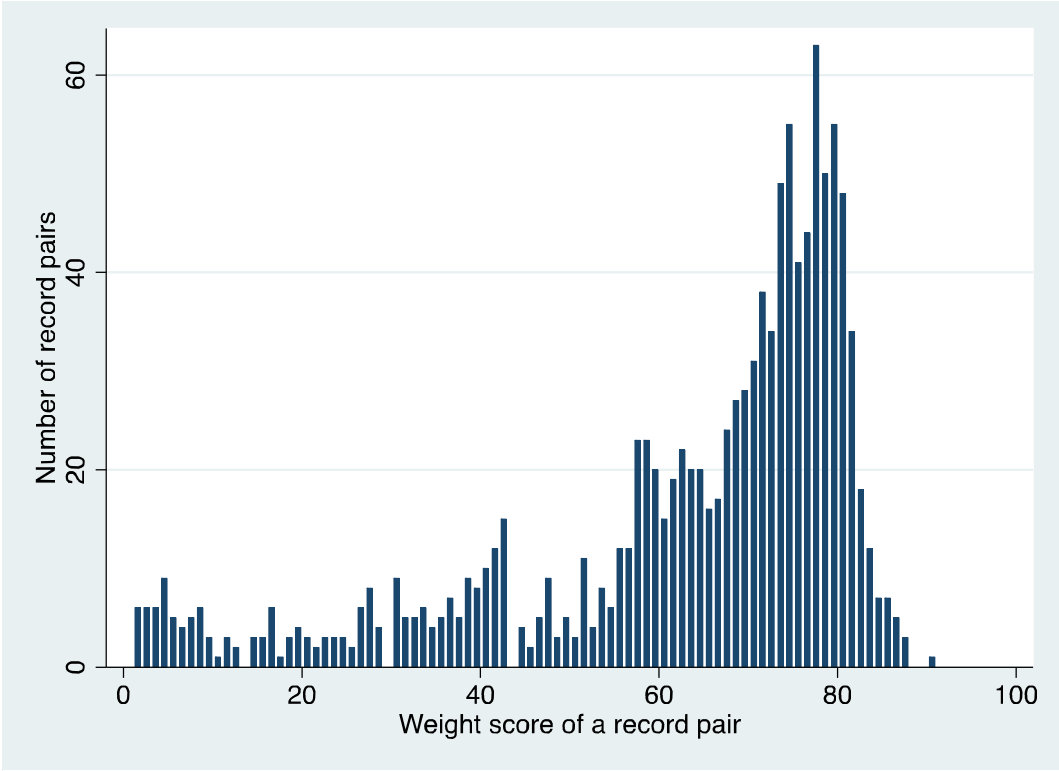

Figure 18. Number of pairs by total weight score, without manual review or de-duplication and not including NHS number. Only pairs with a total weight score greater than zero are presented.

4.5 Discussion

The Enhanced Matching System uses probabilistic linkage that was found to have high accuracy compared to gold standard exact matching based upon NHS number. Probabilistic linkage was most accurate when including address variables and using the automatically generated threshold for determining matches with manual review. Accuracy remained relatively high even after exclusion of address information from the linkage process, in linkage without manual review, and in a case notifications dataset that only included non-UK born individuals. Varying the weight threshold within a plausible range had minimal impact on the sensitivity and specificity of probabilistic linkage.

4.5.1 Strengths and weaknesses of this analysis

The characteristics of records in the case notification and laboratory datasets with missing or invalid NHS numbers were examined. As a result of the lack of demographic data in the laboratory database, the only differences found were for age and isoniazid resistance, which has been associated with an outbreak of tuberculosis in homeless, prison, certain ethnic groups and drug using populations in London.(140) In the case notifications database there were more differences for records with and without NHS numbers. NHS numbers were missing for more records in those aged between 15 and 44, males, ethnic minorities, non-UK born and individuals with social risk factors. It is not surprising that there were greater levels of missing NHS numbers for those with social risk factors, as these individuals tend to have poorer access and usage of NHS services.(141)

The dataset used for main analysis, which only included records with a valid NHS number, comprised of low levels of missing data on linkage variables and is therefore likely to represent the higher end of accuracy achievable by EMS. The datasets used in this study contained a high proportion of individuals not born in the UK, and ethnic minority groups making this analysis relevant to these populations. The sensitivity analysis examining probabilistic linkage in a case notifications dataset that only contained UK born individuals without address information found that the accuracy of EMS remained high. This provides reassurance that a linkage between the IOM pre-entry screening dataset and ETS would have a high level of accuracy.

NHS number is a unique identifying variable that is verifiable and reliable, making it an ideal gold standard comparator. However, it is not always available in public health datasets, demonstrated by the fact that it was available and valid for only 23.3% of laboratory database records, and 67.8% of case notification records. The high level of missing or invalid NHS numbers highlights a strength of probabilistic linkage over exact matching, in addition to the fact that it is able to account for errors and omissions of other data on linkage variables. However, the accuracy of probabilistic matching is still dependent on what identifying variables are available, and the quality of data contained within these variables. The amount of missing data on linkage variables (e.g. first name, surname, date of birth, sex and address) for those records with or without a valid NHS number was similar in the case notifications dataset, but the laboratory dataset had more missing data in those records without an NHS number. This may mean that linkage performs less well on such records, but this issue is complicated by the fact that the laboratory dataset may contain records that should not have an NHS number or address information (e.g. those isolated from animals). It is therefore difficult to generalise with certainty, whether linkage to the laboratory records without a valid NHS number will be lower than those with a valid NHS number.

Not all tuberculosis cases are microbiologically confirmed, and it is therefore not unexpected that the case notification dataset included more individuals than the laboratory database. Despite this, it is still possible that some patients recorded in the laboratory database were not notified as a case of tuberculosis and are therefore not included in the case notifications dataset. There are many reasons for laboratory records not to be found in the case notification dataset including those isolated from animals, cases reported to the case notification surveillance system in subsequent years after laboratory confirmation, cases of non-tuberculosis mycobacterium, samples positive due to laboratory contamination, and samples originating from the Channel Islands which are not notified to the surveillance system.

The accuracy of matching was assessed using a relatively small dataset. Repeating the analysis on a larger dataset may have some effect on the measures of accuracy found in this analysis, depending on the frequency of matches between the datasets. For this analysis, matching was performed between two data sets of the same disease, using several matching variables including names, date of birth and address data that had low

levels of missing data, and therefore a high degree of matching is to be expected. In scenarios where a small sample is being matched into a very large database (such as national records of hospital attendance) then the positive predictive value or sensitivity will drop according to the prevalence of the sample within the larger database. This may lead to overestimation of the frequency of occurrence of the sample within the larger database. If this linked dataset was used to calculate incidence or prevalence of an outcome, such a bias within the linked dataset would result in estimates were higher than the true values.(133)

Manual review of matches introduces subjectivity into the matching process and this may have implications for repeatability of this part of the linkage. Additionally, as the treating tuberculosis clinician does not perform the manual review (as is usually the case), those performing manual review were quite removed from the clinical situation and this may bias results. Such human error is likely to be differentially (and not randomly) biased. Further work should be carried out to assess the impact of the subjectivity of the manual review process, and examine the applicability of developing rules to provide consistent and potentially unbiased results.

4.5.2 Strengths and weaknesses in relation to other studies

When used for epidemiological studies, errors in probabilistic linkage have the potential to impact on findings and conclusions drawn. Linkage is typically used for the generation of outcome data in cohort studies, for example, to determine the vital status for individual participants by matching data into death registries. Assuming there is non-differential misclassification bias of exposure variables, false positive links will bias risk ratios and risk differences towards the null.(133,142,143) Risk ratios will be unaffected by false negative results (assuming there is non-differential misclassification bias); however, risk differences in cohort studies will be biased towards the null. False positive and negative probabilistic links in public health surveillance or outbreak studies will also result in under or over estimation of the number of cases. We are not aware of studies that have analysed this directly, but capture recapture studies attempt to examine this issue and in the UK have demonstrated the utility of the probabilistic linkage for improving data quality and case ascertainment levels.(144)

For this study, there were insufficient numbers of false positive and false negative results to enable examination of the issue of misclassification bias further, however, the fact that there was very high sensitivity and specificity means misclassification should introduce minimal bias. Further research is needed to understand the implications of these misclassification biases, particularly when such analyses are being conducted to estimate disease incidence or prevalence.

Probabilistic linkage has been widely adopted in research and service public health analyses. Several studies have previously examined the accuracy of probabilistic linkage using datasets ranging in size from 250 to 3,131,176 records. (145) Findings in these studies are consistent with the results presented in this analysis, with sensitivities ranging from 86% (database sizes: 250 records with N of second dataset not published (146)) to 99.2% (database sizes: 6,000 records in both(147)), and specificity ranging from 99.4% (database sizes: 6,000 records in both(147)) to 100% (database sizes: 822 and 450(148)). Variation in these results may be due to algorithms used for probabilistic linkage, as well as characteristics of the datasets such as the rates of missing data, errors and omissions which impact on the results.

4.6 Conclusion

EMS was demonstrated to have high accuracy for the linkage of public health datasets. With the establishment of national electronic datasets across health and social care, the accuracy of this software enables previously unanswerable research questions to be tackled.(149–151) Probabilistic linkage has great potential to be used where exact or deterministic linkage isn't possible, including in low-income settings, and for vulnerable populations, where the absence of unique identifiers has historically hindered the ability to identify individuals across separate datasets in order to establish outcomes or exposures as required by many types of epidemiological study design. This analysis provides reassurance that the Enhanced Matching System can be appropriately used to link a dataset of pre-entry screened migrants to the UK national tuberculosis surveillance system using first name, surname, date of birth, nationality and sex as matching variables.

4.7 Publications and presentations arising from this chapter

1. Aldridge RW, Shaji K, Hayward AC, Abubakar I. Accuracy of probabilistic linkage using the Enhanced Matching System for public health and epidemiological studies. Accepted for publication in PLOS ONE.
2. Aldridge RW, Yates TA, Zenner D, White PJ, Abubakar I, Hayward AC. Pre-entry screening of migrants to the UK: using probabilistic matching to identify cases of tuberculosis post-migration. 45th Union World Conference in Barcelona. 30 October 2014.

CHAPTER 5

Incidence of tuberculosis in migrants screened pre-entry after arrival in England, Wales and Northern Ireland: a retrospective cohort study.

5.1 Abstract

Background: UK tuberculosis surveillance data in their current configuration allow only a period prevalence of tuberculosis to be calculated in individuals not born in the UK. Probabilistic matching can be used to identify incident cases of disease and missed prevalent cases, among individuals screened by the pre-entry programme.

Methods: A database of migrants screened pre-entry between 1st January 2006 and 31st December 2012 was probabilistically linked to ETS notifications during the period 1st January 2006 and 31st December 2013. A cohort study was performed, incidence rates were estimated and a multivariable risk factor analysis conducted.

Results: 519,955 migrants entered the cohort. There were 622 cases of bacteriologically confirmed pulmonary tuberculosis with an estimated incidence rate of 65 cases per 100,000 person years at risk (95%CIs: 60, 70). After adjusting for age and sex, there was strong evidence that a history of contact with a case of tuberculosis before migration (IRR 4.9; 95%CIs: 2.5, 9.4; p-value <0.001) and a chest radiograph classified as consistent with tuberculosis at migration (IRR 4.4; 95%Cis: 3.5, 5.5; <0.001) were associated with an increased risk of bacteriologically confirmed pulmonary tuberculosis in the UK.

Conclusion: Several novel findings are presented, including the first time direct estimates of the incidence of tuberculosis in a high-risk population that has been screened for active pulmonary disease prior to arrival.

5.2 Introduction

The UK pre-entry screening of migrants aims to detect prevalent cases of active tuberculosis at the time of migration.(7) Individuals moving to the UK from countries with a prevalence of tuberculosis greater than 40 per 100,000 are likely to remain at higher risk of developing active disease after migrating through reactivation of infection acquired abroad. Migrants are also likely to be at higher risk of acquiring tuberculosis and developing active disease in the UK due to other risk factors that are common in new entrants including poor housing, behavioural and lifestyle factors or immunosuppression due to HIV.(74–76)

UK surveillance data in their current configuration allow only a period prevalence of tuberculosis to be calculated in individuals not born in the UK.(136) Self-reported data on country of birth was known in 97% (7,632/7,892) of tuberculosis cases notified in 2013, and 72% (5,529/7,632) of these individuals were born outside the UK. The notification rate of tuberculosis in migrants was 18 times higher than the rate in the UK born, at 70 per 100,000. Time since entry into the country was known for 91% of cases among individuals not born in the UK, and 44% of cases were diagnosed within five years of entering the country. Data were only available on country of birth for non-UK born migrants, and not the country of migration, which, depending on migration patterns may have an important impact on the subsequent risk of developing active tuberculosis.

Follow-up of migrants for epidemiological and public health purposes after entering the UK has not previously been possible as no dataset existed to facilitate this process. Therefore, it has not been possible until now to calculate the incidence of active tuberculosis post-migration or the risk factors associated with these incident cases. Such data would be helpful for several reasons. Firstly, removing prevalent cases from estimates of incidence after arrival in the UK enables more accurate estimates to be made as well as the ability to explore risk factors for newly developed cases. Secondly, it would enable more accurate estimates of the burden of disease within migrant populations. Instead of reporting data by country of birth, it is possible to estimate rates by country of migration, which more accurately reflects the risk of disease particularly if, for example, an individual was born in a low incidence country, but migrated to the

UK from a high incidence country (or vice versa). Thirdly, incidence data has the potential to improve the pre-entry screening programme, by identifying groups that may be an appropriate target for latent tuberculosis screening. Finally, a linked dataset that enables follow up of pre-entry screened migrants in the UK also has a role in quality assurance by identifying potential "missed prevalent" cases at pre-entry screening that were notified very shortly after migration. Probabilistic matching, as described and validated in the previous chapter, can be used to identify incident cases of disease (and missed prevalent cases) notified in the UK among individuals screened by the pre-entry programme. Linking these two datasets would enable the estimation of incidence of tuberculosis among migrants as well as the risk factors associated with these cases.

5.2.1 Research questions:

The analysis presented in this chapter aims to answer the following research questions:

1. What is the incidence of tuberculosis in migrants screened pre-entry?
2. What are the risk factors for incident cases of tuberculosis notified in migrants screened pre-entry?

5.3 Methods

5.3.1 Study design and setting

This study was a retrospective cohort in migrants to the UK. Between 2005 and April 2013 the International Organisation of Migration (IOM) conducted a pilot pre-entry screening programme in 15 countries with a tuberculosis prevalence of greater than 40 per 100,000 population. Individuals from these 15 countries were screened for active tuberculosis according to the UK technical instructions described in Chapter 3.(152) Individuals undergoing pre-entry screening consented to their data being shared with the UK government agencies including UK immigration authorities, the UK Department of Health, Public Health England and the UK National Health Service.

5.3.2 Outcomes

The primary outcomes for this study were:

1. Incidence of all cases of tuberculosis notified to the Enhanced Tuberculosis Surveillance system.
2. Incidence of bacteriologically confirmed pulmonary tuberculosis.
3. Incidence of bacteriologically confirmed extra-pulmonary tuberculosis.

The secondary outcomes for the study were:

1. Incidence of pulmonary tuberculosis
    a. Confirmed by culture testing on liquid or solid media
    b. Confirmed by microscopy for acid fast bacilli (AFB)
    c. Resistant to at least one first-line drug (isoniazid, rifampicin, pyrazinamide & ethambutol)
2. Incidence of extra-pulmonary tuberculosis
    a. Confirmed by culture testing on liquid or solid media
    b. Resistant to at least one first-line drug (isoniazid, rifampicin, pyrazinamide & ethambutol)

WHO reporting definitions are used to define bacteriologically, pulmonary and extra-pulmonary cases.(109) A bacteriologically confirmed case is one "from whom a biological specimen is positive by smear microscopy, culture or WHO-approved rapid diagnostics (such as Xpert MTB/RIF)". Pulmonary tuberculosis (PTB) is: "any bacteriologically confirmed or clinically diagnosed case of tuberculosis involving the lung parenchyma or the tracheobronchial tree. Miliary tuberculosis is classified as PTB because there are lesions in the lungs. Tuberculosis intra-thoracic lymphadenopathy (mediastinal and/or hilar) or tuberculosis pleural effusion, without radiographic abnormalities in the lungs, constitutes a case of extra-pulmonary tuberculosis. A patient with both pulmonary and extra-pulmonary tuberculosis should be classified as a case of PTB". Finally, extra-pulmonary tuberculosis refers to: "any bacteriologically confirmed or clinically diagnosed case of tuberculosis involving organs other than the lungs, e.g. pleura, lymph nodes, abdomen, genitourinary tract, skin, joints and bones, meninges".

The Enhanced Tuberculosis Surveillance system contains data on all cases notified in Cases in England, Wales and Northern Ireland. All ETS notified cases of tuberculosis meet the following criteria, as defined by Public Health England: "All new tuberculosis cases that meet one of the two following case definitions Culture confirmed case due to

M. tuberculosis complex (including *M. tuberculosis*, *M.bovis*, *M. africanum* or *M.microti*). In the absence of culture confirmation, a case that meets the following criteria: a clinician's judgement that the patient's clinical and/or radiological signs and/or symptoms are compatible with tuberculosis, AND a clinician's decision to treat the patient with a full course of anti- tuberculosis therapy. The requirement to notify applies if there is reasonable ground for suspecting that a patient has died with, but not necessarily from, active tuberculosis (including post mortem diagnoses). Notification requirement applies also to UK residents who are diagnosed abroad but continue with their anti-tuberculosis therapy in the UK and to non-UK residents diagnosed in the UK, even if anti-tuberculosis therapy is not initiated in the UK."(153)

5.3.3 Data sources and linkage

Two data sources were used for this analysis: 1) The IOM database of migrants screened pre-entry between 1st January 2006 and 31st December 2012; and 2) ETS notified cases between 1st January 2006 and 31st December 2013. These two datasets were probabilistically linked using the enhanced matching system described in chapter 4 using first name, surname, date of birth, nationality and sex as matching variables. Cleaning and consistency checking of the final dataset was undertaken by examining the distribution of variables, the range of individual variables, and missing data.

5.3.4 Censoring

Migrants entered the cohort upon receiving a certificate of medical clearance after pre-entry screening at IOM clinics. Individuals were followed up until the first of tuberculosis, death, or emigration. Primary and secondary outcomes were identified through probabilistic linkage of the cohort of migrants screened by IOM to ETS.(136)

5.3.5 Exclusions

Any primary and secondary outcomes occurring within 90 days of the issue of a medical certificate of clearance were assumed to be prevalent cases that were missed by pre-entry screening. These cases were therefore excluded for all primary or secondary outcomes. This assumption was examined further in a sensitivity analysis.

5.3.6 Duplicates

Individuals screened pre-entry may have multiple records on the IOM screening database. Duplicate records were analysed on the basis of whether they occurred within 12 months of each other or not. 12 months was chosen to distinguish duplicates on the basis that this period of time was long enough to capture individuals who were found to have tuberculosis on their first screen, and were undergoing repeat screening for visa clearance, but not too long a time period such that tuberculosis risk factors (as described in the previous section) would have changed significantly between screens.

Duplicates within 12 months:

Rules used to determine whether duplicate records within 12 months were included in the cohort or not are provided in Table 15. Individuals with a positive first and second screen would not have received a medical certificate of clearance, and therefore will never have been able to enter the UK and are excluded from this analysis. Migrants with a positive first screen and negative second screen are likely to have undergone treatment, been rescreened and found to be clear and will then have been able to enter UK. Therefore the first screen is excluded, but the second included in the cohort. Migrants with a negative first screen and positive second are unlikely to have entered the UK due to short time period involved. These records are likely to represent individuals who weren't able to get their visa processed within six months of the initial clearance and on repeat screening were found to be positive. They are therefore all excluded. Records negative on both screens are likely to be individuals who weren't able to get their visa processed within six months of the initial clearance and therefore the first is excluded, but the second is included.

Table 15. Rules for dealing with duplicate screens less than 12 months apart.

		First screen TB result	
		+ve	-ve
Last screen TB result	+ve	Exclude all	Exclude all
	-ve	1st screen: Exclude 2nd screen: Include	1st screen: Exclude 2nd screen: Include

Duplicates greater than 12 months apart:

Rules used to determine whether duplicate records occurring more than 12 months apart were included in the cohort analysis are provided in Table 16. Individuals with a positive first and second screen will not have been given a medical certificate of clearance, and therefore will never have been able to enter the UK. All these records are therefore excluded. Records with a positive first screen, but negative second screen are likely to represent migrants reapplying after treatment because their initial screen was positive. Therefore the first screen is excluded but the second included. Migrants with the negative first screen, but positive second screen are allowed to enter the cohort after the first screen but are censored one month prior to the second screen, and the second screen is excluded. Records with negative first and second screen could represent a variety of different scenarios, including: 1) individuals that have visited the UK and are reapplying for a new visa having returned to their country of origin after their initial visa ran out; and 2) individuals reapplying because their medical clearance ran out in the time they were able to get their visa application processed (clearance certificates are only valid for 6 months). These individuals are therefore allowed to enter the cohort after each screen, but are censored one month prior to the next screen to allow for return to the country of origin.

Table 16. Rules for dealing with duplicate screens greater than 12 months apart.

		First screen TB result	
		+ve	-ve
Last screen TB result	+ve	Exclude all	1st: Include and censor at 1 month prior to 2nd screen 2nd: Exclude
	-ve	Exclude all duplicates apart from the last screen	1st: include and censor at 1 month prior to 2nd screen 2nd: Include

5.3.7 Sample size

Before access to the IOM database was granted, it was assumed there would be 350,000 individuals eligible for this study with an estimated length of follow up 1,040,000 person years. It was estimated that incidence rates in individuals from high, medium and low risk countries (defined as 300, 150 and 40 cases per 100K person years respectively) would be calculated with the following confidence intervals: 300: 95% CIs 294-306; 150: 95% CIs 142-158; and 40: 95% CIs 32-48.

5.3.8 Confounding variables and risk factors

Age and sex were considered a-priori as confounding variables. Risk factor variables from the IOM dataset included: visa category, contact with a case of tuberculosis, CXR classification at pre-entry screening, whether a migrant was screened at a clinic where culture testing of sputum samples was performed, and WHO prevalence estimates of tuberculosis in the country of origin. WHO prevalence was used to stratify countries estimates, rather than incidence estimates, for consistency with Chapters 2 and 4 of this thesis.

5.3.9 Statistical analysis

Baseline descriptive statistics of the cohort were provided using simple counts and proportions. To account for the uncertainty of death, migration to Scotland and emigration, multiple imputation was performed. External data were used to inform the imputation models as described in the following two sections. Ten imputed datasets were created. Individual imputed datasets were analysed using Poisson regression, a

suitable method for modelling rare event data, and estimates of crude and adjusted incidence rates for the primary and secondary outcomes were calculated. A multivariable Poisson regression model was used to identify risk factors for the primary outcomes. All results were adjusted for clustering by individual, to take account of repeated entries by migrants into the cohort. The results of the analyses from individual imputed datasets were then combined using Rubin's rules; these appropriately account for uncertainty in the imputed information. Final results were presented as incidence rate per 100,000 person years at risk, incidence rate ratios, 95% confidence intervals and p-values. Stata v.13 (Statacorp LP, College Station, TX, USA) was used for all statistical analyses.

5.3.10 Migration out of the country

No data were available to be able to determine how long each individual migrant screened pre-entry was able to stay in England, Wales and Northern Ireland. It was therefore necessary to impute data on length of stay, and hence time at risk of being notified as a case of tuberculosis was based on visa category that was collected at the time of pre-entry screening. Lengths of stay were imputed using data on UK entry clearance visas, which details length of stay by type of visa and year of issue.(154) These data have high coverage for all visas issued outside the UK to non-EEA nationals who are subject to immigration control and require a visa to enter the UK. These data includes the number of visas issued, by year and by duration of stay (0-3mths, 3-6mths, 6mths-1yr, 1-2yrs, 2-3yrs, 3-4yrs, 4+yr). For the purposes of imputation, visas were grouped into four categories: total; work; study; or student visitor (which relates to individuals attending short courses for a period of less than one year). The "total" category excluded visitor and transit visas, but included all other categories including work, both types of study, family reunion, dependant's joining or accompanying individuals already in the UK, and other rarer visa types. The proportion of study and student visas that were issued for short courses is detailed in Table 17.

Table 17. Percentage of study and student visas that are for short courses

Year	Percentage
2005	8
2006	10
2007	13
2008	15
2009	11
2010	15
2011	19
2012	25
2013	26

IOM pre-entry screening data had the following visa type categories: family reunion, settlement, work, working holiday, students and other. Table 18 shows how IOM visa categories were matched to visa duration data. Imputation assumed that migrants remained in England, Wales and Northern Ireland for the duration of the visa issued.

Table 18. Rules for imputing visa duration

IOM visa category	Visa duration data category	Imputation rule
Family reunion	Included in "total"	Assume these stay for the duration of the cohort analysis.
Settlement	Included in "total"	Assume these stay for the duration of the cohort analysis.
Students	Study and short study	(i) Select long/short course according to probabilities in Table 17. (ii) Then impute from either the study or short-course study distribution.
Work	Work	Impute from work distribution.
Working holiday	Work	Impute from work distribution, restricting to categories < 2yrs (working holidays are maximum 2yrs)
Other	Included in "total"	Impute from overall distribution

The distributions of visa duration over time provide data for the imputation of length of stay. Figures 19-22 show the distribution of the duration of visas issued over time by visa category, from 2005 to 2012. Overall (Figure 19), fewer long-term (4+yrs) visas

were issued from 2009, with a corresponding increase short-term (3mths-1yr) and mid-long-term visas (3-4yrs). For work visas, fewer long-term visas (4+yrs) have been issued since 2009. The duration of study visas has stayed broadly similar over time, whereas the short course study visas for 0-3mths were not issued in recent years.

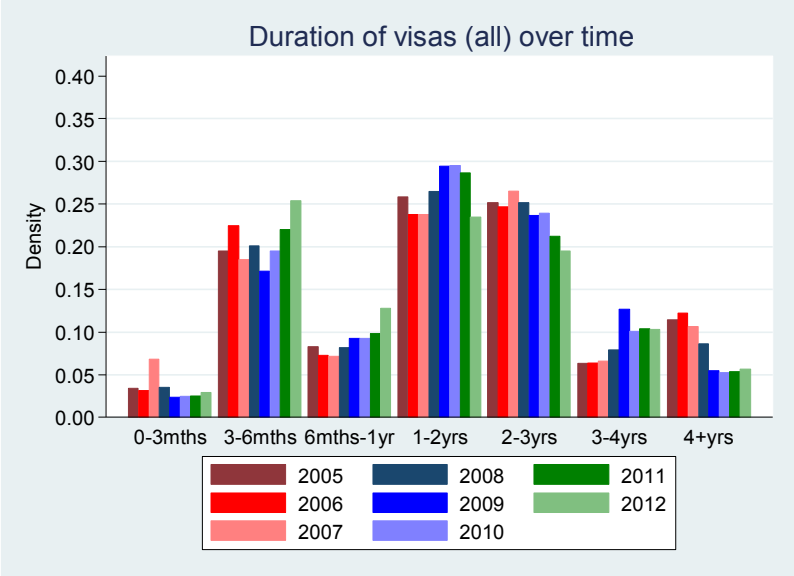

Figure 19. Changes in the distribution of all visa durations over time.

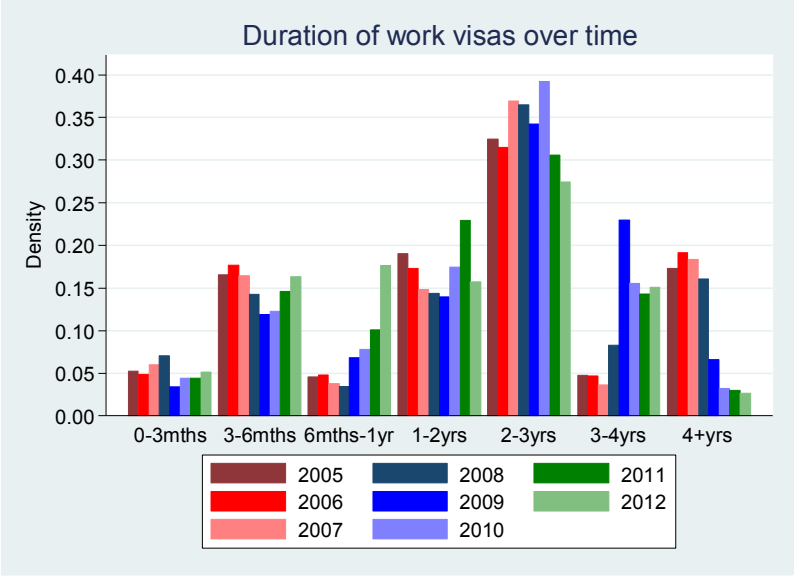

Figure 20. Changes in distribution of work visa duration over time.

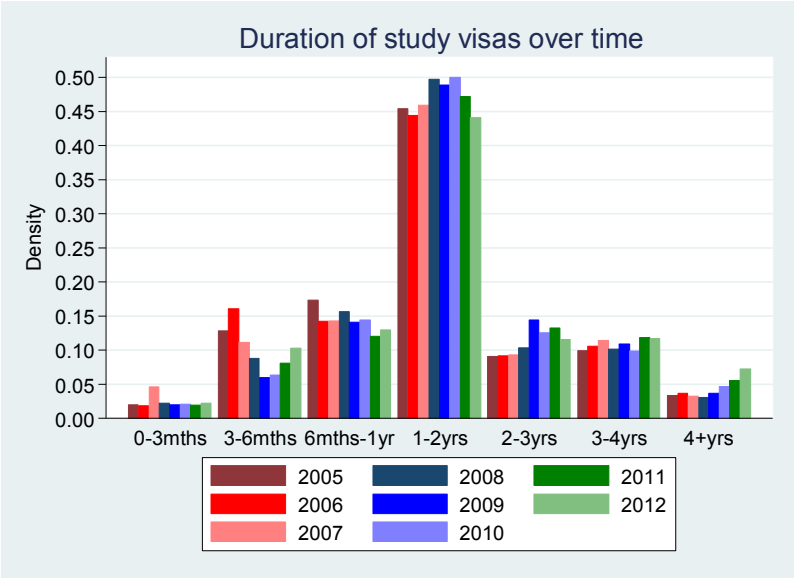

Figure 21. Changes in distribution of study visa duration over time

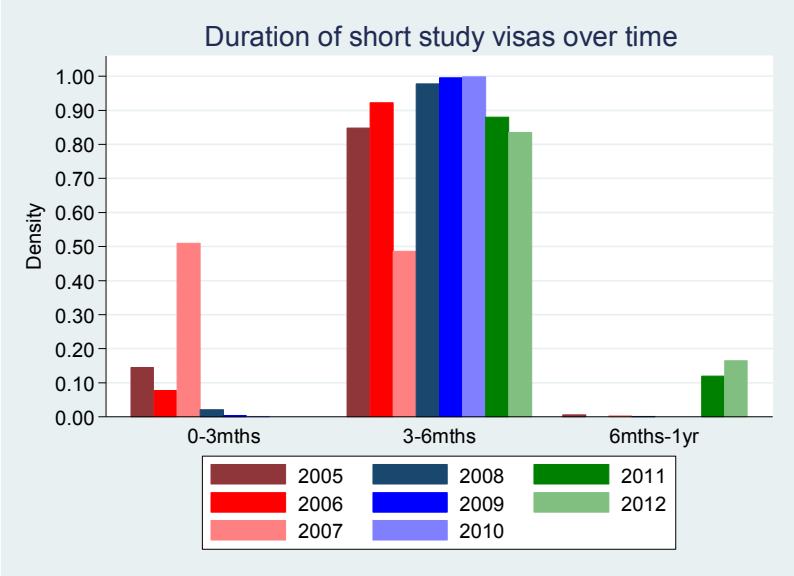

Figure 22. Changes in distribution of short study visa duration over time

The distributions in the visa issue data provide an imputed visa category for length of stay e.g. 1-2 yrs. To impute the exact duration of stay, a time for each migrant within a visa category was chosen by linear interpolation. For example, if a random uniform variable used for this imputation lands 25% of the way between the cumulative probabilities of categories 1-2yrs, the duration that lies 25% of the way between 2yrs and 1yr was used, which is equal to 1.25yrs.

5.3.11 Death:

To estimate survivor functions, Kaplan-Meier product-limit survival probabilities were calculated using the survival rates from 2009 data for England and Wales (Figure 23). These data were created by the Cancer Research UK Cancer Survival Group at LSHTM, and is based upon mortality data provided by ONS.(155)

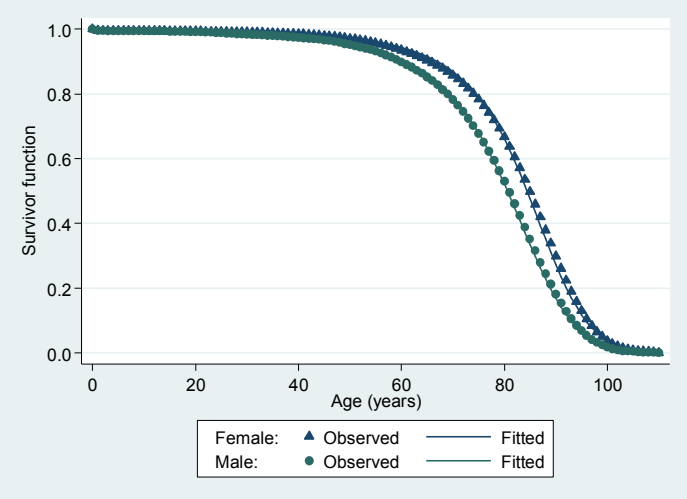

Figure 23. Survival probabilities calculated from England and Wales 2009 data, with fitted line

To model these survival probabilities, the logit transformation of the survival probability was created, and then modelled using linear regression as a cubic function of exponentiated age. The logit transformation provides a fit to the observed data with a smoother and more stable gradient. The exponential transformation of age enables a better fit for older ages, which is where the survival changes rapidly, and is therefore important to represent accurately. The fitted model is given in Equations 5 and 6.

$$\ln\left(\frac{p}{1-p}\right) = 9.231431 - 5.106226 \times e^{age} + 2.250698 \times (e^{age})^2 - 0.7572033 \times (e^{age})^3$$

Equation 5. For men

$$\ln\left(\frac{p}{1-p}\right) = 11.45322 - 8.980079 \times e^{age} + 4.07654 \times (e^{age})^2 - 1.049887 \times (e^{age})^3$$

Equation 6. For women

As illustrated in Figure 24, this function provides a good fit to the observed data. In order to draw from this distribution it was necessary to be able to calculate the age corresponding to a particular fitted value of the survival probability. This was done by solving the cubic equation created by entering a specific value for p in the two equations above. Figure 25 shows the ages calculated from survival probabilities across the interval 0-1, with the fitted survival probabilities generated from the model above overlaid.

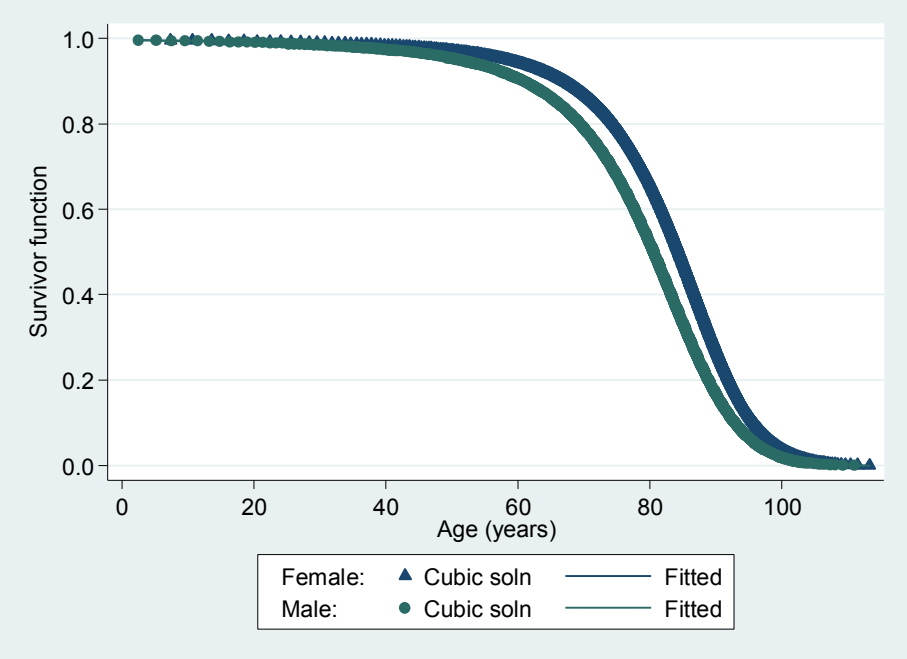

Figure 24. Ages calculated from survival probabilities, with fitted line from

Age at of death for someone entering the cohort at age 0 was imputed by:

(i) Drawing a random value from the standard uniform distribution, U~[0,1].

(ii) Setting the predicted age of death to be the age that corresponds to the survival probability U.

Age at death for someone who is already aged x years, death was imputed by:

(i) Drawing a random value from the standard uniform distribution, U~[0,1].

(ii) Calculated the survival probability for age x, S_x

(iii) Set the predicted age at death to be the age that corresponds to the survival probability U x S_x.

5.3.12 Migrants to Scotland:

Personal identifiable information was not available for tuberculosis notifications in Scotland. Therefore migrants to Scotland could not be identified as incident cases for any of the primary or secondary outcomes. It was therefore necessary to specify that a proportion of migrants given a medical certificate of clearance would not enter the cohort. Using data on long-term international migration produced by ONS, it was estimated that 7.3% of migrants entering the UK between 2006 and 2012 would be resident in Scotland.(156) Therefore, for each imputation, 92.7% of migrants issued with a medical certificate of clearance were randomly selected (from a uniform distribution) to enter the cohort. This issue is examined further in a sensitivity analysis.

5.3.13 Sensitivity analyses:

The assumptions made in this analysis were examined by several sensitivity analyses. The rules used to de-duplicate the cohort were examined in order to assess changes in the primary outcomes. In a sensitivity analysis, all individuals apart from those with a positive screen were allowed to enter the cohort, but censored one month prior to any re-screening. This analysis will provide a more conservative estimate for person time at risk, and therefore reduce the incidence rates to a lower estimate.

Migration out of England, Wales and Northern Ireland was examined. In the first analysis, all migrants were assumed to stay for one and a half years, providing a lower estimate of person time at risk. In a second sensitivity analysis, all migrants were assumed to stay until the end of the study period of 31st December 2013. This is the more conservative assumption, and whilst it unrealistically inflates the denominator, it provides a lower bound for the estimates of incidence.

Death was examined by assuming there were no deaths in the cohort. This is the most conservative assumption and unrealistically inflates the denominator, but also provides a lower bound for the estimates of incidence.

Varying the definition of how incident and prevalent cases were distinguished (i.e. the 90 day cut off between pre-entry screening and notification) was examined. For the main analysis, it was assumed that cases notified within 90 days of pre-entry screening

were ‘missed’ prevalent and not incident cases. The impact of varying this definition (up to 180 days) on the outcomes was assessed.

5.3.14 Ethics approval

Ethical approval was received for this analysis from UCL research ethics committee (3294/002; See Appendix 4 for copy of approval letter). The work was conducted with Public Health England which has Health Research Authority approval to hold and analyse national surveillance data for public health purposes under Section 251 of the NHS Act 2006.

5.4 Results

Between 2006 and 2012 there were 640,808 visa applicants screened for tuberculosis pre-entry by the IOM programme (Figure 25). Records from these migrants were probabilistically matched to the UK Enhanced Tuberculosis Surveillance dataset for the years 2006 to 2013, which contained 83,781 records. Missing data on linkage variables was low and less than 2% for all variables used in both the IOM and ETS dataset (Table 19). After excluding duplicates, individuals migrating to Scotland, and missed prevalent cases, 519,955 visa applicants entered the cohort between 2006-2012, representing 514,968 individual migrants.

A majority of migrants were aged between 16 and 44 (490,806; 94.4%) and were male (346,839; 66.7%; Table 20). Self-report of contact with a case of tuberculosis at pre-entry screening was rare (1,220; 0.2%) and this was the only variable within the dataset with missing data on 2,025 records (0.4%). Student visas were the most common visa type among the migrants in this cohort (307,127; 59.1%) and most applicants had no abnormality on their chest radiograph at the time of pre-entry screening (489,733; 94.2%), with tuberculosis suspected in 21,862 individuals (4.2%). Just under two thirds of migrants were screened at sites where sputa were smear and culture tested for tuberculosis (340,020; 65.4%).

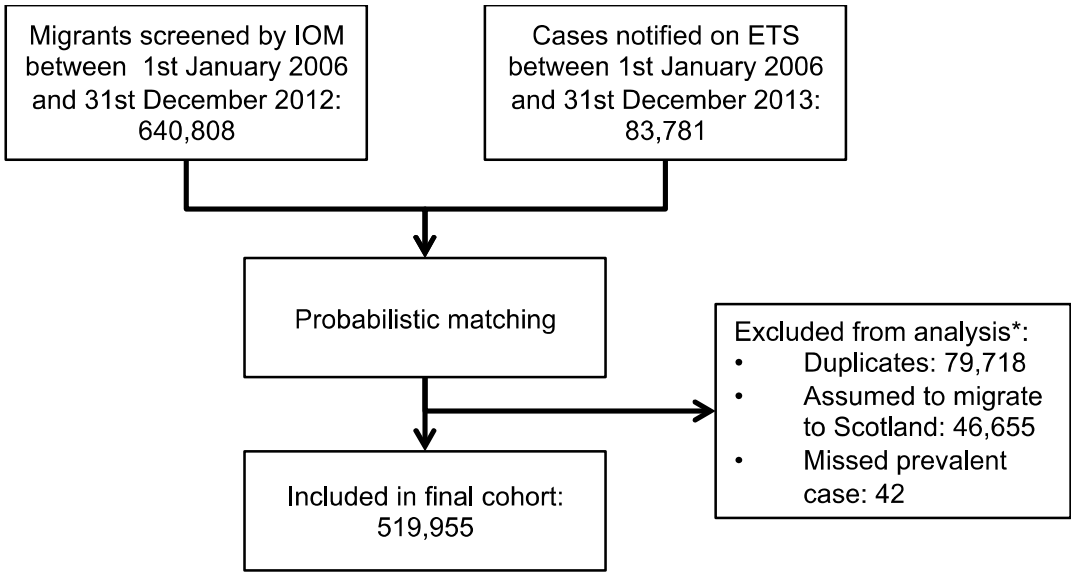

Figure 25. Study participant flow diagram

*Numbers assumed to migrate to Scotland varied by imputation, and sum of those excluded does not equal difference between total visa applicants and the number of migrants included in final cohort, as groups are not mutually exclusive.

Table 19. Description of missing data for linkage variables in ETS and IOM pre-entry screening dataset

	ETS Missing data		IOM Missing data	
Variable	**N**	**%**	**N**	**%**
All	83781	100%	640808	100%
First name	0	0%	13078	2%
Surname	0	0%	4824	1%
Date of Birth	32	0%	910	0%
Sex	186	0%	786	0%
Nationality	1823	2%	796	0%

The total length of follow up for the cohort was 1,038,000 person years at risk with a mean follow up of 2.0 years per person. Considering all cases of tuberculosis notified to ETS, there were 1,863 cases identified in pre-entry screened migrants, with a crude notified incidence rate of 194 per 100,000 person years at risk (95%CIs: 186, 203; Table 20). The crude rates were highest in those with a self-reported history of contact with a case of tuberculosis (662; 95%CIs: 399, 1098) and those with a chest radiograph

classified as consistent with tuberculosis at pre-entry screening (599; 95%CIs: 528, 680). A multivariable risk factor analysis was performed to examine the evidence for risk factors associated with all notified case of tuberculosis, with results presented as incidence rate ratios (IRR). There was strong evidence that contact with a case of tuberculosis (2.9; 95%CIs: 1.8, 4.9; p-value <0.001), chest radiographs classified as suspected tuberculosis (3.4; 95%CIs: 2.9, 3.9; p-value <0.001), settlement and dependant (1.6; 95%CIs: 1.5, 1.8; p-value <0.001) and family reunion (5.3; 95%CIs: 4.0, 7.1; p-value <0.001) visa categories were at increased risk of being notified as a case of tuberculosis in the UK after adjusting for age and sex. There was also strong evidence that migrants from countries with a WHO prevalence of 40-149 (0.2; 95%CIs: 0.2, 0.3; p-value <0.001), and those from countries with a prevalence between 150-349 (0.6; 95%CIs: 0.5, 0.7; p-value <0.001), migrants screened at sites where culture testing was routinely performed (0.6; 95%CIs: 0.5, 0.6; p-value <0.001), and working holiday visa holders (0.4; 95%CIs: 0.3, 0.7; p-value <0.001) were at lower risk after adjusting for age and sex.

There were 622 cases of bacteriologically confirmed pulmonary tuberculosis with an estimated incidence rate of 65 cases per 100,000 person years at risk (95%CIs: 60, 70; Table 21). The incidence of bacteriologically confirmed pulmonary tuberculosis was highest in those reporting a history of contact with a case of tuberculosis (398; 95%CIs: 207, 766) and lowest in those on a working holiday visa (6; 95%CIs: 2, 25). Incidence was also high in those with a chest radiograph classified as suspected tuberculosis at pre-entry screening, but not found to have tuberculosis at the time (249; 95%CIs: 205, 303).

After adjusting for age and sex in a multivariable analysis, there was strong evidence that several risk factors were associated with being detected as a case of bacteriologically confirmed pulmonary tuberculosis, including a history of contact with a case of tuberculosis before migration (IRR 4.9; 95%CI: 2.5, 9.4; p-value <0.001) and those with a chest radiograph classified as consistent with tuberculosis disease (IRR 4.4; 95%CI 3.5, 5.5; p-value <0.001). There was strong evidence that individuals screened pre-entry at clinics where culture testing was performed had a lower incidence of bacteriologically confirmed pulmonary tuberculosis post entry than those screened at sites where this was not conducted (IRR 0.6; 95%CI 0.5, 0.7; p-value <0.001). Migrants

on settlement and dependent, and family reunion visas had increased incidence of bacteriologically confirmed pulmonary tuberculosis (IRR 1.4; 95%CI 1.2, 1.7; and 3.0; 95%CI 1.6, 5.3 respectively). Compared to migrants from countries with a WHO prevalence of greater than 350 per 100,000, bacteriologically confirmed pulmonary tuberculosis post migration was lower in those from countries with an incidence of 150-349 (IRR 0.7; 95%CIs: 0.6, 0.9; p-value 0.01) and 40-149 (IRR 0.4; 95%CIs: 0.2, 0.7; p-value <0.001).

There were 674 incident cases of bacteriologically confirmed extra-pulmonary tuberculosis in migrants screened pre-entry, with an incidence rate of 70 per 100,000 person years at risk (95%CIs: 65, 76; Table 22). The rate of bacteriologically confirmed extra-pulmonary tuberculosis was highest in migrants on a family reunion visa (201 per 100,000 person years at risk; 95%CIs: 123, 328). After adjusting for age and sex, there was no evidence that being in contact with a case of tuberculosis pre-entry increased the incidence of bacteriologically confirmed extra-pulmonary tuberculosis (IRR 1.9; 95%CIs: 0.6, 5.9; p-value 0.28). Migrants had a lower incidence of bacteriologically confirmed extra-pulmonary disease if they were from countries with a WHO prevalence of tuberculosis between 40-149 per 100,000 population (IRR 0.1; 95%CIs: 0.1, 0.3; p-value <0.001) and 150-349 (IRR 0.6; 95%CIs 0.5, 0.8; p-value <0.001) compared to those from a prevalence of greater than 350 per 100,000 population.

There were 619 cases of culture confirmed pulmonary tuberculosis in the cohort, with an incidence rate of 64 per 100,000 person years at risk (95%CIs: 59, 70; Table 23), and therefore the majority of bacteriologically confirmed cases were also culture confirmed (619/627; 99%). As a result the multivariable analysis was also very similar to that for bacteriologically confirmed cases (Table 21).

There were fewer cases of smear positive pulmonary tuberculosis with 228 in total. The estimated incidence of 24 per 100,000 person years at risk (95%CIs: 21, 27; Table 24). Higher rates were seen in those with a reported history of contact with a case of tuberculosis at pre-entry screening (160 per 100,000 person years at risk; 95%CIs: 60, 427). After adjusting for age and sex there was strong evidence that self-reported history of contact with a case of tuberculosis at pre-entry screening was associated with an increased incidence of AFB positive pulmonary tuberculosis (IRR 5.7; 95%CIs 2.2, 15.3; p-value <0.001). There was also strong evidence that individuals screened pre-

entry at clinics where culture testing had a lower incidence of AFB positive pulmonary tuberculosis than those screened at sites where this was not conducted (IRR 0.6; 95%CI 0.4, 0.7; p-value <0.001).

There were 41 cases of pulmonary tuberculosis with resistance to at least one first-line drug (isoniazid, rifampicin, pyrazinamide or ethambutol). There were no cases of extra-pulmonary tuberculosis with resistance to at least one first-line drug. As a result of these low numbers no univariable or multivariable risk factor analysis was undertaken for these estimates.

Table 20. Baseline characteristics, univariate and multivariate analysis of incidence rates for all ETS notified cases of tuberculosis in migrants screened pre-entry (2006-2012) and notified in ETS (2006-2013)

Risk Factor	Migrants contributing (%)	Episodes	Person years at risk (1000)	Rate per 100,000 person years (95% CI)	Univariable IRR (95% CI)	Multivariable IRR (95%CI)	p-value
All	519955 (100.0%)	1863	1038	194 (186, 203)	-	-	-
Age							
0-15	15468 (3%)	54	31	188 (144, 245)	1 (0.7, 1.3)	0.7 (0.5, 1.0)	0.02
16-44	490806 (94.4%)	1743	979	193 (184, 202)	1.0	1.0	
45-64	11243 (2.2%)	50	23	238 (180, 314)	1.2 (1.0, 1.6)	1 (0.7, 1.3)	0.89
65+	2438 (0.5%)	16	5	354 (217, 577)	2 (1.2, 3.3)	1 (0.6, 1.6)	0.92
Sex							
Female	173116 (33.3%)	713	347	223 (207, 240)	1.0		
Male	346839 (66.7%)	1150	691	180 (170, 191)	0.8 (0.7, 0.9)	1 (0.9, 1.1)	0.81
Contact with case TB							
No	516710 (99.4%)	1841	1030	193 (184, 202)	1.0		
Yes	1220 (0.2%)	15	3	662 (399, 1098)	3.4 (2.0, 5.6)	2.9 (1.8, 4.9)	<0.001
Missing	2025 (0.4%)	7	5				
Visa							
Students	307127 (59.1%)	904	609	160 (150, 171)	1.0		
Settlement and Dependents	159986 (30.8%)	790	320	267 (249, 287)	1.7 (1.5, 1.9)	1.6 (1.5, 1.8)	<0.001
Work	21140 (4.1%)	75	45	183 (146, 229)	1.1 (0.9, 1.5)	1 (0.8, 1.3)	0.99
Working Holiday Maker	17526 (3.4%)	20	35	63 (40, 97)	0.4 (0.2, 0.6)	0.4 (0.3, 0.7)	<0.001
Family Reunion	3989 (0.8%)	53	8	720 (550, 943)	4.5 (3.4, 5.9)	5.3 (4.0, 7.1)	<0.001
Other	10187 (2.0%)	21	21	107 (69, 163)	0.7 (0.4, 1.0)	1 (0.6, 1.5)	0.88

CXR							
No abnormality	489733 (94.2%)	1605	977	178 (169, 187)	1.0		
TB suspected	21862 (4.2%)	240	45	599 (528, 680)	3.4 (2.9, 3.9)	3.4 (2.9, 3.9)	<0.001
Abnormality not TB	8360 (1.6%)	18	17	116 (73, 184)	0.6 (0.4, 1.0)	0.8 (0.5, 1.4)	0.48
WHO category							
40-149	29143 (5.6%)	24	59	44 (30, 66)	0.2 (0.2, 0.3)	0.2 (0.2, 0.3)	<0.001
150-349	75294 (14.5%)	231	151	166 (146, 189)	0.9 (0.8, 1.0)	0.6 (0.5, 0.7)	<0.001
350+	415518 (79.9%)	1608	828	210 (200, 220)	1.0	1.0	
Sputum culture testing							
No	179935 (34.6%)	867	353	268 (251, 286)	1.0		
Yes	340020 (65.4%)	996	685	157 (147, 167)	0.6 (0.5, 0.6)	0.6 (0.5, 0.6)	<0.001

Table 21. Baseline characteristics, univariate and multivariate analysis of incidence rates for bacteriological confirmed pulmonary tuberculosis in migrants screened pre-entry (2006-2012) and notified in ETS (2006-2013)

Risk Factor	Migrants contributing (%)	Episodes	Person years at risk (1000)	Rate per 100,000 person years (95% CI)	Univariable IRR (95% CI)	Multivariable IRR (95%CI)	p-value
All	519955 (100%)	622	1038	65 (60, 70)	-	-	-
Age							
0-15	15468 (3%)	20	31	70 (45, 108)	1.1 (0.7, 1.7)	0.9 (0.5, 1.4)	0.51
16-44	490806 (94.4%)	583	979	64 (59, 70)	1.0	1.0	
45-64	11243 (2.2%)	12	23	57 (32, 101)	0.9 (0.5, 1.6)	0.7 (0.4, 1.2)	0.23
65+	2438 (0.5%)	7	5	155 (74, 325)	2.7 (1.3, 5.9)	1.4 (0.6, 3.1)	0.42
Sex							
Female	173116 (33.3%)	240	347	75 (66, 85)	1.0		
Male	346839 (66.7%)	382	691	60 (54, 66)	0.8 (0.7, 0.9)	0.9 (0.8, 1.1)	0.49
Contact with case TB							
No	516710 (99.4%)	610	1030	64 (59, 69)	1.0		
Yes	1220 (0.2%)	9	3	398 (207, 766)	6.2 (3.2, 11.9)	4.9 (2.5, 9.4)	<0.001
Missing	2025 (0.4%)	3	5	60 (54, 66)			
Visa							
Students	307127 (59.1%)	323	609	57 (51, 64)	1.0		
Settlement and Dependents	159986 (30.8%)	255	320	86 (76, 98)	1.5 (1.3, 1.8)	1.4 (1.2, 1.7)	<0.001
Work	21140 (4.1%)	23	45	56 (37, 84)	1 (0.6, 1.5)	0.9 (0.6, 1.3)	0.48
Working Holiday Maker	17526 (3.4%)	2	35	6 (2, 25)	0.1 (0, 0.4)	0.1 (0, 0.5)	<0.001
Family Reunion	3989 (0.8%)	12	8	163 (93, 288)	2.8 (1.6, 5.1)	3.0 (1.6, 5.3)	<0.001
Other	10187 (2%)	7	21	35 (17, 74)	0.6 (0.3, 1.3)	0.8 (0.4, 1.7)	0.61

CXR							
No abnormality	489733 (94.2%)	516	977	57 (52, 62)	1.0		
TB suspected	21862 (4.2%)	100	45	249 (205, 303)	4.4 (3.6, 5.5)	4.4 (3.5, 5.5)	<0.001
Abnormality not TB	8360 (1.6%)	6	17	39 (17, 86)	0.7 (0.3, 1.5)	0.8 (0.4, 1.9)	0.64
WHO category							
40-149	29143 (5.6%)	10	59	18 (10, 34)	0.3 (0.1, 0.5)	0.4 (0.2, 0.7)	<0.001
150-349	75294 (14.5%)	86	151	62 (50, 76)	0.9 (0.7, 1.2)	0.7 (0.6, 0.9)	0.01
350+	415518 (79.9%)	526	828	69 (63, 75)	1.0	1.0	
Sputum culture testing							
No	179935 (34.6%)	282	353	87 (77, 98)	1.0		
Yes	340020 (65.4%)	340	685	54 (48, 60)	0.6 (0.5, 0.7)	0.6 (0.5, 0.7)	<0.001

Table 22. Baseline characteristics, univariate and multivariate analysis of incidence rates for bacteriological confirmed extra-pulmonary tuberculosis in migrants screened pre-entry (2006-2012) and notified in ETS (2006-2013)

Risk Factor	Migrants contributing (%)	Episodes	Person years at risk (1000)	Rate per 100,000 person years (95% CI)	Univariable IRR (95% CI)	Multivariable IRR (95%CI)	p-value
All	519955 (100.0%)	674	1038	70 (65, 76)	-	-	-
Age							
0-15	15468 (3.0%)	11	31	35 (19, 63)	0.5 (0.3, 1.0)	0.4 (0.2, 0.7)	0.003
16-44	490806 (94.4%)	639	979	66 (61, 71)	1.0	1.0	
45-64	11243 (2.2%)	21	23	97 (64, 147)	1.4 (0.9, 2.2)	1.3 (0.8, 2.0)	0.26
65+	2438 (0.5%)	3	5	61 (20, 190)	0.9 (0.3, 2.9)	0.6 (0.2, 1.8)	0.35
Sex							
Female	173116 (33.3%)	244	347	71 (63, 81)	1.0		
Male	346839 (66.7%)	430	691	63 (57, 69)	0.9 (0.8, 1.0)	1.1 (0.9, 1.3)	0.58
Contact with case TB							
No	516710 (99.4%)	668	1030	65 (61, 70)	1.0		
Yes	1220 (0.2%)	3	3	120 (39, 372)	1.9 (0.6, 5.9)	1.9 (0.6, 5.9)	0.28
Missing	2025 (0.4%)	3	5	60 (54, 66)			
Visa							
Students	307127 (59.1%)	343	609	57 (51, 63)	1.0		
Settlement and Dependents	159986 (30.8%)	271	320	85 (76, 96)	1.5 (1.3, 1.8)	1.6 (1.3, 1.9)	<0.001
Work	21140 (4.1%)	27	45	63 (43, 91)	1.1 (0.7, 1.6)	1 (0.6, 1.4)	0.81
Working Holiday Maker	17526 (3.4%)	11	35	32 (18, 57)	0.6 (0.3, 1.0)	0.7 (0.4, 1.3)	0.27
Family Reunion	3989 (0.8%)	16	8	201 (123, 328)	3.6 (2.2, 5.9)	4.9 (2.9, 8.3)	<0.001
Other	10187 (2.0%)	6	21	33 (16, 69)	0.5 (0.2, 1.1)	0.8 (0.4, 1.8)	0.6

CXR							
No abnormality	489733 (94.2%)	615	977	63 (59, 69)	1.0		
TB suspected	21862 (4.2%)	51	45	118 (90, 155)	1.9 (1.4, 2.5)	1.9 (1.4, 2.5)	<0.001
Abnormality not TB	8360 (1.6%)	8	17	48 (24, 96)	0.8 (0.4, 1.5)	1.1 (0.5, 2.1)	0.86
WHO category							
40-149	29143 (5.6%)	6	59	12 (6, 25)	0.1 (0.1, 0.3)	0.1 (0.1, 0.3)	<0.001
150-349	75294 (14.5%)	75	151	50 (40, 62)	0.7 (0.6, 0.9)	0.6 (0.5, 0.8)	<0.001
350+	415518 (79.9%)	593	828	72 (66, 78)	1.0	1.0	
Sputum culture testing							
No	179935 (34.6%)	303	353	86 (77, 96)	1.0		
Yes	340020 (65.4%)	371	685	55 (50, 61)	0.6 (0.5, 0.7)	0.6 (0.5, 0.7)	<0.001

Table 23. Baseline characteristics, univariate and multivariate analysis of incidence rates for culture confirmed pulmonary tuberculosis in migrants screened pre-entry (2006-2012) and notified in ETS (2006-2013)

Risk Factor	Migrants contributing (%)	Episodes	Person years at risk (1000)	Rate per 100,000 person years (95% CI)	Univariable IRR (95% CI)	Multivariable IRR (95%CI)	p-value
All	519955 (100.0%)	619	1038	64 (59, 70)	-	-	-
Age							
0-15	15468 (3.0%)	20	31	64 (41, 99)	1.0 (0.7, 1.6)	0.9 (0.5, 1.4)	0.53
16-44	490806 (94.4%)	582	979	60 (56, 66)	1.0	1.0	
45-64	11243 (2.2%)	11	23	48 (27, 88)	0.8 (0.4, 1.4)	0.6 (0.4, 1.2)	0.16
65+	2438 (0.5%)	6	5	122 (55, 272)	1.9 (0.9, 4.2)	1.1 (0.5, 2.4)	0.88
Sex							
Female	173116 (33.3%)	240	347	71 (62, 80)	1.0		
Male	346839 (66.7%)	379	691	56 (50, 61)	0.8 (0.7, 0.9)	0.9 (0.8, 1.1)	0.5
Contact with case TB							
No	516710 (99.4%)	607	1030	60 (55, 65)	1.0		
Yes	1220 (0.2%)	9	3	400 (215, 744)	6.2 (3.2, 12)	5 (2.6, 9.6)	<0.001
Missing	2025 (0.4%)	3	5	60 (54, 66)			
Visa							
Students	307127 (59.1%)	322	609	54 (48, 60)	1.0		
Settlement and Dependents	159986 (30.8%)	253	320	81 (71, 91)	1.5 (1.3, 1.8)	1.4 (1.2, 1.7)	0.001
Work	21140 (4.1%)	23	45	52 (34, 78)	1.0 (0.6, 1.5)	0.9 (0.6, 1.3)	0.52
Working Holiday Maker	17526 (3.4%)	2	35	6 (1, 23)	0.1 (0, 0.4)	0.1 (0.0, 0.5)	0.003
Family Reunion	3989 (0.8%)	12	8	151 (86, 265)	2.9 (1.6, 5.1)	3 (1.7, 5.4)	<0.001
Other	10187 (2.0%)	7	21	33 (16, 69)	0.6 (0.3, 1.3)	0.8 (0.4, 1.8)	0.65

CXR							
No abnormality	489733 (94.2%)	516	977	53 (49, 58)	1.0		
TB suspected	21862 (4.2%)	97	45	237 (196, 287)	4.2 (3.4, 5.3)	4.2 (3.4, 5.3)	<0.001
Abnormality not TB	8360 (1.6%)	6	17	36 (16, 80)	0.7 (0.3, 1.5)	0.8 (0.4, 1.9)	0.69
WHO category							
40-149	29143 (5.6%)	10	59	19 (10, 34)	0.3 (0.2, 0.6)	0.4 (0.2, 0.7)	0.001
150-349	75294 (14.5%)	88	151	60 (49, 73)	1 (0.8, 1.3)	0.7 (0.6, 0.9)	0.02
350+	415518 (79.9%)	521	828	64 (58, 69)	1.0	1.0	
Sputum culture testing							
No	179935 (34.6%)	279	353	81 (72, 91)	1.0		
Yes	340020 (65.4%)	340	685	50 (45, 56)	0.6 (0.5, 0.7)	0.6 (0.5, 0.7)	<0.001

Table 24. Baseline characteristics, univariate and multivariate analysis of incidence rates for smear positive pulmonary tuberculosis disease in migrants screened pre-entry (2006-2012) and notified in ETS (2006-2013)

Risk Factor	Migrants contributing (%)	Episodes	Person years at risk (1000)	Rate per 100,000 person years (95% CI)	Univariable IRR (95% CI)	Multivariable IRR (95%CI)	p-value
All	519955 (100.0%)	228	1038	24 (21, 27)	-	-	-
Age							
0-15	15468 (3.0%)	7	31	22 (11, 47)	1 (0.5, 2.2)	0.9 (0.4, 1.8)	0.72
16-44	490806 (94.4%)	215	979	22 (19, 25)	1.0	1.0	
45-64	11243 (2.2%)	3	23	13 (4, 41)	0.6 (0.2, 1.9)	0.5 (0.2, 1.5)	0.21
65+	2438 (0.5%)	3	5	41 (10, 163)	2.8 (0.6, 12.1)	1.4 (0.3, 6.5)	0.65
Sex							
Female	173116 (33.3%)	94	347	27 (22, 33)	1.0		
Male	346839 (66.7%)	134	691	19 (16, 23)	0.7 (0.5, 0.9)	0.8 (0.6, 1.1)	0.24
Contact with case TB							
No	516710 (99.4%)	224	1030	22 (19, 25)	1.0		
Yes	1220 (0.2%)	4	3	160 (60, 427)	7.5 (2.8, 20.1)	5.7 (2.2, 15.3)	<0.001
Missing	2025 (0.4%)	0	5				
Visa							
Students	307127 (59.1%)	117	609	21 (17, 25)	1.0		
Settlement and Dependents	159986 (30.8%)	97	320	32 (26, 39)	1.6 (1.2, 2.1)	1.4 (1, 1.9)	0.03
Work	21140 (4.1%)	9	45	22 (11, 42)	1.1 (0.5, 2.1)	0.9 (0.5, 1.8)	0.84
Working Holiday Maker	17526 (3.4%)	1	35	3 (0, 22)	0.2 (0.0, 1.1)	0.2 (0, 1.2)	0.07
Family Reunion	3989 (0.8%)	4	8	54 (20, 145)	2.6 (1.0, 7.1)	2.5 (0.9, 7)	0.07
Other	10187 (2.0%)	0	21	-	-	-	<0.001

CXR							
No abnormality	489733 (94.2%)	187	977	21 (18, 24)	1.0		
TB suspected	21862 (4.2%)	39	45	95 (69, 130)	4.7 (3.3, 6.7)	4.6 (3.2, 6.6)	<0.001
Abnormality not TB	8360 (1.6%)	2	17	13 (3, 52)	0.6 (0.2, 2.5)	0.8 (0.2, 3.1)	0.69
WHO category							
40-149	29143 (5.6%)	3	59	6 (2, 17)	0.2 (0.1, 0.7)	0.3 (0.1, 1.1)	0.07
150-349	75294 (14.5%)	34	151	24 (17, 34)	1 (0.7, 1.4)	0.8 (0.5, 1.1)	0.17
350+	415518 (79.9%)	191	828	25 (21, 28)	1.0	1.0	
Sputum culture testing							
No	179935 (34.6%)	104	353	32 (26, 38)	1.0		
Yes	340020 (65.4%)	124	685	20 (16, 23)	0.6 (0.5, 0.8)	0.6 (0.4, 0.7)	<0.001

The peak in bacteriologically confirmed pulmonary and non-pulmonary tuberculosis cases notified in ETS occurred one year after migration (Figure 26). There is a gradual decline in cases up to the maximum of 7 years after pre-entry screening. Adding in the prevalent cases detected by pre-entry screening (439 as described in Chapter 3) and post-entry prevalent cases (42 cases notified within 90 days of pre-entry screening) there is a clear decline in the number of cases notified post-migration.

Figure 27 illustrates the contribution of all ETS notified incident cases screened pre-entry (1863) in the context of all UK born and non-UK born tuberculosis cases in the UK. This figure demonstrates that an increasing number of migrants screened pre-entry are notified in the more recent years, but these remain a small proportion of all cases (578/7892; 7.3%), and of those not born in the UK (578/5529; 10.4%) in 2013.

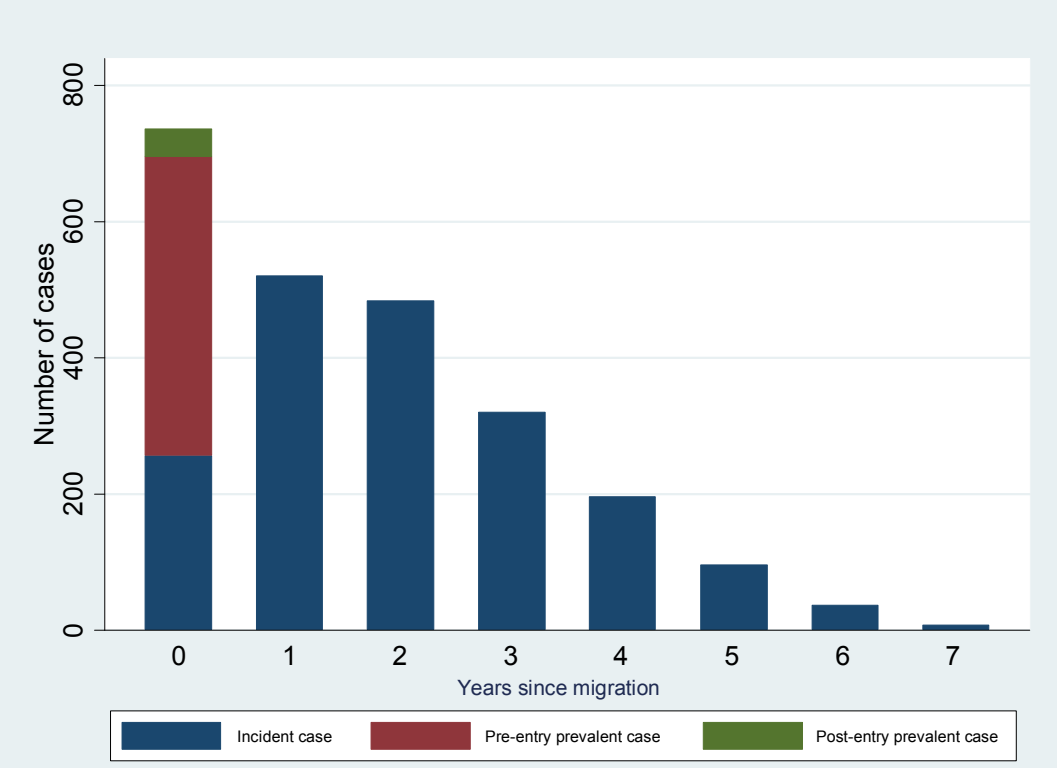

Figure 26. Pre-entry prevalent, post-entry prevalent (cases notified 90 days post migration) and incident bacteriologically confirmed (pulmonary and non-pulmonary) tuberculosis cases notified in the UK among migrant by year since migration.

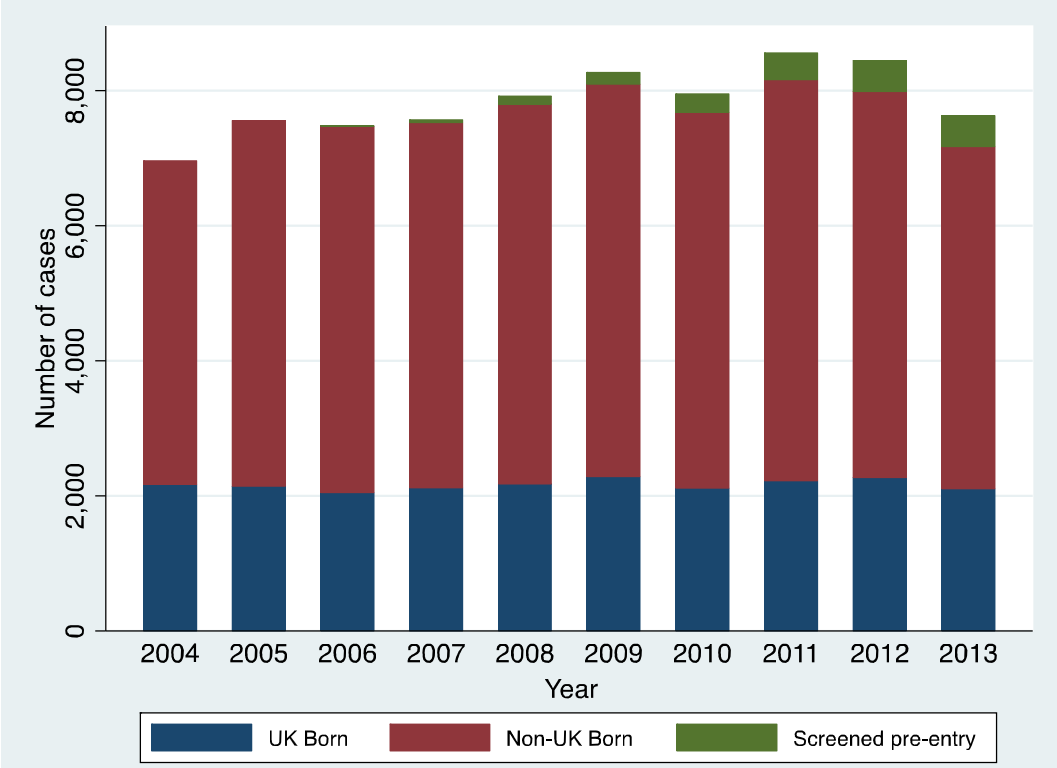

Figure 27. Tuberculosis case reports by place of birth and whether screened pre-entry or not (all ETS notified cases of tuberculosis in migrants screened pre-entry), 2004-2013, UK.

Several assumptions were made in estimating the incidence rates presented in this analysis. In order to examine the influence of these assumptions on the estimates and risk factors for bacteriologically confirmed pulmonary tuberculosis, several sensitivity analyses were carried out. Varying the assumption of prevalent cases up to 180 days (from 90 days) post migration made very little difference to crude incidence rates and the multivariable analyses (Figures 28 & 29). Several assumptions on when to censor migrants in the cohort were also tested. Assuming no deaths, no migrants leaving England, Wales and Northern Ireland until the end of study follow up (31st December 2013) and reducing the length of stay to a mean of 1.5 years for all visa categories, had minimal effect on incidence rates of bacteriologically confirmed pulmonary tuberculosis and the results of multivariable analyses. Increasing the number of migrants assumed to be living in Scotland increased the incidence rates slightly, but had no discernible effect on the multivariable analysis. Including all duplicates except those with a positive screen pre-entry had minimal effect on the incidence estimates, but did change the magnitude of some of the multivariable associations. Unlike the baseline analysis, when

including these duplicates there was evidence that those under the age of 15 were at lower risk of bacteriologically confirmed pulmonary tuberculosis (IRR 0.4; 95%CIs: 0.2, 0.7) and there was no longer evidence that those on working holiday makers were at lower risk of bacteriologically confirmed pulmonary tuberculosis (IRR 0.7; 95%CIs: 0.4, 1.3).

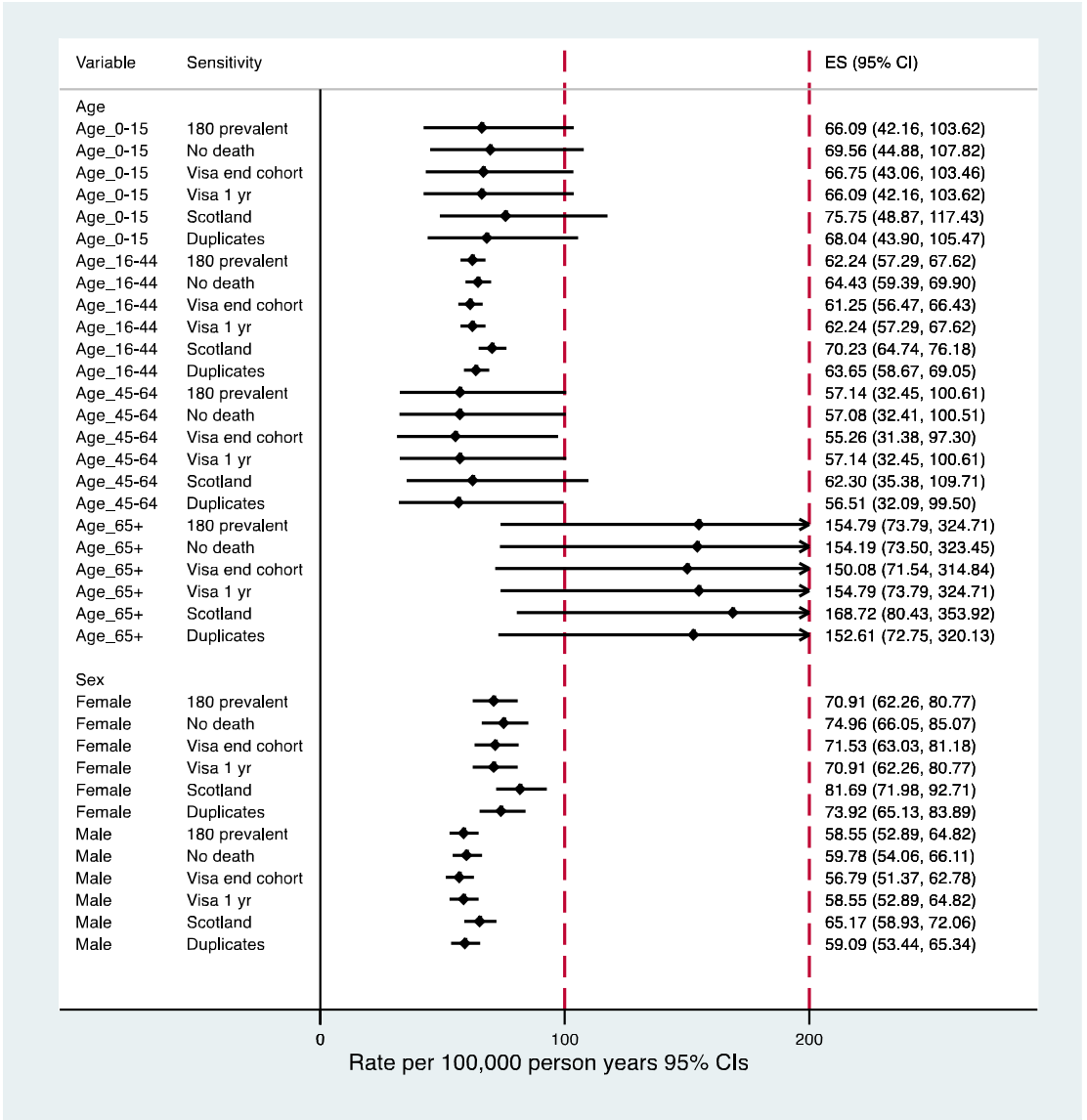

Figure 28. Sensitivity analysis of estimates bacteriologically confirmed cases of pulmonary tuberculosis incidence rates by age and sex under different model assumptions.

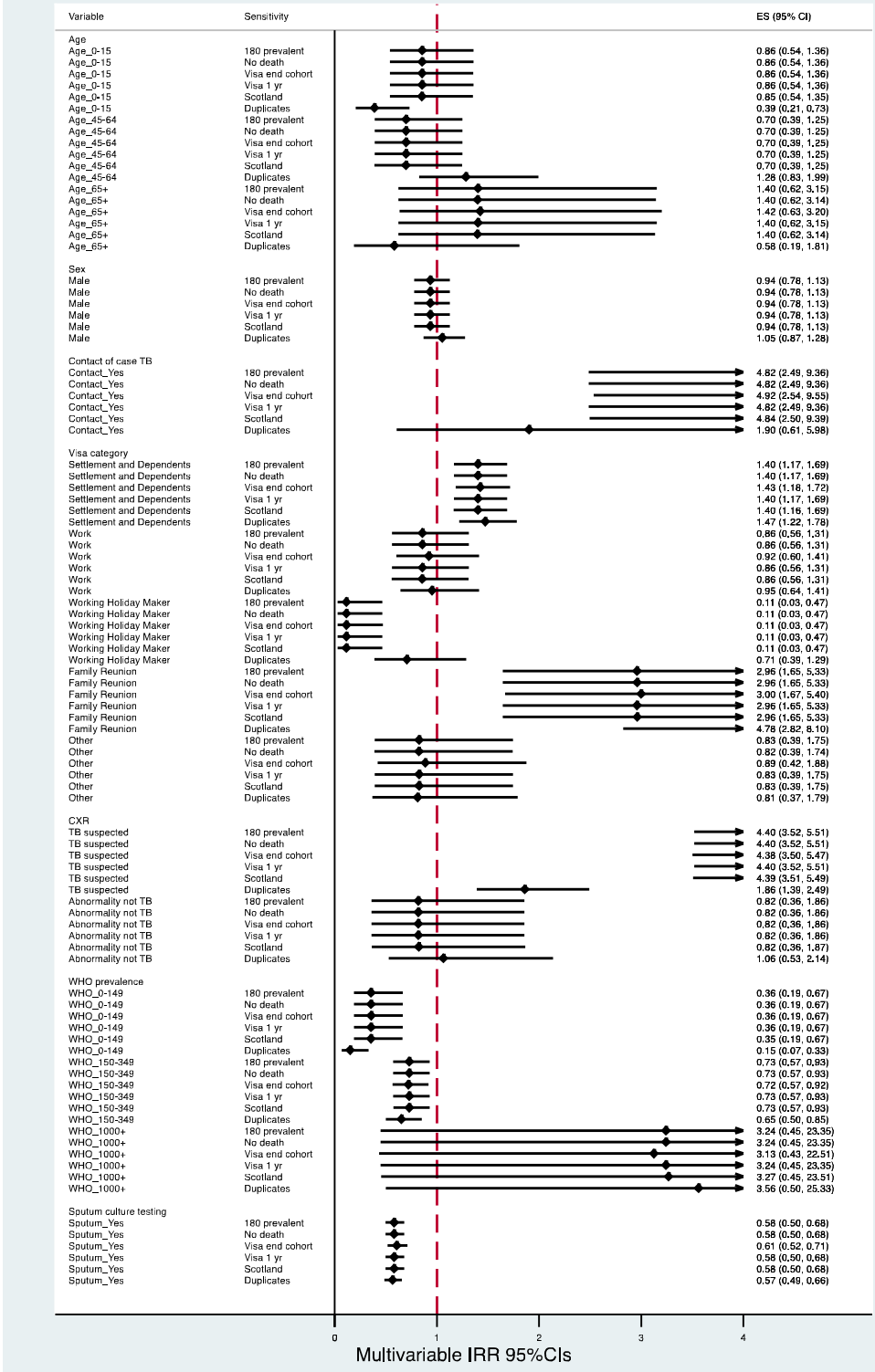

Figure 29. Sensitivity analysis of multivariable risk factor analysis for bacteriologically confirmed cases of pulmonary tuberculosis under different model assumptions.

5.5 Discussion

The analysis presents data on 640,808 visa applications, representing 514,968 migrants from 15 high incidence countries between 2006 and 2012 and over a million person years of follow up. The overall incidence of all notified cases was 194 per 100,000 person years at risk. The incidence of bacteriologically confirmed pulmonary tuberculosis was 65 per 100,000 person years at risk, which was slightly lower than the incidence of extra-pulmonary tuberculosis at 70 per 100,000 person years. Migrants with the highest crude incidence of bacteriologically confirmed pulmonary tuberculosis were those with a history of contact with a case of tuberculosis prior to migration, and those with a chest radiograph suggestive of active pulmonary disease at pre-entry screening. These two factors not only had the highest crude incidence rates of bacteriologically confirmed pulmonary tuberculosis, but they were also associated with the highest incidence rate ratio in a multivariable risk factor analysis adjusting for age and sex. Migrants screened at locations where sputum samples underwent culture testing had a lower incidence rate ratio for bacteriologically confirmed pulmonary and extra-pulmonary tuberculosis after migration in a multivariable risk factor analysis.

Self-reported history of contact with a case of tuberculosis was associated with an increased risk of bacteriological confirmed tuberculosis after arrival, and therefore this may be a useful group to target health improvement interventions at. The fact that there was no corresponding increased risk for extra-pulmonary disease is potentially explained by the small number of cases, but could also represent the longer time period between exposure and disease outcome, and the potential for a recall bias in the extra-pulmonary disease group.(157,158) Unlike the analysis in chapter 3, this study found an increasing risk of incident tuberculosis with migration from higher prevalence countries. This finding not only provides some additional evidence that the results of the analysis are robust, but also suggests that risk after migration is, to some extent, determined by historical exposures prior to migration, although it raises questions about why migrants from these countries when screened at pre-entry do not have the highest prevalence of active disease. This finding will be explored further in the next chapter by examining the incidence of tuberculosis reactivation and transmission in migrants in England, Wales and Northern Ireland, using additional strain typing data available in ETS.

5.5.1 Strengths and weaknesses of the study

All visa applicants from 15 countries taking part in the UK pre-entry screening programme between 2006-2012 were included in this analysis which as a result is highly representative of migrants from these locations applying to stay for 6 months or longer. Levels of missing data were low on matching variables and therefore probabilistic linkage accuracy should be high and consistent with findings presented in chapter 4. Migrants entering the cohort had between one and seven years of follow up, with a mean of two years. The large sample size and follow up time minimize statistical chance as an alternative explanation for the majority of findings. The primary outcomes were based upon WHO case definitions for bacteriologically proven pulmonary and extra-pulmonary tuberculosis, enabling international comparisons. The analysis presented in chapter 4 of the thesis established the ability of probabilistic matching algorithms to be used to determine outcomes in such a cohort with a high level of accuracy. Chapter 4 also highlighted the possibility that linkages performed between large datasets with a low match rate between them could lead to overestimates in incidence rates when the combined dataset was used for such purposes. This was therefore a possibility with this analysis, however, the estimate of incidence rates for all ETS cases was 194 per 100,000 person years at risk (95%CIs: 186, 203), which is lower that the incidence rate in non-UK born cases of Pakistani ethnicity in ETS which was 286 per 100,000 population in 2014. Whilst this does not definitely rule out the possibility of over estimation due to a linkage bias, it provides some reassurance that the estimates are within an appropriate range.

Whilst there was a high level of certainty that migrants receiving a visa for entry into the UK do migrate, there is less certainty about when and whether these same individuals leave after their visa expires. The statistical approach taken in this analysis attempts to account for the unknown duration of stay and death rates through the use of imputation that was based on historical visa length data and national death rates. Several sensitivity analyses were undertaken to examine the reliability of these assumptions.

To provide conservative estimates of incidence rates, by using a maximum potential person time at risk for the cohort, one sensitivity analysis assumed that all migrants stayed until the end of the cohort (31st December 2013). This assumption had little effect on both the crude incidence rates, and the risk factor analysis. Assuming no

deaths in the cohort, again to provide a more conservative person time at compared to the baseline imputation, and had minimal impact on crude incidence rates and the results of a multivariable analysis. Mortality rates in migrants may be higher than general population from which these estimates were derived, but given the fact that the population included in this analysis was young, with the majority of individuals within the 16-44 age range, it is unlikely to lead to a substantial bias and it is also not surprising that assuming no deaths had little effect on the results.

The IOM pre-entry screening database was linked to ETS, which contains personal identifiable information for all cases in England, Ireland and Wales, but not for Scotland. It was therefore necessary to account for the fact that the probabilistic matching would not identify tuberculosis cases in Scotland, and to reduce the denominator used in estimating incidence rates to account for this issue. This assumption was examined by increasing the number of migrants in the cohort assumed to reside in Scotland. This sensitivity analysis resulted in an increase in the incidence rates for bacteriologically confirmed pulmonary tuberculosis, but had minimal effect on the multivariable risk factor analysis.

The baseline analysis assumed that cases of tuberculosis notified within 90 days of a medical clearance certificate were missed prevalent cases and therefore should therefore be excluded from the estimates of incidence. Increasing this time period to 180 days had very little difference to incidence rates and the results of the multivariable analyses.

The assumptions used to exclude duplicates from the analysis were also examined. In the baseline analysis, individuals were excluded according to the rules described in Tables 15 & 16. A sensitivity analysis, which included all duplicates except for those classified as active tuberculosis had little impact on the incidence rates, but did influence the multivariable analysis. Under this new assumption, migrants under the age of 15 were found to have evidence for being at lower risk of bacteriologically confirmed pulmonary tuberculosis (there was no evidence in the base case analysis) and migrants on working holiday visas were no longer at lower risk. These two groups were both small categories within the analysis, and although the number of duplicates for these two groups was similar to others (13.5% for 0-15 year olds and 9.8% for working holiday visas) the absolute number of individuals within the analysis was smaller. The changes with regards to de-duplication could therefore be explained by several factors

including chance (due to the small numbers), confounding by age and sex, or bias by created by inappropriately including the duplicates in this new analysis. Overall the sensitivity analyses provide reassurance that the main results of this analysis were reliable despite changes in some of the underlying assumptions.

Whilst this analysis has several strengths, some important limitations should be considered when interpreting these results. These data are highly representative of migrants to the UK who are intending to stay for month than six months, but they do not cover undocumented migrants, asylum seekers or those intending to stay in the UK for less than 6 months.(79) Asylum seekers and undocumented migrants are likely to have the highest risk tuberculosis, particularly when compared with visa applicants in this analysis who will be of a higher socio-economic status.(111–116,159–161) These two groups are therefore likely to have a higher risk of tuberculosis than migrants included in this study, but due to the small numbers, will account for a smaller proportion of the total cases in ETS.

It has not been possible to account for visits back to the country of origin, where the risk of being exposed to a case of tuberculosis will be increased compared with those in the UK.(76) However, whilst these trips will increase the likelihood of exposure to a case of infectious tuberculosis, the decrease in case numbers over time since migration (Figure 26) suggests that the impact of return visits is likely to account for only a small proportion of all cases notified in migrants screened pre-entry.

The use of probabilistic linkage to determine outcomes in migrants screened pre-entry introduces several potential biases in the study, including false positive and false negative matches. False positive links will bias incidence rate ratios towards the null, and false negatives will have no impact on incidence rate ratios, assuming there is non-differential misclassification bias of risk factor variables.(133,142,143) Chapter 4 provides some reassurance that the number of false positive and false negative results will be negligible, however, it is hard to provide a more robust estimate of the likely impact of these issues using the current dataset which lacks the ability to confirm with certainty whether matches are true or not.

Several important risk factors for tuberculosis were included within this analysis, but there were no data on socio-economic status, clinical conditions associated with an

increased risk of tuberculosis (such as HIV) or lifestyle and behavioural risk factors such as smoking, problem drug and alcohol use, and a history of imprisonment, all of which have been shown to be associated with an increased risk of tuberculosis in the UK.(35,74,75,162–165) The limited duration of follow up prevents looking at incidence after first few years of migration.

5.5.2 Strengths and weaknesses in relation to other studies

Several previous studies have examined the incidence of tuberculosis in migrant populations screened pre-entry after arrival in a host country.(166–169) None of these studies included all migrants from pre-entry screening programmes, and included only migrants to certain geographical locations in the destination country. One study used probabilistic matching to identify migrants in a national case notification database.(166) Consistent with the data presented here, two studies found a reduction in incidence of tuberculosis post-arrival in those migrants screened using a protocol that included culture testing of sputum samples, compared to one that only used smear testing.(166,168) One of these studies was not able to link records between the pre-entry screening programme and notification data post arrival, and therefore the analysis was based on a description of incidence rates in non-US born migrants in the periods before and after the introduction of culture testing within the pre-entry screening protocol. A second study also presented data on a small high-risk population (Hmong refugees screened at a camp in Thailand) which will be at much higher risk of tuberculosis than the population in this cohort analysis.(168) Therefore the results presented in this analysis provide stronger evidence of the ability of the culture testing to reduce the incidence of tuberculosis after migration from a large representative sample of longer-term migrants compared to the existing literature.

One US study used probabilistic matching to identify tuberculosis cases in California-bound Filipino migrants, and found that those migrants with an abnormal chest radiograph when screened pre-entry had a higher incidence of tuberculosis post-arrival.(170) This study also found that the incidence of tuberculosis decreased with time since follow up, both results are therefore consistent with the findings presented in this chapter. This US study had additional risk factor data on migrants, including self-reported data on current or historical smoking and clinical risk factors such as diabetes

and malignancy, but unfortunately did not include these within a risk factor analysis to examine which were associated with an increased risk of tuberculosis.

5.6 Conclusion

This is the first comprehensive study to examine the incidence of tuberculosis in migrants from high to low incidence countries, after they have been screened for active tuberculosis as part of a visa application system. The large sample size, mix of countries involved, and representativeness of the study for migrants intending to stay for longer than six months mean that the study is appropriately powered and representative. Several novel findings are presented, including for the first time true estimates of the incidence of tuberculosis in a high-risk population that have been screened for active pulmonary disease prior to arrival. The risk factors identified can be used to improve existing health programmes and policy in this area. The public health implications of this analysis, recommendations, and directions of future research are discussed in the final chapter of the thesis.

CHAPTER 6

Molecular epidemiology of tuberculosis cases detected in England, Wales and Northern Ireland in migrants screened pre-entry: A cross-sectional and cohort study.

6.1 Abstract

Background: A better understanding of the amount of tuberculosis in migrants screened pre-entry that leads to transmission in the UK, and the burden that is due to reactivation from infection acquired abroad, would guide screening policy and the development of health improvement interventions in this high-risk group.

Methods: Strain typing data from ETS notifications were linked to pre-entry data. First cases in a cluster and cases of reactivation were examined in a cross-sectional study comparing migrants screened pre-entry with non-UK born individuals. A cohort study was also performed in pre-entry screened migrants. Incidence rates were estimated and a multivariable risk factor analysis conducted.

Results: In the cross-sectional analysis, after adjusting for age and sex, there was strong evidence that having been screened pre-entry was associated with a lower odds of being the first case in a cluster (OR 0.6; 95%CIs: 0.5, 0.8; p-value <0.001). In the cohort analysis, there were 35 migrants who were the first case of tuberculosis in a cluster with an estimated crude incidence rate of 6 per 100,000 person years at risk (95%CIs 4, 8). There were 301 cases of reactivation with a crude incidence rate estimated at 38 per 100,000 person years at risk (95%CIs: 33, 43).

Conclusion: These results will inform the effective strategies to improve the health of migrants and can be used to parameterise health economic models examining the most cost effective strategies to reduce burden of disease in this population.

6.2 Introduction

As chapter 3 demonstrated, pre-entry screening for tuberculosis in migrants detects cases of active disease prior to entering the UK with a prevalence of culture confirmed cases consistent with other published international studies. Prevalent cases detected at pre-entry screening undergo treatment and re-screening before they are given medical clearance for entry. Removing these prevalent cases of tuberculosis in migrants prior to entering the UK may mean they are less likely to result in clusters of transmission than migrants who have not been screened. Chapter 5 provides evidence that after arrival in the UK, the incidence rate of tuberculosis among migrants screened pre-entry from high burden countries remains higher than UK born population in England, Wales and Northern Ireland. This higher incidence post migration compared to the UK born population is likely to be due to several reasons, including an increased risk of latent tuberculosis reactivation (from exposure to cases of tuberculosis before migration), or through contact with active tuberculosis cases after arrival.

Molecular epidemiology data can be used to understand whether migrants screened pre-entry are at lower risk of reactivation and less likely than non-UK born individuals to transmit tuberculosis (using the proxy marker of being the first case in a cluster of tuberculosis cases). Clustered cases indicate transmission, and first in cluster cases arise when transmission occurs from a reactivation case followed by progression to disease in secondary cases. Therefore, the occurrence of first in cluster cases indicates opportunities to improve tuberculosis control, such as promoting early diagnosis through awareness raising activities and through the provision of better access to healthcare, reducing time at risk for exposing other individuals. Estimating the incidence of reactivation cases is important as these cases are potentially preventable through screening and treatment for latent infection, but currently there is poor information on risk of progression in new migrants.

The molecular epidemiology data required for these analyses are available from the UK tuberculosis strain typing service that was set up to reduce misdiagnoses and transmission by improving detection of active and latent cases through improved targeting of outbreak investigations.(171) Several molecular epidemiological techniques have been developed in order to help better understand the transmission of tuberculosis,

several of which are summarized in image 2, including: IS6110-RFLP (Image 2: 1); IS6110-Based polymerase chain reaction (PCR) Fingerprinting (Image 2: 5, 6 & 7); Spoligotyping (Image 2: 2) and Mycobacterial Interspersed Repetitive Units (MIRU) strain typing (Image 2: 3). MIRU based on 24 loci has been used by the UK tuberculosis strain typing service on all culture confirmed cases since 2010 (45). MIRU strain typing is based on a PCR technique that targets specific regions on the *M.tuberculosis* chromosome that contain repeated sequences of DNA. The technique uses primers ("a strand of short nucleic acid sequences that serves as a starting point for DNA synthesis"(3)) that are specific to regions on either side of these repeated sequences, which are then amplified using PCR. The size of the products resulting from this PCR amplification (yellow horizontal lines in image 2) are estimated to deduce the number of repeats in each individual locus, which provides the basis for classifying the strain type using this technique.(172)

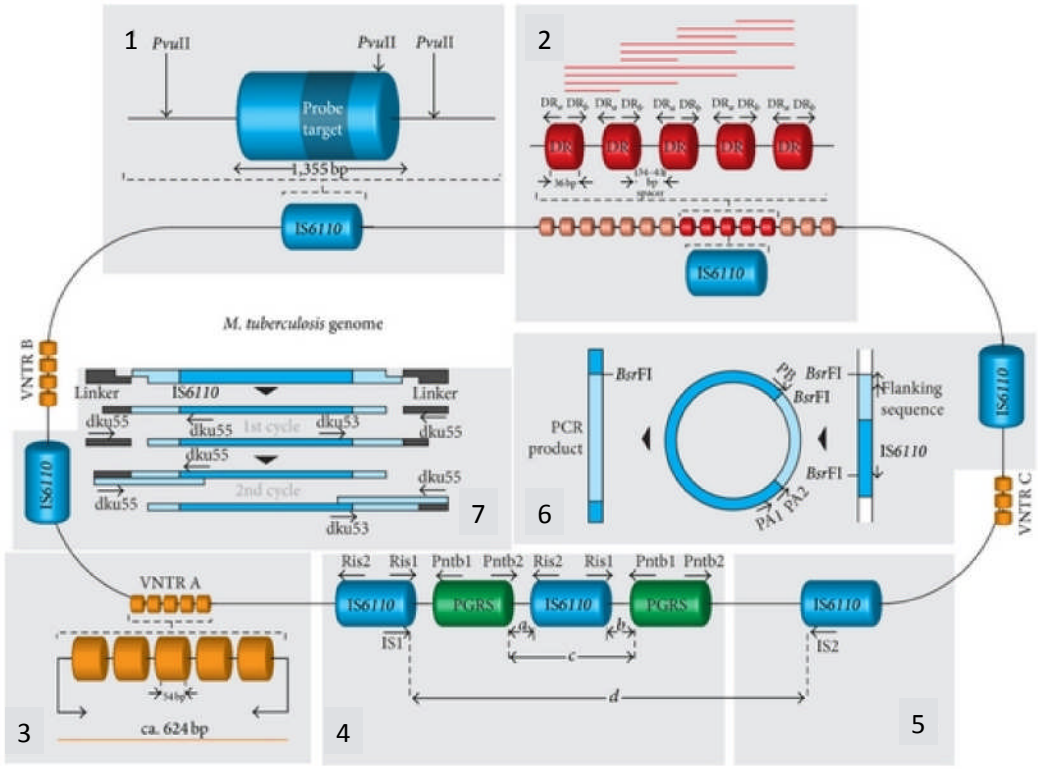

Image 2. **Schematic representation of the chromosome of a hypothetical M.tuberculosis complex isolate with marked repetitive elements as targets for different typing methods.**

Source: Jagielski T et al. Current Methods in the Molecular Typing of Mycobacterium tuberculosis and Other Mycobacteria. BioMed Res Int. 2014 Jan 5;2014:e645802. Reproduced with permission under creative commons.**(172)**

Several studies have examined the extent of tuberculosis transmission in migrants from high incidence countries to a host population. Using data from 1995-2006, a study in the Netherlands estimated that 38% of tuberculosis cases were a result of infection acquired abroad, and 36% resulted from recent transmission in the Netherlands, 18% resulting from longer term infection in the Netherlands, and 9% undetermined.(173) A study in Barcelona using restriction fragment length polymorphism (RFLP)-IS6110 and MIRU12 as a secondary typing method, concluded that there was evidence of transmission occurring between Spanish-born and migrant populations (in both directions), but that over half of the cases in migrants occurred within two years of arrival were a result of reactivation.(174) Outside of Europe, a study in Canada

examined lineage-specific trends and found that phylo-geographic strain type lineage was highly conserved in migrants up to five years after migration, indicating either reactivation of strains acquired in the country of origin, or transmission between people originating from the same geographical area post migration.(175) A study of molecular epidemiology data from cases of tuberculosis reported in San Francisco between 1 January 1991 and 31 December 2003 found that the rate of clustering decreased during this time period from 11.4 to 3.1 cases per 100,000 population.(176)

This chapter will create a dataset that links pre-entry screening data to strain typing data in the UK case notification system to enable the investigation of whether a pre-entry screened population is less likely to be the first case in a transmission cluster, compared to migrants not screened, as well as the ability to examine whether they are at a lower risk of tuberculosis reactivation as a result of screening. The chapter also estimates the incidence of first in cluster cases of tuberculosis and the incidence of reactivation in pre-entry screened migrants. Evidence that migrants screened pre-entry for active tuberculosis were at lower risk of being the first case in a cluster, would support the effectiveness of pre-entry screening. Additionally, a better understanding of the incidence of first in cluster cases and reactivation, and the risk factors associated with each of these outcomes, is of public health importance as it could be used to guide screening policy and the development of health improvement interventions in this high risk group.

6.2.1 Research questions:

The analysis presented in this chapter aims to answer the following research questions:

1. Are migrants screened pre-entry at a lower risk of being the first case in a cluster of transmission compared to non-UK born individuals not screened pre-entry?
2. Are migrants screened pre-entry at a lower risk of reactivation of tuberculosis compared to non-UK born individuals not screened pre-entry?
3. What is the incidence of first cases in a cluster of transmission and reactivation tuberculosis in migrants screened pre-entry to England, Wales and Northern Ireland?

4. What are the risk factors for being the first case in a cluster of transmission, and of reactivation notified in in migrants screened pre-entry to England, Wales and Northern Ireland?

6.3 Methods

6.3.1 Study design, participants and settings

Cross-sectional study

Two different study designs were used in this analysis. The first was a classical analysis of risk factors for clustering, which could be described as a cross-sectional analysis, and included all Enhanced Tuberculosis Surveillance system cases notified between 1st January 2010 and 31st December 2013. Migrants screened pre-entry between 1st January 2009 and 31st December 2012 were included in this analysis and identified through probabilistic linkage. In this cross-sectional study, those migrants screened pre-entry were compared to other non-UK born individuals notified in the same time period, who were able to arrive at the UK at any time point. It should be noted that for the rest of this chapter the comparator population of individuals not screened pre-entry will be called non-UK born as per convention in ETS, from where they have been identified. This population is not called migrants as they will include refugees and asylum seekers and no details are available from where they migrated to the UK - instead data are available as to which country they were born in. This study will be referred to as the cross-sectional study for the remainder of this chapter.

Cohort study

The second study was a retrospective cohort of migrants screened pre-entry by IOM. Migrants included in this cohort were screened pre-entry between 1st January 2009 and 31st December 2012. Tuberculosis outcomes were identified in the ETS between 1st January 2010 and 31st December 2013 using the probabilistic linkage methods described in chapter 4. This second part of the study will be referred to as the cohort analysis for the remainder of the chapter.

6.3.2 Strain typing

Prospective strain typing on culture confirmed cases of tuberculosis diagnosed in the UK was undertaken from January 2010 until 31st December 2013 using 24 loci MIRU.(177) Strain typing was carried at regional centres for Mycobacteriology with the aim of typing over 95% of all initial *M.tuberculosis* isolates as part of the routine service.(178)

6.3.3 Outcomes

For the cross-sectional study the primary outcomes were:

1. The first case of tuberculosis in a cluster.
2. Reactivation cases of tuberculosis.

Two primary outcomes were considered for the cohort study:

1. Incidence of first in cluster cases of tuberculosis.
2. Incidence of reactivation tuberculosis cases.

6.3.4 Definitions of a cluster, first case of tuberculosis in a cluster, and reactivation

Cases of tuberculosis classified as first case of tuberculosis in a cluster or reactivation must have been culture confirmed and have MIRU profiles with at least 22 complete loci notified in ETS between 1st January 2010 and 31st December 2013.(179) As per the definition used by Public Health England, a cluster was defined as: "two or more cases with indistinguishable 24 MIRU strain types with at least one case with a complete 24 loci profile".(136,179,180) Patients were clustered irrespective of geographical area of residence in England, Wales and Northern Ireland. This first case in a cluster was identified by recorded date of notification, and used the primary outcome for the cross-sectional and cohort study.

Reactivation cases were all cases with a unique 24 MIRU strain type between 1st January 2010 and 31st December 2013, and the first reported case in a cluster with an indistinguishable MIRU strain type to others as this first case in a cluster was assumed to be due to reactivation.(169)

6.3.5 Exposures

Cross-sectional study

Age and sex were considered a-priori as confounding variables. Risk factor analysis for incident disease included the following variables from ETS: BCG vaccination status; at least one social risk factor (drug use, homelessness, alcohol misuse/ abuse, prison); time since entry in the UK; whether a case had been screened pre-entry (identified through probabilistic linkage as described in next section); and WHO prevalence estimates of tuberculosis in the country of origin at the time of migration. WHO prevalence, rather than incidence data was used to stratify countries estimates in order to remain consistent with results presented in Chapters 2, 4 and 5.

Cohort study

Age and sex were considered a-priori as confounding variables. Risk factor analysis for incident disease included the following variables from the IOM dataset: visa category, contact with a case of tuberculosis, chest radiograph classification at pre-entry screening, whether a migrant was screened at a centre with culture testing of sputum samples, and WHO prevalence estimates of tuberculosis in the country of origin.

6.3.6 Statistical methods

Cross-sectional study

Baseline descriptive statistics were provided using simple counts and proportions. A multivariable logistic regression model was used to identify risk factors for primary outcomes. Final results were presented as odds ratios, 95% confidence intervals and p-values. Missing data were examined with a descriptive analysis, and variables with missing data levels greater than 1% were included as a separate "missing" category in the final multivariable regression analysis.

Cohort study

Baseline descriptive statistics of the cohort were provided using simple counts and proportions. To account for the uncertainty of death, migration to Scotland and emigration, multiple imputation was performed. External data were used to inform the imputation models. Ten imputed datasets were created. Individual imputed datasets

were analysed using Poisson regression, a suitable method for modelling rare event data, and estimates of crude and adjusted incidence rates for the primary and secondary outcomes were calculated. A multivariable Poisson regression model was used to identify risk factors for the primary outcomes. All results were adjusted for clustering by individual, to take account of repeated entries by migrants into the cohort. The results of the analyses from individual imputed datasets were then combined using Rubin's rules; these appropriately account for uncertainty in the imputed information. Final results were presented as incidence rate per 100,000 person years at risk, incidence rate ratios, 95% confidence intervals and p-values. Stata v.13 (Statacorp LP, College Station, TX, USA) was used for all statistical analyses.

6.3.7 Data sources

The datasets for both analyses were created using two sources: 1) The IOM database of migrants screened in their country of origin between 1st January 2009 and 31st December 2012; and 2) Enhanced Tuberculosis Surveillance system for cases notified between 1st January 2010 and 31st December 2013. These two datasets were probabilistically linked using the enhanced matching system described in chapter 4 using first name, surname, date of birth, nationality and sex as identifying variables. Cleaning and consistency checking of the final dataset was undertaken by looking at the distribution of variables, the range of individual variables, and missing data.

6.3.8 Exclusions:

Cross-sectional study

The dataset for the cross-sectional study excluded migrants screened pre-entry that were not notified as a case of tuberculosis, and included all other ETS cases notified during 1st January 2010 and 31st December 2013 for baseline descriptions of the data. The analysis to determine risk factors for first in cluster cases and reactivation excluded UK born individuals.

Cohort study

The dataset for the cohort study excluded all ETS cases that had not been screened pre-entry. Any primary outcomes occurring within 90 days of issue of a medical certificate

of clearance by IOM were assumed to be prevalent cases that were missed by pre-entry screening and were not included as a primary outcome. This assumption was examined further in a sensitivity analysis.

6.3.9 Sensitivity analyses

The assumptions made in these analyses were examined in several sensitivity analyses.

Cross-sectional study

In the primary analysis all cases in non-UK born migrants notified during 1st January 2010 and 31st December 2013 were included. Many of these individuals will have migrated to the UK before 2005 when the first pre-entry case would have arrived. The primary analysis included a variable for time since arrival in the analysis to control for this issue. However, as time since infection is important factor that distinguishes infection versus reactivation, a multivariable analysis to determine risk factors for the primary outcomes was performed on a dataset restricted to non-UK born cases that arrived after 2005 to make this population more comparable (in terms of time since entry to the UK) with the pre-entry screened migrants.

Cohort study

Varying the definition of how incident and prevalent cases were distinguished (i.e. a 90 day cut off between pre-entry screening and notification) was examined. For the main analysis, it was assumed that cases notified within 90 days of pre-entry screening were 'missed' prevalent and not incident cases. The impact of varying this definition to include all 'missed' prevalent cases, and increasing it up to 180 days was assessed. A sensitivity analysis was conducted to include only migrants screened under the routine pre-entry culture testing protocol as this is likely provide a better indication of those who were genuinely disease free at time of entry due to its increased sensitivity.

6.3.10 Cohort study specific methods:

Censoring

Migrants entered the cohort upon receiving a certificate of medical clearance for tuberculosis after screening pre-entry by IOM. Individuals were followed up until the

first of tuberculosis, death, or emigration. Death and migration out of the country were imputed as described in the baseline scenario for Chapter 5.

Duplicates and migrants to Scotland

Duplicates were removed from the analysis according to the rules described in Tables 15 & 16 in chapter 5. It was also assumed that 92.7% of migrants issued with a medical certificate would reside in Scotland, and therefore would not have primary outcomes identified as part of the probabilistic matching process. Therefore for each imputation 92.7% of the total population screened were randomly selected not enter the cohort.

Estimating rates in the non-UK born population

Estimates of the first in cluster cases of tuberculosis and reactivation from this study were compared to rates found in all non-UK born cases not screened pre-entry. To undertake this analysis, estimates of population denominator for non-UK born individuals entering the UK between 1st January 2009 and 31st December 2013 were required. The denominator population was estimated using data on long-term migrants (staying more than 12 months) and short term migrants (staying longer than 3 months).(181,182) The number of migrants screened pre-entry was removed from this denominator estimate. For long-term migrants, person time at risk was assumed to be the same as the mean estimated through imputation for migrants included in the cohort analysis. Short-term migrants (who stay less than 12 months) were assumed to stay for nine months.

Numerator cases for first in cluster cases of tuberculosis and reactivation were identified from the ETS dataset, using the definitions for the primary outcomes as described above, but excluding migrants screened pre-entry with these outcomes. Year of migration was not known for 10.9% (2,606/23,911) of non-UK born cases in ETS. For migrants with year of migration known, 30% migrated after 1st January 2009. It was assumed that there were no biases in missing data for year since migration, and therefore 30% were assumed to migrate after 2009 and these cases were included in the numerator estimates.

6.3.11 Ethics approval

Ethical approval was received for this analysis from UCL research ethics committee (3294/002; See Appendix 4 for copy of approval letter). The work was conducted with Public Health England which has Health Research Authority approval to hold and analyse national surveillance data for public health purposes under Section 251 of the NHS Act 2006.

6.4 Results

Between 1st January 2009 and 31st December 2012 there were 402,053 visa applicants screened for tuberculosis pre-entry by the IOM programme (Figure 30). Records from these applications were probabilistically matched to ETS for the period 1st January 2010 and 31st December 2013, which contained 33,942 records. Probabilistic linkage identified 1,590 migrants screened pre-entry who were included in the cross-sectional study. After excluding duplicates, individuals migrating to Scotland, and missed prevalent cases, 318,983 visa applicants (315,631 individuals) entered the cohort analysis.

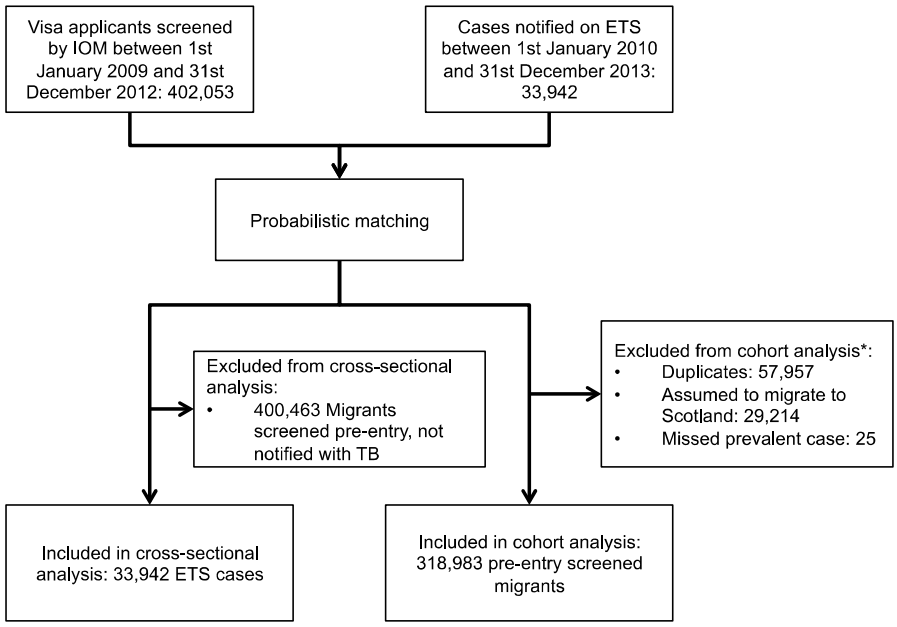

Figure 30. Study participant flow diagram

*Numbers assumed to migrate to Scotland varied by imputation. Sum of those excluded does not equal difference between total visa applicants and the number of migrants included in final cohort, as groups are not mutually exclusive.

6.4.1 Cross-sectional analysis

For ETS cases notified between 2010 and 2013, 80.7% (16,602/20,560) of all culture confirmed cases had at least 23 loci strain typing performed, compared with 79.5% (836/1,051) in migrants screened pre-entry (Table 25). Clustering occurred in 53.5% of all ETS cases (8,890/16,602), with a total of 1,854 molecular clusters. There were 1,605 clusters including at least one non-UK born individual not screened pre-entry, and 247 involving at least one pre-entry screened case. For all UK notified cases, unique strain types were found in 46.5% (7,712/16,602), 50.3% (5,575/11,090) for migrants not pre-entry screened, and 53.8% (450/836) for pre-entry screened migrants, respectively.

Of those cases in clusters, the largest proportion of all UK notified cases were in clusters that involved more than 10 others of the same strain type (0.20; 3,396/16,602; Figure 31). The proportion of migrants screened pre-entry that were in clusters involving more than 10 other individuals was 0.18 (146/836), and it was also 0.18 (1,966/5,575) for migrants not screened pre-entry. The 95% confidence intervals for estimates of unique cases in migrants screened pre-entry, compared to non-UK born migrants overlap, providing no statistical evidence for a difference in these two populations.

When considering the proportion of all clusters by size, 0.46 (858/1854) of all UK clusters involved only 2 cases with the same strain type (Figure 32). Clusters involving two cases with at least one of these being an individual screened pre-entry, accounted for 0.32 of all clusters (80/247), compared to 0.45 in all ETS clusters (724/1,605) that involved at least one non-UK born individual not screened pre-entry.

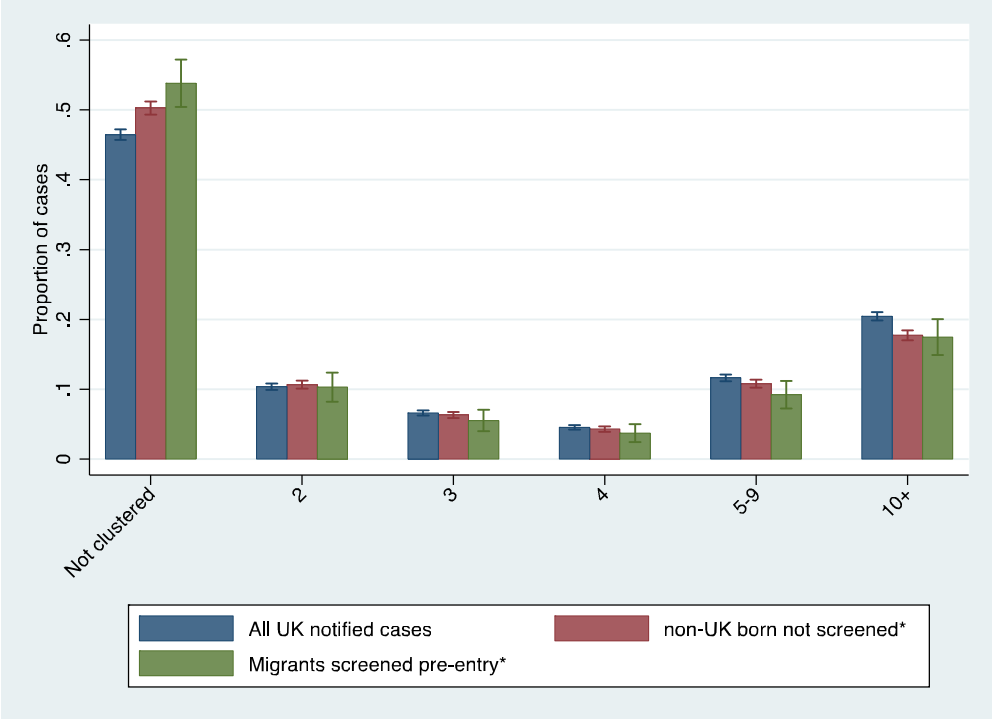

Figure 31. Proportion of all cases by unique strain or cluster size.

*Clusters with at least one non-UK born, or pre-entry screen person in the cluster

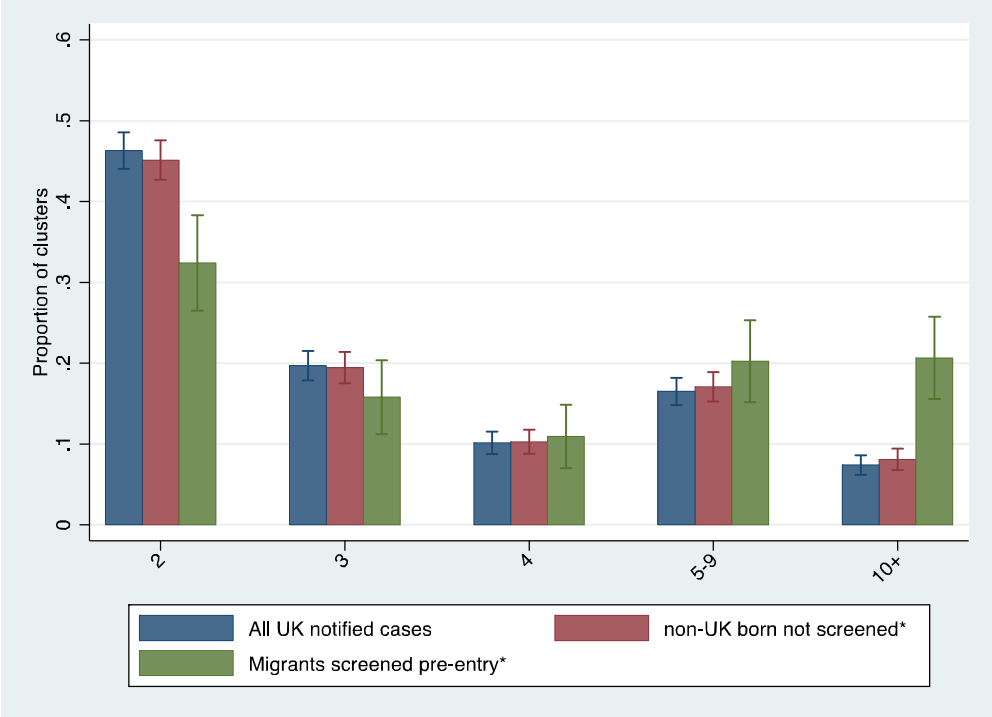

Figure 32. Proportion of clusters by size, 2010-2013.

*Clusters with at least one non-UK born, or pre-entry screen person in the cluster

Place of birth was missing for 1,354 individuals notified as a case of tuberculosis between 2010-2013, of which 44.7% (95%CIs: 42.0, 47.4) were culture confirmed cases with a strain type, compared to 49.9% (95%CIs: 49.2, 50.5) in non-UK born individuals and 46.9% (95%CIs: 45.9, 48.0) in UK born individuals. There were a total of 11,926 culture confirmed cases with a strain type in individuals not born in the UK between 2010 and 2013 (Table 26). These records were included in the cross-sectional analysis to identify risk factors for being a case of tuberculosis reactivation in migrants.

The majority of cases were men (59.8%; 7,130/11,926) and 71.7% of cases (8,547/11,926) were aged between 16-64 (Table 26). Over half of all non-UK born individuals had been BCG vaccinated (51.6%; 6,156/11,926) and 44.8% (5,346/11,926) came from countries with a WHO prevalence of greater than 350 per 100,000 population. A total of 836 (7%) of cases were identified as having been screened pre-entry, with a median time since entry of two years compared to six years for migrants not screened pre-entry.

There were 1,316 (11.0%) first cases in a cluster in non-UK born individuals. Those migrating within three to five years had a higher percentage of first cases in a cluster (11.5%; 238/2,076). Those screened pre-entry were at lowest risk of being the first case in a cluster at 8.0% (67/836). A multivariable logistic regression was performed to identify risk factors for being the first case in a cluster. After adjusting for age and sex, there was strong evidence that having been screened pre-entry was associated with a lower odds of being the first case in a cluster (OR 0.6; 95%CIs: 0.5, 0.8; p-value <0.001). There was evidence that compared to individuals from countries with a WHO prevalence of tuberculosis greater than 350 per 100,000 population, those from countries with a prevalence between 150 and 349 were at reduced odds (OR 0.8; 95%CIs: 0.7, 1.0; p-value 0.03) after adjusting for age and sex. No other risk factors were identified.

There were 6,025 (50.5%) cases of reactivation in non-UK born individuals between 2010-2013 (Table 27). The highest percentage of reactivation was found in those migrants screened pre-entry (53.8%; 450/836) compared to 50.3% (5,575/11,090) in those not screened. The lowest percentage of reactivation was found in non-UK born individuals with a social risk factor (drug use, homelessness, alcohol misuse/ abuse, prison) at 42.1% (334/794).

A multivariable logistic regression was performed to identify risk factors for being a case of tuberculosis reactivation in migrants and non-UK born individuals notified between 2010-2013. After adjusting for age and sex, there was strong evidence that having a social risk factor was associated with a lower odds of being notified as a case of reactivation (OR 0.7; 95%CIs: 0.6, 0.9; p-value <0.001). There was no evidence that being screened pre-entry was associated with lower odds of being a case of reactivation compared to non-UK born individuals not screened pre-entry (OR 1.1; 95%CIs: 0.9, 1.3; p-value 0.39). Compared to individuals from countries with a WHO prevalence of tuberculosis greater than 350 per 100,000 population, there was weak evidence that those from countries with a prevalence between 40 and 149 were at increased odds of reactivation (OR 1.0; 95%CIs: 1.0, 1.3; p-value 0.08) and that those from a country with a prevalence of less than 39 were at reduced risk (OR 0.8; 95% CIs: 0.7, 0.9; p-value 0.01) after adjusting for age and sex. Migrants between the aged between 45 and 64 had an increased risk of reactivation (OR 1.1; 95%CIs: 1.0, 1.3; p-value 0.01), as did those over the age of 65 (OR 1.4; 95%CIs: 1.2, 1.6; p-value <0.001).

Table 25. Number of tuberculosis cases and proportion of clustering stratified by place of birth and whether pre-entry screened, 2010-2013

	Notified cases	Culture confirmed cases		Culture confirmed cases with a strain type*		Number of cases with a unique strain		Number of cases clustered		Number of clusters
	N	N	%	N	%	N	%	N	%	N
All UK notified cases	33,942	20,560	60.6%	16,602	80.7%	7,712	46.5%	8,890	53.5%	1,854
Non-UK born not screened	22,321	13,820	61.9%	11,090	80.2%	5,575	50.3%	5,515	49.7%	1,605
Non-UK born screened*	1,590	1,051	66.1%	836	79.5%	450	53.8%	386	46.2%	247

Table 26. Baseline characteristics, univariable and multivariable logistic regression to examine risk factors for first in cluster cases of tuberculosis in non-UK born individuals notified between 2010-2013.

Risk Factor	Migrants contributing (%)	First in cluster (%row)	Univariable OR (95% CI)	Multivariable OR (95%CI)	p-value
All	11926 (100%)	1316 (11.0%)			
Age					
0-15	125 (1.1%)	24 (19.2%)	1.9 (1.2, 2.9)	1.8 (1.2, 2.9)	0.01
16-44	8547 (71.7%)	959 (11.2%)	1.0	1.0	
45-64	2101 (17.6%)	221 (10.5%)	0.9 (0.8, 1.1)	0.9 (0.8, 1.1)	0.40
65+	1152 (9.7%)	112 (9.7%)	0.9 (0.7, 1.0)	0.9 (0.7, 1.1)	0.22
Missing	1 (0.0%)	0 (0.0%)			
Sex					
Female	4786 (40.1%)	520 (10.9%)	1.0		
Male	7130 (59.8%)	795 (11.2%)	1.0 (0.9, 1.2)	1 (0.9, 1.2)	0.51
Missing	10 (0.1%)	1 (10.0%)			
WHO prevalence					
0-39	614 (5.2%)	62 (10.1%)	0.9 (0.7, 1.2)	0.8 (0.6, 1.1)	0.17
40-149	771 (6.5%)	87 (11.3%)	1.0 (0.8, 1.3)	0.9 (0.7, 1.2)	0.46
150-349	3713 (31.1%)	402 (10.8%)	1.0 (0.9, 1.1)	0.8 (0.7, 1.0)	0.03
350+	5346 (44.8%)	590 (11.0%)	1.0	1.0	
Not known	1482 (12.4%)	175 (11.8%)	1.1 (0.9, 1.3)	1.0 (0.7, 1.5)	0.93
BCG vaccinated					
No	2273 (19.1%)	235 (10.3%)	1.0	1.0	
Yes	6156 (51.6%)	696 (11.3%)	1.1 (0.9, 1.3)	1.1 (0.9, 1.3)	0.30
Missing	3497 (29.3%)	385 (11.0%)	1.1 (0.9, 1.3)	1.0 (0.9, 1.2)	0.59
Social risk factor					
No	9189 (77.1%)	997 (10.8%)	1.0	1.0	
Yes	794 (6.7%)	87 (11.0%)	1.0 (0.8, 1.3)	1.0 (0.8, 1.3)	0.98
Missing	1943 (16.3%)	232 (11.9%)	1.1 (1.0, 1.3)	1.1 (0.9, 1.3)	0.27
Time since entry to UK					
0-2	3403 (28.5%)	385 (11.3%)	1.0	1.0	
3-5	2076 (17.4%)	238 (11.5%)	1.0 (0.9, 1.2)	1.0 (0.8, 1.2)	0.78
6-10	2148 (18.0%)	235 (10.9%)	1.0 (0.8, 1.1)	0.9 (0.7, 1)	0.13
11+	2937 (24.6%)	301 (10.2%)	0.9 (0.8, 1.1)	0.8 (0.7, 1)	0.07
Not known	1362 (11.4%)	157 (11.5%)	1.0 (0.8, 1.2)	0.9 (0.6, 1.3)	0.57
Screened pre-entry					
No	11090 (9.03%)	1249 (11.3%)	1.0	1.0	
Yes	836 (7.0%)	67 (8.0%)	0.7 (0.5, 0.9)	0.6 (0.5, 0.8)	<0.001

Table 27. Baseline characteristics, univariable and multivariable logistic regression to examine risk factors for being a case of tuberculosis reactivation in reactivation in non-UK born individuals notified between 2010-2013.

Risk Factor	Migrants contributing (%)	Reactivation tuberculosis cases (%row)	Univariable OR (95% CI)	Multivariable OR (95%CI)	p-value
All	11926 (100%)	6025 (50.5%)			
Age					
0-15	125 (1.1%)	53 (42.4%)	0.7 (0.5, 1.0)	0.7 (0.5, 1.0)	0.08
16-44	8547 (71.7%)	4286 (50.1%)	1.0	1.0	
45-64	2101 (17.6%)	1057 (50.3%)	1.0 (0.9, 1.1)	1.1 (1.0, 1.3)	0.01
65+	1152 (9.7%)	628 (54.5%)	1.2 (1.1, 1.3)	1.4 (1.2, 1.6)	<0.001
Missing	1 (0.0%)	1 (100.0%)			
Sex					
Female	4786 (40.1%)	2468 (51.6%)	1.0		
Male	7130 (59.8%)	3550 (49.8%)	0.9 (0.9, 1.0)	0.9 (0.9, 1.0)	0.12
Missing	10 (0.1%)	7 (70.0%)			
WHO prevalence					
0-39	614 (5.2%)	267 (43.5%)	0.8 (0.6, 0.9)	0.8 (0.7, 0.9)	0.01
40-149	771 (6.5%)	411 (53.3%)	1.1 (1.0, 1.3)	1.1 (1.0, 1.3)	0.08
150-349	3713 (31.1%)	1915 (51.6%)	1.1 (1.0, 1.1)	1.0 (0.9, 1.1)	0.84
350+	5346 (44.8%)	2685 (50.2%)	1.0	1.0	
Not known	1482 (12.4%)	747 (50.4%)	1.0 (0.9, 1.1)	0.9 (0.7, 1.2)	0.51
BCG vaccinated					
No	2273 (19.1%)	1140 (50.2%)	1.0	1.0	
Yes	6156 (51.6%)	3055 (49.6%)	1.0 (0.9, 1.1)	1.0 (0.9, 1.1)	1.0
Not known	3497 (29.3%)	1830 (52.3%)	1.1 (1.0, 1.2)	1.1 (1.0, 1.2)	0.14
Social risk factor					
No	9189 (77.1%)	4670 (50.8%)	1.0	1.0	
Yes	794 (6.7%)	334 (42.1%)	0.7 (0.6, 0.8)	0.7 (0.6, 0.9)	<0.001
Not known	1943 (16.3%)	1021 (52.5%)	1.1 (1.0, 1.2)	1.1 (1, 1.2)	0.28
Time since entry to UK					
0-2	3403 (28.5%)	1803 (53.0%)	1.0	1.0	
3-5	2076 (17.4%)	1075 (51.8%)	1.0 (0.9, 1.1)	1 (0.9, 1.1)	0.47
6-10	2148 (18%)	1060 (49.3%)	0.9 (0.8, 1.0)	0.9 (0.8, 1.0)	0.02
11+	2937 (24.6%)	1391 (47.4%)	0.8 (0.7, 0.9)	0.7 (0.7, 0.8)	<0.001
Not known	1362 (11.4%)	696 (51.1%)	0.9 (0.8, 1.1)	0.9 (0.7, 1.2)	0.62
Screened pre-entry					
No	11090 (93.0%)	5575 (50.3%)	1.0	1.0	
Yes	836 (7.0%)	450 (53.8%)	1.2 (1.0, 1.3)	1.1 (0.9, 1.3)	0.39

6.4.2 Cohort analysis

A cohort analysis was undertaken to examine the incidence of being the first case of tuberculosis in a cluster and reactivation of tuberculosis. The majority of migrants in this cohort analysis were aged between 16 and 44 (301,358; 94.5%) and male (217,268; 68.1%; Table 28). Self-reported contact with a case of tuberculosis at the time of pre-entry screening was uncommon (857; 0.3%). Most applicants had no abnormality on their chest radiograph at the time of pre-entry screening (302,364; 94.8%), with 12,304 (3.9%) classified as tuberculosis suspected. The majority of migrants were screened in countries with a WHO prevalence of greater than 350 per 100,000 population (267,294; 83.8%).

There was a total of 598,000 person years at risk within the cohort and a mean follow time of 1.87 years per migrant. There were 35 migrants who were the first case of tuberculosis in a cluster, providing an estimated crude rate of 6 per 100,000 person years at risk (95%CIs 4, 8). The crude incidence rate for the first case of tuberculosis in a cluster was highest in those with chest radiographs suggestive of active tuberculosis (39; 95%CIs: 20, 74). A multivariable Poisson regression analysis was performed to identify risk factors associated with being the first case of tuberculosis in a cluster, with results presented as incidence rate ratios (IRR). After adjusting for age and sex, there was strong evidence that a chest radiograph classified as consistent with tuberculosis disease (IRR 9.6; 95%CI 4.5, 20.8; <0.001) was associated with being the first case of tuberculosis in a cluster. No other risk factors were identified.

There were 301 cases of reactivation with a crude incidence rate estimated at 38 per 100,000 person years at risk (95%CIs: 33, 43; Table 29). The highest rate of reactivation was found in those with a chest radiograph classified as suspected tuberculosis at pre-entry screening at 168 per 100,000 person years at risk (95%CIs: 123, 230). Rates of reactivation increased with increasing prevalence in the country in which screening was conducted from 8 per 100,000 person years at risk (95%CIs: 2, 34) in countries with a prevalence of 40-149 per 100,000 population, to 57 per 100,000 person years at risk (95%CIs: 50, 64) in countries with a prevalence greater than 350 per 100,000 population.

After adjusting for age and sex, there was evidence that compared with migrants from countries with prevalence greater than 350 per 100,000 population, those from countries with a prevalence of 40-149 were at lower risk of reactivation (IRR 0.2; 95%CI 0.1, 0.9; p-value 0.03) as were those from countries with a prevalence of 150-349 (IRR 0.3; 95%CI 0.2, 0.6; p-value <0.001). There was also strong evidence that migrants with a chest radiograph classified as suspected tuberculosis at pre-entry screening were at increased of reactivation (3.9; 95%CI 2.8, 5.5; p-value <0.001) after adjusting for age and sex.

Table 28. Baseline characteristics, univariable and multivariate analysis of incidence rates for first in cluster cases of tuberculosis in migrants screened pre-entry (2009-2012) and notified in ETS (2010-2013)

Risk Factor	Migrants contributing (%)	Episodes	Person years at risk (1000)	Rate per 100,000 person years (95% CI)	Univariable IRR (95% CI)	Multivariable IRR (95%CI)	p-value
All	318983 (100%)	35	598	6 (4, 8)			
Age							
0-15	9542 (3.0%)	1	18	6 (1, 39)	0.9 (0.1, 6.7)	1.2 (0.2, 8.3)	0.86
16-44	301358 (94.5%)	34	565	6 (4, 8)	1.0	1.0	
45-64	6466 (2.0%)	0	12	-			
65+	1617 (0.5%)	0	3	-			
Sex							
Female	101715 (31.9%)	6	192	3 (1, 7)	1.0		
Male	217268 (68.1%)	29	406	7 (5, 10)	2.3 (0.9, 5.5)	2 (0.8, 4.7)	0.12
Contact with case TB							
No	318126 (99.7%)	35	596	6 (4, 8)			
Yes	857 (0.3%)	0	2	-			
Visa							
Students	206142 (64.6%)	25	385	6 (4, 10)	1.0		
Settlement and Dependents	94118 (29.5%)	9	177	5 (3, 10)	0.8 (0.4, 1.7)	1.1 (0.5, 2.6)	0.74
Work	10578 (3.3%)	1	21	5 (1, 35)	0.7 (0.1, 5.5)	0.8 (0.1, 5.3)	0.83
Working Holiday Maker	861 (0.3%)	0	2	-			
Family Reunion	2335 (0.7%)	0	4	-			
Other	4949 (1.6%)	0	10	-			

CXR							
No abnormality	302364 (94.8%)	26	566	5 (3, 7)	1.0		
TB suspected	12304 (3.9%)	9	23	39 (20, 74)	8.4 (4.0, 18)	9.6 (4.5, 20.8)	<0.001
Abnormality not TB	4315 (1.4%)	0	8	-			
WHO category							
40-149	12402 (3.9%)	1	24	-	0.6 (0.1, 4.7)	1.2 (0.2, 7.7)	0.88
150-349	39287 (12.3%)	1	75	1 (0, 10)	0.2 (0.0, 1.5)	0.2 (0.0, 1.7)	0.15
350+	267294 (83.8%)	33	500	7 (5, 9)	1.0	1.0	
Sputum culture testing							
No	11570 (3.6%)	1	22	5 (1, 33)	1.0	1.0	
Yes	307413 (96.4%)	34	576	6 (4, 8)	1.3 (0.2, 9.2)	1.1 (0.2, 7.2)	0.90

Table 29. Baseline characteristics, univariable and multivariate analysis of incidence rates for reactivation cases of tuberculosis in migrants screened pre-entry (2009-2012) and notified in ETS (2010-2013)

Risk Factor	Migrants contributing (%)	Episodes	Person years at risk (1000)	Rate per 100,000 person years (95% CI)	Univariable IRR (95% CI)	Multivariable IRR (95%CI)	p-value
All	318983 (100%)	301	598	38 (33, 43)			
Age							
0-15	9542 (3.0%)	3	18	17 (5, 52)	0.3 (0.1, 1.0)	0.4 (0.1, 1.1)	0.07
16-44	301358 (94.5%)	292	565	52 (46, 58)	1.0	1.0	
45-64	6466 (2.0%)	5	12	41 (17, 98)	0.8 (0.3, 1.9)	0.9 (0.4, 2.3)	0.89
65+	1617 (0.5%)	1	3	33 (5, 232)	0.6 (0.1, 4.5)	0.4 (0.1, 3.3)	0.44
Sex							
Female	101715 (31.9%)	78	192	41 (33, 51)	1.0	1.0	
Male	217268 (68.1%)	223	406	55 (48, 63)	1.3 (1.0, 1.7)	1.2 (0.9, 1.6)	0.35
Contact with case TB							
No	318126 (99.7%)	300	596	50 (45, 56)	1.0	1.0	
Yes	857 (0.3%)	1	2	63 (9, 446)	1.2 (0.2, 8.8)	1.2 (0.2, 8.9)	0.83
Visa							
Students	206142 (64.6%)	209	385	54 (47, 62)	1.0	1.0	
Settlement and Dependents	94118 (29.5%)	81	177	46 (37, 57)	0.8 (0.6, 1.1)	1 (0.7, 1.3)	1.0
Work	10578 (3.3%)	6	21	29 (13, 65)	0.5 (0.2, 1.2)	0.5 (0.2, 1.2)	0.12
Working Holiday Maker	861 (0.3%)	0	2	-	-	-	-
Family Reunion	2335 (0.7%)	5	4	114 (48, 275)	2.1 (0.9, 5.1)	4.9 (2.0, 11.9)	<0.001
Other	4949 (1.6%)	0	10	-	-	-	-

CXR							
No abnormality	302364 (94.8%)	261	566	46 (41, 52)	1.0	1.0	
TB suspected	12304 (3.9%)	39	23	168 (123, 230)	3.7 (2.6, 5.2)	3.9 (2.8, 5.5)	<0.001
Abnormality not TB	4315 (1.4%)	1	8	12 (2, 87)	0.3 (0.0, 1.9)	0.5 (0.1, 3.7)	0.52
WHO category							
40-149	12402 (3.9%)	2	24	8 (2, 34)	0.2 (0.0, 0.6)	0.2 (0.1, 0.9)	0.03
150-349	39287 (12.3%)	16	75	21 (13, 35)	0.4 (0.2, 0.6)	0.3 (0.2, 0.6)	<0.001
350+	267294 (83.8%)	283	500	57 (50, 64)	1.0	1.0	
Sputum culture testing							
No	11570 (3.6%)	14	22	65 (38, 109)	1.0	1.0	
Yes	307413 (96.4%)	287	576	50 (44, 56)	0.8 (0.4, 1.3)	0.7 (0.4, 1.2)	0.23

The incidence rates for being the first in a cluster and reactivation were compared between migrants screened pre-entry during the period 1st January 2009 and 31st December 2012 and non-UK born individuals not screened, but entering the UK between 1st January 2009 and 31st December 2013 (Figure 33). The incidence of being the first case of tuberculosis in a cluster in migrants screened pre-entry was 6 per 100,000 person years at risk (95%CI 4, 8) compared to 6 per 100,000 person years at risk (95%CI 5, 7) in non-UK born individuals not screened entering the UK during the same period. The incidence of tuberculosis reactivation in migrants screened pre-entry was 38 per 100,000 person years at risk (95%CI 33, 43) compared to 33 per 100,000 person years at risk (95%CI 32, 35) in non-UK born individuals not screened. As the 95% confidence intervals overlapped for both of these comparisons there was no statistical evidence that the incidence rates were different between the two groups.

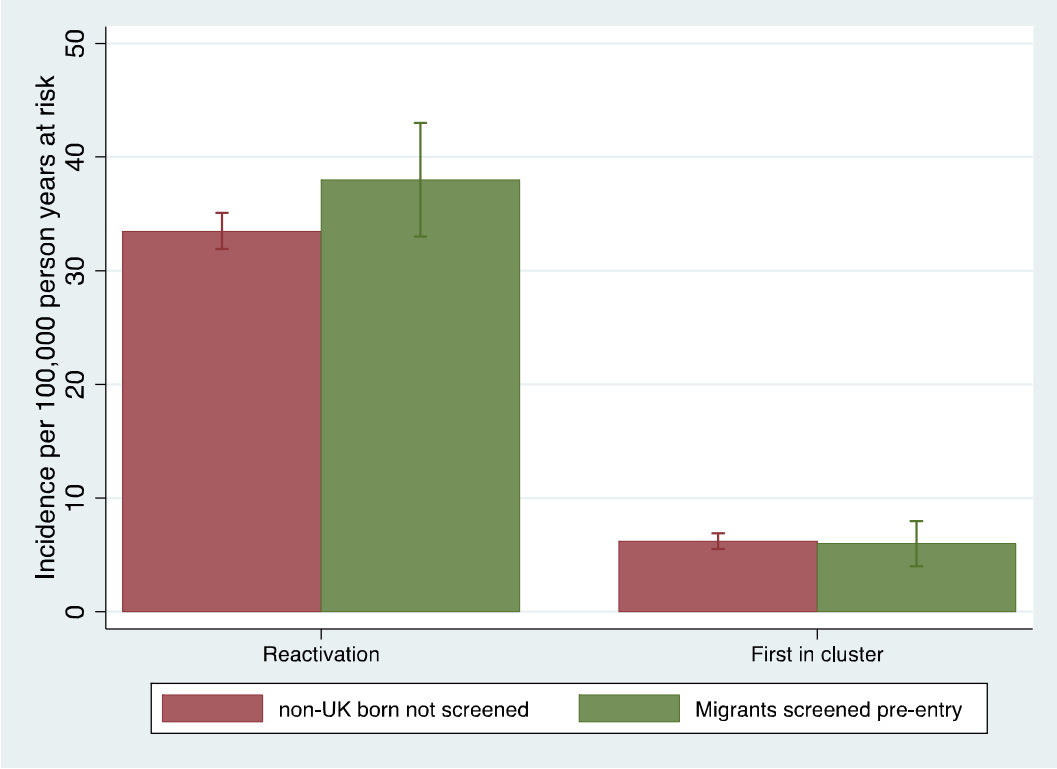

Figure 33. Rates of reactivation and first in cluster cases of tuberculosis in all notified cases of tuberculosis in non-UK born individuals not screened and migrants screened pre-entry, 2010-2013.

A series of sensitivity analyses were carried out to examine the effect varying the assumptions made in the cohort analyses. Varying the definition of a prevalent case from zero to 180 days after issuing a medical certificate of clearance (from the baseline assumption of 90 days) had little impact on the estimates of the incidence of first in

cluster or reactivation (Figures 34 & 35). Incidence rates increased slightly when reducing the prevalent case definition to zero days and when including only cases screened using the culture confirmation protocol, but reduced when increasing the prevalent case definition to 180 days. The 95% confidence intervals for the incidence rates overlapped for each sensitivity analysis, providing no evidence that any of these estimates were statistically different to the baseline scenario.

Varying these same assumptions in multivariable analyses to look at risk factors for the primary outcomes also had very little effect on the results. There was strong evidence in each sensitivity analysis that having a chest radiograph classified as suspected tuberculosis was associated with an increased risk of being the first case in a cluster (Figure 36). In the baseline analysis this was the only group associated with an increased risk, and this remained the case in all scenarios in the sensitivity analysis, but due to the fact that there were no first cases in a cluster for some subcategories, several variables had null results in sensitivity analysis (e.g. work and other visa categories, and WHO prevalence countries between 40 and 149). In the sensitivity analysis for reactivation as the primary outcome, there remained strong evidence that coming from a country with WHO prevalence of less than 350 per 100,000 population resulted in a reduced risk of reactivation (Figure 37). All other results in the sensitivity analysis for this outcome remained consistent with the baseline assumptions.

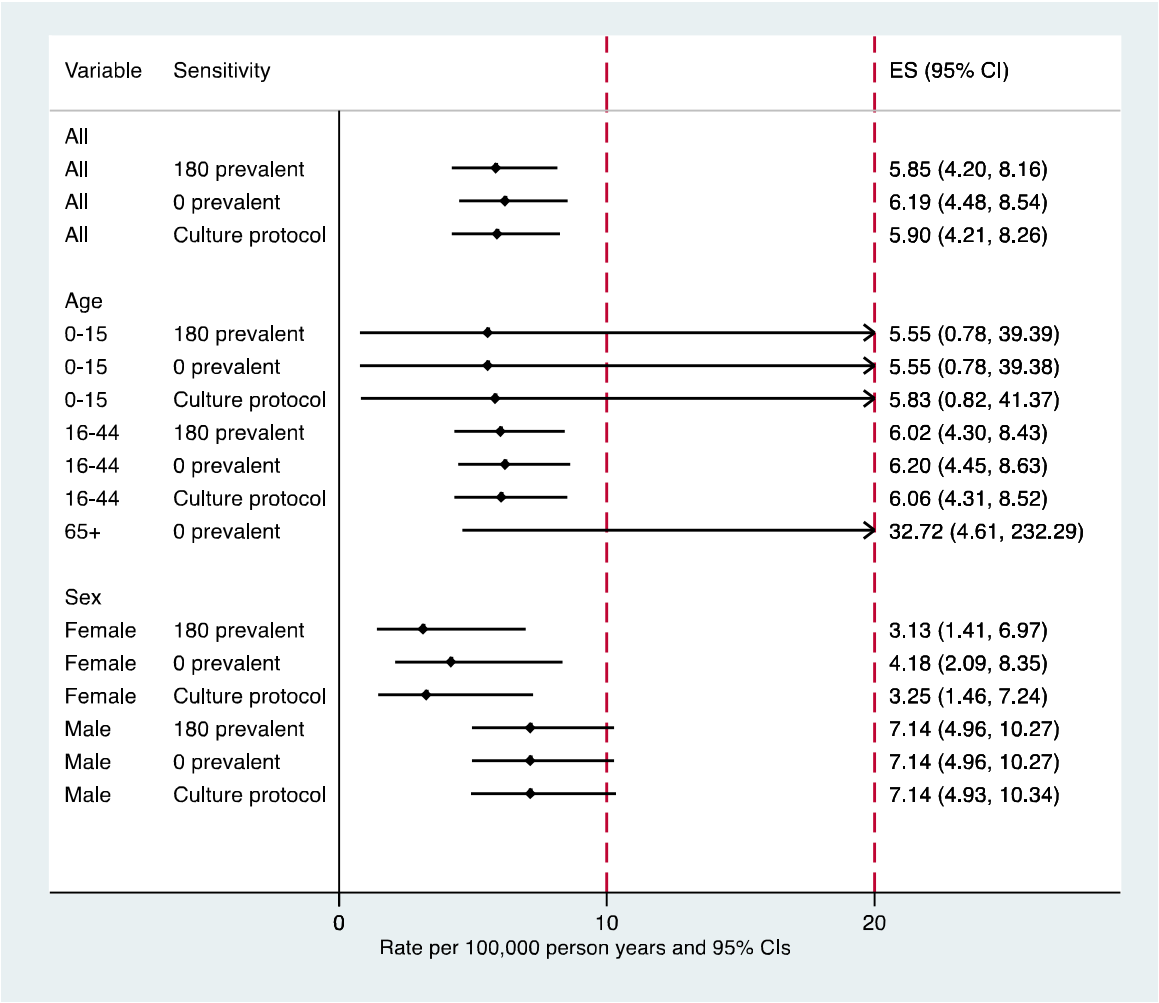

Figure 34. Sensitivity analysis of incidence rates for first in cluster cases by age and sex under different model assumptions.

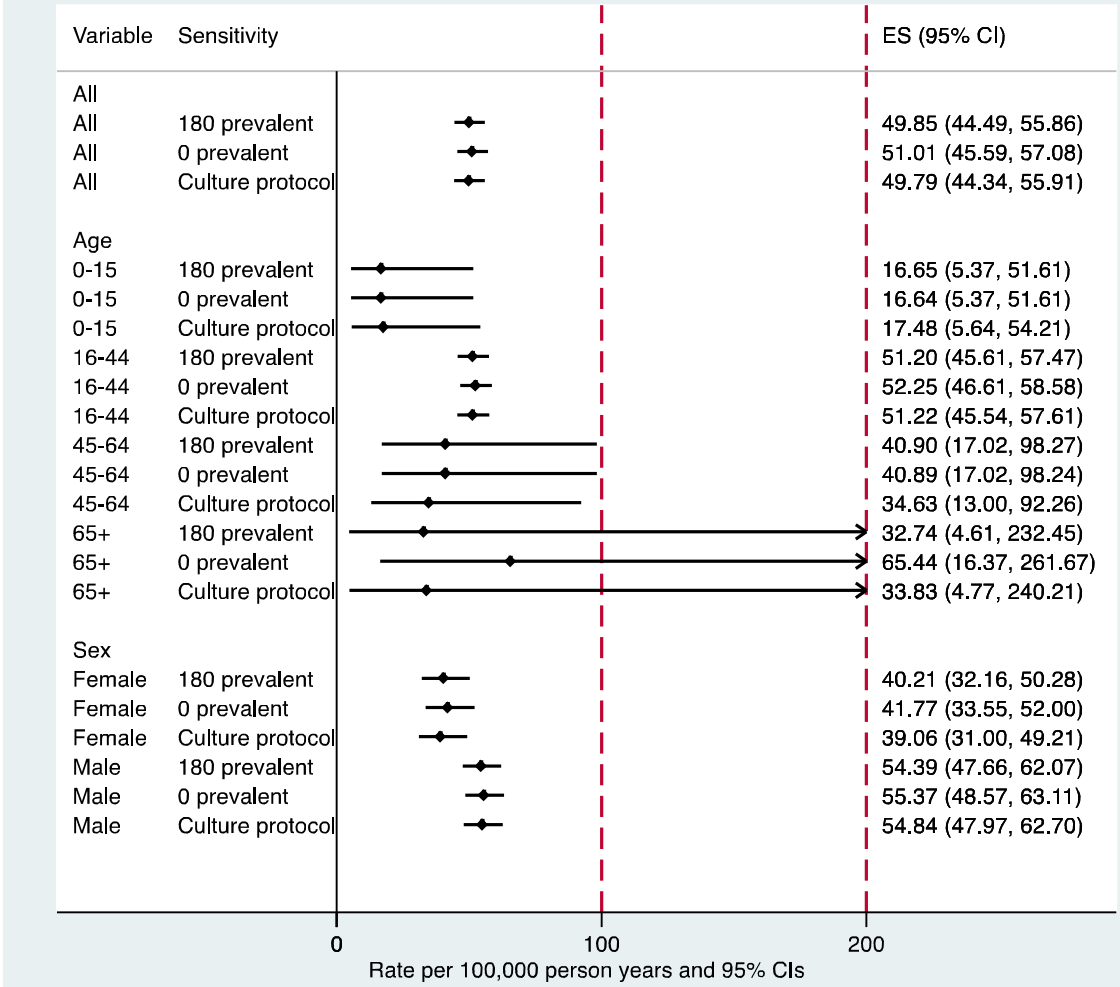

Figure 35. Sensitivity analysis of incidence rates for reactivation by age and sex under different model assumptions.

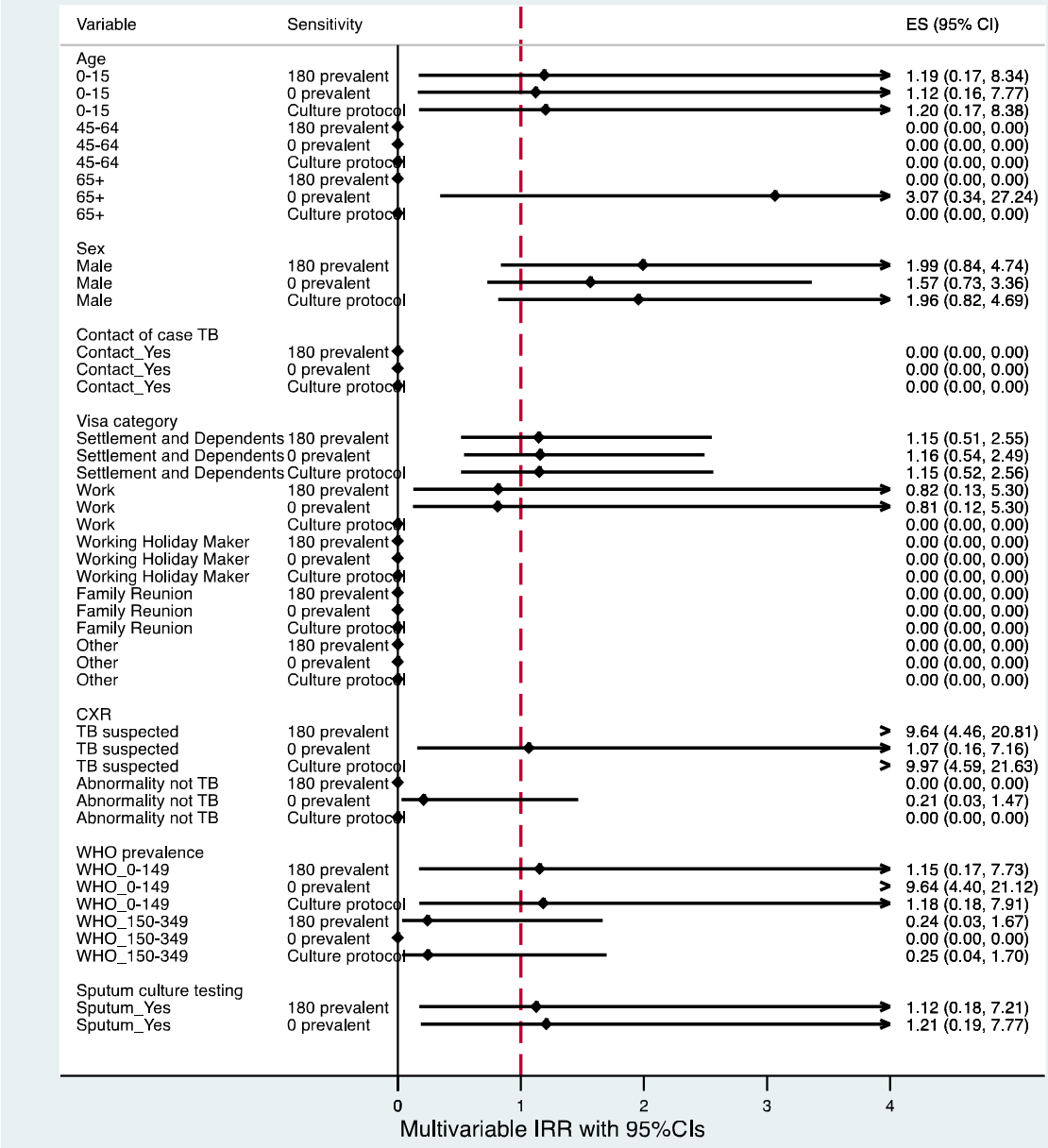

Figure 36. Sensitivity analysis of multivariable risk factor analysis of incidence rates for first in cluster cases under different model assumptions.

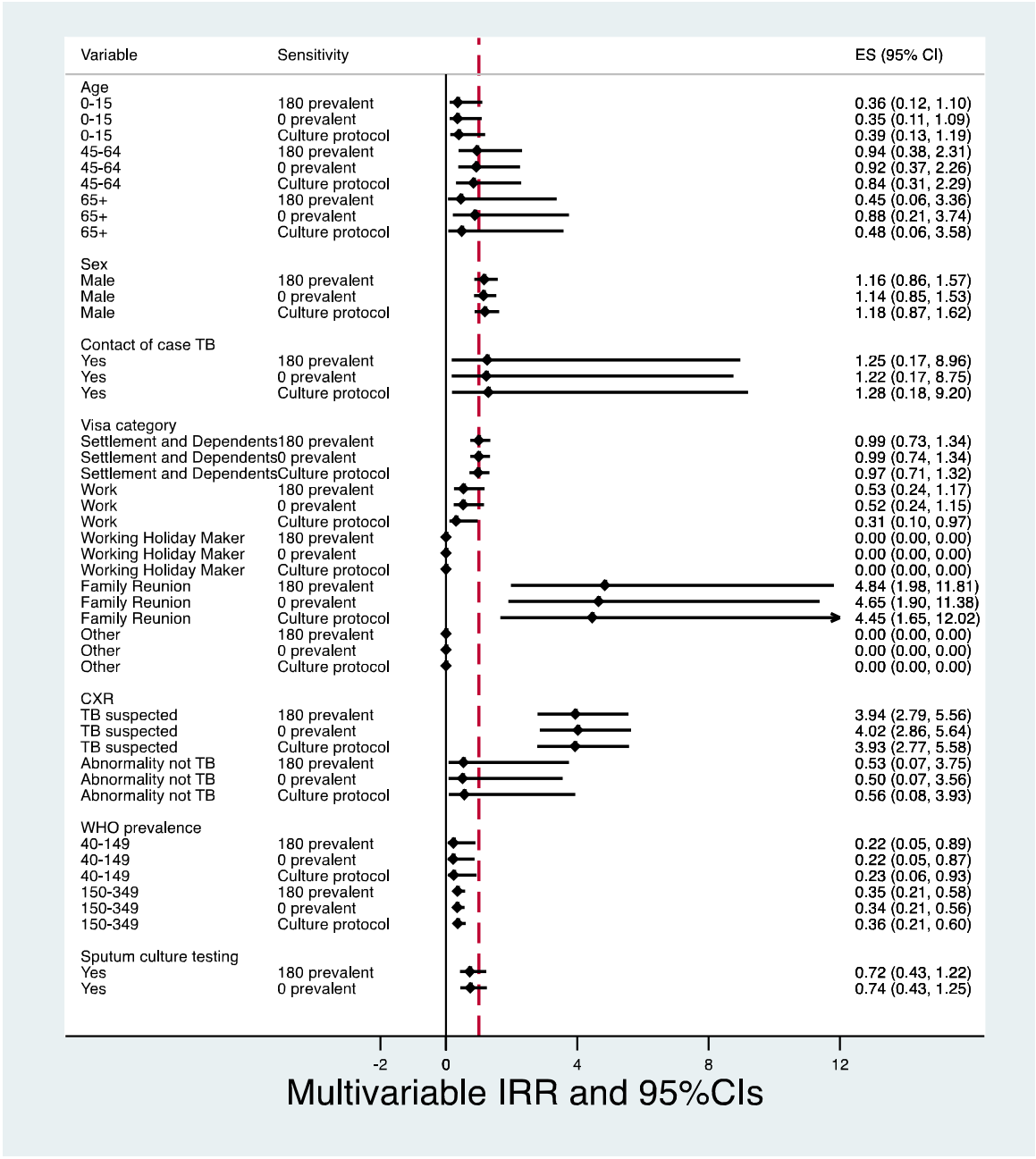

Figure 37. Sensitivity analysis of multivariable risk factor analysis of incidence rates for reactivation under different model assumptions.

In the primary analysis all cases in non-UK born migrants notified during 1st January 2010 and 31st December 2013 were included regardless of when they arrived in the UK. A separate sensitivity analysis of the cross-sectional study was conducted which included all pre-entry screened cases (from 2005-2012) and non-UK born individuals for the same time period. The full results are presented in Appendix 7, Tables 34 and 35. The number of pre-entry screened migrants in the study is reduced to 5,998, but

there remains strong evidence that pre-entry screening is associated with a reduced odds (0.6; 95%CIs: 0.4, 0.8; p-value <0.001) of being the first case in a cluster. There was no longer strong evidence that coming from a country with WHO prevalence less than 350 per 100,000 was associated with a lower risk of reactivation, but the direction of effect remained constant (Table 35).

A final sensitivity analysis was performed to examine the effect of the time window used to determine whether cases were clustered or not (Figure 38). As the time window increased from one year of strain typing data (2010) up to four years of strain typing data (2010-2013) the proportion of cases clustered increased from 48.3% (95%CIs: 46.6, 50.0; 3,494 cases included) to 53.5% (95%CIs: 52.8, 54.3; 16,602 cases included). There was no evidence that clustering increased when including four years of data (2010-2013) compared to three years of data (2010-2012) when it was estimated that 53.1% of cases were clustered (95%CIs: 52.3, 54.0; 12,764 cases included).

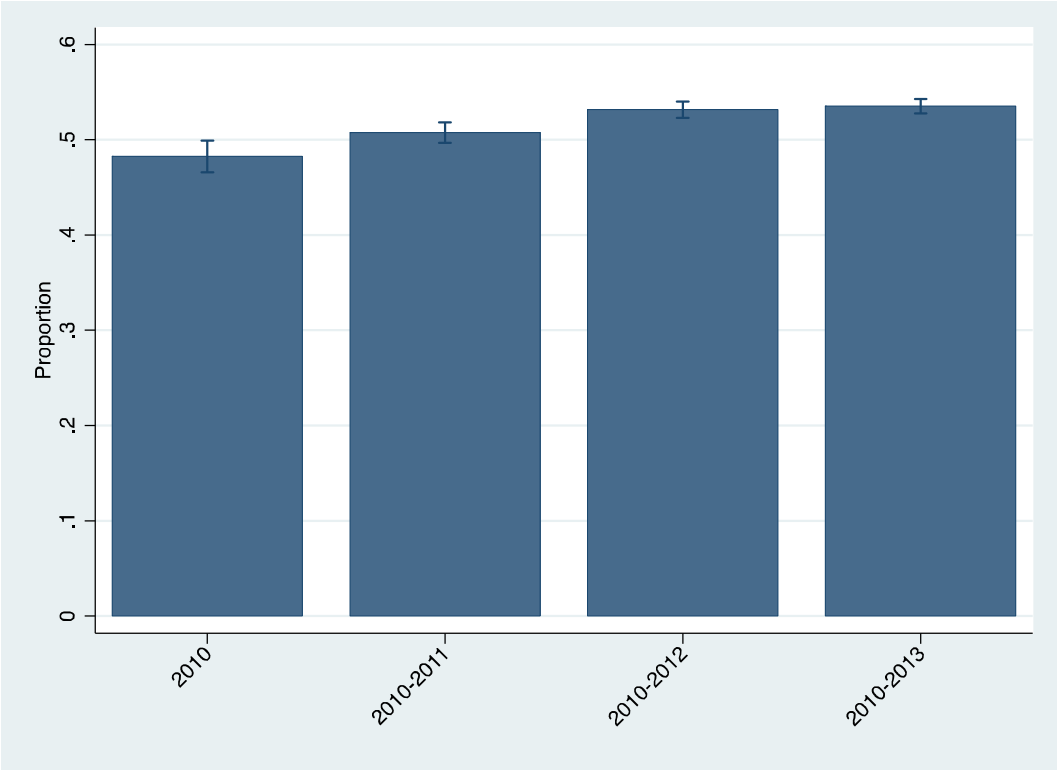

Figure 38. Proportion of all cases clustered by time window.

6.5 Discussion

This analysis examines molecular epidemiology data for ETS notified cases of tuberculosis and found strong evidence that migrants screened pre-entry were less likely than non-UK born individuals not screened to be the first case in a cluster of tuberculosis. There was no evidence that migrants screened pre-entry were less likely to be a reactivation case of tuberculosis. In a multivariable analysis conducted in a cohort of migrants screened pre-entry, there was evidence that having a chest radiograph classified as active tuberculosis was associated with increased risk of being both the first case in a cluster, and for tuberculosis reactivation. Coming from a lower prevalence country was associated with a lower risk of reactivation, but sputum culture testing was not associated with either an increased or decreased odds of reactivation.

6.5.1 Strengths and weaknesses of the study

This study provides the first description of strain typing data in migrants that were screened for active tuberculosis by the UK pre-entry programme. The study includes all migrants from 15 high incidence countries intending to stay for more than six months, and is the first time strain typing data has been analysed in migrants whilst accounting for person time at risk. Unlike existing data on tuberculosis notifications in the UK, this study was able to identify migrants screened pre-entry, and not just those self-reporting that they were born outside the UK. The work builds on the previous chapter examining the incidence of tuberculosis in pre-entry screened migrants by including strain typing data to investigate how much of the tuberculosis in this population is due to reactivation of disease acquired in the country of origin and how many cases may lead to chains of local transmission.

As described in Chapter 4, the probabilistic matching used to identify primary outcomes has a high level of accuracy, and the outcomes used in the study were clearly defined. Whilst there is relatively high certainty that those migrants who were issued with a medical certificate travel to the UK, there is less certainty about when they might emigrate or die. The analysis presented in this chapter therefore builds upon the methodology described in chapter 5 to censor migrants within the cohort to account for duration of stay and death. The assumptions made in estimating person time at risk were consistent under several sensitivity analyses.

Limitations with the primary outcomes

The definition of reactivation depends upon several assumptions. Reactivation cases were identified as those with a unique strain type and the first case in a cluster. There are several alternative explanations for an individual having a unique strain, other than this being due to reactivation. Firstly, the unique strain may be involved in a cluster with another case that has not yet been reported. For example, a case currently classified as reactivation in the dataset in ETS 2013, may in fact be re-classified as transmission in ETS 2014 if an associated case were notified. However, the sensitivity analysis examining the proportion of cases clustered over time in this chapter (Figure 38) suggests that after four years of data there is a plateau in the number of cases clustered. Therefore whilst further years of follow up will provide more data for cases clustered (reducing the number of reactivation cases) it is likely to only have minimal effect. Secondly, the rate of mutation of the strain types may lead to cases being classified as reactivation, when transmission had occurred, along with a strain type mutation. Both of these misclassification biases would therefore lead to an over-estimation in the number of cases of reactivation.

Cross-sectional study

The study is limited by the fact that strain typing was only performed on UK notified cases since 2010. These results therefore provide helpful estimates of what happens in the first four years after migration, but may not be representative of time periods before or after this. Undocumented migrants, and asylum seekers were not included in the pre-entry screened dataset and these individuals will be at higher risk of tuberculosis reactivation compared to those migrating on visas for longer than 6 months who are included.(166) This selection bias is a potential alternative explanation for the finding that pre-entry screening reduces the risk of being the first case of transmission, as the two groups being compared may not have comparable risks.

Cohort analysis

There are several reasons why the incidence rates estimated in the cohort analysis will differ from the true values. Firstly, the time period for identifying cases of reactivation was short, and whilst time at risk was included in the denominator, estimates over a longer time period will change. Secondly, as approximately 60% of cases are culture confirmed, and only 80% of these are strain typed, numerator data for the in the

incidence rates will be missing for approximately 50% of all cases. Assuming that cases without missing data were representative of those not strain typed then the true estimates would be twice those presented here, i.e. 12 per 100,000 person years at risk for first in cluster cases of tuberculosis, and 76 per 100,000 person years at risk for reactivation. Thirdly, due to the limitations of strain typing data (which is based on MIRU), it is possible that cases will be assigned to a cluster when they should not be. This could result in cases that truly are reactivation being misclassified in a cluster and if they are not deemed "first in cluster" then they are assumed not to be reactivation cases. Finally, the person years at risk used to calculate both incidence rates is an overestimate of the true value as it includes migrants from 2009, when no outcome data were available. The decision to include these cases was taken, because as Figure 26 in chapter 5 shows, the greatest number of cases occurs one year after migration. This means that someone migrating on 1st January 2009 will contribute 12 months of person time at risk when it would not be possible to find any associated primary outcomes for them. However, it does mean that a migrant entering the UK on 1st June 2009 with a case notification between 1st June 2010 and 31st December 2010 would be found – six months of which is the time period when the analysis in Chapter 5 shows a large number of cases do occur. Excluding all cases from 2009 would therefore miss such cases.

Several important risk factors for first in cluster and reactivation were not included in the cohort analysis as they are not recorded by IOM, including socio-economic status, clinical conditions associated with an increased risk of tuberculosis reactivation (such as HIV and immunosuppression with biological drugs) and lifestyle risk factors such as smoking. These factors have been shown to be associated with an increased risk of reactivation of tuberculosis both in the UK and internationally and these unmeasured risk factors may be stronger risk factors for reactivation than those included in this analysis.(74,162,183,184)

Comparing incidence estimates for first in cluster and reactivation between those not born in the UK with those pre-entry screened (both groups migrating between 2009 and 2012), found no evidence that rates of in pre-entry screened migrants are lower than the rest of the non-UK born cases (Figure 33). These data should be interpreted in the context of the limitations discussed above for the primary outcomes, but also considering several limitations for the population denominator estimates used for non-

UK born individuals. For the national estimates of non-UK born populations it was assumed that non-UK born individuals stayed in the UK for the same length of as those screened pre-entry. This may be an over-simplification given the greater range of countries that non-UK born individuals in the comparison group, and could bias the estimates of incidence in either direction. The comparison of rates have an additional limitation as the denominator for the non-UK born population had to include short term migrants planning to stay less than 6 months as it was not possible to identify and exclude these in ETS, but the pre-entry screened population did not contain a comparable group of short term migrants. Given the heterogeneity of short-term visa applicants it is very difficult to state whether this is under or overestimate the person time at risk, but it is likely to be somewhat different to that used in this analysis. The cross-sectional analysis provides some support for the assumptions made, as consistent with the cohort analysis, it also found no evidence for pre-entry screening on the risk of reactivation, but nonetheless it seems appropriate to interpret these results with some caution. For these reasons, it is felt that results from the cross-sectional analysis provide a stronger basis on which to understand whether migrants screened pre-entry were at a lower risk of tuberculosis after arrival in the UK.

6.5.2 Strengths and weaknesses in relation to other studies

A key strength of this study, in comparison to the existing analyses of strain typing data in the UK, is that it attempts to estimate the incidence of the first case of tuberculosis in a cluster and reactivation, in addition to the proportion of cases clusters in non-UK born individuals. The proportion of cases clustered does not account for person time at risk for each population, which may be very different for the UK and non-UK born groups. Incidence rates also provide measure of disease frequency rather than just a measure of first case in a cluster and reactivation among all strain typed cases.(136)

Studies conducted in the US have estimated incidence rates of reactivation in migrants screened pre-entry in 2008 to be 98 per 100,000 person years at risk (95%CIs: 96, 100), which is higher than the estimated incidence in this study.(169) Direct comparisons between these studies are difficult for several reasons. Overall in the US, 80.1% of cases were due to reactivation (including US and non-US born), compared to 46.5% in the UK, representing a significantly different epidemiological situation in the whole population. In 2008 just under 80% of US notified cases were culture confirmed, and

80% of these were strain typed, therefore the coverage of strain typing for all notified cases was significantly higher than in the UK. The US study may over-estimate the number of cases of reactivation as it attempts to take account of epidemiological links between identical strains, and only classifying cases into a cluster if they were resident within the same local geographical area. This analytical strategy will serve to increase the number of cases classified as reactivation compared to those presented in this chapter where geographical location was not included in this analysis. Therefore, unlike this US study, and others examining reactivation and transmission in migrants, this study presents data for all cases in England, Wales and Northern Ireland over a four year period of time.(170,173,185) Conversely, compared to whole genome sequencing, 24 loci MIRU misclassifies cases as clustered due to the lack of resolution, particularly compared to whole genome sequencing which if it had been used universally, the number of cases of reactivation may have been higher, and more consistent with US estimates.(186–189)

6.6 Conclusion

This analysis examines strain typing data for ETS notified cases of tuberculosis in migrants screened pre-entry, estimating for the first time the incidence of first in cluster cases and reactivation, and the risk factors associated with these incident cases. It provides some evidence that pre-entry screening may reduce the chance of being a first in cluster case of tuberculosis, compared to non-UK born individuals that have not been screened, suggesting that pre-entry screening may have a role in reducing tuberculosis transmission in the UK. These results are crucial to be able to inform the most effective ways of improving the health of this population and can be used with health economic models to determine cost effective strategies aimed at reducing the burden of disease in this group through the use of latent tuberculosis screening. The full public health implications of this analysis, recommendations, and directions of future research are discussed in the final chapter of the thesis.

CHAPTER 7

Summary of research and main findings, recommendations for evidence-based migrant tuberculosis screening, and final conclusions

7.1 Summary of research and main findings

The global epidemiology of tuberculosis in high-income countries is changing with a higher proportion of cases occurring in individuals born outside their country of residence.(32,40) Screening new migrants for tuberculosis has therefore become a high public health priority, and there has been a renewed interest in the approaches that can be taken to tackle the burden of disease in this vulnerable group.(55,61,73,77) Historically the UK used a combination of upon- and post-entry screening of migrants, but the upon-entry system was not cost-effective.(59) A decision was therefore taken to discontinue the upon-entry programme and fully roll out a pre-entry screening system to all countries with a prevalence of tuberculosis greater than 40 per 100,000 population.(59,190) The pre-entry screening programme was operating in 101 countries as of 31st March 2014.

This PhD aimed to inform the development of evidence-based migrant screening for tuberculosis by examining the epidemiology of tuberculosis in migrants. Table 30 summarises the findings from chapters 2, 3, 5 and 6 of this thesis in relation to improving our understanding of the epidemiology of tuberculosis in migrants.

Despite the long history of migrant screening for tuberculosis, pre-entry screening is comparatively new and also had a relatively weak evidence-base compared to upon- and post-entry screening of migrants.(55,128) The first part of this thesis set out to systematically review published literature on this topic and describe the data from the 15 countries taking part in the pilot UK pre-entry screening programme. The systematic

review presented in chapter 2 identified 15 unique studies with data on nearly 4 million migrants that had been screened for tuberculosis. Restricting a meta-analysis to include only studies that used culture confirmation of cases found that the prevalence of tuberculosis detected by pre-entry screening increased with WHO prevalence of tuberculosis in the country of origin.

Table 30. Summary of the main thesis findings from Chapters 2, 3, 5 and 6

	Systematic review – prevalence on pre-entry screening (chapter 2)	**Prevalence on pre-entry screening by UK programme (chapter 3)**	**Incidence post-migration (chapter 5)**		**First in cluster (chapter 6)**		**Reactivation (chapter 6)**	
			ETS	Bacteriologically confirmed	Cross-sectional	Cohort	Cross-sectional	Cohort
Older age	-	Increased risk	Increased risk	None	None (increased risk in children)	-	Increased risk	Increased risk
Sex	-	None	None	None	None	None	None	None
TB contact	-	Increased risk	Increased risk	Increased risk	-	-	-	None
Visa category	-	Increased: Family reunion	Increased: - Family reunion and settlement dependant	Increased: - Settlement dependant and family reunion Decreased: - Working holiday -	-	-	-	Increased: - Family reunion Decreased: - Students
Positive chest radiograph	-	-	Increased risk	Increased risk	-	Increased risk	-	Increased risk
Low WHO prevalence	Reduced risk	Reduced risk	Reduced risk	Reduced risk	Reduced risk	-	Reduced risk	Reduced risk
Sputum culture	-	Increased risk	Reduced risk	Reduced risk	-	-	-	None
BCG	-	-	-		None	-	None	-
Social risk factor	-	-	-		None	-	None	-
Longer time in country	-	-	-		Reduced risk	-	Reduced risk	-
Pre-entry screening	-	-	-		Reduced risk	-	None	-

The systematic review and meta-analysis was the first study to identify and synthesis all published pre-entry screening data and established the state of the existing literature on this topic. To reduce bias in the review process, empirically based systematic review and meta-analysis guidelines were used, and double screening of identified papers was performed. There was a great deal of variation in the protocols being used to screen migrants and this was likely to be part of the explanation for the high levels of heterogeneity found in the prevalence of tuberculosis reported by each screening programme. A further limitation of the study was the fact that data were not entirely

representative of global migrant flows and therefore may not be generalizable to all pre-entry screening programmes, particularly new ones set up for countries without previously published data.

No data had been previously published using the 15 countries taking part in the pilot UK pre-entry programme and chapter 3 set out to examine these data. The crude prevalence of bacteriologically confirmed (culture or smear positive) tuberculosis was 92 per 100,000 population screened. Compared to studies identified by the systematic review in chapter 2, this prevalence was the second lowest, but this finding is likely to be explained by differences in the populations screened, the screening protocols used, and rates of sputum culture testing (rather than smear microscopy) for suspected cases of tuberculosis. A multivariable regression model was built to determine risk factors for being identified with bacteriologically confirmed tuberculosis at pre-entry screening. After adjusting for age and sex, migrants reporting a history of contact with a case of tuberculosis, and those from higher prevalence countries were at increased risk.

This study was the first comprehensive analysis of UK data on pre-entry screening of migrants. The results were highly representative of migrants from the 15 countries included in the pilot programme due to the compulsory nature of the screening process. Standardised definitions were used for the primary outcomes, and the UK technical instructions for pre-entry screening were implemented at all sites conducting the screening, both factors therefore reduce issues of measurement error and misclassification bias. Despite the study being highly representative of migrants taking part in the pilot programme, no data were included on refugees or asylum seekers, which are likely to be at higher risk tuberculosis. The findings are therefore likely to under estimate prevalence compared to what would be expected in these higher-risk populations. The data used in the analysis did not include important additional risk factors for tuberculosis such as HIV status, other medical conditions, and no information on socio-economic status. These unmeasured confounding factors may explain some of the differences found between estimates of prevalence across countries.

In 2013 there were 7,892 cases of tuberculosis in the UK and 5,529 of these cases were in individuals not born in the UK. The existing tuberculosis surveillance system does not routinely link data from pre-entry screened migrants to notified cases in the UK, partly because of the lack of a single unique identifying variable between the two datasets. Chapter 4 therefore set out to examine the accuracy of a probabilistic matching

algorithm to identify individuals in the two datasets based on demographic variables such as first name, surname, date of birth and sex, and without the use of a single unique identifying variable. Compared to the gold standard variable of NHS number, the Enhanced Matching System was able to identify individuals with a high sensitivity and specificity. Accuracy remained high in an analysis linking to non-UK born individuals without manual review and only using first name, surname, date of birth, nationality and sex as matching variables. The results of this study therefore provided reassurance that this system can correctly identify individuals across datasets using a minimal number of identifying variables, and whilst manual review had improved accuracy further, its effect was minimal. Given the potential to introduce bias through manual review it may be appropriate to not conduct this step, particularly for large datasets where those conducting the analysis have little or no knowledge of the individuals being linked.

This validation analysis of probabilistic matching using the Enhanced Matching System then formed the basis of the analysis presented in chapters 5 and 6 by linking the pre-entry screening dataset of migrants screened in 15 countries between 2005 and 2012 to the UK tuberculosis register. Using this linked dataset, it was possible to undertake a cohort analysis to examine the incidence of tuberculosis in the UK among migrants screened pre-entry and identify risk factors for these cases. The overall incidence rate for all notified cases was 194 per 100,000 person years at risk and 65 per 100,000 for bacteriologically confirmed pulmonary cases. After adjusting for age and sex, individuals reporting contact with a case of tuberculosis before migration, those with a chest radiograph classified as suspected tuberculosis, migrants on settlement and dependant or family reunion visas, and those from higher prevalence countries were all at a higher risk of tuberculosis after migration. Those migrants screened under a protocol that included culture confirmation of sputum samples were at lower risk of being a case of tuberculosis after arrival in the UK.

This cohort analysis was the first international study to comprehensively link records from migrants screened pre-entry for tuberculosis to a national tuberculosis disease register. It provided unique insights into the burden of disease in this group after pre-entry screening had removed the majority of prevalent cases, and the risk factors for these incident cases. The analysis contained many assumptions, particularly in relation to person time at risk used in the survival analysis, but varying these assumptions in a

series of sensitivity analyses had minimal effect on the estimates of incidence and risk factors identified in the multivariable analyses. Similar to the analysis of prevalent cases detected at pre-entry screening, this cohort analysis was not representative of some higher risk groups such as asylum seekers and refugees. Data were not analysed by country due to the fact that some countries had a low number of cases, and therefore whilst the findings are likely to generalizable by WHO prevalence category, there still may be some issues when comparing individual countries, particularly if the socio-economic mix of migrants is different, or if there are large difference in unmeasured variables such as HIV, compared to those included in this analysis.

Since 2010 the UK surveillance system has collected data on the strain type for culture confirmed cases of tuberculosis. In chapter 6, molecular epidemiology was used to make inferences about whether tuberculosis cases in migrants arose due to reactivation of disease acquired abroad (assumed to be likely in cases with unique molecular strain types and in those who were the first case in a cluster). First cases in a cluster were also assumed more likely than others to be the originator of a local chain of transmission and in migrants was used as a proxy for imported disease that subsequently transmitted or imported infection that reactivated and subsequently transmitted. Chapter 6 included these additional data into two separate analyses. A cross-sectional analysis was conducted to examine the evidence for whether migrants screened pre-entry were at lower risk of being the first case in a cluster of transmission or of being reactivation case, compared to non-UK born individuals. Compared to non-UK born individuals, migrants screened pre-entry were less likely to be the first case in a cluster of tuberculosis. There was also evidence that those who had been in the country for greater than five years were at lower risk of being the first in a cluster as were those from lower prevalence countries. The cross-sectional analysis also examined risk factors for cases of tuberculosis reactivation (i.e. those who might have benefitted from screening for and treatment of latent infection) and found that those over the age of 45 were at higher risk. Migrants from countries with a prevalence less than 40 per 100,000 population, those with a social risk factor (problem drug or alcohol use, homelessness or history of imprisonment), and individuals having entered the country more than 5 years previously were at lower risk.

The final analysis was conducted in a cohort of pre-entry screened migrants alone who entered the UK after 2009. This study examined the incidence and risk factors for being

first case in a cluster of tuberculosis and for being a case of reactivation. The incidence rate for first in cluster cases of tuberculosis was low at 6 per 100,000 person years at risk, and due to small number of cases the only risk factor identified was having a chest radiograph at pre-entry screening classified as suspected tuberculosis. The incidence rate for reactivation cases of tuberculosis was 38 per 100,000 person years at risk. Migrants from lower prevalence countries were at lower risk of reactivation. There was no evidence that pre-entry screening for active tuberculosis reduced the incidence of reactivation after migration to the UK.

This final chapter is the first international study that has comprehensively compared the outcomes of migrants screened pre-entry to a large population that had not been screened. The results were consistent with chapter 5 and provide unique insights into the benefits (or not) of pre-entry screening and form the basis of several policy recommendations. Similar to the analysis in chapter 5, the assumptions made in the cohort analysis were largely unaffected by the various sensitivity analyses conducted, and whilst the duration of strain typing data were limited to four years, an analysis looking at the likely importance of this suggested that this is likely to have little effect on the overall conclusions drawn. The incidence estimates provide a lower bound for the true values due to the fact that only 60% of cases are culture confirmed, and only 80% of these are strain typed. Therefore the true estimates of incidence could be up double those presented in this chapter. An important limitation with the cross-sectional analysis is that the pre-entry screening variable may contain a potentially important bias, as those not screened pre-entry will include refugees and asylum seekers, whereas those screened pre-entry will not. Finally unmeasured confounding variables such as HIV and socio-economic status could change some of the results of the risk factor analysis in both the cross-sectional and the cohort analysis if they were available.

7.2 Recommendations for evidence-based migrant tuberculosis screening

The findings across the chapters in this thesis can inform evidence-based recommendations for UK technical instructions for pre-entry screening.(7) This work has been conducted in close collaboration with the tuberculosis pre-entry screening unit at Public Health England. These results have, and will continue, to inform their work on

pre-entry screening and the technical instructions. Evidence-based recommendations from the chapters in this thesis are outlined in the following sections, along with a set of research questions that should be answered as a priority.

7.2.1 Recommendations and research questions for active tuberculosis pre-entry screening

Pre-entry screening programmes may require the migrant to bear the bulk of costs of testing and treatment, but they still may not be entirely cost-neutral for the receiving country as a result of the governance and oversight required to appropriately run these programmes.(59) The systematic review found a paucity of cost-effectiveness data on these schemes which should be addressed as a priority as there remains uncertainty as to the value of pre-entry screening compared to other tuberculosis control activities. There was a great deal of heterogeneity in the results published by different screening countries which should be investigated further to better understand the strengths and limitations of the different screening approaches.

- **Recommendation:** To continue to ensure that countries conducting pre-entry screening are able to share best practice and further understand the strengths and limitations of the different approaches to pre-entry screening, epidemiological data on these screening programmes should continue to be collected and published internationally.
- **Research question:** What is the cost-effectiveness of pre-entry screening from a receiving country, migrant and wider societal perspective?

Culture testing sputum samples in suspected cases of tuberculosis

The findings from this analysis provide strong support for the previous change to UK technical instructions for the inclusion of culture testing in the screening protocol. Those screened at sites where culture testing was performed were more likely to be found as a prevalent case of tuberculosis, but were less likely to be notified as an incidence case after arrival in the UK. The reduced risk of being an incident case is likely to be a result of appropriate treatment after detection at pre-entry screening, particularly for those cases with a low bacillary load in the sputum that would have not been detected by sputum smear.

Many of the individuals screened pre-entry are entering the UK to take up places in higher education or employment as described by the high number of migrants with these visa categories in these analyses. There is a potential concern that delays introduced by the requirement for culture testing, which can take a minimum of six weeks in liquid media and eight weeks in solid media, could mean that students miss the beginning of the academic year or that employers do not have the skilled migrants they require. New rapid tests with high sensitivity are available and could potentially reduce these delays, but these should be tested in the operational setting of pre-entry screening and compared to traditional culture methods.(65,66)

- **Recommendation:** UK technical instructions should continue to include culture testing of sputum.
- **Research question:** What is the sensitivity of rapid molecular tests compared to traditional solid culture when used in a pre-entry screening setting, and what is the cost effectiveness of these tests?

Migrants with chest radiographs classified as suspected tuberculosis

Across all the chapters in the study, those with a positive chest radiograph had the highest risk of tuberculosis. They were more likely to be an incident case of tuberculosis after arrival in the UK, as well being at increased risk of being the first case in a cluster. This important high-risk group, identified by these analyses, provides support for the current chest radiograph classification system specified in the UK technical instructions. It may be possible to identify categories of chest radiograph (as defined in Table 5) that are at particularly high risk, and as these data are available in the pre-entry dataset this analysis could be relatively easily and quickly undertaken.

- **Recommendation:** UK technical instructions should continue to use the classification system for chest radiographs, although this may be improved further with additional analysis to identify particularly high-risk groups.
- **Recommendation:** Migrants with suspected tuberculosis on chest radiograph, but negative sputum smears and cultures should be assessed for latent tuberculosis infection and offered treatment where appropriate.
- **Recommendation:** After migration to the UK, those migrants with suspected tuberculosis on chest radiograph, but negative sputum smears and cultures should be given additional health improvement advice in order to reduce their

risk of tuberculosis, such as smoking cessation and appropriate treatment for medical conditions that increase risk of tuberculosis.

- **Recommendation**: Migrants with suspected tuberculosis on chest radiograph, but negative sputum smears and cultures should be targeted with information that will inform them about how to seek medical help in a timely manner if they develop tuberculosis systems, including information about how to access the health service.
- **Recommendation:** Barriers to accessing health services that may result in a delay in presentation among migrants should be carefully examined as these could result in increased risk of transmission in those migrants with tuberculosis.
- **Research question:** Are any sub-categories of suspected tuberculosis on radiological classification (as defined in the UK technical instructions) associated with an increased risk of incident cases of tuberculosis after arrival in the UK.

Screening in lower WHO prevalence countries

Consistently across all chapters of the thesis, including the systematic review and meta-analysis of studies from other countries, migrants from low prevalence countries were at the lowest risk tuberculosis. A total of 1,863 cases of tuberculosis were notified in migrants screened pre-entry and 1% (24/1,863) of these were migrants from countries with a prevalence of less than 150 per 100,000 population despite the fact that they accounted for 6% (29,143/519,955) of all migrants screened. These individuals had an incidence of tuberculosis of 44 per 100,000 person years at risk compared 210 in those migrants from countries with a WHO prevalence greater than 350 per 100,000 person years. The UK government invests a substantial amount of money into the quality assurance of pre-entry screening, cost-effectiveness analyses should therefore be undertaken to examine the cost effectiveness of screening countries with a prevalence of less than 150 per 100,000 population.

- **Research question:** What is the cost-effectiveness of pre-entry screening in countries with a prevalence of less than 150 per 100,000 population?

Visa categories and socio-economic factors among migrants

As the largest group of visa applicants, students were used as the baseline comparator in all of the analyses presented in the thesis. In the prevalence analysis in Chapters 3, 5

and 6, migrants on a family reunion or settlement and dependant visas were at increased risk compared to students, and in Chapter 5 working holiday visas were at a lower risk. Although students were at a lower risk than family reunion or settlement and dependant visas, they accounted for a large number of the incident cases, and had a reasonably high incidence of tuberculosis post entry. Therefore it would not seem appropriate to consider exempting this group from screening, but health improvement interventions could be targeted efficiently at this group through higher education institutions.

- **Research question:** What evidence-based interventions could be targeted at family reunion or settlement and dependant visas and students to ensure they receive quick and appropriate diagnosis and treatment if they develop tuberculosis after arrival in the UK?

Enlightened self-interest approach to tuberculosis control

Emerging evidence suggests that domestic returns for investment in tuberculosis control programmes overseas may make them cost effective, and policy-makers may wish to consider implementation alongside pre-entry screening programmes.(191,192) Such an "enlightened self-interest" approach to global tuberculosis control may be not only more cost effective, but could overcome screening-induced inequalities, so that a greater number of individuals in need benefit from treatment, not just those in a position to leave their country of origin. This broader view would enhance global collaboration in efforts to eliminate tuberculosis.

- **Research question:** What is the cost-effectiveness of an enlightened self-interest approach to global tuberculosis control when considering migration patterns to the UK?

7.3 Recommendations for latent tuberculosis screening and tuberculosis control in the UK and globally

In January 2015 Public Health England published a collaborative tuberculosis strategy for England that included ten evidence-based recommendations that are described in Box 3.(64) Several members of the team who produced this strategy (including Professor Ibrahim Abubakar, secondary supervisor of this PhD thesis and head of the tuberculosis at Public Health England, and Dominik Zenner, head of the tuberculosis screening section) have been directly involved in the analyses presented in this thesis.

Additionally, the author of this thesis has an honorary appointment with the tuberculosis section at Public Health England and has had access to many key decision makers for tuberculosis screening policy. Further work will be undertaken to support or refine recommendations in the strategy to ensure results from this thesis are translated into evidence-based public health policy. Several findings from the thesis relate to the ten priorities in the collaborative tuberculosis strategy for England and are outlined below under the appropriate heading from the strategy.

Box 3. Recommendations from collaborative tuberculosis strategy for England.(64)

Ten evidence-based recommendations from collaborative tuberculosis strategy for England:

1. Improve access to services and ensure early diagnosis
2. Provide universal access to high quality diagnostics
3. Improve treatment and care services
4. Ensure comprehensive contact tracing
5. Improve BCG vaccination uptake
6. Reduce drug-resistant TB
7. Tackle TB in under-served populations
8. Systematically implement new entrant latent TB screening
9. Strengthen surveillance and monitoring
10. Ensure an appropriate workforce to deliver TB control

7.3.1 Ensure comprehensive contact tracing

The cohort analysis examining the incidence of first in cluster cases of tuberculosis in chapter 6 found that children under the age of 16 were at higher risk. Many studies report that children are less likely to be infectious than adults, possibly as a result of being less likely to have cavitary lesions, and producing smaller numbers of droplets as a result of a weaker cough compared to adults.(193–195) It is therefore more likely that this finding is consistent with misclassification of the first in cluster case, with an adult family member being the source, but getting notified as a case of tuberculosis later than

the child. This potential explanation supports a continued focus on household contact tracing, and particularly rigorous efforts should be made for each paediatric case found.

- **Recommendation:** Tuberculosis services should continue to ensure robust contact tracing mechanisms are in place particularly for incidents involving young children.

7.3.2 Systematically implement new entrant latent tuberculosis screening

The tuberculosis strategy for England recognised that many cases of tuberculosis in migrants to the UK were a result of reactivation and recommended latent infection screening for new entrants from high incidence areas. The specific recommendations relating to the roll out of new entrant latent tuberculosis screening in the strategy are outlined in Box 4.

Box 4. Collaborative tuberculosis strategy for England – actions relating to systematic implementation of new entrant latent tuberculosis screening.(64)

Actions in relation to latent tuberculosis screening in new entrants:

- Establish co-ordinated LTBI screening for new entrants from areas of the world with high incidence living in England and ensure TB control board support to implement systematic LTBI screening nationally and as a high priority intervention in high burden areas (areas with an incidence of TB over 20 per 100,000)
- Offer LTBI screening to new entrants who were born or lived in Sub Saharan Africa or countries with an estimated TB incidence of greater than 150 per 100,000 and who arrived in the UK within the last five years
- Ensure robust policies for LTBI screening for other high risk population groups, where this is NICE recommended (such as in patients with immunosuppression)
- Work with local authorities, communities and third sector organisations to raise awareness and improve health education regarding LTBI screening
- Ensure local LTBI screening is well resourced, co-ordinated and quality assured and as appropriate embedded in local health check procedures for other illnesses such as hepatitis or HIV

Several findings from Chapter 6 could inform additional recommendations for the targeted roll out of a new systematic new entrant latent tuberculosis screening programme. Given the evidence that those migrants who had been in the UK for more than 5 years were at lower risk of reactivation, little or no resource should be put in to any sort of 'catch up' screening programme targeted at this longer term group due to the unlikely benefit this would have. The cross-sectional analysis in chapter 6 found that those over the age of 45 remained at higher risk of reactivation. NICE guidelines currently recommend limiting screening to those under the age of 35, reflecting the increased risk hepatotoxicity from the drugs given as chemoprophylaxis to those with latent infection. (196) The findings from this chapter therefore suggest that the risk and benefit of screening and treatment should be re-examined, and this new information should be provided to those identified at risk using appropriate decision making tools.

Those with an abnormal chest radiograph at pre-entry screening were at increased risk of reactivation compared to other migrants, and therefore this group should be a high priority for the offer of treatment for latent infection. As these individuals were negative by sputum culture, smear or both, the high likelihood is that the chest radiograph finding is the result of infection and then reactivation. It may therefore be appropriate to offer these individuals treatment for infection, with or without the additional offer of a test for latent infection. In the cohort analysis, migrants from countries with a prevalence of less than 350 per 100,000 population were at lower risk of reactivation with an incidence rate of 21 per 100,000 person years at risk compared to those from countries greater than 350 per 100,000 at 57 per 100,000 person years at risk. Therefore the cost effectiveness of offering migrants from countries with a prevalence of between 150-349 per 100,000 population should be carefully examined to ensure this is an appropriate use of time and resource.

- **Research question:** What are the risks and benefits of providing treatment for latent infection in migrants to the UK over the age of 35?
- **Research question:** What decision making tools could be used to help patients make an informed choice about whether to take treatment for latent infection based on the risks and benefits outlined in this thesis?

- **Research question:** Should migrants with a chest radiograph classified as suspected tuberculosis, but negative sputum smears and cultures, be offered treatment for latent infection without further testing?
- **Research question:** What is the cost effectiveness of offering latent tuberculosis infection treatment to migrants with a chest radiograph classified as suspected tuberculosis and negative sputum smears and cultures?
- **Research question:** What is the cost-effectiveness of latent tuberculosis screening and treatment in migrants from countries with a WHO prevalence between 149-350 per 100,000 population?

7.3.3 Strengthening surveillance and monitoring

This analysis was only possible because of the high quality surveillance data collected by Public Health England and the pre-entry screening dataset managed by the International Organisation for Migration. These are both excellent resources supported by dedicated teams of public health consultants and epidemiologists. A number of actions from the tuberculosis strategy for England aim to improve the UK surveillance system further, and some additional specific recommendations can be made on the basis of the strengths and limitations discussed in the previous chapters.

As the number of countries where pre-entry screening is performed increases, the power to conduct even more detailed epidemiological analyses will improve. The study to examine the accuracy of the probabilistic linkage algorithms was conducted in a relatively small dataset, and through restricting the analysis to non-UK born individuals. A possible way of increasing the size and accuracy of this analysis would be to use passport number as a unique identifying variable, or to undertake a series of manual reviews with follow up of individual cases to check that links were appropriately made, or not. This would help ensure that on-going linkages remained accurate and free from bias.

A potential alternative explanation for some of the findings in this thesis are unmeasured confounding factors in particular these include socio-economic factors for migrants and HIV. Efforts should therefore be made to collect these data in an appropriately sensitive manor that met information governance and public health legislation. Such data would not only provide additional reassurance to the findings of

the thesis, but it would enable additional health policy and improvement recommendations to be made accounting for these potential confounding factors.

- **Recommendation:** Attempts should be made to collect additional socio-economic and clinical risk factors for migrants screened pre-entry.
- **Research question:** Are other study designs able to further validate the accuracy of probabilistic matching in a large dataset that exclusively contains migrants or non-UK born individuals?

7.4 Conclusion

This thesis has generated new knowledge that improves our understanding of the epidemiology of tuberculosis in migrants to the UK. The thesis has established the current state of the published literature around pre-entry screening for tuberculosis, and undertaken the first comprehensive analysis of the pilot pre-entry screening programme in migrants to the UK. It has developed and validated new methodologies that will enable future research into these vulnerable groups both for tuberculosis and other diseases. The studies presented estimate for the first time the incidence of tuberculosis in migrants to the UK after screening for active tuberculosis disease and identified the risk factors for several different tuberculosis outcomes. Areas for further research are identified, and several of the findings have important public health implications. Working closely with the tuberculosis section of Public Health England, NHS England, and the Home Office, these evidence-based recommendations will be acted upon and implemented in a timely manner.

References

1. Koch R. Die Atiologic der Tuberkulose. Berl Klin Wochenschr. 1882;15:221–30.

2. Nobelprize.org. Robert Koch and Tuberculosis: Koch's Famous Lecture [Internet]. 2015 [cited 2015 Sep 5]. Available from: http://www.nobelprize.org/educational/medicine/tuberculosis/readmore.html

3. Bynum H. Spitting Blood: The history of tuberculosis. 1 edition. Oxford: Oxford University Press; 2012. 256 p.

4. Kaufmann SHE. Tuberculosis vaccines—a new kid on the block. Nat Med. 2011 Feb;17(2):159–60.

5. World Health Organization. Guidelines on the Management of Latent Tuberculosis Infection. Geneva: World Health Organization; 2015.

6. CDC | TB | Fact Sheets - Latent TB Infection vs. TB Disease [Internet]. [cited 2015 May 5]. Available from: http://www.cdc.gov/tb/publications/factsheets/general/ltbiandactivetb.htm

7. UK tuberculosis technical instructions - Publications - GOV.UK [Internet]. [cited 2014 Sep 15]. Available from: https://www.gov.uk/government/publications/uk-tuberculosis-technical-instructions

8. Canadian Lung Association, Canadian Thoracic Society, Public Health Agency of Canada, Centre for Communicable Diseases and Infection Control (Canada). Canadian tuberculosis standards. [Internet]. 2014 [cited 2015 Mar 27]. Available from: http://epe.lac-bac.gc.ca/100/201/301/weekly_checklist/2014/internet/w14-14-U-E.html/collections/collection_2014/aspc-phac/HP40-18-2014-eng.pdf

9. Nienhaus A, Schablon A, Costa JT, Diel R. Systematic review of cost and cost-effectiveness of different TB-screening strategies. BMC Health Serv Res. 2011;11:247.

10. CDC | TB | LTBI - Diagnosis of Latent TB Infection [Internet]. [cited 2015 May 17]. Available from: http://www.cdc.gov/tb/publications/ltbi/diagnosis.htm

11. Pai M, Zwerling A, Menzies D. Systematic review: T-cell-based assays for the diagnosis of latent tuberculosis infection: an update. Ann Intern Med. 2008 Aug 5;149(3):177–84.

12. Ewer K, Deeks J, Alvarez L, Bryant G, Waller S, Andersen P, et al. Comparison of T-cell-based assay with tuberculin skin test for diagnosis of Mycobacterium tuberculosis infection in a school tuberculosis outbreak. The Lancet. 2003 Apr 5;361(9364):1168–73.

13. Selwyn PA, Hartel D, Lewis VA, Schoenbaum EE, Vermund SH, Klein RS, et al. A prospective study of the risk of tuberculosis among intravenous drug users with human immunodeficiency virus infection. N Engl J Med. 1989 Mar 2;320(9):545–50.

14. Akolo C, Adetifa I, Shepperd S, Volmink J. Treatment of latent tuberculosis infection in HIV infected persons. Cochrane Database Syst Rev. 2010;(1):CD000171.

15. McShane H. Co-infection with HIV and TB: double trouble. Int J STD AIDS. 2005 Feb;16(2):95–100; quiz 101.

16. Williams BG, Dye C. Antiretroviral drugs for tuberculosis control in the era of HIV/AIDS. Science. 2003 Sep 12;301(5639):1535–7.

17. Lawn SD, Zumla AI. Tuberculosis. The Lancet. 2011 Jul 2;378(9785):57–72.

18. Smieja MJ, Marchetti CA, Cook DJ, Smaill FM. Isoniazid for preventing tuberculosis in non-HIV infected persons. Cochrane Database Syst Rev Online. 2000;(2):CD001363.

19. Ziakas PD, Mylonakis E. 4 months of rifampin compared with 9 months of isoniazid for the management of latent tuberculosis infection: a meta-analysis and cost-effectiveness study that focuses on compliance and liver toxicity. Clin Infect Dis Off Publ Infect Dis Soc Am. 2009 Dec 15;49(12):1883–9.

20. Dye C, Williams BG. Eliminating human tuberculosis in the twenty-first century. J R Soc Interface. 2008 Jun 6;5(23):653–62.

21. Styblo K. The relationship between the risk of tuberculous infection and the risk of developing infectious tuberculosis. Bull Int Union Tuberc Lung Dis. 1985;60(3-4):117–9.

22. Van Leth F, Van der Werf MJ, Borgdorff MW. Prevalence of tuberculous infection and incidence of tuberculosis: a re-assessment of the Styblo rule. Bull World Health Organ. 2008;86(1):20–6.

23. Tuberculosis | Guidance and guidelines | NICE [Internet]. [cited 2014 Nov 25]. Available from: http://www.nice.org.uk/guidance/cg117

24. Frieden T. Toman's Tuberculosis: Case Detection, Treatment and Monitoring. 2nd ed. World Health Organization; 2004.

25. Van Deun A, Roorda FA, Chambugonj N, Hye A, Hossain A. Reproducibility of sputum smear examination for acid-fast bacilli: practical problems met during cross-checking. Int J Tuberc Lung Dis Off J Int Union Tuberc Lung Dis. 1999 Sep;3(9):823–9.

26. Johnson JL, Vjecha MJ, Okwera A, Hatanga E, Byekwaso F, Wolski K, et al. Impact of human immunodeficiency virus type-1 infection on the initial bacteriologic and radiographic manifestations of pulmonary tuberculosis in Uganda. Makerere University-Case Western Reserve University Research Collaboration. Int J Tuberc Lung Dis Off J Int Union Tuberc Lung Dis. 1998 May;2(5):397–404.

27. Narain R, Nair SS, Naganna K, Chandrasekhar P, Rao GR, Lal P. Problems in defining a "case" of pulmonary tuberculosis in prevalence surveys. Bull World Health Organ. 1968;39(5):701–29.

28. Tuberculosis in a rural population of South India: a five-year epidemiological study. Bull World Health Organ. 1974;51(5):473–88.

29. World Health Organization, Stop TB Initiative (World Health Organization), editors. Treatment of tuberculosis: guidelines. 4th ed. Geneva: World Health Organization; 2010. 147 p.

30. World Health Organization. Global tuberculosis report 2014. Geneva: World Health Organization; 2014.

31. Murray CJ, Vos T, Lozano R, Naghavi M, Flaxman AD, Michaud C, et al. Disability-adjusted life years (DALYs) for 291 diseases and injuries in 21 regions, 1990–2010: a systematic analysis for the Global Burden of Disease Study 2010. The Lancet. 2013;380(9859):2197–223.

32. Murray CJL, Ortblad KF, Guinovart C, Lim SS, Wolock TM, Roberts DA, et al. Global, regional, and national incidence and mortality for HIV, tuberculosis, and malaria during 1990–2013: a systematic analysis for the Global Burden of Disease Study 2013. The Lancet. 2014 Sep;384(9947):1005–70.

33. Dodd PJ, Gardiner E, Coghlan R, Seddon JA. Burden of childhood tuberculosis in 22 high-burden countries: a mathematical modelling study. Lancet Glob Health. 2014 Aug;2(8):e453–9.

34. Jenkins HE, Tolman AW, Yuen CM, Parr JB, Keshavjee S, Pérez-Vélez CM, et al. Incidence of multidrug-resistant tuberculosis disease in children: systematic review and global estimates. Lancet. 2014 May 3;383(9928):1572–9.

35. Public Health England. Tuberculosis (TB) in the UK: annual report - 2014 [Internet]. [cited 2015 Apr 6]. Available from: https://www.gov.uk/government/publications/tuberculosis-tb-in-the-uk

36. Schaaf HS, Collins A, Bekker A, Davies PDO. Tuberculosis at extremes of age. Respirol Carlton Vic. 2010 Jul;15(5):747–63.

37. Gallant CJ, Cobat A, Simkin L, Black GF, Stanley K, Hughes J, et al. Impact of age and sex on mycobacterial immunity in an area of high tuberculosis incidence. Int J Tuberc Lung Dis Off J Int Union Tuberc Lung Dis. 2010 Aug;14(8):952–9.

38. Hinks TSC, Varsani N, Godsiff DT, Bull TC, Nash KL, McLuckie L, et al. High background rates of positive tuberculosis-specific interferon-γ release assays in a low prevalence region of UK: a surveillance study. BMC Infect Dis. 2012;12:339.

39. Davies PD. Tuberculosis in the elderly. Epidemiology and optimal management. Drugs Aging. 1996 Jun;8(6):436–44.

40. De Vries G, Aldridge RW, Cayla JA, Haas WH, Sandgren A, van Hest NA, et al. Epidemiology of tuberculosis in big cities of the European Union and European Economic Area countries. Euro Surveill Bull Eur Sur Mal Transm Eur Commun Dis Bull. 2014;19(9).

41. Van Hest NA, Aldridge RW, de Vries G, Sandgren A, Hauer B, Hayward A, et al. Tuberculosis control in big cities and urban risk groups in the European Union: a consensus statement. Euro Surveill Bull Eur Sur Mal Transm Eur Commun Dis Bull. 2014;19(9).

42. WHO | Global tuberculosis report 2012 [Internet]. WHO. [cited 2013 Aug 28]. Available from: http://www.who.int/tb/publications/global_report/en/

43. Centers for Disease Control and Prevention (CDC). Trends in tuberculosis--United States, 2012. MMWR Morb Mortal Wkly Rep. 2013 Mar 22;62(11):201–5.

44. European Centre for Disease Prevention and Control/WHO Regional, Office for Europe. Tuberculosis surveillance and monitoring in Europe 2013. Stockholm: European Centre for Disease Prevention and Control; 2013.

45. Agency HP. Tuberculosis in the UK: 2013 report [Internet]. [cited 2014 Apr 9]. Available from: http://www.hpa.org.uk/webw/HPAweb&HPAwebStandard/HPAweb_C/1317139689732

46. United Nations Population Division - Department of Economic and Social Affairs. International Migration Flows to and from Selected Countries: The 2010 Revision [Internet]. [cited 2013 Aug 28]. Available from: http://esa.un.org/MigFlows/Notes.html

47. United Nations Population Division | Department of Economic and Social Affairs. International Migration Report 2013 [Internet]. New: United Nations Population Division | Department of Economic and Social Affairs; 2013 [cited 2014 Apr 9]. Available from: http://www.un.org/en/development/desa/population/publications/migration/migration-report-2013.shtml

48. International Organization for Migration. World Migration Report 2011: Communicating Effectively about Migration [Internet]. Geneva: International Organization for Migration; 2011 [cited 2014 Sep 15]. Available from: publications.iom.int/bookstore/free/WMR2011_English.pdf

49. International Organization for Migration. The World Migration Report 2013: Migrant Well-being and Development [Internet]. Geneva: International Organization for Migration; 2013 [cited 2014 Sep 15]. Available from: http://publications.iom.int/bookstore/free/WMR2013_EN.pdf

50. Office for National Statistics. Migration Statistics Quarterly Report [Internet]. Office for National Statistics. [cited 2015 May 5]. Available from: http://www.ons.gov.uk/ons/rel/migration1/migration-statistics-quarterly-report/index.html

51. Statham P, Geddes A. Elites and the "organised public": Who drives British immigration politics and in which direction? West Eur Polit. 2006 Mar 1;29(2):248–69.

52. Lahav G. Public Opinion Toward Immigration in the European Union Does It Matter? Comp Polit Stud. 2004 Dec 1;37(10):1151–83.

53. Office for National Statistics. International migration: a recent history. Office for National Statistics. 2015.

54. Dara M, Gushulak BD, Posey DL, Zellweger J-P, Migliori GB. The history and evolution of immigration medical screening for tuberculosis. Expert Rev Anti Infect Ther. 2013 Feb;11(2):137–46.

55. Klinkenberg E, Manissero D, Semenza JC, Verver S. Migrant tuberculosis screening in the EU/EEA: yield, coverage and limitations. Eur Respir J Off J Eur Soc Clin Respir Physiol. 2009 Nov;34(5):1180–9.

56. Pareek M, Baussano I, Abubakar I, Dye C, Lalvani A. Evaluation of immigrant tuberculosis screening in industrialized countries. Emerg Infect Dis. 2012 Sep;18(9):1422–9.

57. R.J. Coker, A. Bell, R. Pitman, A. Hayward, J. Watson. Screening programmes for tuberculosis in new entrants across Europe. Int J Tuberc Lung Dis. 2004 Aug;8(8):1022–6.

58. Coker R, Bell A, Pitman R, Zellweger J-P, Heldal E, Hayward A, et al. Tuberculosis screening in migrants in selected European countries shows wide disparities. Eur Respir J Off J Eur Soc Clin Respir Physiol. 2006 Apr;27(4):801–7.

59. UK Border Agency - Central Policy Unit. Screening for tuberculosis and immigration control [Internet]. London: UK Border Agency; 2012 Oct [cited 2015 Feb 3]. Available from: https://www.gov.uk/government/publications/immigration-pre-entry-tuberculosis-testing

60. International Organization for Migration. International Organization for Migration [Internet]. International Organization for Migration. 2014 [cited 2015 Mar 2]. Available from: http://unitedkingdom.iom.int/about-us

61. Lalvani A, Pareek M. Immigrant screening for TB: a missed opportunity to improve TB control in the United Kingdom. Pathog Glob Health. 2012 Mar;106(1):5–7.

62. Pareek M, Abubakar I, White PJ, Garnett GP, Lalvani A. Tuberculosis screening of migrants to low-burden nations: insights from evaluation of UK practice. Eur Respir J Off J Eur Soc Clin Respir Physiol. 2011 May;37(5):1175–82.

63. Pareek M, Abubakar I, White PJ, Garnett GP, Lalvani A. U.K. immigrant screening is inversely related to regional tuberculosis burden. Thorax. 2011 Nov;66(11):1010.

64. Tuberculosis (TB): collaborative strategy for England - Publications - GOV.UK [Internet]. [cited 2015 May 8]. Available from: https://www.gov.uk/government/publications/collaborative-tuberculosis-strategy-for-england

65. Pai M, Schito M. Tuberculosis diagnostics in 2015: landscape, priorities, needs, and prospects. J Infect Dis. 2015 Apr 1;211 Suppl 2:S21–8.

66. Lawn SD, Mwaba P, Bates M, Piatek A, Alexander H, Marais BJ, et al. Advances in tuberculosis diagnostics: the Xpert MTB/RIF assay and future prospects for a point-of-care test. Lancet Infect Dis. 2013 Apr;13(4):349–61.

67. Boehme CC, Nabeta P, Hillemann D, Nicol MP, Shenai S, Krapp F, et al. Rapid Molecular Detection of Tuberculosis and Rifampin Resistance. N Engl J Med. 2010 Sep 9;363(11):1005–15.

68. Boehme CC, Nicol MP, Nabeta P, Michael JS, Gotuzzo E, Tahirli R, et al. Feasibility, diagnostic accuracy, and effectiveness of decentralised use of the Xpert MTB/RIF test for diagnosis of tuberculosis and multidrug resistance: a multicentre implementation study. The Lancet. 2011 Apr 30;377(9776):1495–505.

69. Penz E, Boffa J, Roberts DJ, Fisher D, Cooper R, Ronksley PE, et al. Diagnostic accuracy of the Xpert® MTB/RIF assay for extra-pulmonary tuberculosis: a meta-analysis. Int J Tuberc Lung Dis Off J Int Union Tuberc Lung Dis. 2015 Mar;19(3):278–84, i – iii.

70. VYNNYCKY E, FINE PEM. The Natural History of Tuberculosis: The Implications of Age-Dependent Risks of Disease and the Role of Reinfection. Epidemiol Infect. 1997;119(02):183–201.

71. Campbell JR, Krot J, Elwood K, Cook V, Marra F. A systematic review on TST and IGRA tests used for diagnosis of LTBI in immigrants. Mol Diagn Ther. 2015 Feb;19(1):9–24.

72. Rangaka MX, Wilkinson KA, Glynn JR, Ling D, Menzies D, Mwansa-Kambafwile J, et al. Predictive value of interferon-γ release assays for incident active tuberculosis: a systematic review and meta-analysis. Lancet Infect Dis. 2012 Jan;12(1):45–55.

73. Pareek M, Watson JP, Ormerod LP, Kon OM, Woltmann G, White PJ, et al. Screening of immigrants in the UK for imported latent tuberculosis: a multicentre cohort study and cost-effectiveness analysis. Lancet Infect Dis. 2011 Jun;11(6):435–44.

74. Davies PDO, Yew WW, Ganguly D, Davidow AL, Reichman LB, Dheda K, et al. Smoking and tuberculosis: the epidemiological association and immunopathogenesis. Trans R Soc Trop Med Hyg. 2006 Apr 1;100(4):291–8.

75. Tocque K, Doherty MJ, Bellis MA, Spence DP, Williams CS, Davies PD. Tuberculosis notifications in England: the relative effects of deprivation and immigration. Int J Tuberc Lung Dis Off J Int Union Tuberc Lung Dis. 1998 Mar;2(3):213–8.

76. Ormerod LP, McCarthy OR, Paul EA. Changing pattern of respiratory tuberculosis in the UK in adult patients from the Indian subcontinent. Thorax. 1997 Sep;52(9):802–4.

77. Alvarez GG, Gushulak B, Abu Rumman K, Altpeter E, Chemtob D, Douglas P, et al. A comparative examination of tuberculosis immigration medical screening programs from selected countries with high immigration and low tuberculosis incidence rates. BMC Infect Dis. 2011;11:3.

78. Alvarez GG, Clark M, Altpeter E, Douglas P, Jones J, Paty M-C, et al. Pediatric tuberculosis immigration screening in high-immigration, low-incidence countries. Int J Tuberc Lung Dis Off J Int Union Tuberc Lung Dis. 2010 Dec;14(12):1530–7.

79. Rieder HL, Zellweger JP, Raviglione MC, Keizer ST, Migliori GB. Tuberculosis control in Europe and international migration. Eur Respir J Off J Eur Soc Clin Respir Physiol. 1994 Aug;7(8):1545–53.

80. European Centre for Disease Prevention and Contro. Progressing towards TB elimination: A follow up to the Framework Action Plan to Fight Tuberculosis [Internet]. Stockholm: European Centre for Disease Prevention and Contro; 2010 Nov [cited 2014 Sep 15]. Available from: http://ecdc.europa.eu/en/publications/publications/101111_spr_progressing_towards_tb_elimination.pdf

81. Moher D, Liberati A, Tetzlaff J, Altman DG, for the PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. BMJ. 2009 Jul 21;339(jul21 1):b2535–b2535.

82. Roy Rosenzweig Center for History and New Media. Zotero [Internet]. Virginia, USA: Roy Rosenzweig Center for History and New Media; 2014. Available from: https://www.zotero.org/

83. Institute of Education, University of London, UK. EPPI-Reviewer 4 [Internet]. London: University College London; 2014. Available from: http://eppi.ioe.ac.uk/cms/Default.aspx?alias=eppi.ioe.ac.uk/cms/er4

84. Freeman MF, Tukey JW. Transformations Related to the Angular and the Square Root. Ann Math Stat. 1950 Dec;21(4):607–11.

85. Barendregt JJ, Doi SA, Lee YY, Norman RE, Vos T. Meta-analysis of prevalence. J Epidemiol Community Health. 2013 Aug 20;jech – 2013–203104.

86. Higgins JPT, Thompson SG. Quantifying heterogeneity in a meta-analysis. Stat Med. 2002 Jun 15;21(11):1539–58.

87. Guyatt G, Oxman AD, Akl EA, Kunz R, Vist G, Brozek J, et al. GRADE guidelines: 1. Introduction—GRADE evidence profiles and summary of findings tables. J Clin Epidemiol. 2011 Apr;64(4):383–94.

88. Bollini P, Pampallona S, O'Rourke T, Jones W, Truong TB, Nguyen HP. Outcome of second-line tuberculosis treatment in migrants from Vietnam. International Organization for Migration (IOM) Tuberculosis Working Group. Trop Med Int Health TM IH. 1998 Dec;3(12):975–80.

89. Gorbacheva O, Mishra AK, Shapovalov D, Sudtasay S. Prevalence of bacteriologically confirmed pulmonary tuberculosis in the Bhutanese refugees in Nepal. Results of active case finding. Int J Infect Dis. 2010 Mar 1;14:e150.

90. Dasgupta K, Schwartzman K, Marchand R, Tennenbaum TN, Brassard P, Menzies D. Comparison of cost-effectiveness of tuberculosis screening of close contacts and foreign-born populations. Am J Respir Crit Care Med. 2000 Dec;162(6):2079–86.

91. King K, Douglas PJ, Beath K. Is premigration health screening for tuberculosis worthwhile? Med J Aust. 2011 Nov 7;195(9):534–7.

92. Lange WR, Warnock-Eckhart E, Bean ME. Mycobacterium tuberculosis infection in foreign born adoptees. Pediatr Infect Dis J. 1989 Sep;8(9):625–9.

93. Liu Y, Weinberg MS, Ortega LS, Painter JA, Maloney SA. Overseas screening for tuberculosis in US-bound immigrants and refugees. N Engl J Med. 2009;360(23):2406–15.

94. Malone JL, Paparello SF, Malone JD, Hill HE, Conrad KA, Myers JW, et al. Drug susceptibility of Mycobacterium tuberculosis isolates from recent Haitian migrants: correlation with clinical response. Clin Infect Dis Off Publ Infect Dis Soc Am. 1994 Nov;19(5):938–40.

95. Maloney SA, Fielding KL, Laserson KF, Jones W, Nguyen TNY, Dang QA, et al. Assessing the performance of overseas tuberculosis screening programs: a study among US-bound immigrants in Vietnam. Arch Intern Med. 2006 Jan 23;166(2):234–40.

96. Mor Z, Leventhal A, Weiler-Ravell D, Peled N, Lerman Y. Chest radiography validity in screening pulmonary tuberculosis in immigrants from a high-burden country. Respir Care. 2012 Jul;57(7):1137–44.

97. Oeltmann JE, Varma JK, Ortega L, Liu Y, O'Rourke T, Cano M, et al. Multidrug-resistant tuberculosis outbreak among US-bound Hmong refugees, Thailand, 2005. Emerg Infect Dis. 2008 Nov;14(11):1715–21.

98. Painter JA, Graviss EA, Hai HH, Nhung DTC, Nga TTT, Ha NP, et al. Tuberculosis screening by tuberculosis skin test or QuantiFERON-TB Gold In-Tube Assay among an immigrant population with a high prevalence of tuberculosis and BCG vaccination. PloS One. 2013;8(12):e82727.

99. Plant AJ, Watkins RE, Motus N, Jones W, O'Rourke T, Streeton J, et al. Results of tuberculosis screening in applicants for migration in Vietnam and Cambodia. Int J Tuberc Lung Dis Off J Int Union Tuberc Lung Dis. 2005 Feb;9(2):157–63.

100. Wang JS, Allen EA, Enarson DA, Grzybowski S. Tuberculosis in recent Asian immigrants to British Columbia, Canada: 1982-1985. Tubercle. 1991 Dec;72(4):277–83.

101. Watkins RE, Plant AJ, Sang D, O'Rourke TF, Eltom AA, Streeton J, et al. The association between subjective and clinical indicators of health in prospective Vietnamese migrants. Asia-Pac J Public Health Asia-Pac Acad Consort Public Health. 2005;17(1):46–50.

102. Yanni EA, Naoum M, Odeh N, Han P, Coleman M, Burke H. The health profile and chronic diseases comorbidities of US-bound Iraqi refugees screened by the International Organization for Migration in Jordan: 2007-2009. J Immigr Minor Health Cent Minor Public Health. 2013 Feb;15(1):1–9.

103. Keane VP, O'Rourke TF, Bollini P, Pampallona S, Siem H. Prevalence of tuberculosis in Vietnamese migrants: the experience of the Orderly Departure Program. Southeast Asian J Trop Med Public Health. 1995 Dec;26(4):642–7.

104. Centers for Disease Control and Prevention (CDC). Health of resettled Iraqi refugees --- San Diego County, California, October 2007-September 2009. MMWR Morb Mortal Wkly Rep. 2010 Dec 17;59(49):1614–8.

105. Centers for Disease Control and Prevention (CDC). Multidrug-resistant tuberculosis in Hmong refugees resettling from Thailand into the United States, 2004-2005. MMWR Morb Mortal Wkly Rep. 2005 Aug 5;54(30):741–4.

106. Graham S, Das GK, Hidvegi RJ, Hanson R, Kosiuk J, Al ZK, et al. Chest radiograph abnormalities associated with tuberculosis: reproducibility and yield of active cases. Int J Tuberc Lung Dis Off J Int Union Tuberc Lung Dis. 2002 Feb;6(2):137–42.

107. Mor Z, Lerman Y, Leventhal A. Pre-immigration screening process and pulmonary tuberculosis among Ethiopian migrants in Israel. Eur Respir J Off J Eur Soc Clin Respir Physiol. 2008 Aug;32(2):413–8.

108. WHO | Policy guidance on drug-susceptibility testing (DST) of second-line antituberculosis drugs [Internet]. WHO. [cited 2015 Mar 27]. Available from: http://www.who.int/tb/publications/2008/whohtmtb_2008_392/en/

109. WHO | Definitions and reporting framework for tuberculosis [Internet]. WHO. [cited 2015 Feb 17]. Available from: http://www.who.int/tb/publications/definitions/en/

110. WHO | Global tuberculosis report 2014 [Internet]. WHO. [cited 2015 Mar 26]. Available from: http://www.who.int/tb/publications/global_report/en/

111. Kimbrough W, Saliba V, Dahab M, Haskew C, Checchi F. The burden of tuberculosis in crisis-affected populations: a systematic review. Lancet Infect Dis. 2012 Dec;12(12):950–65.

112. Bhatia S, Dranyi T, Rowley D. Tuberculosis among Tibetan refugees in India. Soc Sci Med 1982. 2002 Feb;54(3):423–32.

113. Mercer A. Mortality and morbidity in refugee camps in eastern Sudan: 1985-90. Disasters. 1992 Mar;16(1):28–42.

114. Nelson LJ, Naik Y, Tsering K, Cegielski JP. Population-based risk factors for tuberculosis and adverse outcomes among Tibetan refugees in India, 1994-1996. Int J Tuberc Lung Dis Off J Int Union Tuberc Lung Dis. 2005 Sep;9(9):1018–26.

115. Cegielski JP, McMurray DN. The relationship between malnutrition and tuberculosis: evidence from studies in humans and experimental animals. Int J Tuberc Lung Dis Off J Int Union Tuberc Lung Dis. 2004 Mar;8(3):286–98.

116. Beggs CB, Noakes CJ, Sleigh PA, Fletcher LA, Siddiqi K. The transmission of tuberculosis in confined spaces: an analytical review of alternative epidemiological models. Int J Tuberc Lung Dis Off J Int Union Tuberc Lung Dis. 2003 Nov;7(11):1015–26.

117. Boerma T, Mathers CD. The World Health Organization and global health estimates: improving collaboration and capacity. BMC Med. 2015 Mar 10;13(1):50.

118. Cantini F, Niccoli L, Goletti D. Adalimumab, etanercept, infliximab, and the risk of tuberculosis: data from clinical trials, national registries, and postmarketing surveillance. J Rheumatol Suppl. 2014 May;91:47–55.

119. Lorenzetti R, Zullo A, Ridola L, Diamanti AP, Laganà B, Gatta L, et al. Higher risk of tuberculosis reactivation when anti-TNF is combined with immunosuppressive agents: a systematic review of randomized controlled trials. Ann Med. 2014 Nov;46(7):547–54.

120. Barter DM, Agboola SO, Murray MB, Bärnighausen T. Tuberculosis and poverty: the contribution of patient costs in sub-Saharan Africa--a systematic review. BMC Public Health. 2012;12:980.

121. Bhutta ZA, Sommerfeld J, Lassi ZS, Salam RA, Das JK. Global burden, distribution, and interventions for infectious diseases of poverty. Infect Dis Poverty. 2014;3:21.

122. Mesfin YM, Hailemariam D, Biadgilign S, Biadglign S, Kibret KT. Association between HIV/AIDS and multi-drug resistance tuberculosis: a systematic review and meta-analysis. PloS One. 2014;9(1):e82235.

123. Beijer U, Wolf A, Fazel S. Prevalence of tuberculosis, hepatitis C virus, and HIV in homeless people: a systematic review and meta-analysis. Lancet Infect Dis. 2012 Nov;12(11):859–70.

124. Kufa T, Mabuto T, Muchiri E, Charalambous S, Rosillon D, Churchyard G, et al. Incidence of HIV-associated tuberculosis among individuals taking combination antiretroviral therapy: a systematic review and meta-analysis. PloS One. 2014;9(11):e111209.

125. Ejaz M, Siddiqui AR, Rafiq Y, Malik F, Channa A, Mangi R, et al. Prevalence of multi-drug resistant tuberculosis in Karachi, Pakistan: identification of at risk groups. Trans R Soc Trop Med Hyg. 2010 Aug;104(8):511–7.

126. Kazi AM, Shah SA, Jenkins CA, Shepherd BE, Vermund SH. Risk factors and prevalence of tuberculosis, human immunodeficiency virus, syphilis, hepatitis B virus, and hepatitis C virus among prisoners in Pakistan. Int J Infect Dis IJID Off Publ Int Soc Infect Dis. 2010 Sep;14 Suppl 3:e60–6.

127. Akhtar S, Rathi SK. Multilevel modeling of household contextual determinants of tuberculin skin test positivity among contacts of infectious tuberculosis patients, Umerkot, Pakistan. Am J Trop Med Hyg. 2009 Mar;80(3):351–8.

128. Aldridge RW, Yates TA, Zenner D, White PJ, Abubakar I, Hayward AC. Pre-entry screening programmes for tuberculosis in migrants to low-incidence countries: a systematic review and meta-analysis. Lancet Infect Dis. 2014 Dec;14(12):1240–9.

129. Gill M, Goldacre MJ, Yeates DGR. Changes in safety on England's roads: analysis of hospital statistics. BMJ. 2006 Jul 8;333(7558):73.

130. Trends in mortality rates comparing underlying-cause and multiple-cause coding in an english population 1979-1998. J Public Health Med. 2003;25:249–53.

131. Jit M, Stagg HR, Aldridge RW, White PJ, Abubakar I, For the Find and Treat Evaluation Team. Dedicated outreach service for hard to reach patients with tuberculosis in London: observational study and economic evaluation. BMJ. 2011 Sep;343(sep13 5):d5376–d5376.

132. WHO | Assessing tuberculosis under-reporting through inventory studies [Internet]. WHO. [cited 2015 Jan 28]. Available from: http://www.who.int/tb/publications/inventory_studies/en/

133. Blakely T, Salmond C. Probabilistic record linkage and a method to calculate the positive predictive value. Int J Epidemiol. 2002 Dec 1;31(6):1246–52.

134. Newcombe HB. Handbook of Record Linkage: Methods for Health and Statistical Studies, Administration and Business. Oxford University Press; 1988. 212 p.

135. Potz N, Powell D, Pebody R, Lamagni T, Bridger D, Duckworth G. Development of a method to link infection and mortality data. In: Health Protection Agency Annual Conference [Internet]. Warwick: Health Protection Agency; 2006 [cited 2015 Feb 1]. Available from: http://webarchive.nationalarchives.gov.uk/20140714084352/http:/www.hpa.org.uk/webc/HPAwebFile/HPAweb_C/1200660056163

136. Tuberculosis (TB) in the UK: annual report - Publications - GOV.UK [Internet]. [cited 2014 Nov 3]. Available from: https://www.gov.uk/government/publications/tuberculosis-tb-in-the-uk

137. Soundex [Internet]. Wikipedia, the free encyclopedia. 2014 [cited 2014 Nov 7]. Available from: http://en.wikipedia.org/w/index.php?title=Soundex&oldid=603659797

138. Fellegi IP, Sunter AB. A theory for record linkage. J Am Stat Assoc. 1969;64(328):1183–210.

139. Gill L. Methods for automatic record-matching and linkage and their use in National Statistics [Internet]. Office for National Statistics. 2011 [cited 2015 May 18]. Available from: http://www.ons.gov.uk/ons/guide-method/method-quality/specific/gss-methodology-series/index.html

140. Maguire H, Brailsford S, Carless J, Yates M, Altass L, Yates S, et al. Large outbreak of isoniazid-monoresistant tuberculosis in London, 1995 to 2006: case-control study and recommendations. Euro Surveill Bull Eur Sur Mal Transm Eur Commun Dis Bull [Internet]. 2011 [cited 2012 Aug 15];16(13). Available from: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=medl&AN=21489373

141. Story A, Aldridge RW, Gray T, Burridge S, Hayward AC. Influenza vaccination, inverse care and homelessness: cross-sectional survey of eligibility and uptake during the 2011/12 season in London. BMC Public Health. 2014 Jan 16;14(1):44.

142. Copeland KT, Checkoway H, McMICHAEL AJ, Holbrook RH. Bias Due to Misclassification in the Estimation of Relative Risk. Am J Epidemiol. 1977 May 1;105(5):488–95.

143. Howe GR. Use of Computerized Record Linkage in Cohort Studies. Epidemiol Rev. 1998 Jan 1;20(1):112–21.

144. VAN Hest NAH, Story A, Grant AD, Antoine D, Crofts JP, Watson JM. Record-linkage and capture-recapture analysis to estimate the incidence and completeness of reporting of tuberculosis in England 1999-2002. Epidemiol Infect. 2008 Dec;136(12):1606–16.

145. Silveira DP da, Artmann E. Accuracy of probabilistic record linkage applied to health databases: systematic review. Rev Saúde Pública. 2009 Oct;43(5):875–82.

146. Coutinho ESF, Coeli CM. [Accuracy of the probabilistic record linkage methodology to ascertain deaths in survival studies]. Cad Saúde Pública. 2006 Oct;22(10):2249–52.

147. Grannis SJ, Overhage JM, Hui S, McDonald CJ. Analysis of a Probabilistic Record Linkage Technique without Human Review. AMIA Annu Symp Proc. 2003;2003:259–63.

148. Nagle CM, Purdie DM, Webb PM, Green AC, Bain CJ. Searching for cancer deaths in Australia: National Death Index vs. cancer registries. Asian Pac J Cancer Prev APJCP. 2006 Mar;7(1):41–5.

149. Callaway E. UK push to open up patients' data. Nature. 2013 Oct 15;502(7471):283–283.

150. BMJ Blogs: The BMJ » Blog Archive » Clare Gerada: Why care.data will be transformational to the NHS [Internet]. [cited 2014 Sep 4]. Available from: http://blogs.bmj.com/bmj/2014/02/12/clare-gerada-on-why-care-data-will-be-transformational-to-the-nhs/

151. Lyons RA, Ford DV, Moore L, Rodgers SE. Use of data linkage to measure the population health effect of non-health-care interventions. The Lancet. 2014 Apr;383(9927):1517–9.

152. Abubakar I, Irvine L, Aldus CF, Wyatt GM, Fordham R, Schelenz S, et al. A systematic review of the clinical, public health and cost-effectiveness of rapid diagnostic tests for the detection and identification of bacterial intestinal pathogens in faeces and food. Health Technol Assess Winch Engl. 2007 Sep;11(36):1–216.

153. Tuberculosis: notifying cases - Publications - GOV.UK [Internet]. [cited 2015 Apr 1]. Available from: https://www.gov.uk/government/publications/tuberculosis-notifying-cases

154. Entry clearance visas by length - GOV.UK [Internet]. [cited 2015 Mar 28]. Available from: https://www.gov.uk/government/publications/entry-clearance-visas-by-length/entry-clearance-visas-by-length

155. Cancer Research UK Cancer Survival Group. Tools for Cancer Survival Analysis [Internet]. 2015 [cited 2015 Mar 27]. Available from: http://www.lshtm.ac.uk/eph/ncde/cancersurvival/tools/index.html#

156. ONS. Long Term International Migrants Area of Destination or Origin within the UK, 1991-2013 [Internet]. 2010 [cited 2015 Mar 28]. Available from: http://www.ons.gov.uk/ons/publications/re-reference-tables.html?edition=tcm%3A77-346438

157. Hennekens CH, Buring JE. Epidemiology in Medicine. Lippincott Williams and Wilkins; 1987. 400 p.

158. Porta MS, Greenland S, Hernán M, Silva I dos S, Last JM, International Epidemiological Association, editors. A dictionary of epidemiology. Six edition. Oxford: Oxford University Press; 2014. 343 p.

159. Clark RC, Mytton J. Estimating infectious disease in UK asylum seekers and refugees: a systematic review of prevalence studies. J Public Health Oxf Engl. 2007 Dec;29(4):420–8.

160. Callister MEJ, Barringer J, Thanabalasingam ST, Gair R, Davidson RN. Pulmonary tuberculosis among political asylum seekers screened at Heathrow Airport, London, 1995-9. Thorax. 2002 Feb;57(2):152–6.

161. Harling R, Pearce M, Chandrakumar M, Mueller K, Hayward A. Tuberculosis screening of asylum seekers: 1 years' experience at the Dover Induction Centres. Public Health. 2007 Nov;121(11):822–7.

162. Story A, Murad S, Roberts W, Verheyen M, Hayward AC. Tuberculosis in London: the importance of homelessness, problem drug use and prison. Thorax. 2007 Aug;62(8):667–71.

163. Ruddy MC, Davies AP, Yates MD, Yates S, Balasegaram S, Drabu Y, et al. Outbreak of isoniazid resistant tuberculosis in north London. Thorax. 2004;59:279–85.

164. Anderson C, Story A, Brown T, Drobniewski F, Abubakar I. Tuberculosis in UK prisoners: a challenge for control. J Epidemiol Community Health. 2010 Apr;64(4):373–6.

165. Story A, Bothamley G, Hayward A. Crack cocaine and infectious tuberculosis. Emerg Infect Dis. 2008 Sep;14(9):1466–9.

166. Liu Y, Posey DL, Cetron MS, Painter JA. Effect of a Culture-Based Screening Algorithm on Tuberculosis Incidence in Immigrants and Refugees Bound for the United StatesA Population-Based Cross-sectional StudyTB Screening in Immigrants and Refugees Bound for the United States. Ann Intern Med. 2015 Mar 17;162(6):420–8.

167. Lowenthal P, Westenhouse J, Moore M, Posey DL, Watt JP, Flood J. Reduced importation of tuberculosis after the implementation of an enhanced pre-immigration screening protocol. Int J Tuberc Lung Dis. 2011;15(6):761–6.

168. Oeltmann J E. Multidrug-resistant tuberculosis in Hmong refugees resettling from Thailand into the United States, 2004-2005 [Internet]. MMWR. Morbidity and mortality weekly report. 2005 [cited 2014 May 1]. Available from: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=emed7&NEWS=N&AN=16079738

169. Shea KM, Kammerer JS, Winston CA, Navin TR, C. Robert Horsburgh. Estimated Rate of Reactivation of Latent Tuberculosis Infection in the United States, Overall and by Population Subgroup. Am J Epidemiol. 2014 Jan 15;179(2):216–25.

170. Walter ND, Painter J, Parker M, Lowenthal P, Flood J, Fu Y, et al. Persistent latent tuberculosis reactivation risk in United States immigrants. Am J Respir Crit Care Med. 2014;189(1):88–95.

171. Mears J, Vynnycky E, Lord J, Borgdorff M, Cohen T, Abubakar I, et al. Evaluation of the Tuberculosis Strain Typing Service (TB-STS) in England. The Lancet. 2013 Nov;382:S73.

172. Jagielski T, van Ingen J, Rastogi N, Dziadek J, aw, Mazur P, et al. Current Methods in the Molecular Typing of *Mycobacterium tuberculosis* and Other Mycobacteria. BioMed Res Int. 2014 Jan 5;2014:e645802.

173. De Vries G, Baars HWM, Sebek MMGG, van Hest NAH, Richardus JH. Transmission classification model to determine place and time of infection of tuberculosis cases in an urban area. J Clin Microbiol. 2008 Dec;46(12):3924–30.

174. Borrell S, Español M, Orcau A, Tudó G, March F, Caylà JA, et al. Tuberculosis transmission patterns among Spanish-born and foreign-born populations in the city of Barcelona. Clin Microbiol Infect Off Publ Eur Soc Clin Microbiol Infect Dis. 2010 Jun;16(6):568–74.

175. Reed MB, Pichler VK, McIntosh F, Mattia A, Fallow A, Masala S, et al. Major Mycobacterium tuberculosis Lineages Associate with Patient Country of Origin. J Clin Microbiol. 2009 Apr;47(4):1119–28.

176. Cattamanchi A, Hopewell PC, Gonzalez LC, Osmond DH, Masae Kawamura L, Daley CL, et al. A 13-year molecular epidemiological analysis of tuberculosis in San Francisco. Int J Tuberc Lung Dis Off J Int Union Tuberc Lung Dis. 2006 Mar;10(3):297–304.

177. Supply P, Allix C, Lesjean S, Cardoso-Oelemann M, Rüsch-Gerdes S, Willery E, et al. Proposal for Standardization of Optimized Mycobacterial Interspersed Repetitive Unit-Variable-Number Tandem Repeat Typing of Mycobacterium tuberculosis. J Clin Microbiol. 2006 Dec 1;44(12):4498–510.

178. Brown TJ, Nikolayevskyy VN, Drobniewski FA. Typing Mycobacterium tuberculosis using variable number tandem repeat analysis. Methods Mol Biol Clifton NJ. 2009;465:371–94.

179. Gibson A, Brown T, Baker L, Drobniewski F. Can 15-locus mycobacterial interspersed repetitive unit-variable-number tandem repeat analysis provide insight into the evolution of Mycobacterium tuberculosis? Appl Environ Microbiol. 2005 Dec;71(12):8207–13.

180. Public Health England. TB Strain Typing and Cluster Investigation Handbook [Internet]. London, UK: Public Health England; 2014 Feb [cited 2015 Feb 17]. Available from: http://webarchive.nationalarchives.gov.uk/20140714084352/http:/www.hpa.org.uk/webc/HPAwebFile/HPAweb_C/1317140774833

181. ONS. Provisional Long-Term International Migration (LTIM) estimates, Year Ending September 2014 [Internet]. Office for National Statistics. 2010 [cited 2015 Apr 5]. Available from: http://www.ons.gov.uk/ons/taxonomy/index.html?nscl=Long-term+Migrants#tab-data-tables

182. ONS. Flows by Main Reason for Migration, Mid 2004 to Mid 2012 (STIM.01) [Internet]. 2010 [cited 2015 Apr 5]. Available from: http://www.ons.gov.uk/ons/publications/re-reference-tables.html?edition=tcm%3A77-338675

183. Gardam MA, Keystone EC, Menzies R, Manners S, Skamene E, Long R, et al. Anti-tumour necrosis factor agents and tuberculosis risk: mechanisms of action and clinical management. Lancet Infect Dis. 2003 Mar;3(3):148–55.

184. Volmink J, Woldehanna S. Treatment of latent tuberculosis infection in HIV infected persons. In: The Cochrane Collaboration, Woldehanna S, editors. Cochrane Database of Systematic Reviews [Internet]. Chichester, UK: John Wiley & Sons, Ltd; 2004 [cited 2011 Jul 2]. Available from: http://onlinelibrary.wiley.com/o/cochrane/clsysrev/articles/rel0002/CD000171/abstract.html

185. Horsburgh CR, O'Donnell M, Chamblee S, Moreland JL, Johnson J, Marsh BJ, et al. Revisiting rates of reactivation tuberculosis: a population-based approach. Am J Respir Crit Care Med. 2010 Aug 1;182(3):420–5.

186. Walker TM, Ip CLC, Harrell RH, Evans JT, Kapatai G, Dedicoat MJ, et al. Whole-genome sequencing to delineate Mycobacterium tuberculosis outbreaks: a retrospective observational study. Lancet Infect Dis. 2013 Feb;13(2):137–46.

187. Walker TM, Lalor MK, Broda A, Ortega LS, Morgan M, Parker L, et al. Assessment of Mycobacterium tuberculosis transmission in Oxfordshire, UK, 2007–12, with whole pathogen genome sequences: an observational study. Lancet Respir Med. 2014 Apr;2(4):285–92.

188. Gardy JL, Johnston JC, Ho Sui SJ, Cook VJ, Shah L, Brodkin E, et al. Whole-genome sequencing and social-network analysis of a tuberculosis outbreak. N Engl J Med. 2011 Feb 24;364(8):730–9.

189. Roetzer A, Diel R, Kohl TA, Rückert C, Nübel U, Blom J, et al. Whole genome sequencing versus traditional genotyping for investigation of a Mycobacterium tuberculosis outbreak: a longitudinal molecular epidemiological study. PLoS Med. 2013;10(2):e1001387.

190. UK plans pre-visa TB screening. BBC [Internet]. 2012 May 21 [cited 2015 Mar 2]; Available from: http://www.bbc.co.uk/news/uk-politics-18149713

191. Schwartzman K, Oxlade O, Barr RG, Grimard F, Acosta I, Baez J, et al. Domestic returns from investment in the control of tuberculosis in other countries. N Engl J Med. 2005 Sep 8;353(10):1008–20.

192. Bloom BR, Salomon JA. Enlightened self-interest and the control of tuberculosis. N Engl J Med. 2005 Sep 8;353(10):1057–9.

193. Piccini P, Chiappini E, Tortoli E, Martino M de, Galli L. Clinical peculiarities of tuberculosis. BMC Infect Dis. 2014 Jan 8;14(Suppl 1):S4.

194. Marais BJ, Gie RP, Schaaf HS, Beyers N, Donald PR, Starke JR. Childhood pulmonary tuberculosis: old wisdom and new challenges. Am J Respir Crit Care Med. 2006 May 15;173(10):1078–90.

195. Singh M, Mynak ML, Kumar L, Mathew JL, Jindal SK. Prevalence and risk factors for transmission of infection among children in household contact with adults having pulmonary tuberculosis. Arch Dis Child. 2005 Jun;90(6):624–8.

196. Kunst H, Khan KS. Age-related risk of hepatotoxicity in the treatment of latent tuberculosis infection: a systematic review. Int J Tuberc Lung Dis Off J Int Union Tuberc Lung Dis. 2010 Nov;14(11):1374–81.

APPENDICES

APPENDIX 1

Supporting information for chapter 2: Detailed search terms used in systematic review

1. Migrant$.mp. [mp=title, abstract, subject headings, heading word, drug trade name, original title, device manufacturer, drug manufacturer, device trade name, keyword]
2. exp migration/
3. "Transients and Migrants"/
4. Expatriate$.mp. [mp=title, abstract, subject headings, heading word, drug trade name, original title, device manufacturer, drug manufacturer, device trade name, keyword]
5. Refugee$.mp. [mp=title, abstract, subject headings, heading word, drug trade name, original title, device manufacturer, drug manufacturer, device trade name, keyword]
6. exp Refugees/
7. Departee$.mp. [mp=title, abstract, subject headings, heading word, drug trade name, original title, device manufacturer, drug manufacturer, device trade name, keyword]
8. Emigrant$.mp. [mp=title, abstract, subject headings, heading word, drug trade name, original title, device manufacturer, drug manufacturer, device trade name, keyword]
9. exp immigrant/
10. Immigrant$.mp. [mp=title, abstract, subject headings, heading word, drug trade name, original title, device manufacturer, drug manufacturer, device trade name, keyword]
11. "Emigrants and Immigrants"/
12. Asylum seeker$.mp. [mp=title, abstract, subject headings, heading word, drug trade name, original title, device manufacturer, drug manufacturer, device trade name, keyword]
13. exp refugee/
14. Asylum.mp. [mp=title, abstract, subject headings, heading word, drug trade name, original title, device manufacturer, drug manufacturer, device trade name, keyword]
15. Foreign-born.mp. [mp=title, abstract, subject headings, heading word, drug trade name, original title, device manufacturer, drug manufacturer, device trade name, keyword]
16. exp foreign worker/
17. exp foreign student/
18. Entrant$.mp. [mp=title, abstract, subject headings, heading word, drug trade name, original title, device manufacturer, drug manufacturer, device trade name, keyword]
19. 1 or 2 or 3 or 4 or 5 or 6 or 7 or 8 or 9 or 10 or 11 or 12 or 13 or 14 or 15 or 16 or 17 or 18
20. Pre*entry screening.mp. [mp=title, abstract, subject headings, heading word, drug trade name, original title, device manufacturer, drug manufacturer, device trade name, keyword]

21. Pre$entry screening.mp. [mp=title, abstract, subject headings, heading word, drug trade name, original title, device manufacturer, drug manufacturer, device trade name, keyword]
22. Preentry screening.mp. [mp=title, abstract, subject headings, heading word, drug trade name, original title, device manufacturer, drug manufacturer, device trade name, keyword]
23. Pre?entry screening.mp. [mp=title, abstract, subject headings, heading word, drug trade name, original title, device manufacturer, drug manufacturer, device trade name, keyword]
24. Pre-entry screening.mp. [mp=title, abstract, subject headings, heading word, drug trade name, original title, device manufacturer, drug manufacturer, device trade name, keyword]
25. Pre*entry.mp. [mp=title, abstract, subject headings, heading word, drug trade name, original title, device manufacturer, drug manufacturer, device trade name, keyword]
26. Pre$entry.mp. [mp=title, abstract, subject headings, heading word, drug trade name, original title, device manufacturer, drug manufacturer, device trade name, keyword]
27. Preentry.mp. [mp=title, abstract, subject headings, heading word, drug trade name, original title, device manufacturer, drug manufacturer, device trade name, keyword]
28. Pre*entry.mp. [mp=title, abstract, subject headings, heading word, drug trade name, original title, device manufacturer, drug manufacturer, device trade name, keyword]
29. Pre-entry.mp. [mp=title, abstract, subject headings, heading word, drug trade name, original title, device manufacturer, drug manufacturer, device trade name, keyword]
30. Pre*screening.mp. [mp=title, abstract, subject headings, heading word, drug trade name, original title, device manufacturer, drug manufacturer, device trade name, keyword]
31. Pre$screening.mp. [mp=title, abstract, subject headings, heading word, drug trade name, original title, device manufacturer, drug manufacturer, device trade name, keyword]
32. Prescreening.mp. [mp=title, abstract, subject headings, heading word, drug trade name, original title, device manufacturer, drug manufacturer, device trade name, keyword]
33. Pre?screening.mp. [mp=title, abstract, subject headings, heading word, drug trade name, original title, device manufacturer, drug manufacturer, device trade name, keyword]
34. Pre-screening.mp. [mp=title, abstract, subject headings, heading word, drug trade name, original title, device manufacturer, drug manufacturer, device trade name, keyword]
35. Pre*immigration.mp. [mp=title, abstract, subject headings, heading word, drug trade name, original title, device manufacturer, drug manufacturer, device trade name, keyword]
36. pre$immigration.mp. [mp=title, abstract, subject headings, heading word, drug trade name, original title, device manufacturer, drug manufacturer, device trade name, keyword]
37. preimmigration.mp. [mp=title, abstract, subject headings, heading word, drug trade name, original title, device manufacturer, drug manufacturer, device trade name, keyword]
38. pre?immigration.mp. [mp=title, abstract, subject headings, heading word, drug trade name, original title, device manufacturer, drug manufacturer, device trade name, keyword]
39. pre-immigration.mp. [mp=title, abstract, subject headings, heading word, drug trade name, original title, device manufacturer, drug manufacturer, device trade name, keyword]
40. oversea$.mp. [mp=title, abstract, subject headings, heading word, drug trade name, original title, device manufacturer, drug manufacturer, device trade name, keyword]

41. Screening.mp. [mp=title, abstract, subject headings, heading word, drug trade name, original title, device manufacturer, drug manufacturer, device trade name, keyword]
42. exp mass screening/
43. Mass Screening.mp. [mp=title, abstract, subject headings, heading word, drug trade name, original title, device manufacturer, drug manufacturer, device trade name, keyword]
44. Screen$.mp. [mp=title, abstract, subject headings, heading word, drug trade name, original title, device manufacturer, drug manufacturer, device trade name, keyword]
45. 20 or 21 or 22 or 23 or 24 or 25 or 26 or 28 or 29 or 30 or 31 or 32 or 33 or 34 or 35 or 36 or 37 or 38 or 39 or 40 or 41 or 42 or 43 or 44
46. exp Latent Tuberculosis/
47. exp Tuberculosis/
48. exp Mycobacterium tuberculosis/
49. Tuberculosis.mp. [mp=title, abstract, subject headings, heading word, drug trade name, original title, device manufacturer, drug manufacturer, device trade name, keyword]
50. Tuberculosis, Pulmonary/
51. TB.mp. [mp=title, abstract, subject headings, heading word, drug trade name, original title, device manufacturer, drug manufacturer, device trade name, keyword]
52. 46 or 47 or 48 or 49 or 50 or 51
53. 19 and 45 and 52

APPENDIX 2

Supporting information for chapter 2: Detailed characteristics of the studies included in review

Study	Bollini(88)
Population Screened	Migrants
Method of screening	X-ray. If compatible with tuberculosis, sputum smear samples were taken on three consecutive days
Principal case definition	One or more positive sample by sputum smear.
Country of Origin	Vietnam
Country where screening took place	Vietnam
Years screened	1992-94
Notes	Case definition extrapolated from study by Keane, which reports data from same IOM screening locations. Single smear positive finding considered positive.

Study	Das Gupta(90)
Population Screened	Migrants
Method of screening	Initially a chest radiograph for active TB. If radiographic abnormalities are detected, the affected individual is referred, usually to a chest specialist, for further evaluation, including additional radiographic, microbiologic (e.g., sputum acid-fast bacilli), and tuberculin tests when judged appropriate.

Principal case definition	The presence of cultures positive for *Mycobacterium tuberculosis* or radiographic improvement after at least 2 months of therapy for active disease.
Country of Origin	Multiple
Country where screening took place	Multiple
Years screened	1996-97
Notes	Primary aim of the study was to examine "the efficiency and cost-effectiveness of immigrant applicant screening and of surveillance of newly arrived immigrants were compared with investigation of close contacts active cases of TB, and all three methods were com- pared with a policy of passive case detection". Immigrant applicant screening was a pre-entry screening programme.

Study	Gobacheva(89)
Population Screened	Refugees
Method of screening	X-ray, clinical examination, history and TST. Three sputum specimens in those with findings suggestive of tuberculosis
Principal case definition	One or more positive sample by sputum smear and/or culture.
Country of Origin	Bhutan
Country where screening took place	Nepal
Years screened	2007-09
Notes	Data on cases found were calculated from prevalence rates as raw numbers not provided. Exact criteria for selecting patients for microbiological investigation not specified "Suspected cases were referred for microbiological examination of three sputum sample by both acid-fast bacilli and liquid culture for TB". Screening was conducted for several receiving countries (USA, Canada, Australia, New Zeland, Denmark and Norway) and exact protocol is therefore likely to vary as per the country technical instructions.

Study	King(91)
Population Screened	Migrants
Method of screening	X-ray. If compatible with tuberculosis, sputum smear and culture testing.
Principal case definition	Clinical cases or one or more positive sample by sputum smear and/or culture.
Country of Origin	Multiple
Country where screening took place	Multiple
Years screened	2009-10
Notes	A positive case was defined on the basis of one or more sputum smear or culture results. Limitations with culture testing data were acknowledged by study authors as this was not uniformly performed across all sites for all cases and not available for all individuals.

Study	Lange(92)
Population Screened	Adoptees aged between 2.5 months and 12 years, with median age of 4 months.
Method of screening	Intradermal injection of 5 tuberculin units of purified protein derivative.
Principal case definition	Latent infection defined as those with 10mm induration after PPD.
Country of Origin	South Korea
Country where screening took place	South Korea
Years screened	1985-88
Notes	Analysis based on data from a case note review, not prospective collection.

Study	Lui(93)
Population Screened	Refugees and Migrants
Method of screening	Based on 1991 Technical Instructions for Panel Physicians: X-ray, clinical examination, and if compatible with tuberculosis, sputum smear samples taken on three consecutive days.
Principal case definition	The principal case definitions for this study were smear-negative and inactive tuberculosis. Smear-negative tuberculosis was defined as: "if the chest radiograph was suggestive of active tuberculosis and sputum smears were negative for acid-fast bacilli on 3 consecutive days". Inactive tuberculosis was defined as: "if the chest radiograph was suggestive of tuberculosis that was not clinically active (e.g., showing fibro- sis, scarring, pleural thickening, diaphragmatic tenting, or blunting of costophrenic angles)". For this systematic review we considered smear-negative tuberculosis as the principal outcome.
Country of Origin	Multiple
Country where screening took place	Multiple
Years screened	1999-2005
Notes	No mycobacterial cultures were obtained during the study period and data for smear positive tuberculosis cases were not presented in the manuscript.

Study	Malone(94)
Population Screened	Migrants
Method of screening	X-ray and physical examination. If compatible with tuberculosis, sputum smear and culture testing on three consecutive samples
Principal case definition	Presumptive active tuberculosis (definition not provided and authors not contactable)
Country of Origin	Haiti
Country where screening took place	U.S. Naval Base in Guantanamo Bay, Cuba
Years screened	1991-93

Notes	Authors report that denominator data for population screened is approximate. Three consecutive sputum specimens were obtained in the morning; these specimens were treated with Kinyoun carbolfuchsin stain and sent to the laboratory at the Naval Hospital in Portsmouth, Virginia. At this laboratory the specimens were digested with sodium hydroxide and then inoculated onto Lowenstein-Jensen culture media and Middlebrook 7HIO agar.

Study	Maloney(95)
Population Screened	Migrants
Method of screening	X-ray. If compatible with tuberculosis, sputum smear and culture testing on three consecutive samples
Principal case definition	"Participants with 1 or more AFB-positive smear results were designated as AFB smear positive, and those with 3 AFB-negative smear results were designated as AFB smear negative. Any participant with at least 1 M tuberculosis–positive culture result was designated as M tuberculosis culture positive and was therefore determined to have PTB. Participants with no M tuberculosis–positive culture results were divided into 4 separate categories: (1) those with 3 negative culture results were designated as being M tuberculosis culture negative; (2) those with 3 contaminated culture results were designated as having contaminated cultures; (3) those with 3 M avium-intracellulare complex– positive results were designated as being M avium-intracellulare complex culture positive; and (4) those with at least 1 but fewer than 3 negative culture results were designated as being M tuberculosis culture indeterminate (i.e., the 2 additional culture results were some combination of negative, contaminated, or M avium-intracellulare complex)."
Country of Origin	Vietnam
Country where screening took place	Vietnam
Years screened	1998-99
Notes	Cases only reported for individuals over the age of 18 whereas all individuals over the age of 15 screened by X-ray. Screening was compulsory, but participation in this study (with reporting of outcomes) was voluntary, with a 95.3% participation rate.

Study	Mor(96)

Population Screened	Migrants
Method of screening	X-ray, clinical examination, history and TST. Three sputum specimens in those with findings suggestive of tuberculosis.
Principal case definition	Active pulmonary TB (Symptomatic patient with pulmonary disease and confirmed Mycobacterium tuberculosis complex culture)
Country of Origin	Ethiopia
Country where screening took place	Ethiopia
Years screened	2001-05
Notes	The primary aim of this study was: "to determine the validity of CXR screening in detecting radiological findings compatible with active PTB or with old healed tuberculosis".

Study	Oeltman(168)
Population Screened	Refugees
Method of screening	X-ray, clinical examination, history. Three sputum specimens in those with findings suggestive of tuberculosis
Principal case definition	Clinical and, or AFB sputum smear positive cases.
Country of Origin	Lao People's Democratic Republic
Country where screening took place	Thailand
Years screened	2004-05
Notes	The primary aim of this study was to present data from an outbreak of MDR TB among US-bound Hmong Refugees and the results of enhanced screening and treatment protocol among this population after initial detection of the outbreak. Latent TB detected using tuberculin skin tests (TSTs) with induration >5 mm considered a positive test result.

Study	Painter(98)
Population Screened	Migrants
Method of screening	Migrants were screened according to 2009 technical instructions published by CDC: X-ray, clinical examination, and if compatible with tuberculosis, sputum smear samples taken on three consecutive days. "Following the results of chest radiograph applicants were invited to participate in a study of TST and QFT for which they would be provided the results, but the result of which would not affect their visa application."
Principal case definition	QuantiFERON ®-TB Gold In-Tube Assay (QFT) and TST (varying size) positive cases.
Country of Origin	Vietnam
Country where screening took place	Vietnam
Years screened	2008-10
Notes	The primary aims of this study were to: 1. Compare the sensitivity of QuantiFERON ®-TB Gold In-Tube Assay (QFT) and TST for culture-positive pulmonary TB. 2. Compare the age-specific and overall prevalence of positive TST and QFT among applicants with normal and abnormal CXR. The study also reports data on 20,100 visa applicants screened over the age of 15 as part of their visa medical exam and that 211 had culture-confirmed pulmonary tuberculosis.

Study	Plant(99)
Population Screened	Migrants
Method of screening	X-ray, clinical examination and history. Three sputum specimens in those with findings suggestive of tuberculosis
Case definition	AFB sputum smear and/or culture positive cases.

Country of Origin	Vietnam
Country where screening took place	Vietnam
Years screened	1997-01
Notes	Screening was compulsory but participation in this study was voluntary (with reporting of outcomes).

Study	Wang(100)
Population Screened	Migrants
Method of screening	X-ray followed by three sputum cultures in those with findings suggestive of tuberculosis
Case definition	Inactive tuberculosis defined by authors as: "radiograph shows evidence of tuberculosis, it is repeated at a minimum interval of 3 months to confirm stability of the lesion. In addition, 3 sputum cultures, incubated for 7-8 weeks, taken at least 24h apart, are required to be negative."
Country of Origin	Multiple
Country where screening took place	Multiple
Years screened	1982-85
Notes	Data reported for migrants from China, Hong Kong, Taiwan, Korea, India, Philippines and Japan.

Study	Watkins(101)
Population Screened	Migrants

Method of screening	X-ray
Principal case definition	X-ray positive cases
Country of Origin	Vietnam
Country where screening took place	Vietnam
Years screened	Not stated
Notes	Primary focus of study was on subjective and clinical indicators of health status, not tuberculosis screening, therefore data were extracted from secondary analyses within the study.

Study	Yanni(102)
Population Screened	Refugees
Method of screening	Corresponding author confirmed that the study used CDC Division of Global Migration and Quarantine 2007 TB Screening guidance: X-ray, clinical examination, history and sputum testing for M. tuberculosis.
Principal case definition	One or more positive sample by sputum smear and/or culture.
Country of Origin	Iraq
Country where screening took place	Jordan
Years screened	2007-09
Notes	Primary aim of study was to provide a health profile of Iraqi refugees therefore data were extracted from secondary analyses presented in this study. Latent TB cases were defined as: “Positive TST C 10 mm with normal chest X-ray, and negative smears and culture”

APPENDIX 3

Supporting information for chapter 2: Additional Forrest plots performed as part of the meta-analysis

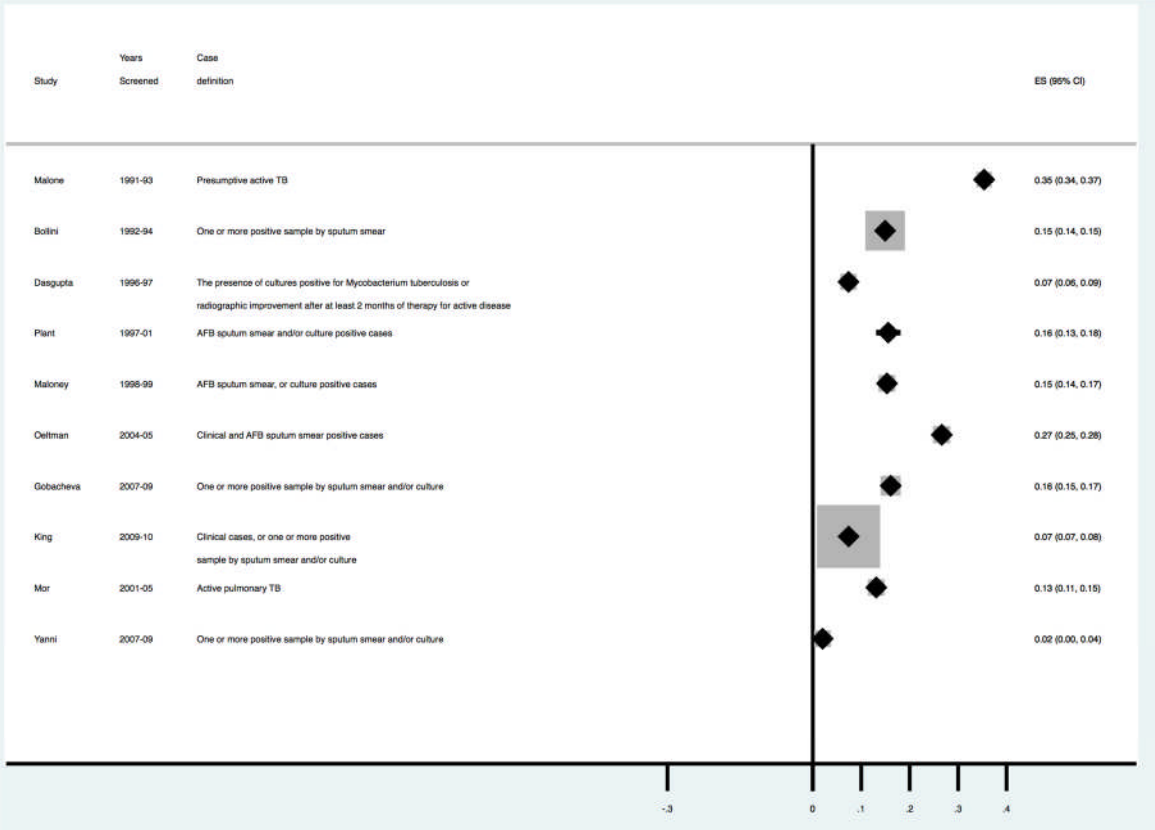

Figure 39. Forrest plot of yield for principal outcome of active tuberculosis cases found by each study (case definition varies between studies, sorted by year of publication).

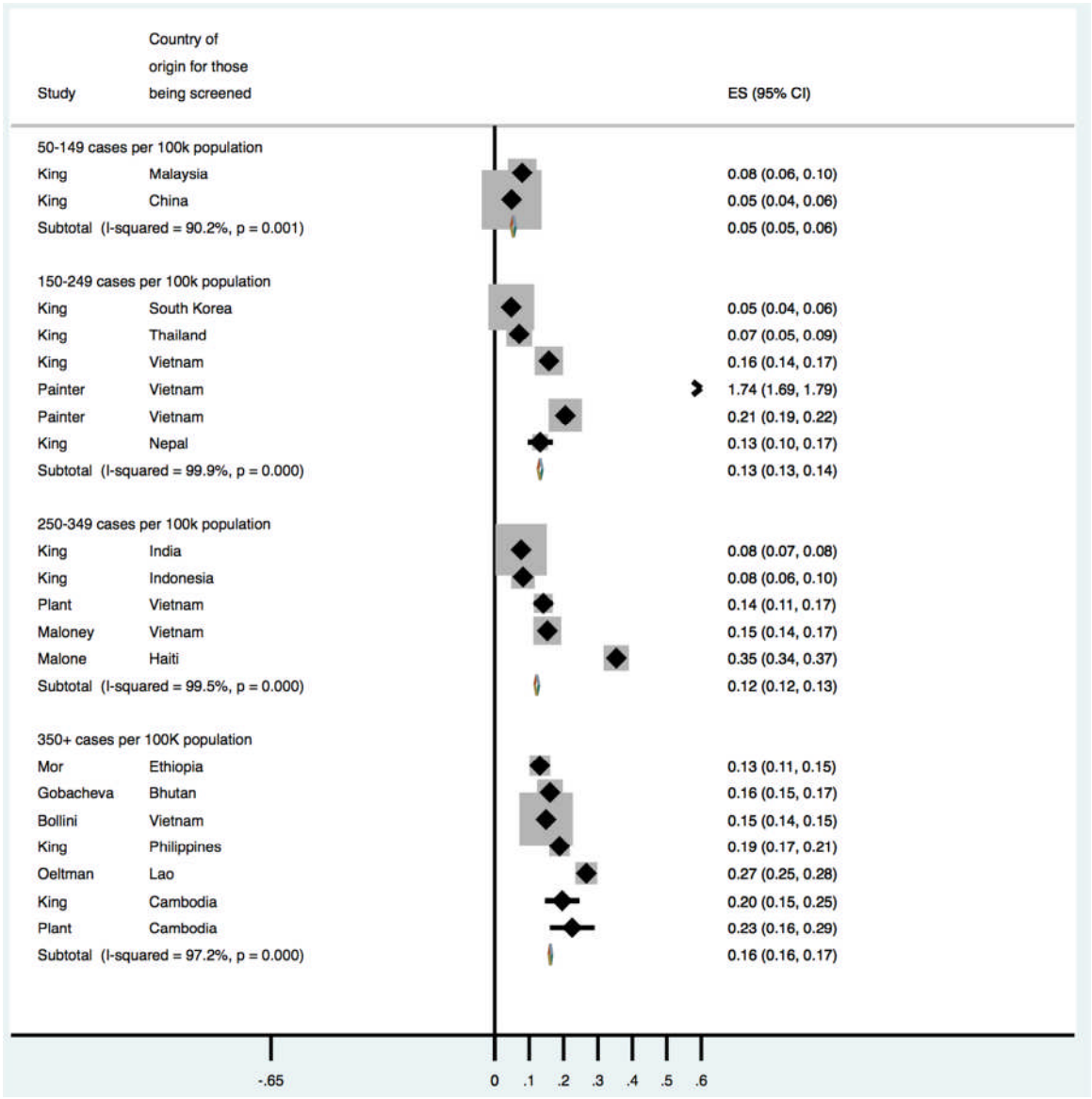

Figure 40. Forrest plot of yield for principal outcome of active tuberculosis cases found by each study (case definition varies between studies, sorted by year of publication) stratified by prevalence in the country of origin (Freeman-Tukey transformed data).

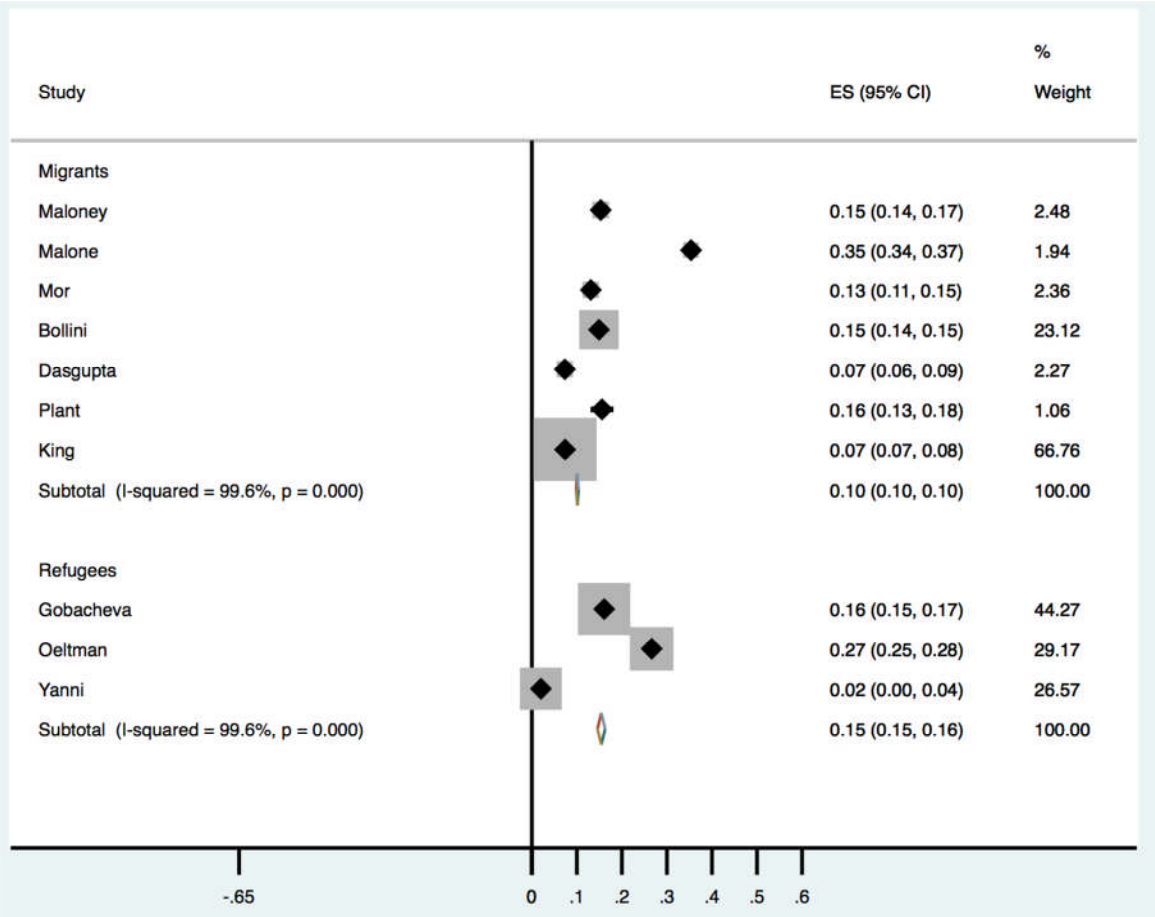

Figure 41. Forrest plot of yield for principal outcome of active tuberculosis cases found by each study (case definition varies between studies, sorted by year of publication) stratified by population screened (Freeman-Tukey transformed data).

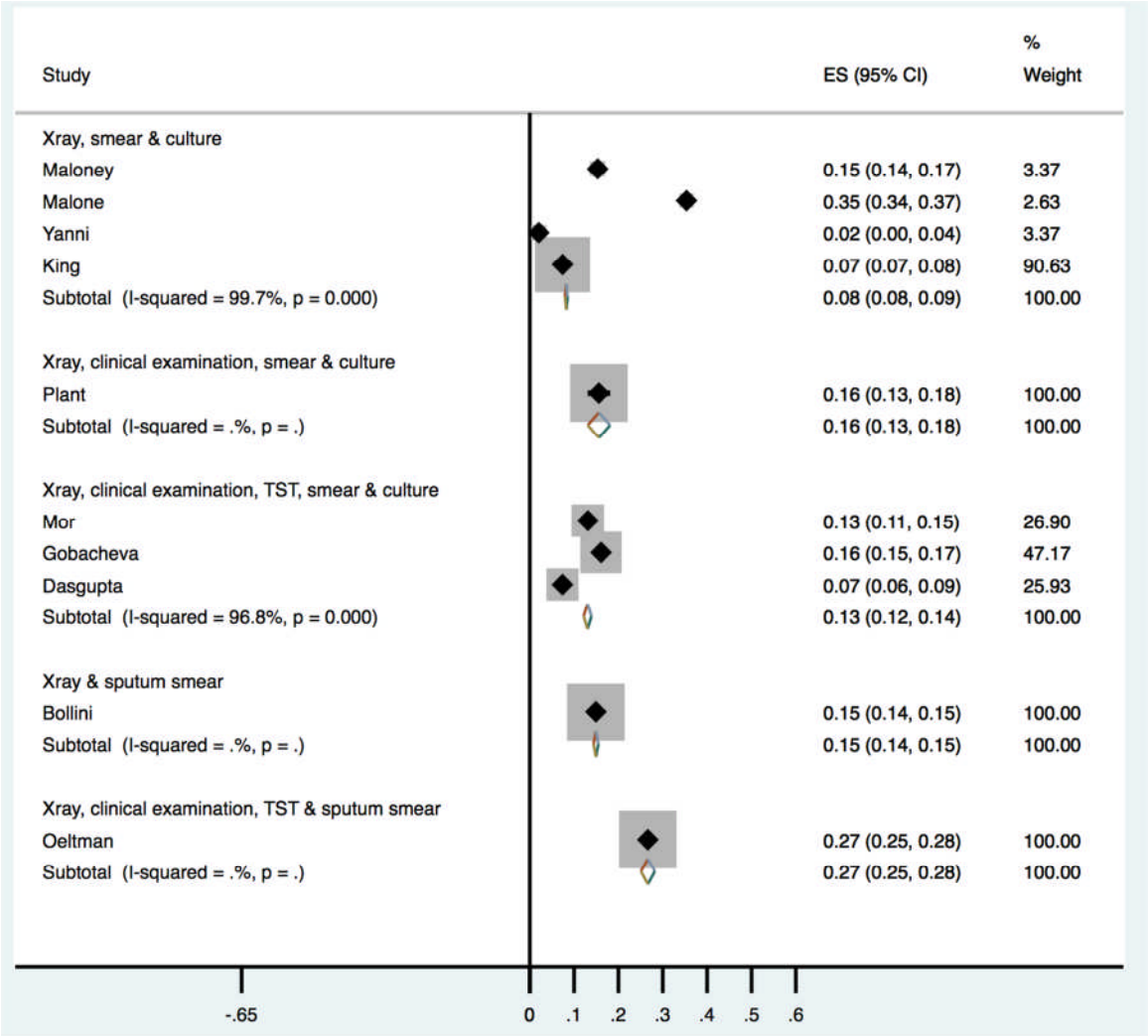

Figure 42. Forrest plot of yield for principal outcome of active tuberculosis cases found by each study (case definition varies between studies, sorted by year of publication) stratified by screening method (Freeman-Tukey transformed data)

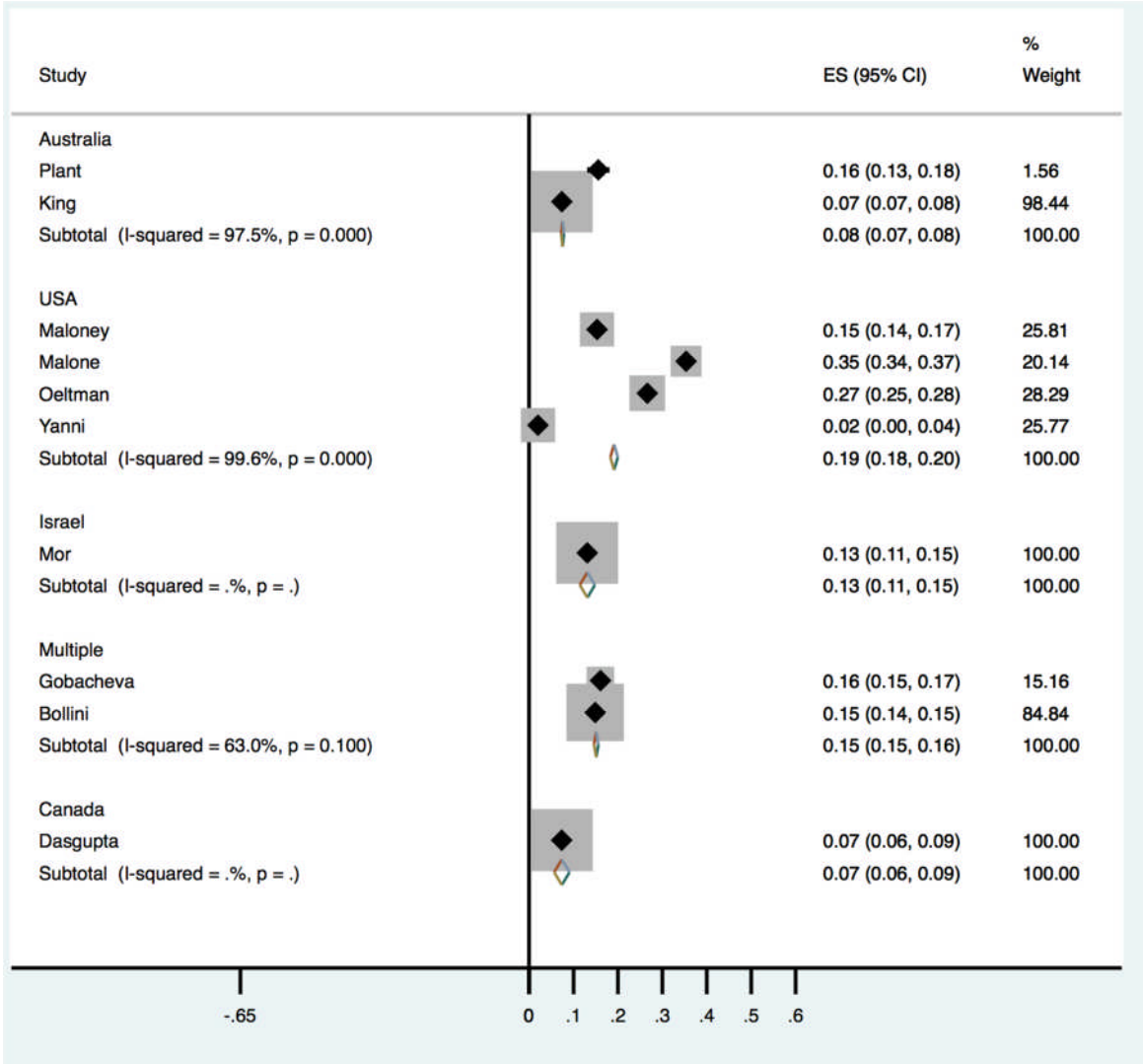

Figure 43. Forrest plot of yield for principal outcome of active tuberculosis cases found by each study (case definition varies between studies, sorted by year of publication) stratified by receiving country (Freeman-Tukey transformed data).

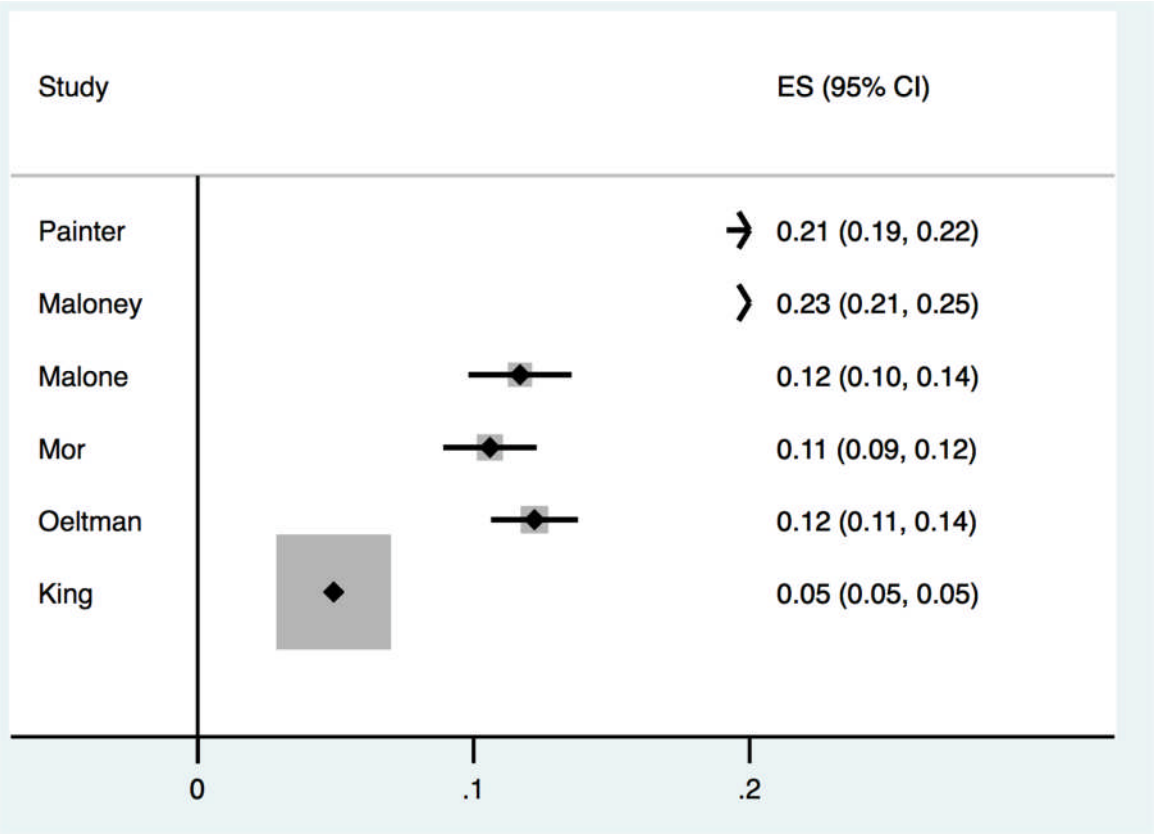

Figure 44. Forrest plot of yield for culture confirmed cases found by each study (Freeman-Tukey transformed data).

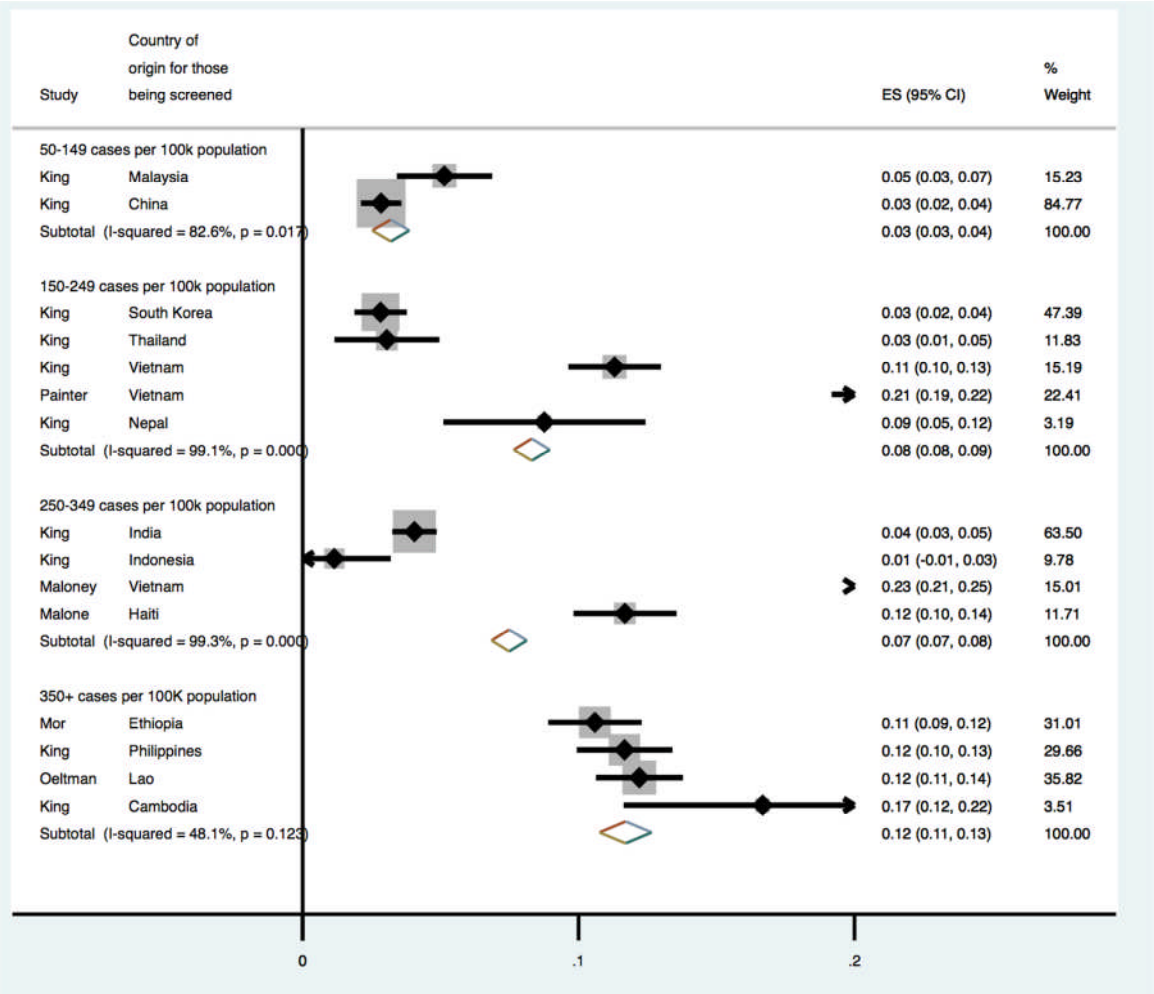

Figure 45. Forrest plot of yield for culture confirmed cases found by each study, stratified by prevalence in the country of origin (Freeman-Tukey transformed data).

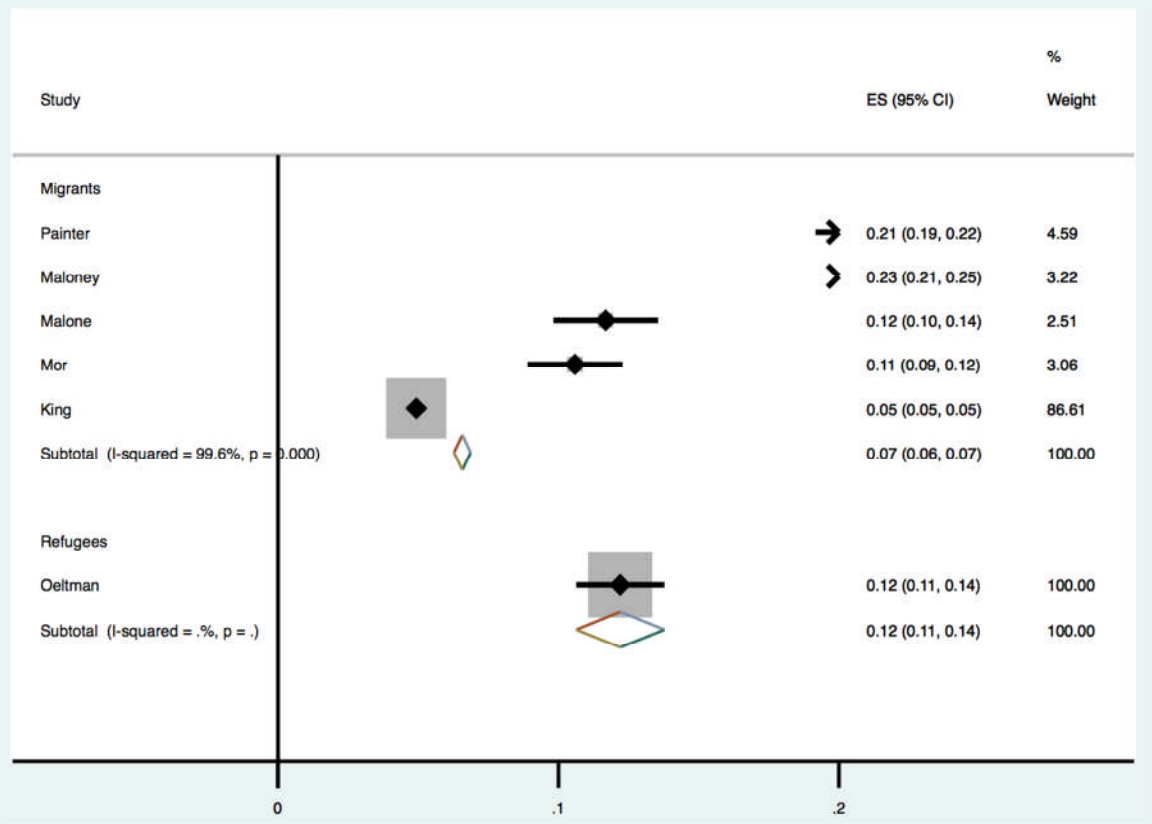

Figure 46. Forrest plot of yield for culture confirmed cases found by each study, stratified by population screened (Freeman-Tukey transformed data).

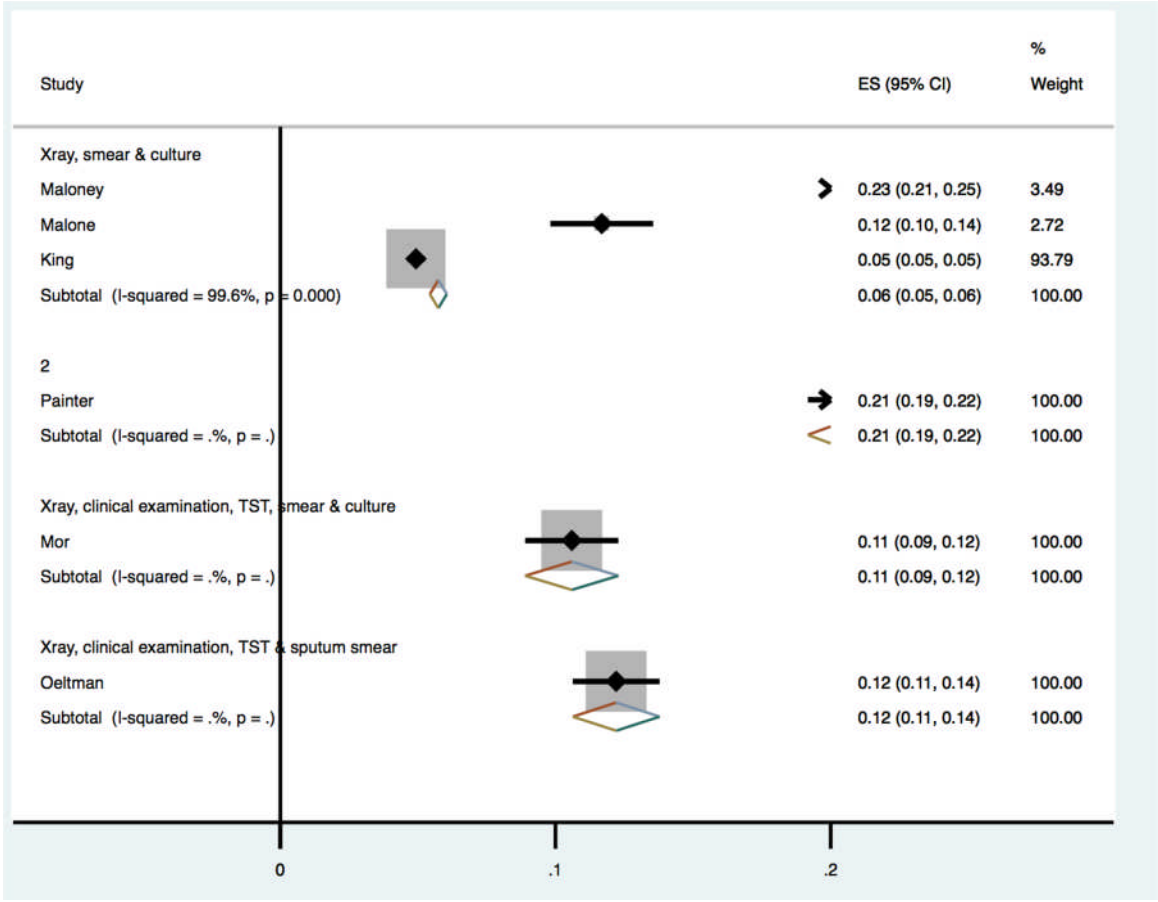

Figure 47. Forrest plot of yield for culture confirmed cases found by each study, stratified by screening method (Freeman-Tukey transformed data).

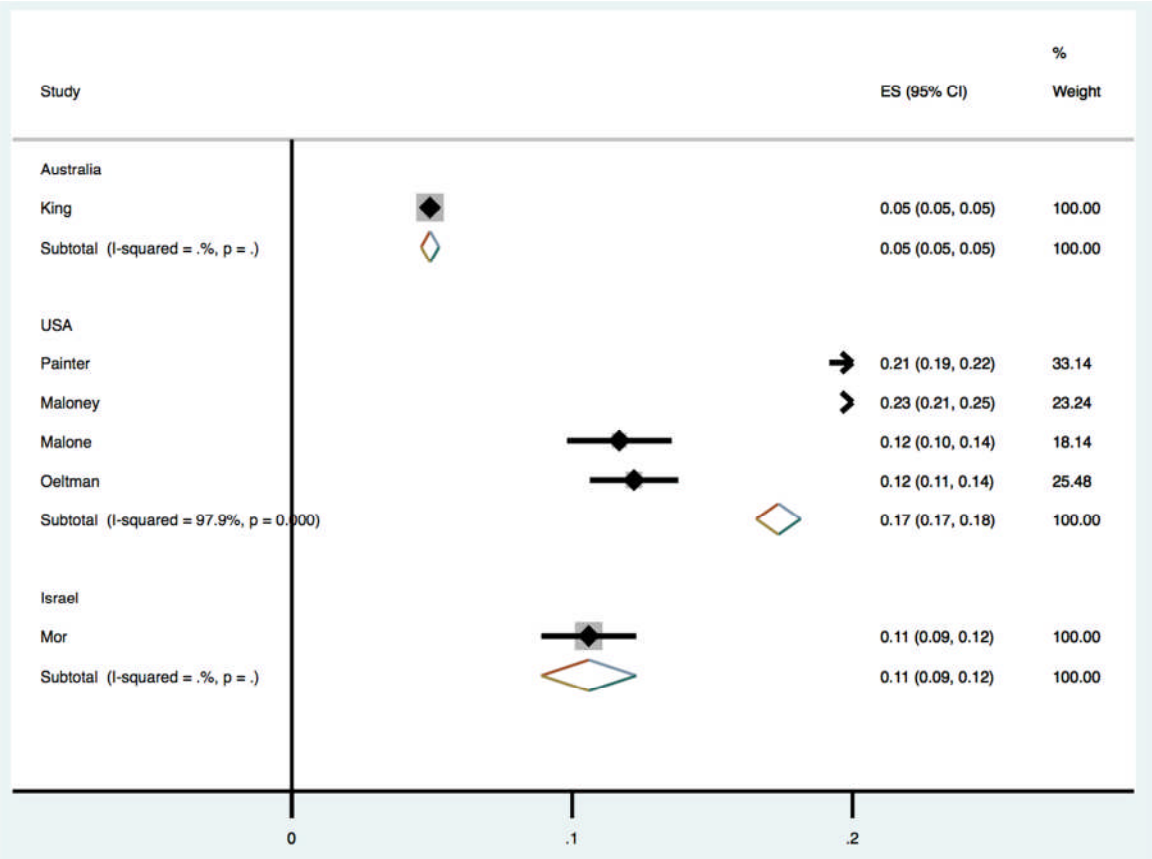

Figure 48. Forrest plot of yield for culture confirmed cases found by each study, stratified by receiving country (Freeman-Tukey transformed data).

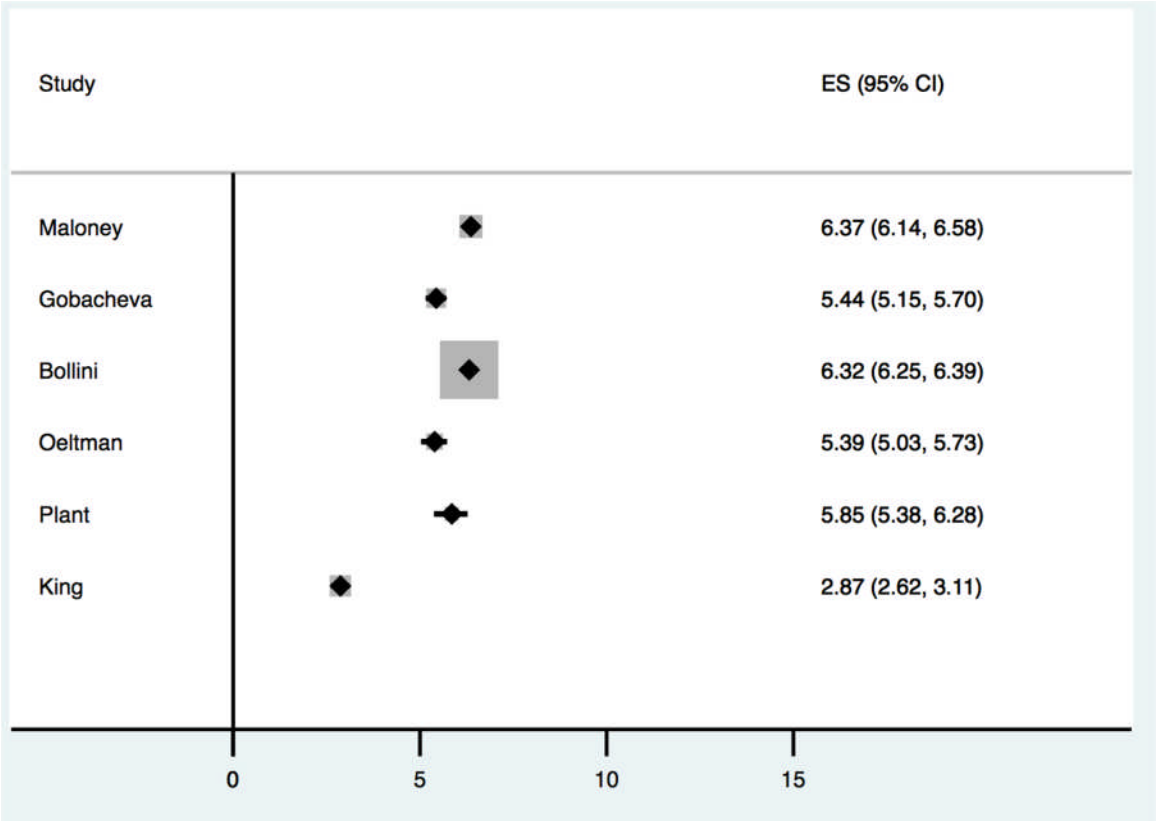

Figure 49. Forrest plot of yield for AFB confirmed cases found by each study (Freeman-Tukey transformed data).

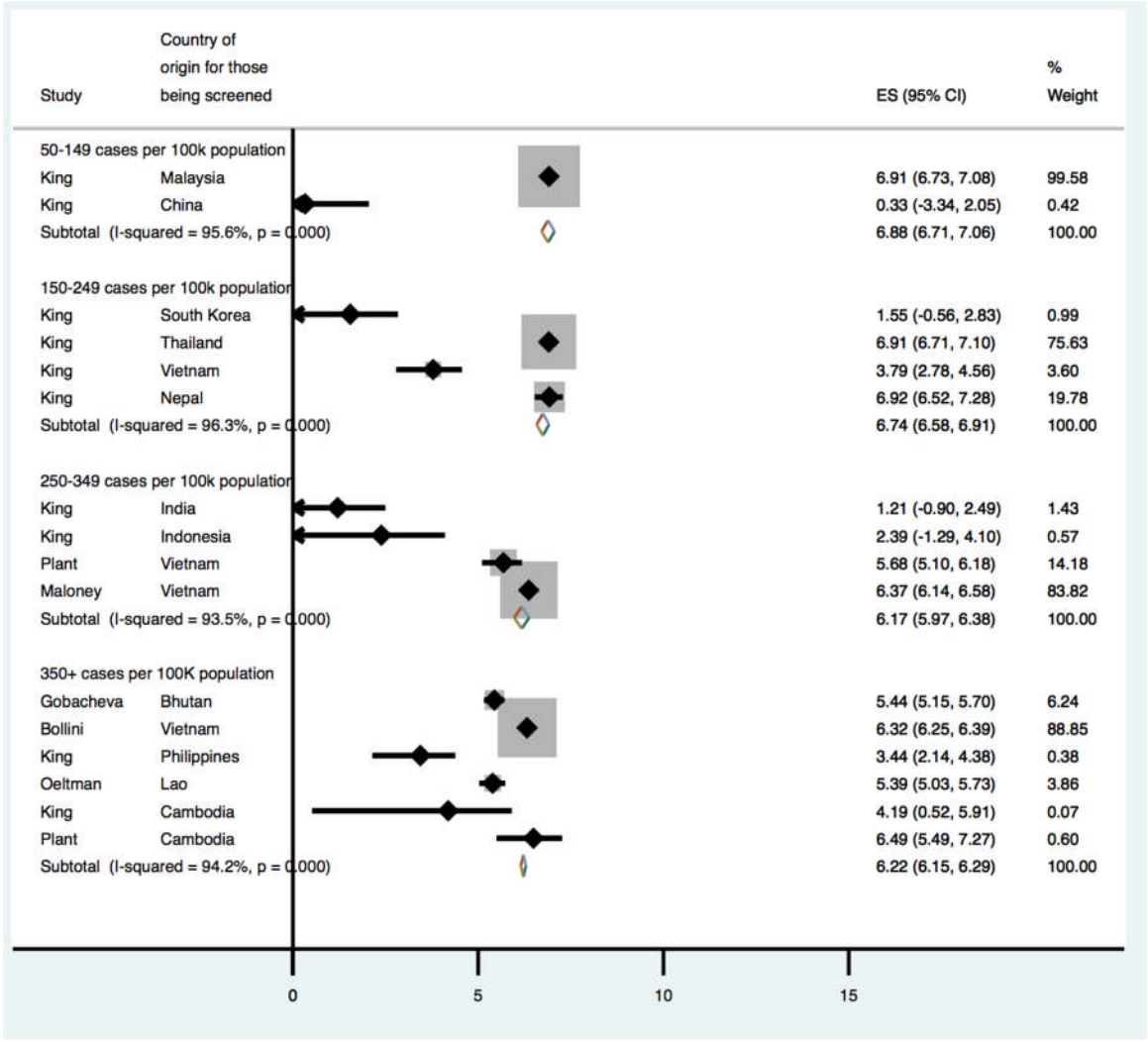

Figure 50. Forrest plot of yield for AFB confirmed cases found by each study, stratified by prevalence in the country of origin (Freeman-Tukey transformed data).

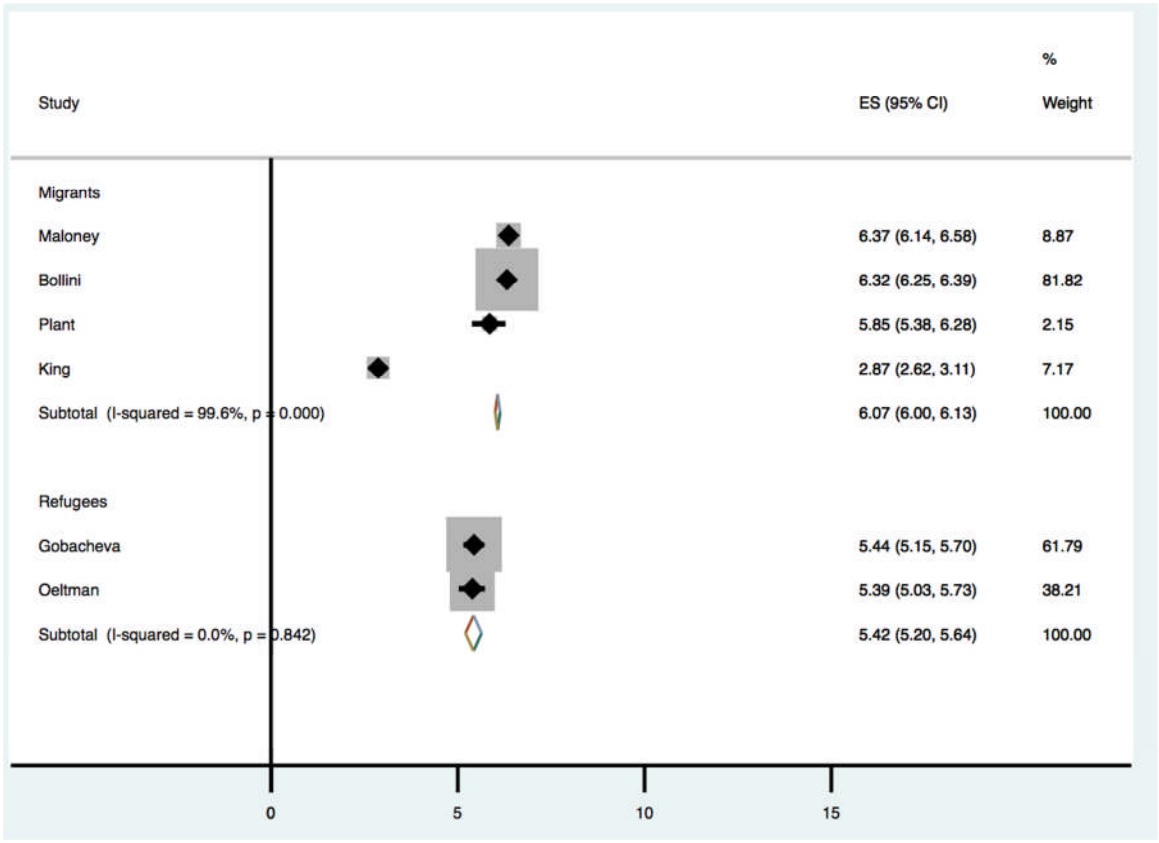

Figure 51. Forrest plot of yield for AFB confirmed cases found by each study, stratified by, population screened (Freeman-Tukey transformed data).

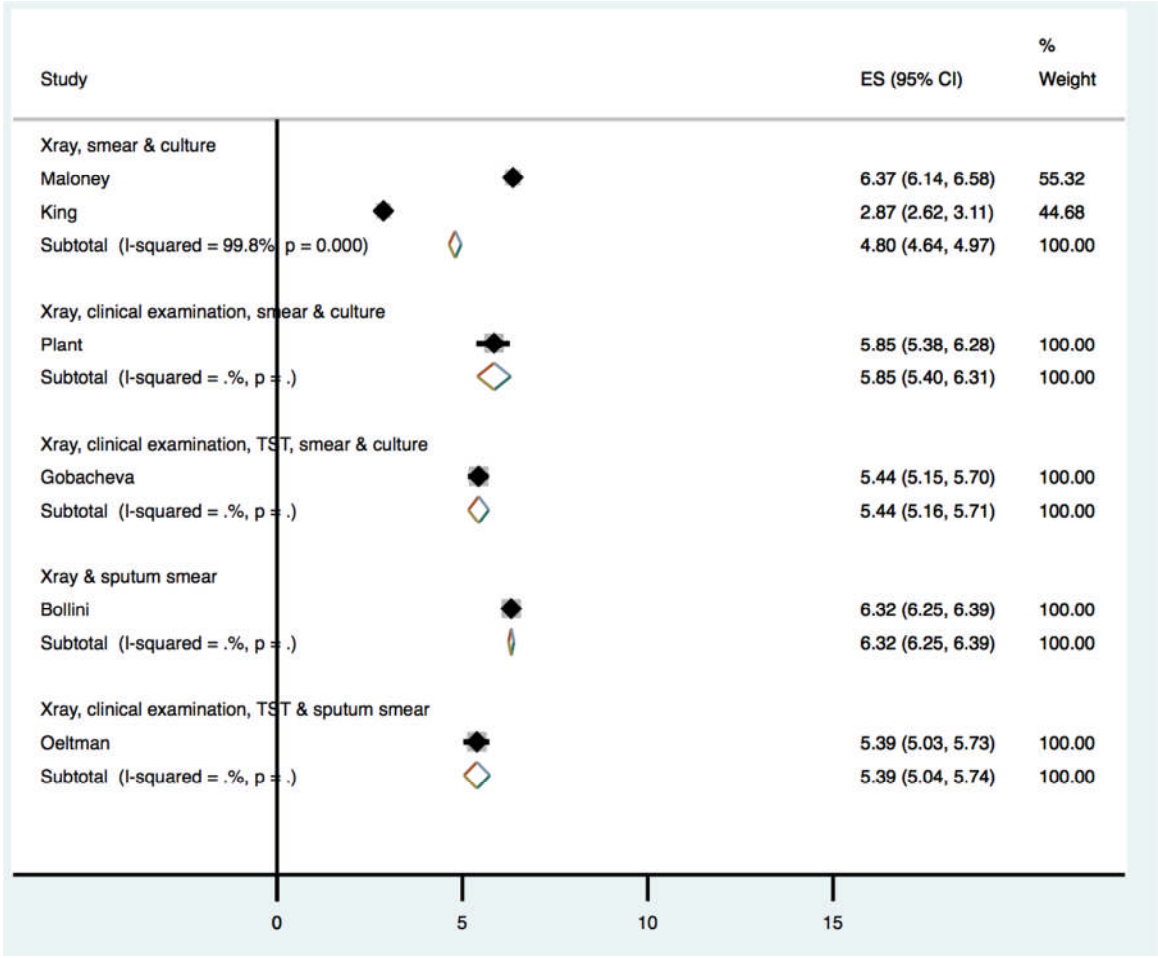

Figure 52. Forrest plot of yield for AFB confirmed cases found by each study, stratified by screening method (Freeman-Tukey transformed data).

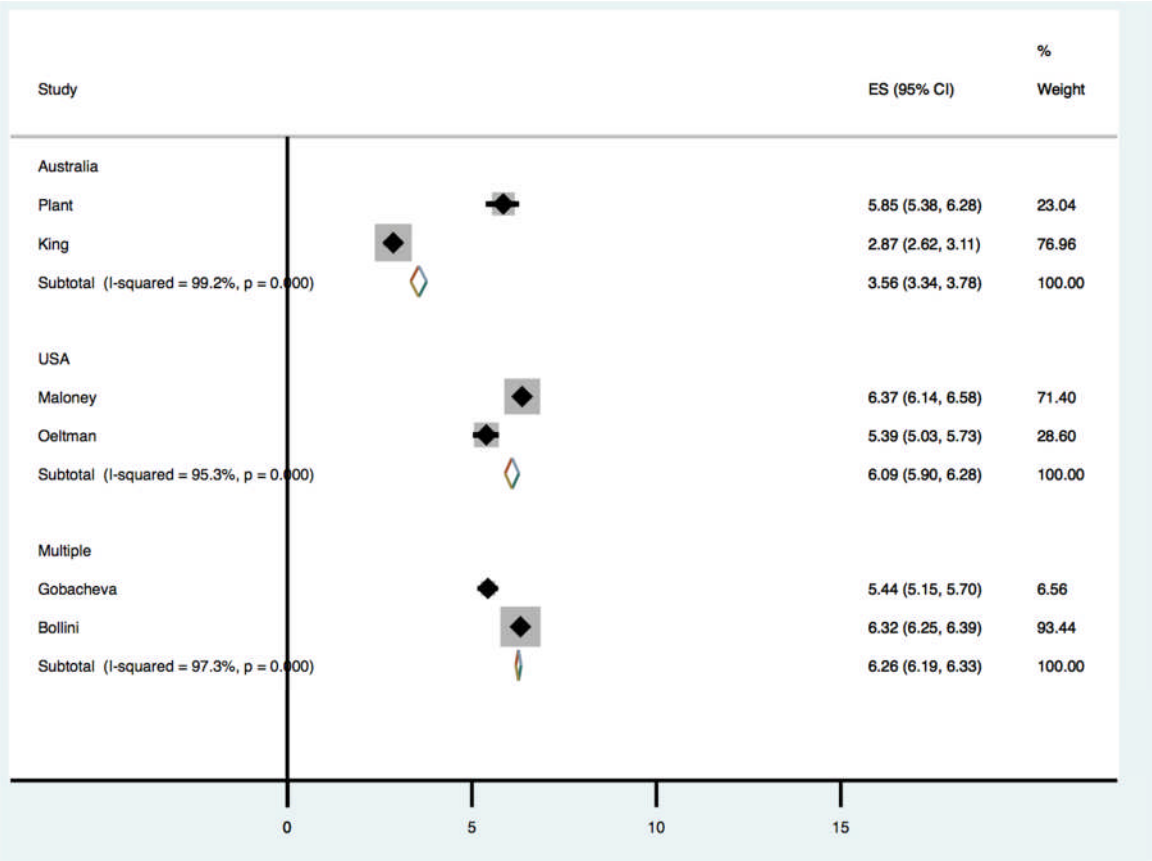

Figure 53. Forrest plot of yield for AFB confirmed cases found by each study, stratified by receiving country (Freeman-Tukey transformed data).

APPENDIX 4

Supporting information for chapter 3: Consent form for migrants undergoing screening for the UK per-entry tuberculosis screening programme and UCL ethics approval.

Annex E Consent form

UNITED KINGDOM PRE ENTRY TUBERCULOSIS SCREENING PROGRAMME

Name:
Date of birth
Clinic location:

Applicant's Declaration:
I understand that:

- I am required to undergo testing for pulmonary tuberculosis (TB), involving an X-ray and possibly sputum tests, prior to applying for entry clearance to go to the UK;
- If my chest X-ray is abnormal, I will receive individual counselling and an explanation of the further testing procedures.
- If my chest X-ray is abnormal, and changes are suggestive of tuberculosis, regardless of whether these changes are old or new, or if there are other clinical reasons to suspect TB, I will have to provide three sputum samples which will be tested for TB with smear and culture. I understand that the results of sputum cultures may take up to ten weeks
- If sputum samples are necessary, I will be required to return for sputum collection on three consecutive mornings starting within seven (7) days of my chest X-ray. If I fail to return within seven days, I will forfeit the opportunity to obtain a TB Certificate.
- If the smear or culture shows the presence of TB bacteria, I will be referred for TB treatment. Treatment shall be at my own expense; I will inform the TB treatment facility that I have close family contacts, who may need evaluation for TB.
- I have the right to refuse to undergo the TB assessment procedure and TB treatment, but accept such a refusal may adversely impact on my UK visa application.
- I understand that the physician has the final decision about whether I receive a Certificate

Female applicants.
All female applicants will be asked about their last menstrual period to identify applicants who possibly may be pregnant:

- If I could be pregnant, I will be offered several alternatives; 1) a chest X-ray with protective shield; 2). I can postpone the CXR (and TB clearance) until after delivery or 3) I can opt to provide three sputum samples for laboratory examination.
- I acknowledge that a CXR can carry a risk for the unborn child, but that this risk is quite small in the second and third trimester. I am therefore advised to consult the panel physician and may wish to consult my gynaecologist to understand the risks before I take a chest X-ray. If I decide to submit to an X-ray, this shall be at my own risk.

I hereby:

- consent to undergo TB testing;
- authorise you and your designated laboratory to store all relevant personal information collected during the assessment process, including health records and chest X-ray;
- authorise you and your designated clinics to share my personal details and assessment results with the UK immigration authorities, the UK Department of Health, Public Health England and the UK National Health Service.
- I authorise you to share my assessment results with the health authorities of my country of residence, where this is required by my country's laws.
- I release and hold harmless the UK Government and you from any liability for loss, injury suffered or other harm during, or as a result of, the TB assessment procedures

I have read this consent form, or had translated for me. I was invited to ask questions to clarify what was not clear to me. I understand the content of this declaration. Applicant's signature Date Please print your name

For children, or adults without the mental capacity to give consent, I confirm that I am the parent or legal guardian of the applicant and confirm that I give my consent For adults who are not able to physically sign the form, I confirm that I am an independent witness and the applicant has given their consent orally or by other non-verbal means Signature Date Please print your name Relationship to applicant

Statement of interpreter (if required); I have translated the content of this document for the applicant to the best of my ability and in a way in which I believe s/he can understand. Signed Date ..... Please print your name)

For female applicants who might be pregnant; I confirm that I have had the risks of having a chest X-ray in pregnancy explained to me and I wish to carry on with the chest X-ray. Signed Date Please print your name

Statement of Physician (if required); I have explained the content of this document to the applicant and confirm that the applicant has declined to go ahead with the assessment. Signed Date Please print your name

UCL RESEARCH ETHICS COMMITTEE
GRADUATE SCHOOL OFFICE

Dr Andrew Hayward
Department of Infection and Population Health
RF & UC Medical School
UCL
London
NW3 5RG

19 June 2013

Dear Dr Hayward

Notification of Ethical Approval
Project ID: 3294/002: Investigating the epidemiology of tuberculosis and the cost effectiveness of novel diagnostic screening pathways in migrants to the UK

I am pleased to confirm that in my capacity as Chair of the UCL Research Ethics Committee I have approved your study for the duration of the project i.e. until June 2015.

Approval is subject to the following conditions:

1. You must seek Chair's approval for proposed amendments to the research for which this approval has been given. Ethical approval is specific to this project and must not be treated as applicable to research of a similar nature. Each research project is reviewed separately and if there are significant changes to the research protocol you should seek confirmation of continued ethical approval by completing the 'Amendment Approval Request Form'.

The form identified above can be accessed by logging on to the ethics website homepage: http://www.grad.ucl.ac.uk/ethics/ and clicking on the button marked 'Key Responsibilities of the Researcher Following Approval'.

2. It is your responsibility to report to the Committee any unanticipated problems or adverse events involving risks to participants or others. Both non-serious and serious adverse events must be reported.

   Reporting Non-Serious Adverse Events
   For non-serious adverse events you will need to inform Helen Dougal, Ethics Committee Administrator (ethics@ucl.ac.uk), within ten days of an adverse incident occurring and provide a full written report that should include any amendments to the participant information sheet and study protocol. The Chair or Vice-Chair of the Ethics Committee will confirm that the incident is non-serious and report to the Committee at the next meeting. The final view of the Committee will be communicated to you.

   Reporting Serious Adverse Events
   The Ethics Committee should be notified of all serious adverse events via the Ethics Committee Administrator immediately the incident occurs. Where the adverse incident is unexpected and serious, the Chair or Vice-Chair will decide whether the study should be terminated pending the opinion of an independent expert. The adverse event will be considered at the next Committee meeting and a decision will be made on the need to change the information leaflet and/or study protocol.

On completion of the research you must submit a brief report (a maximum of two sides of A4) of your findings/concluding comments to the Committee, which includes in particular issues relating to the ethical implications of the research.

With best wishes for the research.

Yours sincerely

Professor John Foreman
Chair of the UCL Research Ethics Committee

Cc: Robert Aldridge Applicant

UCL Research Ethics Committee, c/o The Graduate School, North Cloisters, Wilkins Building
University College London Gower Street London WC1E 6BT
Tel: +44 (0)20 7679 7844 Fax: +44 (0)20 7679 7043
ethics@ucl.ac.uk
www.ucl.ac.uk/gradschool

APPENDIX 5

Supporting information for chapter 3: sensitivity analysis of primary outcomes including data on all migrants screened pre-entry

Table 31. Baseline characteristics of applicants screened for tuberculosis and prevalence of primary and secondary outcomes per 100,000 individuals screened.

	N (%)	Bacteriologically confirmed (95%CIs)	Culture positive (95%CIs)	Smear positive (95%CIs)
All	692232 (100.0%)	75 (69, 82)	58 (52, 63)	52 (47, 58)
Age group				
0-15	25555 (3.7%)	27 (13, 57)	27 (13, 57)	8 (2, 31)
16-44	647178 (93.5%)	75 (69, 82)	57 (52, 64)	50 (45, 56)
45-64	15829 (2.3%)	107 (67, 173)	76 (43, 133)	152 (102, 226)
>65	3670 (0.5%)	300 (166, 540)	218 (109, 435)	245 (128, 471)
Sex				
Female	242592 (35.0%)	94 (83, 108)	75 (64, 86)	73 (63, 85)
Male	449640 (65.0%)	65 (58, 73)	48 (42, 55)	40 (35, 47)
Family contact with infectious case of TB				
No	688296 (99.8%)	73 (67, 80)	56 (50, 62)	50 (45, 56)
Yes	1604 (0.2%)	1185 (757, 1853)	935 (565, 1548)	686 (380, 1236)
Visa Type				
Student	390803 (56.5%)	70 (62, 79)	55 (48, 63)	48 (41, 55)
Settlement and Dependent	233015 (33.7%)	87 (76, 100)	69 (59, 81)	56 (47, 67)
Work	26823 (3.9%)	78 (51, 120)	37 (20, 69)	0 (0, 0)
Working Holiday Maker	21043 (3.0%)	52 (29, 94)	29 (13, 63)	29 (13, 63)
Family Reunion	5639 (0.8%)	124 (59, 260)	35 (9, 142)	0 (0, 0)
Other	14909 (2.2%)	40 (18, 90)	34 (14, 81)	34 (14, 81)
CXR				
No abnormality	652313 (94.2%)	0 (0, 0)	0 (0, 0)	0 (0, 0)
TB suspected	28552 (4.1%)	1821 (1672, 1983)	1394 (1264, 1537)	1229 (1108, 1364)
Abnormality not TB	11367 (1.6%)	0 (0, 0)	0 (0, 0)	0 (0, 0)

WHO prevalence of TB in country of migration				
40-149	39060 (5.6%)	15 (7, 34)	5 (1, 20)	13 (5, 31)
150-349	99114 (14.3%)	181 (156, 209)	153 (131, 180)	168 (145, 196)
350+	554058 (80.0%)	61 (55, 67)	44 (39, 50)	34 (29, 39)
Year of examination				
2005	994 (0.1%)	0 (0, 0)	0 (0, 0)	0 (0, 0)
2006	31266 (4.5%)	0 (0, 0)	0 (0, 0)	0 (0, 0)
2007	97828 (14.1%)	52 (40, 69)	8 (4, 16)	0 (0, 0)
2008	109604 (15.8%)	67 (53, 84)	48 (37, 63)	0 (0, 0)
2009	132816 (19.2%)	83 (69, 100)	64 (52, 79)	63 (51, 78)
2010	109356 (15.8%)	68 (54, 85)	56 (43, 72)	47 (35, 61)
2011	97455 (14.1%)	87 (71, 108)	82 (66, 102)	33 (23, 46)
2012	62338 (9.0%)	106 (83, 135)	103 (80, 131)	59 (43, 82)
2013	50575 (7.3%)	93 (70, 124)	93 (70, 124)	32 (19, 52)
Country of screening				
Burkina Faso	117 (0.0%)	0 (0, 0)	0 (0, 0)	0 (0, 0)
Bangladesh	180612 (26.1%)	80 (68, 94)	61 (51, 73)	44 (36, 55)
Cambodia	741 (0.1%)	135 (19, 958)	135 (19, 958)	0 (0, 0)
Cote D'Ivoire	1531 (0.2%)	0 (0, 0)	0 (0, 0)	0 (0, 0)
Eritrea	238 (0.0%)	420 (59, 2989)	0 (0, 0)	420 (59, 2989)
Ghana	38644 (5.6%)	16 (7, 35)	5 (1, 21)	13 (5, 31)
Kenya	16868 (2.4%)	119 (77, 184)	77 (45, 133)	77 (45, 133)
Laos	229 (0.0%)	0 (0, 0)	0 (0, 0)	0 (0, 0)
Niger	82 (0.0%)	0 (0, 0)	0 (0, 0)	0 (0, 0)
Pakistan	369194 (53.3%)	49 (42, 57)	35 (29, 42)	27 (22, 33)
Sudan	7596 (1.1%)	26 (7, 105)	13 (2, 93)	0 (0, 0)
Somalia	3282 (0.5%)	274 (143, 526)	122 (46, 324)	213 (102, 447)
Togo	299 (0.0%)	0 (0, 0)	0 (0, 0)	0 (0, 0)
Tanzania	10859 (1.6%)	55 (25, 123)	46 (19, 111)	37 (14, 98)
Thailand	61940 (9.0%)	242 (206, 285)	215 (181, 255)	239 (203, 281)
Sputum culture testing				
Yes	215777 (31.2%)	92 (84, 101)	83 (75, 92)	55 (49, 62)
No	476455 (68.8%)	38 (31, 47)	0 (0, 0)	38 (31, 47)

Table 32. Multivariable analysis examining risk factors for bacteriologically confirmed tuberculosis

Risk Factor	Univariable IRR (95% CIs)	Multivariable IRR (95%CIs)	p-value
Age			
0-15	0.4 (0.2, 0.8)	0.3 (0.1, 0.6)	<0.001
16-44	1.0	1.0	
45-64	1.4 (0.9, 2.3)	1.2 (0.7, 1.9)	0.49
>65	4 (2.2, 7.3)	3.3 (1.8, 6.1)	<0.001
Sex			
Female	1.0	1.0	
Male	0.7 (0.6, 0.8)	1.0 (0.9, 1.3)	0.75
Contact with case TB			
No	1.0	1.0	
Yes	16.4 (10.4, 26)	11.4 (7.0, 18.4)	<0.001
Visa			
Students	1.0	1.0	
Settlement and Dependents	1.2 (1.0, 1.5)	1.3 (1.1, 1.6)	0.01
Work	1.1 (0.7, 1.7)	1.1 (0.7, 1.8)	0.60
Working Holiday Maker	0.7 (0.4, 1.4)	1.5 (0.8, 2.8)	0.18
Family Reunion	1.8 (0.8, 3.8)	1.1 (0.5, 2.5)	0.73
Other	0.6 (0.3, 1.3)	0.8 (0.3, 1.8)	0.54
WHO category			
40-149	0.1 (0.0, 0.2)	0.1 (0.0, 0.2)	<0.001
150-349	1.0	1.0	
350+	0.3 (0.3, 0.4)	0.3 (0.3, 0.4)	<0.001
Sputum culture testing			
No	1.0	1.0	
Yes	2.4 (1.9, 3.1)	2.4 (1.9, 3.0)	<0.001

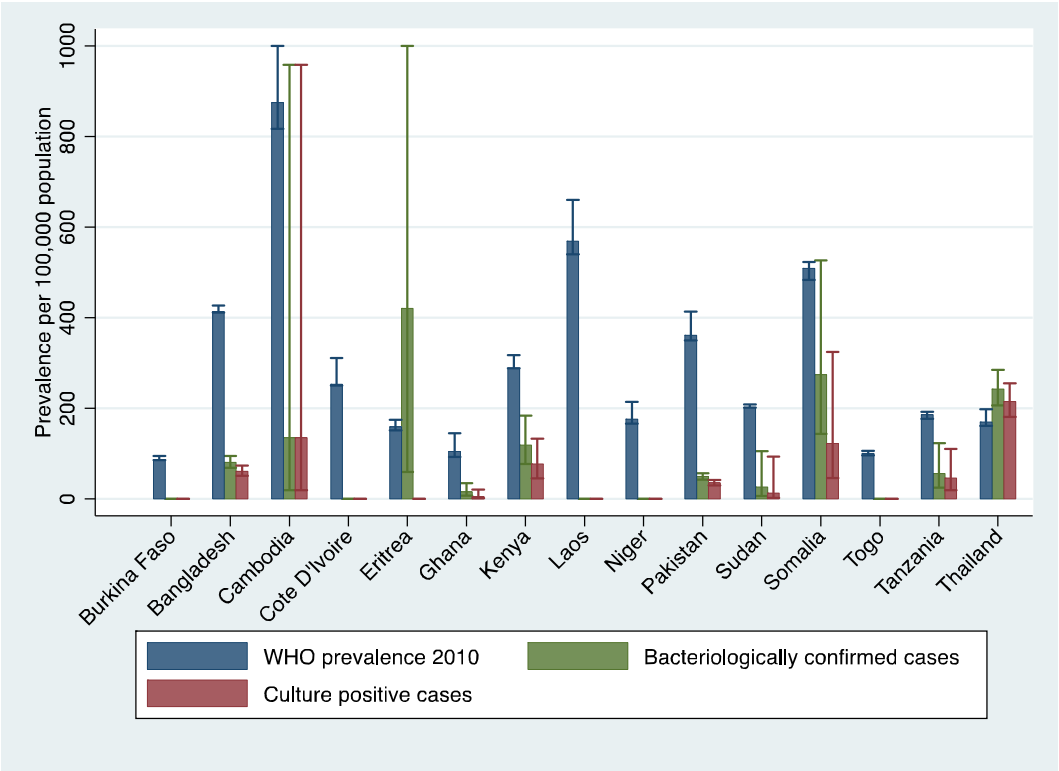

Figure 54. Crude prevalence of bacteriological and culture confirmed TB diagnosed at pre-entry screening compared to 2010 WHO country prevalence estimates.

Note: Error bars on bacteriological and culture confirmed tuberculosis estimates are 95%confidence intervals. Error bars on WHO 2010 prevalence country estimates are highest and lowest prevalence estimate for each country between 2007 and 2013. Confidence intervals limited to a maximum of 1,000 per 100,000 population.

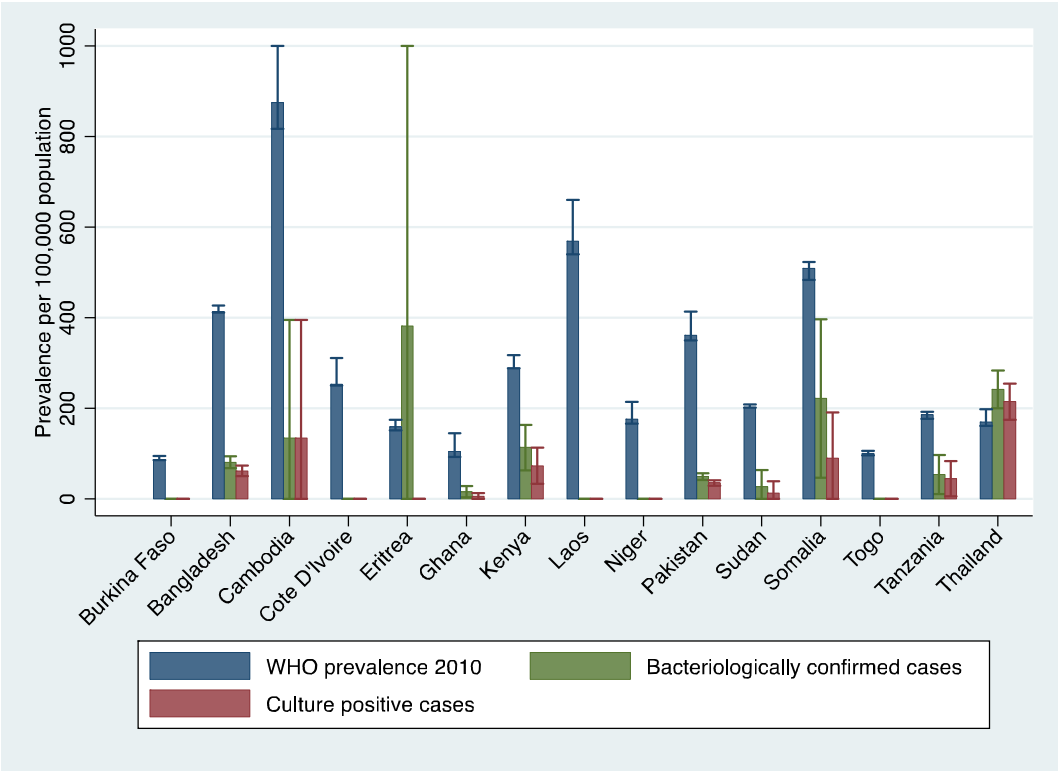

Figure 55. Age and Sex adjusted prevalence of bacteriological and culture confirmed TB diagnosed at pre-entry screening compared to 2010 WHO country prevalence estimates.

Note: Error bars on bacteriological and culture confirmed tuberculosis estimates are 95%confidence intervals. Error bars on WHO 2010 prevalence country estimates are highest and lowest prevalence estimate for each country between 2007 and 2013. Confidence intervals limited to a maximum of 1,000 per 100,000 population.

APPENDIX 6

Supporting information for chapter 4: additional details on how primary outcomes were calculated.

Table 33. Description of how sensitivity, specificity, positive and negative predictive values were calculated.

		Deterministic (NHS Number)		
		+ve	-ve	Total
Probabilistic (EMS)	+ve	a	b	a+b
	-ve	c	d	c+d
	Total	a+c	b+d	a+b+c+d

Sensitivity = a/a+c

Specificity = d/b+d

Positive predictive value= a/a+b

Negative predictive value = d/(c+d)

APPENDIX 7

Supporting information for chapter 6: Sensitivity analysis on cross-sectional study restricted to non-UK migrants arriving after 2005.

Table 34. Baseline characteristics, univariable and multivariable logistic regression to examine risk factors for first cases in a cluster of tuberculosis in non-UK born individuals who arrived in the UK after 2005 and were notified as a case of tuberculosis between 2010-2013.

Risk Factor	Migrants contributing (%)	First in cluster (%row)	Univariable OR (95% CI)	Multivariable OR (95%CI)	p-value
All	5998 (100%)	623 (10.4%)			
Age					
0-15	85 (1.4%)	18 (21.2%)	2.3 (1.4, 3.9)	2.2 (1.3, 3.8)	<0.001
16-44	5276 (88.0%)	554 (10.5%)	1.0	1.0	
45-64	490 (8.2%)	38 (7.8%)	0.8 (0.5, 1.5)	0.7 (0.5, 1.0)	0.09
65+	146 (2.4%)	13 (8.9%)	0.0 (0.0, 0.0)	0.8 (0.5, 1.5)	0.54
Missing	1 (0.0%)	0 (0.0%)			
Sex					
Female	2314 (38.6%)	225 (9.7%)	1.0		
Male	3683 (61.4%)	398 (10.8%)	1.1 (0.9, 1.3)	1.1 (1.0, 1.4)	0.12
Missing	1 (0.0%)	0 (0.0%)			
WHO prevalence					
0-39	304 (5.1%)	34 (11.2%)	1.1 (0.8, 1.6)	0.9 (0.6, 1.4)	0.75
40-149	494 (8.2%)	55 (11.1%)	1.1 (0.8, 1.5)	0.9 (0.7, 1.3)	0.60
150-349	3023 (50.4%)	313 (10.4%)	1 (0.9, 1.2)	0.8 (0.7, 1)	0.05
350+	2099 (35.0%)	212 (10.1%)	1.0	1.0	
Not known	78 (1.3%)	9 (11.5%)	1.2 (0.6, 2.4)	1.0 (0.5, 2)	0.95

BCG vaccinated					
No	1102 (18.4%)	105 (9.5%)	1.0	1.0	
Yes	3291 (54.9%)	357 (10.8%)	1.2 (0.9, 1.5)	1.2 (0.9, 1.5)	0.14
Missing	1605 (26.8%)	161 (10.0%)	1.1 (0.8, 1.4)	1.1 (0.8, 1.4)	0.57
Social risk factor					
No	4701 (78.4%)	488 (10.4%)	1.0	1.0	
Yes	357 (6.0%)	37 (10.4%)	1.0 (0.7, 1.4)	1.0 (0.7, 1.4)	0.95
Missing	940 (15.7%)	98 (10.4%)	1.0 (0.8, 1.3)	1.0 (0.8, 1.2)	0.89
Time since entry to UK					
0-2	3403 (56.7%)	385 (11.3%)	1.0	1.0	
3-5	2025 (33.8%)	218 (10.8%)	0.9 (0.8, 1.1)	0.9 (0.8, 1.1)	0.31
6-10	570 (9.5%)	20 (3.5%)	0.3 (0.2, 0.5)	0.3 (0.2, 0.4)	<0.001
Screened pre-entry					
No	5273 (87.9%)	566 (10.7%)	1.0	1.0	
Yes	725 (12.1%)	57 (7.9%)	0.7 (0.5, 0.9)	0.6 (0.4, 0.8)	<0.001

Table 35. Baseline characteristics, univariable and multivariable logistic regression to examine risk factors for being a case of tuberculosis reactivation in reactivation in non-UK born individuals who arrived in the UK after 2005 and were notified as a case of tuberculosis between 2010-2013.

Risk Factor	Migrants contributing (%)	Unique strain type (%row)	Univariable OR (95% CI)	Multivariable OR (95%CI)	p-value
All	5998 (100%)	3158 (52.7%)			
Age					
0-15	85 (1.4%)	36 (42.4%)	0.7 (0.4, 1.0)	0.7 (0.4, 1.0)	0.06
16-44	5276 (88.0%)	2780 (52.7%)	1.0	1.0	
45-64	490 (8.2%)	259 (52.9%)	1.0 (0.8, 1.2)	1.0 (0.8, 1.2)	0.83
65+	146 (2.4%)	82 (56.2%)	1.2 (0.8, 1.6)	1.1 (0.8, 1.6)	0.45
Missing	1 (0.0%)	1 (100.0%)			
Sex					
Female	2314 (38.6%)	1246 (53.8%)	1.0		
Male	3683 (61.4%)	1912 (51.9%)	0.9 (0.8, 1.0)	0.9 (0.8, 1.0)	0.14
Missing	1 (0.0%)	0 (0.0%)			
WHO prevalence					
0-39	304 (5.1%)	152 (50.0%)	0.9 (0.7, 1.1)	0.9 (0.7, 1.2)	0.59
40-149	494 (8.2%)	251 (50.8%)	0.9 (0.8, 1.1)	0.9 (0.8, 1.2)	0.59
150-349	3023 (50.4%)	1604 (53.1%)	1.0 (0.9, 1.1)	1.0 (0.9, 1.2)	0.63
350+	2099 (35.0%)	1112 (53.0%)	1.0	1.0	
Not known	78 (1.3%)	39 (50.0%)	0.9 (0.6, 1.4)	0.9 (0.6, 1.5)	0.75
BCG vaccinated					
No	1102 (18.4%)	560 (50.8%)	1.0	1.0	
Yes	3291 (54.9%)	1725 (52.4%)	1.1 (0.9, 1.2)	1.1 (0.9, 1.2)	0.36
Not known	1605 (26.8%)	873 (54.4%)	1.2 (1.0, 1.3)	1.1 (1.0, 1.3)	0.11
Social risk factor					
No	4701 (78.4%)	2463 (52.4%)	1.0	1.0	
Yes	357 (6%)	175 (49.0%)	0.9 (0.7, 1.1)	0.9 (0.7, 1.1)	0.36
Not known	940 (15.7%)	520 (55.3%)	1.1 (1.0, 1.3)	1.1 (1.0, 1.3)	0.15

Time since entry to UK					
0-2	3403 (56.7%)	1803 (53.0%)	1.0	1.0	
3-5	2025 (33.8%)	1052 (52.0%)	1.0 (0.9, 1.1)	1.0 (0.9, 1.1)	0.52
6-10	570 (9.5%)	303 (53.2%)	1.0 (0.8, 1.2)	1.0 (0.9, 1.2)	0.75
Screened pre-entry					
No	5273 (87.9%)	2763 (52.4%)	1.0	1.0	
Yes	725 (12.1%)	395 (54.5%)	1.1 (0.9, 1.3)	1.1 (0.9, 1.3)	0.33