The evidence for and utility of current contact tracing methods for tuberculosis in congregate locations in low incidence settings

Darryl Alexander Braganca Menezes

A thesis submitted for the degree of Doctor of Philosophy

College of Medical and Dental Sciences

Graduate School

University of Birmingham

July 2022

Abstract

Introduction

Tuberculosis is a global threat

While numbers are lower in the UK, disease persists and in order to achieve WHO targets of elimination, latent infection needs to be aggressively sought after and treated. In order to facilitate this, we need an effective contact tracing approach. I aim to examine the evidence behind current contact tracing approaches and examine their utility in the current low incidence setting, with a particular focus on congregate settings.

Methods

Data was gathered through the interrogation of Public Health England (PHE) statutory data collection augmented by local chest clinic data which are both collected from confirmed TB patients and subsequent contact tracing efforts. Literature reviews were conducted using established systematic review processes. Further analytical work was conducted as needed using statistical software (STATA/SE 16).

Results

Chapter 1 provides the context for this thesis, discussing tuberculosis as a clinical entity. Chapter 2 demonstrates the evolution of contact tracing and the highlighting how the process was established in a high prevalence environment and how this may no longer apply to the lower incidence setting. Chapter 3 and 4 examine the evidence based behind risk factors in congregate settings as well as the comparative evidence for current contact tracing methods and the availability and utility of others. Chapter 5 summarises the data set,

providing outcome data for large scale contact investigations. In chapter 6 I examine the longer-term impact of our current contact tracing approach employing recently available whole genome sequencing data. Using the data set, I derive a prediction model in order to predict TB disease with the variables to hand. Finally, in chapter 8 I address the key aims of this thesis and to what extent I have answered these questions.

Discussion

Our current contact tracing approach while effective in a higher disease prevalence setting, does not allow us to effectively curtail the on-going TB epidemic in a low incidence, congregate setting environment. This approach is resource intensive and we can see that on-going propagation of transmission chains occur. This is largely due to the lack of identified contacts as well as the inability to effectively engage them in the contact tracing process. There is little evidence to support the use of other contact tracing processes. Largely due to the lack of this area having been examined in relation to TB spread in low incidence settings. In addition, the predictive modelling suggests that congregate settings are in themselves, heterogeneous locations that do not possess the same risk factors for individuals who acquire disease. Moving forward, teams involved in Tb prevention and contact tracing need to adopt an approach which can demonstrate relationships and contact status to a higher resolution and this will likely involve a form of social network analysis. More work on inputting prediction model outputs into a risk assignment score can help preventative services prioritise cases for follow up preventing individuals propagating disease chains.

Dedication

To my wife for being my constant support and motivation

To my father, for my daughter:

Super humeros tuos i oriri

Acknowledgments

My gratitude goes to Professor Tom Marshall and Emmanouil Tranos, who despite the challenges of COVID and interruptions to the programme managed to constantly maintain input and focus my ambition, guiding me towards the finish line.

To Dr Martin Dedicoat, my mentor and supervisor whose vision and knowledge helped build this thesis and who remains a constant guide and reference for me.

To the Cochrane community and team that helped with the review process -
I am grateful to our affiliated institutions and organizations, and thank the referees and editors for their comments and encouragement especially the academic editor Professor Mical Paul. I am grateful to the Liverpool School of Tropical Medicine, UK for their help with the Cochrane review as well as the Department for International Development (DFID), UK for Cochranes funding.

Thanks to Hanna Kaur, Lead TB nurse at the TB chest clinic and service, Birmingham, for supporting this project and helping with my access and understanding of the data and collection.
Thanks also to Dr Esther Robinson and Dr Grace Smith whose pioneering work on whole genome sequencing helped with the conception of this project and improvement in TB care.

Table of Contents

Table of Figures:

Table of Tables:

Chapter 1 - Introduction

1.1 What is tuberculosis

Tuberculosis, is a multi-organ clinical disease affecting humans and other mammalian species. Mycobacterium tuberculosis (Mtb) is an infectious bacterium that is a causative agent of tuberculosis. [1] Clinical TB disease presents in two broad configurations; pulmonary and extra-pulmonary disease.

1.2 Symptoms

There is no pathognomonic symptom of TB. A combination of clinical features can however guide the clinician to a positive diagnosis.

TB can present with constitutional, non-specific symptoms and more local symptoms depending on the organ system affected.

Constitutional features are common in TB and include malaise and fatigue generally over months given the slow progression of the disease. Weight loss is also a common presentation with patients potentially losing up to a third of their body weight prior to presentation. Another common presentation and one frequently associated with TB are

night sweats, profuse and frequent sweating at night. [2, 3] These features are not unique to TB and can also occur in some malignancies, in particular lymphoma.

1.2.1 Pulmonary

Features of pulmonary disease mimic other causes of respiratory infections. In 80% of cases, a cough is present. This is usually a prolonged 'dry' cough initially with no sputum production. TB infection can however cause a purulent cough over time. In addition to sputum, patients can also cough blood (haemoptysis), this again can occur in other respiratory conditions. TB related disease can cause significant haemoptysis in later cases with progressive lung destruction. This can also result in breathlessness. Chest pain has also been noted in patients with TB, however this is not related to pulmonary disease but rather spread of disease to the nerve containing areas of the chest and is an uncommon presentation. [4, 5]

Pulmonary disease is not restricted to limited lesions. Other presentations include pleural effusions, an accumulation of fluid in the lining of the lungs as a result of inflammatory changes. This occurs in 10% of cases and is more common in adult patients. If TB reaches and is spread through the blood or lymphatics system, it can result in a disseminated pattern in the lungs known as miliary TB, a presentation signalling widespread disease. Other presentations include lung collapse as a result of obstructed airways (lobar collapse) and pneumothoraces (air outside the lungs compressing them). In addition, local invasion by

disease can result in fistulae (abnormal communications between adjacent structures e.g. bronchopleural fistula). [6]

1.2.2 Extra-pulmonary

Disease that is not localised to the pulmonary system is known as extra-pulmonary disease. This usually represents spread of TB from the initial respiratory infection. Extra-pulmonary disease can therefore occur concomitantly with pulmonary disease. Immigrants from high burden settings, younger patients and those with HIV infection represent higher risk groups for extra-pulmonary tuberculosis (EPTB). [7]

EPTB can affect any other site in the body. Common sites for infection include: the lymphatic system (TB lymphadenitis); the central nervous system (brain); bones and joints; the heart (pericardial TB); and the abdominal and genitourinary systems.

Extra-pulmonary TB typically presents slowly over several months given the slow growing nature of TB and the need for significant lesions to be present for symptoms to develop. The exception being TB meningitis.

Pulmonary disease results from an infection anywhere along the respiratory tract. It is pulmonary disease that is considered the infectious form of tuberculosis as it allows expectoration of tubercle bacilli.

It is these pulmonary manifestations that are of particular note clinically, as it is active pulmonary disease that results in further spread through the infected individual coughing and aerosolising Mtb. Furthermore, pulmonary disease is significant not only because of its propensity to spread TB but also because untreated, up to 50% of people will die. [8]

Individuals with infectious disease are usually delayed in presenting to healthcare services. In order for Mtb to spread, the host needs to be in relatively good health while active bacilli replication takes place. Thus, patients with developing tuberculosis remain relatively asymptomatic to begin with. This coupled with the slow rate of progression in most cases and the gradual development of troubling clinical features increase the opportunity for on-going transmission chains.

Close, persistent contact with an infected individual can result in the acquisition of infection. These individuals have a subsequent risk of developing active disease of 5-15% over their lifetime (with 5% of them developing active disease in the first 2 years). This risk is heavily modified by co-existing factors. HIV/AIDs infection for example has a risk of development of active disease of 10% *per year*. [9-14]

1.3 History/ origins

M(Tb) has existed alongside humankind for millennia. Some authors have posited that a cousin of *M. tuberculosis*, *M. ulcerans* may have existed for as long as 150 million years. [15]

Archaeological evidence from mummified sources demonstrated acid-fast bacilli (remnants of Mtb) disease in ancient Egypt as far back as 1000BC. [16] With European archaeological investigations demonstrating TB disease in skeletal remains up to 6000 years old. [17, 18]

Studies on new world ancient human skeletons suggests Mtb existed in Peru millennia before European contact, suggesting TB existed in a pre-migratory, early hominid state which would mean African hominids had infections. [19]

Genetic studies have demonstrated the extreme homogeneity in MTBC organisms, suggestive of a clonal proliferation from a successful ancestral progenitor some 30,000 years ago. [20]

Classical clinical features of this disease have been recognised for thousands of years. Several cutaneous manifestations have even earnt royal intervention. Scrofula, a painful enlargement of the lymph glands was recognised in the middle ages. Beliefs at the time around the divine mandate of royalty resulted in treatments including the laying on of hands from royal patronage. Scrofula became known as the king's evil as a result. [21]

Whilst recognised in various forms, it wasn't until the late 19th century that the causative agent for this disease was discovered by Robert Koch. Following this discovery and subsequent public health and medical developments in detection and treatment, the Mtb epidemic in developed countries began to recede. This correlated with improvements in housing and sanitation as well as interventions such as pasteurisation, improved food production legislation and vaccination. [22]

Whilst there was a steady decline in tuberculosis during this period, it wasn't until the 1940's when mortality of cases began to decline in earnest. This was the era of effective anti-tuberculosis chemotherapy. Over the subsequent decades, the key components of anti-tuberculosis treatment were discovered and subsequent trial and error resulted in a 4-drug combination used to this day. [23], [24]

1.4 Pathology/ Pathogenesis

M(TB) forms part of the mycobacterium tuberculosis complex (MTBC), a grouping of genetically related bacterial organisms known to cause tuberculosis. This complex includes, *M. tuberculosis* as well as *M. bovis, M. africanum and M. microti* among others. These various agents can cause a variety of clinical disease presentations and affect other mammalian hosts. There are other species of mycobacterium that are implicated in human disease. These are unrelated to MTBC bacteria and are known as non-tuberculous mycobacterium (NTM). NTM's have a different epidemiology and pathogenesis and will not be considered further here. Only *M. tuberculosis* will be considered here, given its primacy in causing human infections. [25]

Mycobacterium tuberculosis are acid-fast bacilli, transmitted through expectorated respiratory droplets from an infected individual through coughing or activities generating infectious droplets e.g. singing. These infectious droplets can persist in the environment for several hours. In humans Mtb has 3 stages; primary tuberculosis, post-primary disease and latency. [1]

Infection begins with a susceptible individual inhaling these infectious droplets. Mtb bacterium then spreads to the respiratory system. This is primary tuberculosis. The outcome from primary tuberculosis infection depends on the host immune state. [26]

If an adequate host immune response exists, primary infection is contained in a granuloma (a localised aggregation of host immune cells including, in particular, macrophages and lymphocyte white blood cells). [27-29]

In immunosuppressed individuals, or those with a deficient immune response (e.g. immunomodulating treatments, individuals with acquired immune deficiency syndrome [AIDS]) primary tuberculosis infection can result in disease which may be severe. In the lungs, rapid spread of tuberculosis causes a widespread disease known as miliary tuberculosis. Bacilli can also spread rapidly via blood and lymphatic systems and result in extra-pulmonary disease (including disease of the central nervous system, gastrointestinal tract, bones etc.). [6]

A further outcome is the conversion of this initial disease state to a dormant state – latency. Latent infection states occur in about 90% of individuals infected with primary tuberculosis. This state presents a risk of future conversion to active disease (re-activation). The period of time from latent infection to active disease varies in individuals and is modified by changes in the host immune response. [13]

Respiratory tract infection results in local lymph node enlargement as the immune system attempts to contain the infection. This combination of pulmonary infection with lymph node reaction is known as the Ghon complex. If the immune reaction is successful in containing a new infection, scarring takes place. This is known as the Ghon focus. [26]

Following primary tuberculosis, the post-primary stage can occur. This results after the primary stage and is a result of re-activation of disease or re-infection from a new environmental contact.

The mycobacterium infection in post-primary tuberculosis begins by evading host immune containment and replication and local invasion results in spread of infection. In pulmonary disease, this manifests as local lung invasion in the form of a pneumonia. [26] Local tissue destruction caused by increasing burden of disease results in a central area of tissue destruction (caseation) surrounded by a weakened outer core of immune cells. Eventually this caseated centre is expectorated and a cavity is formed. Cavities represent large areas of mycobacterial organism replication and correlate with increased host infectiousness. [26], [30]

The role the immune system plays in post-primary disease remains a source of investigative research. The link between latent and active disease states increasingly appears to be a spectrum determined by immune and pathogen interactions, rather than a latent state. [31]

1.5 Epidemiology

1.5.1 Global

Tuberculosis infection is a global problem. An estimated one quarter of the world population harbour TB disease, with 10 million people displaying active infection. Up to 1.4 million people a year die as a result of TB. It has been postulated that TB may also be the greatest bacterial killer in history. [32, 33]

Over the past 2 centuries, TB has seen a decline in developed countries. Where once it was a ubiquitous disease across social strata, rates of disease demonstrated a steady decline from the late 1800's through to the 1970s. This followed the discovery in 1884 of the causative organism of TB, the tubercle bacillus and subsequent public health and medical developments including vaccination and effective drug therapy in the 1940's.

During the late 1980's and early 1990's TB disease rates began to increase in developed countries. WHO declared TB a global emergency in 1993. This resurgence had a combination of contributory factors. The concomitant HIV epidemic, increased migration and decreasing TB services have all been implicated in rising rates. In addition, while TB had traditionally been indistinguishable between social classes, it now retreated to at risk demographic pockets becoming a disease of the socially deprived. These include the homeless, incarcerated populations, immigrants from high incidence countries and those with high alcohol and drug abuse behaviour as well as HIV populations.

The strategic development goals set by WHO aim to reduce the burden of TB to less than 1 per million cases by 2050 with a 90% reduction in TB incidence rates compared with 2015. WHO estimates that in order to achieve this ambitious target, services will need to detect 70% of incident cases and cure 85% of them. This is in order to prevent further spread of infection through the early identification and treatment of cases. This will require the use of aggressive contact tracing approaches. Furthermore, treatments for latent TB disease involve fewer agents and therefore less side-effects. They are more likely to be tolerated and do not involve the considerable time and expense necessitated to treat resistant strains. [34, 35]

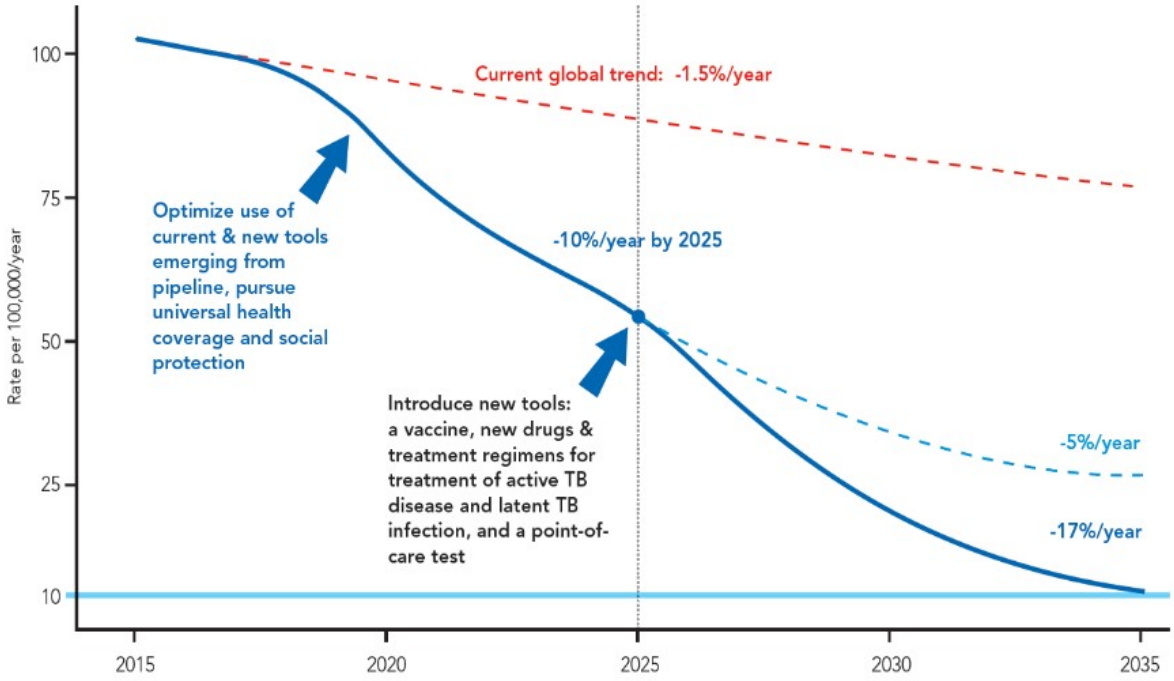

Figure 1 - Desired decline in global TB incidence rates to reach 2035 targets

[36]

1.5.2 National

In the UK, during the 1980's, TB demonstrated a resurgence with incidence increasing annually up to 2011. Since then however, there has been a steady decline in rates. They now stand at the lowest levels recorded (8280 – in 2011 to 4672 – in 2018) representing a 44% decline in new case notifications. Despite this decline, there are key epidemiological changes worth noting. There are still significant discrepancies following the socioeconomic divide with the most deprived 10% of the population having a 7x higher incidence than the most well off 10%. While immigrants from high incidence settings who traditionally had the highest rates of TB disease have demonstrated a declining rate of TB in large part due to the

new-entrant screening programmes, TB still affects those born abroad 13x more than those born in the UK. Rates are now increasing in 2nd and 3rd generation children of immigrants i.e. those born and raised in the UK. In addition, the rates of TB in the white population of UK born individuals with no travel history have not improved and remains a troublesome source of cases. [37, 38]

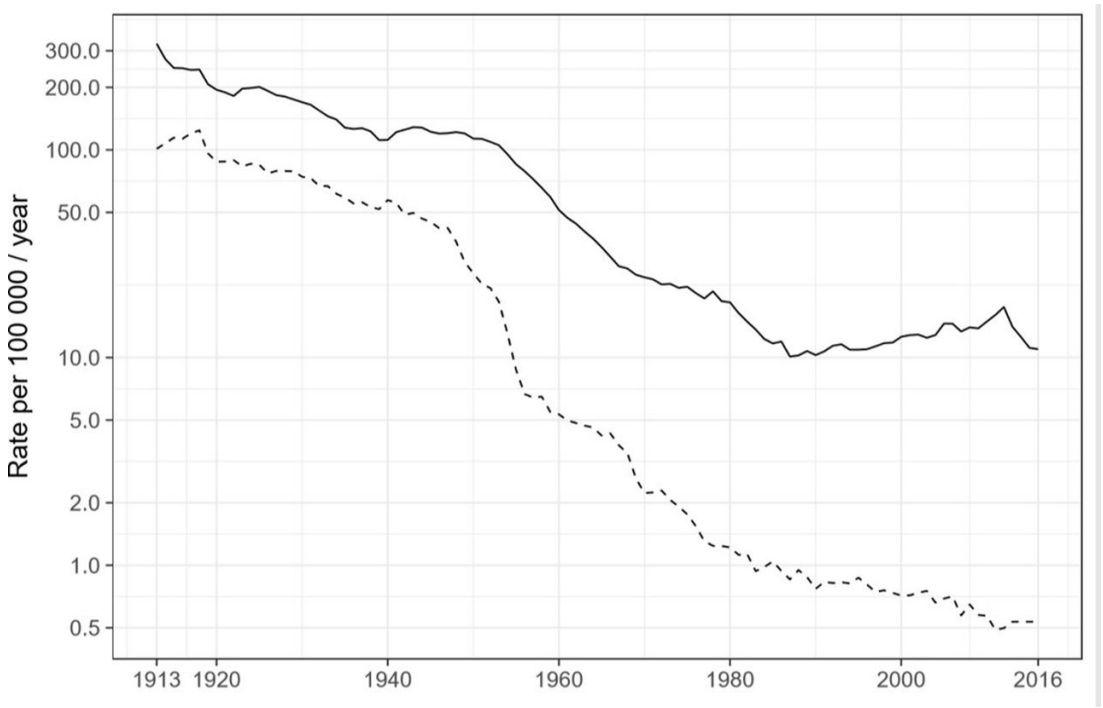

Figure 2 - TB incidence (solid line) and mortality (dashed line) rates per 100,000 population per year in England and Wales (1913-2016)

[39]

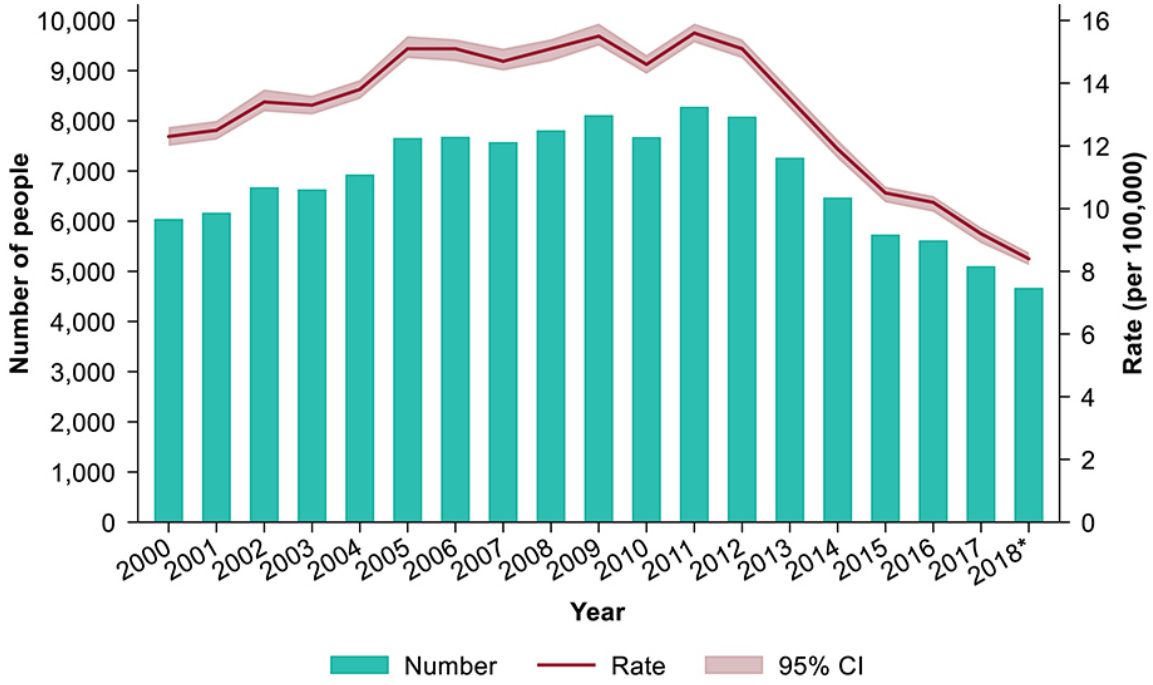

Figure 3 - Number of TB case notifications and rates, England 2000-2018

[40]

TB is however not declining in all groups with the so called underserved population (USP) remaining at risk. USP members are those with one or more social risk factor. Since records began in 2010 (collecting social risk factor data), the highest rate of TB has been recorded in this group, (13% of all cases in 2017). USPs often have multiple, complex needs, and the social risk factors they experience that are known to increase the risk of TB include; homelessness, drug and alcohol misuse, incarcerated population groups, co-existing mental health issues, asylum seekers and unemployment and poverty. [41]

These underserved population groups often display significant social overlap, are less likely to engage with services, more likely to conceal activities and contacts and are less likely to complete treatment when started. This latter point raising the risk of resistant disease development.

TB in the UK is not evenly distributed throughout the country, rather it is localised around metropolitan centres. In particular, London has the highest rate of new TB cases, followed by the midlands and in particular, Birmingham. This closely mirrors the significant ethnic diversity present in these large cities. [42]

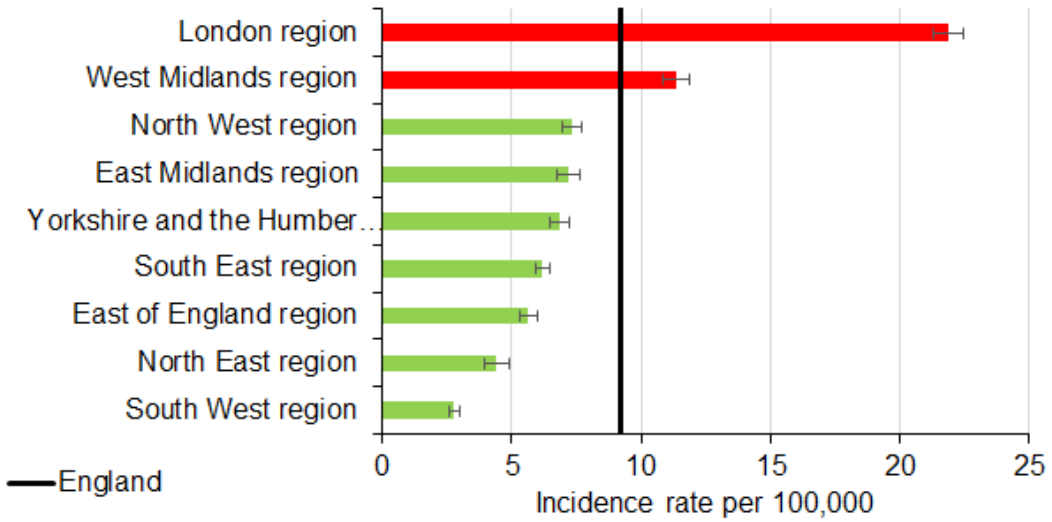

Figure 4 - TB incidence rate, by region, UK (2016-2018)

[43]

1.6 Diagnosis

1.6.1 Microscopy and culture

The ability to visualise the mycobacterium bacillus remains a mainstay of diagnostic techniques and was the basis for the identification of tubercule bacillus by Koch in 1884.

Samples that contain disease are collected and stained appropriately. Given the unique cellular properties of Mtb, this stain is not the commonly used gram stain as for other clinically relevant bacteria, but requires the specialised Ziel-Neelsen stain.

The possible samples collected also follow the clinical importance and ease of collection. In particular sputum collection remains the mainstay of investigation. Apart from reflecting the likely and primary infectious zone of the lungs, sputum positivity also provides a guide to the source patients infectiousness.

Sputum is cultured to allow growth, amplification and subsequent analysis of Mtb. Given the slow growing nature of this organism, this can take up to 10 weeks to grow. If there is a particularly high burden of disease present in collected sputum samples, Mtb may be readily identified through rapid assessment. This initial rapid assessment is the sputum smear test involving staining and applying sputum to a microscopy slide. Subsequent identification of the Mtb organism on smear samples suggests a high degree of infectiousness of the source case versus cases with smear negative results. [44]

Culture results remain the gold standard. Particularly as growth of the organism allows susceptibility to treatment tests to be carried out. This is especially relevant where resistance to common anti-tuberculous agents are known or suspected. These resistance tests allow clinicians to make appropriate treatment decisions allowing for a better clinical outcome and preventing the development of treatment resistance.

1.6.2 Molecular techniques

Although culture remains the gold standard of diagnosis for Mtb, molecular techniques have developed which allow rapid diagnosis (species identification) and limited drug susceptibility testing. Molecular and genetic techniques have been used to determine links for transmission. These techniques include polymerase chain reactions (PCR) and genomic analysis.

PCR can be used to amplify genomic regions to aid in species identification as well as regions coding for resistance genes. [45], [46]

Commercial products now make use of the unique genomic properties of Mtb. They offer off the shelf products that can provide a rapid identification of Mtb as well as providing limited, common drug resistance information to assist the clinician in making rapid treatment choices. These are PCR based products. Examples include Genexpert (Cepheid).

Genomic augmentation of epidemiological investigations by matching isolates from patients has been on-going for over 2 decades. Initially this was done using Insertion Sequence based Restriction Fragment Length Polymorphisms (RFLP) typing. This worked by examining the genome for repetitive sequences, in particular sequence IS6110. [47] This was subsequently replaced by Multiple Interspersed Repetitive Units- Variable number tandem repeat (MIRU-VNTR) typing looking at 15- then 24-loci. MIRU-VNTR typing relies on examining micro-satellite gene regions through nucleic acid amplification. The output from MIRU-VNTR can be readily compared across laboratories which made it a useful tool for public health services. [48]

Comparable MIRU-VNTR of different individuals Mtb samples allowed clinicians and public health teams to assign molecular links and therefore describe the possible spread of TB cases in different groups where multiple infections have occurred concurrently or sequentially. This is particularly relevant for public health teams attempting to link potentially infected individuals (contact tracing).

Patients with differing genotypic disease can be reliably detected and transmission ruled out by standard MIRU-VNTR testing. Where genotypes match, ruling out transmission becomes less certain. [49] In order to do this, a higher level of resolution is needed in order to examine micro-evolution of the genome, i.e. earlier, subtler genomic changes. Whole genome sequencing offers this approach. [50] [51-53] Detecting accumulated genomic changes can provide investigators with a directionality to transmission events, a function not previously available with older techniques. This has been shown to allow for better

understanding of outbreaks and more targeted epidemiological investigations as well as allowing preliminary results for resistance patterns. All UK TB strains are now sent for WGS which is the standard of practice. [54], [55]

1.7 Treatment

1.7.1 Latent infection

Following inhalation of mycobacterium from an infected source, symptomatic (active) disease is contained by immunological processes. These involve macrophages and granulocytes balancing bacterial proliferation and spread. This results in an infected but asymptomatic state. [56] Over a lifetime, between 5-15% of latent infection cases will become active and can spread infection. [12, 13]

Traditionally thought to represent a binary state (i.e. either latent or active), latency is increasingly considered to represent a disease spectrum. Asymptomatic and symptomatic disease states being determined by host-pathogen interactions and their relative predominance. [57-59] [60]

Proving an infection in these pre-infectious, latent cases is not possible given the lack of viable bacteria in sputum to culture. Indirect testing is therefore the accepted method. This relies on testing antigenic responses. Two methods are currently used, the tuberculin skin test (TST) and interferon-gamma release assay (IGRA). [12] These tests work by detecting historical (memory) T-cell responses. They are therefore unable to distinguish between active and latent infection or indeed treated and untreated disease. [61]

Treating latent infection is less complicated than active cases. US Public health services demonstrated the benefit of 12 months of Isoniazid (INH) in the 1950s and 60s. These studies showed that after 9 months of treatment, efficacy plateaued. Therefore, the mainstay of latent treatment has been 9 months of INH therapy. [62, 63] [64] [65] More recently, the 9-month duration has been challenged with a large meta-analysis demonstrating similar risks of progression with 6 versus 12 months of INH therapy. [66] Shorter combination therapies have demonstrated non-inferiority to traditional 9 month regimes and are advocated by the World Health Organization [67], [68], with new ultrashort regimes being evaluated. [69]

Identifying and treating latent infection is an important piece for solving the TB puzzle. The WHO strategic development goals call for TB elimination by 2050. It has been estimated that in order to reduce the global incidence 14 times by 2050, you would need to treat and protect only 8% of patients with latent infection. [70]

1.7.2 Active disease

Over the past 2 centuries since the tubercle bacilli was discovered, therapeutic interventions for Mtb have changed radically. Lacking effective drug therapies until the mid-20th century, initial approaches involved exposure to 'clean air' through the use of geographically located sanatoria. The development of which is considered in the next chapter. Some benefit was gained from this approach and may be related to altitude and vitamin D exposure. [71], [72] In addition, surgical interventions blossomed with targets including collapsing the afflicted lung and removal of the cavities caused by disease. [73]

Between 1943 and 1952, 4 drug treatments had been discovered which had become effective at treating TB; injectable streptomycin, para-aminosalicylate (PAS), isonicotinic acid hydrazide (Isoniazid, INH) and pyrazinamide. Rifampicin was discovered in 1957 followed by ethambutol in 1961. [74, 75] More than half a century later, the latter 4 drugs continue to form the mainstay of TB treatment.

Between 1948 and 1986 trials conducted by the US (public health) and UK (medical research council) services gradually established the optimal formulation of drug combinations and duration of therapy. Current treatment for active disease therefore consists of an induction phase (2 months) using all 4 drug treatments, followed by a maintenance phase (4 months) using isoniazid and rifampicin. Treatment duration was shortened from 18 to 6 months with the use of the 4-drug regime. Subsequent efforts to decrease this duration further have not proven successful. [76] [77] [31, 61, 78]

A 6-month course of multidrug therapy was found to result in cure of TB in the majority of cases, with relapse being unlikely and usually attributed to poor treatment compliance. [57]

One theory for requiring multiple therapeutic agents and longer durations is the concept of different mycobacterial subpopulations. These subpopulations exist in different metabolic states and respond differently, with metabolically active populations dying faster and slower, persistent populations requiring longer time periods and different agents for effective elimination. [79] [80, 81] A further suggestion is the sequestration of mycobacteria

to less drug accessible spaces e.g. thick walled granulomas. Drug therapies have different penetration capabilities and this might further explain the inability to shorten treatment durations with current regimes. [82-84]

TB therapy is challenging, not least because the duration of treatment is longer than for other bacterial infections, but also due to the side effects that occur using multiple drug regimens over long periods. Drug adverse effects are common, resulting in treatment interruption or discontinuation in up to 15% of patients. [85] It is estimated that up to 49% do not complete the treatment regime. [62]

Earlier studies also demonstrated the benefit of multidrug regimes over single agents, in preventing resistance disease development. [76] The development of drug resistance appears to be the result of random genetic mutations. [86]

Multidrug resistance disease (MDR), defined as resistance to at least rifampicin and isoniazid is a growing concern globally. Increasing drug resistance can result in extremely drug resistant disease (XDR) for which there are few therapeutic options available. [42, 87, 88]

1.8 Prevention

Detecting and treating individuals with active disease is seen as the most effective way to prevent on-going transmission and curb the global TB epidemic. However, this approach alone whilst reducing mortality rates, has not been shown to decrease the global TB burden. [89] As a quarter of the world's population is predicted to have latent TB disease. There is therefore a significant burden, and prevention of symptomatic disease in this group is key to reducing the global incidence of TB. [79], [32]

Contact tracing is the method by which public health services identify at risk individuals, those that have been exposed to active cases of Mtb disease. The currently adopted approach utilises a 'stone-in-pond' model, with each 'ripple' representing a social circle with varying degrees of physical proximity to the index case, thus suggesting a way to limit screening sizes in contact tracing scenarios. This involves screening individuals deemed to fall within different risk strata to the patient; thus, family is the most proximal relationship, is therefore seen to have the highest risk, and family members are screened as a priority, followed by close friends deemed to have the next highest exposure risk, casual contacts, and so on. Screening contacts thus proceeds from the innermost (most proximal/perceived highest risk) social circle to the least-related. [90-93]

Isoniazid therapy has been shown to prevent the development of active disease in a latently infected person. [64] This mono-therapeutic approach is preferable to the multidrug regimes used in active disease treatment due to fewer side effects. Contacts of patients with active

disease are at particularly high risk of developing infection. Modelling studies have demonstrated the population level impact that tracing and treating contacts can achieve. [94], [95]

Preventing TB will require a multifaceted approach. This is clear from looking at trends in TB incidence and particularly the correlation with socioeconomic decline and marginalised groups. [96, 97] Of note is the correlation of TB with HIV. HIV positive individuals have a much higher rate of TB than the general population and worse outcomes. HIV is the strongest risk factor for TB acquisition. Despite this, out of the estimated 33 million HIV suffers globally, only 1 million are on preventative therapy. [32, 98, 99]

1.8.1 Vaccination

Effective vaccine therapy could result in significant population benefit by preventing transmission of Mtb. [100] Despite an existing vaccine since the mid-20th century, significant progress on a more effective preventative measure has been lacking. The Bacille Calmette-Geurin (BCG) vaccine has been shown to offer moderate protection against Mtb disease but wanes over time. That said, the primary role was to prevent severe disease in children and this therefore forms the basis of the UK vaccination schedule. [101] The search for novel vaccine targets for Mtb continues [102, 103]

1.9 Thesis aims

The World Health Organisation goals to eliminate TB disease by 2035 requires a change in approach. We know this with given the discrepancy in the current trend of cases versus where we need to be to achieve this ambitious target. [42] Simply, while TB incidence was previously falling, it is not falling fast enough to meet the elimination goal. Furthermore, with newer treatments resulting in a cure for TB disease, identifying at risk individuals earlier, ideally prior to active disease development i.e. contact tracing to identify latent cases, has become more important.[104, 105] In addition, the increasing prevalence of resistant disease means there is a greater emphasis on earlier identification and intervention at a latent stage. This can result in reduction not only of morbidity and mortality but also reduce further spread of resistant strains.

1.9.1 Hypothesis

The current contact tracing approach using the 'stone-in-pond' model is not effective in identifying the largest number of at-risk contacts and is not appropriate for underserved population groups. Additionally, a contact tracing method should take into account the heterogeneous nature of congregate settings and by identifying risk factors in these settings, we can be predictive about TB acquisition risk.

1.9.2 Aims

1. Identify how did we arrive at our current contact tracing approach and whether there any alternatives.

2. Explore whether the current contact tracing approach is effective in congregate settings.

3. Assess whether we can be predictive about risk of TB acquisition to contacts in congregate settings

The first chapter will examine the historical developments and gradual evolution of our current contact tracing methods. I will explore how we arrived at the current contact tracing approach and how contact tracing itself has evolved as a tool to reduce and eliminate TB disease. I will do this by examining the trends of TB epidemiology in the UK over the past 200 years and the factors that contributed to a reduction in prevalence of TB.

By exploring the gradual transition from a high to a low incidence setting, the decreasing impact of historically derived contact tracing methods will be seen. Furthermore, the demographic of affected individuals will be seen to have changed i.e. the higher risk groups are not the population at large but rather demographic pockets of underserved populations in whom current contact tracing approaches has limited effectiveness. In addition, changing social practices and lower prevalence rates of TB increase the emphasis on more active contact tracing approaches. In exploring these themes, I will demonstrate that the current contact tracing approach is not the best method to achieve the WHO aims of TB elimination in our current TB prevalence state.

In the second chapter, I will examine the evidence behind current contact tracing methods and whether there exist any high-quality studies in support of the current methods as well as whether there are any alternatives that have been examined in head to head trials. I will do this systematically using accepted systematic review methodologies. I will demonstrate in this chapter the lack of comparative data or high quality supporting evidence for our current contact tracing approach. This will follow on from the conclusions drawn from the evolutionary aspects of contact tracing namely that they evolved organically due to demographic needs and have not adapted with changing epidemiology, not undergone further evidenced based scrutiny.

The third chapter will look specifically at non-household (congregate) settings and whether these have been examined as having unique characteristics over and above household settings for disease transmission. This is to acknowledge the changing areas of transmission of TB disease and the primary risk areas for the broader population. This better reflects areas for underserved populations to transmit disease.

Given social mixing and the potential for larger groups being exposed and therefore infected, as well as the implications to the health service and limited resources of contact tracing units, it is important to understand if there are risk factors unique to these settings. The risk factors identified will inform further chapter themes.

The fourth chapter will look at the database for my study. In order to demonstrate the broader applicability of this local data to national settings, I will compare our data to

nationally published figures (Public Health England – Annual Reports) and whether there are unique differences. This will enable a description of the scope and scale of the data and the limitations in terms of missing or incomplete data. I will also look for any patterns in the data and how these hold up to current assumptions about transmission and disease outcomes in the congregate (non-household) settings investigated.

The fifth chapter will build further on the data, in particular looking at a subsection of cases that fall into the category of outbreaks (more than 2 active cases in a congregate setting). I will compare prospective (epidemiological) contact tracing approaches (current methods) with subsequent disease development. Utilising WGS data, we will be able to link cases and demonstrate transmission events retrospectively. This will enable an examination of the effectiveness of the current contact tracing methods by examining how many missed cases occurred and consequences in terms of subsequent infections.

The next chapter will utilise the same database to derive a prediction model based on the currently collected data. In doing this, I hope to demonstrate that risk of transmission can be predicted and that this is linked to social relationship status in addition to other well recognised factors such as sputum positivity and HIV status. This will highlight the relevance of data collected by contact tracing services and how this latter process can be improved or augmented for a more optimal outcome i.e. fewer active cases developing by identifying at risk contacts.

Finally, in the conclusions chapter, I will summarise how my aims have been addressed, synthesising the arguments made from the above chapters. Given the significant impact the recent Coronavirus Pandemic-2019 has had on healthcare systems and public health measures, a postscript will also be included in the conclusion. This will comment on the impact the pandemic has had on TB disease in the UK as well as its broader impact on contract tracing conduct and strategies.

Chapter 2: The evolution of tuberculosis control and prevention

In this chapter I aim to trace the origin of contact tracing and TB control and demonstrate how the original intent and effectiveness of contact tracing approaches have become outdated by the changing epidemiology in which contact tracing occurs.

2.1 Introduction

Mycobacterium tuberculosis (MTB) is an infectious agent that has plagued humans for many millennia, with archaeological evidence suggesting prehistoric disease. [22] In the 19th century, infectious diseases and in particular, tuberculosis (TB) were the leading causes of mortality. [106] Today, MTB is estimated to have infected a quarter of the world's population with 10 million new cases and over a million deaths a year. [107]

TB disease is spread via respiratory droplets from an infected person to susceptible individuals. Rarely, different strains of disease can be acquired from animal sources.[108] Once acquired, the infection proceeds through to a latent phase (an asymptomatic state) and finally to active disease. This latent phase varies between individuals and can occur at any point in their lifetime, but usually occurs within 2 years. [109] Given that there is a pre-symptomatic state for which treatment exists, early identification of this state and prompt treatment can prevent new cases developing. The identification and treatment of latent cases is one of the ways that the total incidence of TB disease can be reduced. How we

identify these latent cases and in fact our knowledge of a latent state is the result of gradual accumulation of scientific knowledge over the last 2 centuries.

Contemporary methods for TB control have developed over these past 2 centuries. Current contact tracing practice does not reflect modern societal interactions and relationships, nor the occurrence of cases in congregate settings. This chapter aims to chart the steps over this period and demonstrate the key changes that have influenced control and prevention changes. In doing so, the current contact tracing practices can be seen in their historical context and a more pragmatic approach considered.

Section 2 examines the antiquated understanding of TB and the impact germ theory had on TB control as well as the pragmatic measures that influenced declining rates. Section 3 looks at how a dramatic understanding of latency changed preventative interventions and how vaccination and treatments changed the epidemiology of TB. Finally, section 4 draws together these elements and examines gaps in modern practice as a result of these epidemiological changes with a view to suggesting potential solutions.

2.2 Understanding an ancient threat

Mycobacterium Tuberculosis (TB) has shadowed humanity's development from our early hunter gather days, through early civilisation and the industrial revolution and remains with us today. Potentially causing the most deaths from a microbial pathogen in history, rates of TB have waxed and waned over the many centuries.[22, 106]

Ideas and concepts of TB can be found in ancient medical writings from Hippocrates to Galen. [106]These early practitioners describe the symptoms and appearances of the disease which can still be recognised today. [22] Despite knowledge of the disease's symptoms and manifestations, the cause of TB remained largely unknown.

It was in the late 18th and early 19th centuries when several experiments demonstrated the ability to inoculate creatures with TB from infected lesions and thus cause the disease to develop that a pathogen was suspected.[22, 106] Despite this, some views held that there was a hereditary element to acquisition.[106] This confusion arose as a result of the poorly understood mechanisms of TB. Unlike smallpox and cholera, TB did not result in epidemic spikes, given its latterly identified latent state. This complicated the understanding of TB and therefore its control and prevention.

This early debate about the source and causative elements of TB remained until 1882 when German scientist, Robert Koch, established the causative organism as the tubercle bacillus a discovery that was lauded contemporaneously. The debate over the cause of TB had ended. The concept of germ theory had begun. [110]

2.2.1 Early attempts at prevention and control

Despite the early lack of consensus as to the cause of TB, effective control measures already existed, preceding Koch's discovery. Public protection measures reflected less an

understanding of the nature of the disease being treated and more the pragmatic efforts that had demonstrated past effectiveness. Legislation included a change in the role of authorities in the provision and monitoring of public health such as; the 1872 Public Health Act which advanced sanitary matters and addressed environmental causes; the change of the local government board to the Poor Law board in 1871 aimed at helping the most vulnerable members of society; and the 1889 notification of infectious diseases (consolidating currently existing local acts for infectious disease notification). [111] [112]

Once identified, a patient with TB would be isolated and their belongings quarantined. This approach was reinforced by the emergence of germ theory which influenced legislation against public spitting, advising against human contact when infected and limited the parental role (e.g. no kissing or suckling for mothers). [113, 114]

Apart from individual measures, the food environment also came under scrutiny. A medical officer survey in 1887 demonstrated that up to 85% of medical officers considered raw milk and meat a significant source for TB infection. TB in cattle was described in the early 1800's. It wasn't until 1902 however that *mycobacterium bovis*, the TB strain in cows was detected in children with disease.[115] Where features were suggestive of infection, these animals/ herds were summarily destroyed and pasteurisation of milk ensured added protection. [108, 116] In this way, infection from contaminated meat and milk was reduced, contributing to diminishing rates of infection and potentially, disproportionally reducing rates in younger groups.

Therapeutic options at this early stage were limited. Isolation and improved ventilation and fresh air were seen as the primary interventions. Any case notification usually prompted a home visit by authorities to either remedy ventilation issues or in the case of a death at home from TB, to effect disinfection. [113, 114] Some of the ideas proposed to control TB at this time included: compulsory registration of cases to enable cleaning of abodes; examination of workers in large industrial sites e.g. factories to 'remove infected persons'; and compulsory care in state facilities. [117]

The emergence of facilities to achieve these objectives came in the form of sanatoria. Sanatoria were already being established in most towns in the UK around 1873 and authority for their management was delegated to a medical officer of health. [111] Section CXXIV of the sanitary authority legislation provided for the compulsory detention of a tuberculous patient in a hospital. This was particularly the case for certain groups: those without homes; where the infected individual occupied a room in which there were multiple families; if they were residing on a ship; and where the hospital was willing to admit such patients. Sanatoria care was limited to fresh air and respite, interventions seen to provide the best treatment for TB. [106]

The isolation through detention of an infected individual was seen as the best way to protect society at large as well as providing an environment to improve the health of the infected individual. [118, 119] Patients first had to be identified and notification of cases was called for by clinicians. Initially voluntary, compulsory notification was eventually instituted, initially in New York but soon after in the UK. There were objections to compulsory

notification due to the stigmatising nature of the illness and the impact this could have on socioeconomic status including prospects for marriage and even life insurance.[115, 120] Contributing to this poor social outlook was the widely-held belief that TB was hereditary, a view that was often supported by the observation that infections tended to cluster around family units. [115, 121, 122]

Alongside these changes, before the close of the 19th century, a further key development came in the form of the TB dispensary. The Scottish physician Robert Phillip is credited with the first such facility in 1887, in Bank street, Edinburgh. Designed essentially as an outpatient clinic, patients would attend and be reviewed by clinicians usually resulting in symptomatic treatment (e.g. cough mixtures). Having a greater interest in the control of TB than his colleagues, Phillips expanded the role of the dispensary in several key ways. Firstly, they provided education for patients and relatives, helping the identification of symptoms as well as demonstrating interventions to reduce spread of disease. Dispensaries also provided for home visits. Perhaps the most significant role however came in the form of a novel and highly prescient intervention with the examination of family contacts of TB patients for any evidence of disease. While lacking any systematic approach, this would closely mirror modern day contact tracing practice. [115, 123-125]

Further research conducted at dispensaries delineated the relationship between TB cases and their contacts. This further weakened the idea of hereditary tuberculosis. [126] TB spread was increasingly recognised as spreading through environmental, habitual and

occupational factors. This was reinforced by the high prevalence of TB in dusty or poorly ventilated environments. [112, 127, 128]

The difficulties of addressing TB presenting in public and congregate settings posed a problem at an early stage and the role of medical inspectors was expanded to allow intervention in the early 1900's. The same approaches used for household outbreaks were generally applied to these congregate populations where medical officers worked alongside factory inspectors, with isolation and subsequent disinfection being the mainstay of interventional practice. [129] Without specific approaches to congregate settings, officials adopted the techniques developed in household settings. Potential exposures in congregate settings therefore evolved along the limited and non-analogous lines of household tracing. Laws targeted among other factors; factory inspection, tenement housing and ventilation. [130-132]

Towards the end of the 19th century TB cases were seen to be decreasing.[22, 106] These changes need to be seen in the context of concurrent social developments. The UK was in a maturing industrial revolution. TB rates had initially increased with the influx of population from rural to urban areas. Crowded working conditions, long working hours (and therefore exposure potential) and decreased ventilation likely contributed to infection acquisition. As the industrial revolution matured, average wages increased and workers could afford better quantities and quality of food as well as larger, less crowded, living spaces. Working acts such as the 1912 Public Health Tuberculosis Regulations and the Great Insurance Act 1911 (enacted 1912) also improved the working environments. [131] These resulted in

compulsory inspections with consequential segregation of patients in sanatoria. These positive factors were further assisted by legislative acts such as the National Insurance Act and schemes like the sanatorium benefit which increased health care access for poorer populations. [133] With the decreasing numbers of cases, the epidemiology of TB changed. Control and prevention measures would have to adapt to continue this downward trend.

2.3 Understanding TB states – A new approach to prevention and control

Sanatoria remained the only viable therapeutic option at the turn of the century. When and for how long a patient was confined to sanatoria care were issues commonly encountered. Koch highlighted the issues related to early discharge of advanced disease (given the lack of curative intervention) and the prohibitive cost of care both for the maintenance of the facilities and the direct cost to the patients. Early intervention in the course of the disease began to gain support and the acknowledgement that an earlier intervention made more economic sense in returning the tuberculous patient to be the primary 'bread winner' instead of a prolonged stay in a sanatorium that ultimately ended in a patient's death due to the advanced nature of the disease. [134] It was noted that up to 20% of those treated at an early stage of disease lost the tubercle bacilli from sputum, which was regarded as the only sure measure of treatment success. [115, 121, 122]

The search for other therapies proceeded at pace. Initially designed by none other than Koch himself, the first such failed attempt at a treatment that would subsequently enjoy a different and crucial role was Tuberculin. A form of TB that elicited an immune response

didn't provide the promised cure, however, Tuberculin's utility was seen when attempting to limit the spread of TB, beginning with purifying food and drink sources. [115] When exposed to tuberculin, cows would develop a skin reaction if infected with TB (reactors). [108] Crucially, this was seen even in seemingly healthy herds. 'Reactors' therefore allowed the destruction of herds and subsequent pasteurisation reduced infections in the food chain.

The notion that individuals could harbour the disease but appear well i.e. without symptoms (asymptomatic) representing a latent form of infection was a new and radical point in the attempt to prevent the spread of TB. The term latent infection was first proposed by von Pirquet in 1909 after his experiments with tuberculin in children in the early 20th century. These demonstrated positive reactions in asymptomatic children at a rate of 80%. [106, 135] Authors began to acknowledge the high burden of disease in carriers, contacts who had not yet manifested clinical features of active disease (latent cases). [136] This development cannot be understated as it provided the basis to follow up asymptomatic contacts.

Using Tuberculin as a basis, two scientists, Calmette and Guerin arrived at a formulation from bovine TB that appeared to be protective against TB. Their first test in 1921 in a baby exposed to maternal and grand-maternal TB proved successful. The road to universal vaccination was however far from straightforward and following a disastrous outcome in 1930 in Lubeck, Germany where a contaminated batch of vaccine killed 75 children, there was much less enthusiasm and uptake. [137]

The uptake of BCG vaccine across most of Europe increased steadily despite this, following impressive programs in Scandinavia. Post war Britain slowly began to utilise BCG as part of the welfare initiatives. Even at this stage, it was used to incentivise recruitment to sanatoria (nurses), for those deemed at risk (medical students) and importantly for any contacts of an infected case who did not react to Tuberculin i.e. asymptomatic, tuberculin negative patients.[138, 139]

Vaccination was gradually rolled out to other groups particularly school children in the UK. America did not adopt BCG, choosing instead to focus on individual responsibility for health and the insistence on tried and tested precautions. They believed, after several equivocal trials that it was better not to have the false sense of security provided by a dubious vaccine. [137]

Following on from work with Tuberculin, Opie et al investigated the epidemiology of TB in households and the spread within families. The authors established the contagion of TB in families and how latent TB disease is a transmitted event that can convert to open/ active disease at a future time point. Finally putting to rest a hereditary link. [140, 141]

In a post war world, TB cases increased, stimulating a need for further intervention. [142, 143] This took shape in the form of mobile x-ray units and public health campaigns resulting in population wide screening events. TB epidemiology began to change with fewer younger people presenting and higher rates in predominantly older populations and among at risk groups e.g. prisoners, the homeless and medics. As these mobile screening units (often

portable x-rays on buses towing generators) resulted in a decrease in the number of cases found and total TB cases began to drop, costs increased. Cost-effectiveness arguments resulted in a gradual limitation of large public screenings to perceived at risk population groups. Two methods were employed to identify these cases. These were either 'surveys' of large social groups, in particular high school children and college students or through screening contacts of TB cases. Contact screening was seen as providing a greater yield of cases. [144]

Contacts were not restricted to households. While younger children are likely to have been exposed in the household, older children and other individuals might have acquired TB from outside of the home. Health departments therefore moved away from blanket tuberculin skin testing in children as a means of case finding and instead tested all patients with known TB contacts. [145]

Patients presenting to chest clinics (the modern dispensary) were either contacts of TB cases or known or suspected cases of TB. [146] [147] Regulations approved at this early stage allowed the clinic nurses to visit family homes and examine contacts. With contacts, a tuberculin skin test was used as a screening tool. Children under 12 received a tuberculin skin test and only positive contacts received a chest x-ray. Periodic follow up of contacts and of patients without known contact but with asymptomatic lesions was advocated. [148, 149]. Routine pre-employment screening with x-rays reduced the number of advanced cases of disease ensuring earlier treatment and isolation with concomitant reduced contact time with susceptible individuals. The rudimentary costs of contact tracing and the high number

of early stage disease diagnosed in contact follow up were highlighted as a cost effective way to approach case finding in TB, with reports of early (incipient) TB being two times greater than suspect cases. [150]

2.4 New treatments – A changed outlook

By the 1950s, 3 highly effective agents existed for the treatment of TB, beginning with Streptomycin. These antimicrobials changed the outlook for TB patients, where progression of disease despite sanatoria care usually had an invariably fatal outcome. [151]

As trials demonstrated the effectiveness of treatment and the economic benefit, the relevance of sanatoria came into question. Patients were seen in chest wards and outpatient clinics saw increasing numbers of TB patients. Home treatment began to gather traction with the increasing recognition of the effectiveness of emerging chemotherapeutic agents and the need to integrate patients back into society. [152, 153] This was supported when a comparison study conducted in Madras, India, demonstrated no difference in patients treated in sanatoria compared with those managed in the community with close supervision. [154] Isolation of TB patients was still advocated, however, this now began to take place in their own home with separations in rooms as well as bedding and cutlery. Compliance with therapy was closely supervised, foreshadowing the introduction of directly observed treatment (DOT), a cornerstone of modern TB intervention.

While clinicians understood the utility and effectiveness of anti-TB multidrug regimes, economic realities limited their widespread implementation. The pressure to diagnose TB before it became active and required a multidrug approach or infected others increased further. Chemoprophylaxis in the form of a single drug, Isoniazid, was used to prevent exposed and latent cases from developing active disease. Contact tracing therefore had an effective treatment option. [23] contact tracing of 'the immediate contacts' began to be nationally recommended in the US as early as 1963 as it was noted these individuals represented a high-risk group for TB. [155] Family members were screened using the tuberculin skin test with clear guidance as to what to do with positive results (negative x-rays had follow up for 2-3 years) and negative results (received BCG vaccine). Further guidance began to develop specifically providing advice as to what level of exposure patients with TB could undertake including when and if they could work whilst undertaking their therapeutic regimes based on infectiousness. [24]

By the mid 1970's TB control began to reach its modern-day equivalent. Long term inpatient care in specialist facilities was phased out and a much shorter inpatient stay in general facilities with an early discharge to ambulatory treatment was the recommendation. The American Thoracic Society issued guidance helping patients reintegrate into work and school. [156-158]

2.4.1 Modern age of TB care

Up to 1985 the trend for TB demonstrated a dramatic decline in incidence. This was not universal, with demographic pockets developing as interventions and control improved. At the turn of this decade, the combination of immigration and the acquired immune deficiency syndrome (AIDs) epidemic served to increase TB in certain demographic groups. Caucasian natives were seen to have the lowest rates of TB compared to with other ethnicities. Lower socioeconomic status and social risk factors also conferred a higher risk. [159] This worsened with changes in social welfare programs and population increases in congregate settings such as prisons. The greatest issues were seen in those with complex personal and social circumstances including those with variable living arrangements/ homeless. [160, 161]

Towards the end of the 20th century, TB transmission to proximal contacts was widely accepted as was the disease process existing in a dormant (latent) phase in asymptomatic contacts. Following an outbreak in Norway, a systematised approach to contact tracing was proposed. Using the analogy of the ripples resulting from a stone dropped in a pond, a screening priority for contacts could be assigned with those deemed most proximal having the highest risk and therefore the greatest need for screening. This stone-in-pond model of contact tracing gained approval, tying in with the general concepts of TB transmission and having a logical basis in assessment and treatment as well as an acknowledgement of risk for transmission. Since its publication by Veen et al, this has been the de-facto approach for TB services and remains so albeit with some modification. [90]

While the stone-in-pond approach has been widely adopted, there remain issues with identifying and subsequent tracing of contacts. For one, the ability to identify contacts relies heavily on the infected individual assigning proximity to their contacts. While this subjective approach would work well for household environments, it proves inadequate for wider social networks as well as congregate settings and transient contact episodes. [54] Additionally, TB can spread to individuals not acknowledged or identified as contacts by an index case.

2.5 Conclusion

Efforts to prevent and eliminate tuberculosis have seen many trends and changes. Initial efforts focused on broad public health interventions involving sanitation. The discovery of the tubercle bacillus encouraged health care professionals and governmental bodies to consider the infectious and transmissible nature of the organism and this resulted in steps to reduce factors perceived to propagate spread. This involved isolating the source of spread and sanatoria facilities expanded to contain the large number of infected individuals.

With the discovery of household spread of disease and the rise of dispensaries to monitor patients as well as the widely-accepted notion that early detection and intervention had better outcomes, visiting the homestead and assessing family units for TB became vogue. Diagnostic developments in terms of the tuberculin skin test and x-rays refined this screening process and enabled the earlier detection of pre-symptomatic lesions and generated a new population of exposed but asymptomatic and lesion-less individuals (latent

cases). When combined with new drug treatments, these new approaches led to a dramatic decrease in the cases of TB in the post war era.

This decline shifted the approach of TB prevention and control with more focused screening programs as evidence amassed that assessing individuals in contact with infected cases was more fruitful than screening the general population. Contact tracing and guidance on assessment and treatment of individuals reached its current state in the mid-70s to 80s.

I have described the multifactorial decline of TB over the past 2 centuries. Where once TB in the UK was a high incidence disease, it is now a low incidence problem. [162-164] Despite the advances made, TB persists. As the epidemiology of the disease changed and as new technologies and techniques came to light, interventions adapted to further drive down this endemic disease. As numbers of new TB cases decreased however, success brought about an apathy for new approaches. Questions as to whether past approaches adapted for the modern era still apply should be asked. Gaps with interventions still remain in workplace and congregate settings. Epidemiological changes have continued with international transport facilitating immigration from high risk countries and societal disparities contributing to persistent demographic pockets of disease.

New approaches attempt to bridge this gap by taking into account not just proximity of exposure but a wider geo-temporal component. Social network analysis is one such approach aiming to extract when and where a patient (index case) may have been and then identifying who may have been exposed as a result. [54, 165] In terms of TB contact tracing,

social network analysis is still in its early stages. When combined with novel diagnostic techniques such as next generation whole genome sequencing, clinicians are able to link cases both epidemiologically and through TB strain genetics. This allows an outbreak to be better defined and appropriate resources brought to bear. [54, 166]

Taking these changes into account requires a more nuanced approach at odds with the rigid contact tracing processes currently in practice. Social network analysis provides a link between two key elements of public health, namely that of social factors involved in disease and the clinical aspects of diagnostics and treatment. These factors have often been described as competing causes for a decline in TB, perhaps together they can help eliminate it.

In the next chapter I will look at the factors that are predictive of TB developing in congregate (non-household) settings and whether this has been examined previously.

Chapter 3 - What are the factors associated with an increased risk of disease acquisition in contacts of Tuberculosis patients in congregate environments in low incidence settings – A systematic review

The aim of this chapter is to review the existing evidence for factors associated with TB transmission in congregate (non-household) settings.

3.1 Protocol

3.1.1 Introduction

3.1.1.1 Background

As previously described above, Mtb is a global health threat. We can see from the previous chapter that the cause and nature of the spread of TB was the matter of significant debate prior to the discovery of the tubercle bacilli. It is now acknowledged that TB is transmitted through the air via respiratory droplets from an infectious patient (index patient) to susceptible contacts who share the same space. [167] Exposure to TB can result in no infection, active or latent TB infection (LTBI). The latter demonstrating no clinical symptoms or radiological evidence of disease but representing a risk of developing active disease (approximately 10% of infected cases). [167] Contacts of patients with TB represent a higher relative risk of developing disease when compared to the general population in low incidence settings (<10 cases per 100, 000 population). [168] Effective investigation of tuberculosis (TB) contacts is therefore essential for continued progress toward TB elimination. [34]

3.1.1.2 Transmission

Proximity and duration of exposure to infected patients have long been recognized as significant factors for disease acquisition. [169] While household contacts are certainly at high risk, this risk can change over time, with index cases younger than 15 years of age demonstrating acquisition and transmission in a household setting and older children and adults transmitting disease in congregate settings. [170, 171]

3.1.1.3 Risk factors

Transmission of TB and risk of acquisition has multiple related factors. This can be viewed in separate components; the index case and disease characteristics; the contact and their susceptibility and the geospatial location where transmission occurs.

The index case can have a varying degree of disease burden. Higher burden of disease and therefore greater risk of transmission represented by pulmonary disease and sputum positivity (smear positive disease) as well as shorter time to culture. In addition, untreated cases are at much higher risk of transmitting disease as are index cases with a positive HIV status. [172-175]

The contact may possess relative susceptibilities that impair their immunity and predispose to infection. These include but are not limited to HIV positive status, chronic kidney disease with dialysis dependency increasing risk, use of immunosuppressive therapy and pre-existing pulmonary disease as well as smoking. [176-181] Their proximity to the index case increases their risk, with greater proximity and time of exposure correlating with disease acquisition.

In addition, preventative measures such as previous BCG vaccination will modify risk of acquisition [173, 174, 182, 183]

A further aspect of disease transmission rests with geospatial location. Several studies have demonstrated the risk posed by congregate settings i.e. areas of public gathering. Locations such as schools have been shown to represent a greater risk of transmission than household settings. Given the propensity of low density droplets to circulate in poorly ventilated settings, locations where large numbers of people congregate can increase transmission events. [169, 170, 182, 184, 185]

3.1.1.4 Contact tracing limitations

Identifying the contacts of Mtb patients usually occurs once an index case has been identified. Current methodology follows a 'stone-in-pond' approach. This involves evaluating contacts in concentric rings correlating with proximity to the index case with the closest 'ring' having the highest proximity. The contact tracing process stops once a 'ring does not demonstrate any transmission. The key assumption here is that the highest risk contacts are household contacts representing the closest ring and that relationships between index and contacts are linear and stable. In addition, the identification of contacts relies on subjective reporting by the index case. [90] Ultimately this means that we cannot be certain about the estimation of significant exposure risk in congregate settings.

These traditional methods may not take into account areas of congregation (i.e. non-household settings of social aggregation including work, educational or recreational

facilities). Significantly, households may not represent the key transmission zones in urban settings and traditional contact methods may miss non-household transmission nodes and additional contacts. [54, 169, 170, 185, 186]

In order to further reduce TB incidence in a low incidence urban setting, a better understanding of transmission in non-household settings is required. While traditional risk factors for disease acquisition have been described, there are not any systematic reviews examining the risk for disease acquisition in congregate settings as opposed to household settings. A search conducted on PubMed and the Cochrane and PROSPERO databases found no other systematic reviews (quantitative or qualitative/ narrative) examining this issue. Informing contact tracing processes in low incidence settings in this manner would allow more effective prioritisation of contacts and improve the overall yield of contact tracing.

Reducing the incidence rate of TB is essential to eliminating this disease. [34] In low incidence countries, contacts represent a higher risk group than the general population and non-household (congregate) settings have been described as areas of high transmission. [182]

This systematic review will collate literature in order to establish the risk factors involved in congregate setting transmission zones. Reducing the effective contact rate could be key to decreasing the on-going incidence of TB in low incidence settings [187] and understanding the factors involved in congregate environments could inform methods to effectively

contact trace in these settings. This review will examine if there is a lack of literature on this topic and collate the available evidence.

If this systematic review finds additional risk factors associated with congregate environments, this could positively inform current contact tracing approaches and control and prevention measures in congregate settings. In addition, factors discovered may assist in risk stratifying contacts and the more efficient allocation in resources. It is likely that there will be a lack of good quality data in which case this systematic review will identify where further research is required. In the event that there are few analytical studies examining this topic, our review may inform subsequent analytical study design.

3.1.2 Aim

The aim of this systematic review is therefore to identify the index, exposure, contact and location factors associated with an increased risk of disease acquisition in congregate settings in a low incidence location.

3.1.3 Associated review articles

I searched the Cochrane database and PROSPERO. No systematic review articles covering our study aim were found. In addition, I searched PubMed central under the search terms: “tuberculosis” AND “transmission” AND “review”. 1726 articles were found. No systematic reviews addressing our aim were discovered.

A number of review articles were identified which provide useful background although these did not follow a systematic review method.

Jelip et al examined risk factors for acquisition of TB in health care workers in a medium to high incidence area [188] along with Uden et al. They noted the increased risk of disease acquisition compared with the general population. This supports the notion of increased risk in congregate settings. [189]

Baussano et al examined the risk in prisons for TB transmission and observed an increased relative risk of TB acquisition for prisoners. [190]

Several authors examined the risk of TB transmission on aircraft and public transport. Surucuoglu examined, through modelling, the risk of TB acquisition and theorised the increased risk in poorly ventilated settings e.g. land based transport (buses and trains) versus ventilated transport (aircraft). [191] Kotila [192] and Abubakar [193] looking at TB transmission on aircraft concluded there was relatively low risk of transmission given the ventilation on newer aircraft. Kotila further observed that the highest risk was posed to the closest individuals (2 rows of the index case). [192] Edelson concurred with the previous authors, observing that there is an increased risk with public transport associated transmission events. [194]

Sane Schepisi et al examined tuberculosis transmission in a school based congregate setting amongst children and adolescents. They found a risk of TB disease of 0.03% and of latent

infection of 0.15%. Whilst lower than household settings, this remains a significant risk of transmission that needs to be addressed. The authors review was limited to a single type of congregate setting and did not examine risk factors in detail. [195] Martinez et al conducted a systematic review looking at TB transmission in the household and the community. Their review was around children exposed and unexposed to a household contact. It therefore does not examine the broader population (adults) nor does it examine risk factors specific to congregate settings. [196] Roberts et al also examined TB transmission in children and demonstrated a higher risk of transmission in school congregate settings with a child index case versus an adult index case. They did not however, look at relative risk factors nor examine alternative congregate settings. Additionally they limited their population to children aged 3-11 years of age. [197]

Nava-Aguilera et al examined risk factors associated with transmission of TB. The authors examined a spectrum of risk factors and describe traditional factors associated with TB disease (including HIV, incarceration, homelessness and drug and alcohol use). They did not look at particular risk factors associated with transmission in congregate settings and instead describe general risk factors. [198] Fok et al examined risk factors for TB clustering. The authors looked particularly at clusters of TB, not specifically applicable to congregate transmitted events given the different nature, location and risk factors involved. In addition the cluster events were not limited to low incidence settings and therefore transmission and risk identified would vary from low incidence congregate setting populations. [199] Faustini et al examined the risk factors for multidrug resistant TB (MDR-TB) transmission in low incidence settings. The authors limited their review to multidrug resistance disease which

may have different factors associated with transmission. In addition, the focus was not on congregate settings but rather resistance in disease. [200]

Nardell et al in their review article examined tuberculosis in congregate settings. They identified factors used to assess the risk of disease in these settings. The authors paper was published in 1989 and does not follow a systematic review process. Whilst they examine congregate settings they do not provide the basis for the risks described in a systematic process and given the time lapse, the evidence base may potentially have changed. [201]

3.1.4 Methods

This review followed the methods stated in the Preferred Reporting Items for Systematic Reviews and Meta-analyses (PRISMA) statement. [202]

3.1.4.1 Participants/population

The study population consisted of contacts in non-household (congregate) settings of patients (index cases) with tuberculosis infection in low incidence countries (i.e. countries with a TB incident rate of less than 30 per 100, 000 population). [32] These were to include males and females of all ages. No other limits were used for setting.

Tuberculosis infection in this context is confirmed pulmonary disease (to include laryngeal, bronchial, pleural and miliary disease) in the index case.

3.1.4.2 Exposures

Characteristics of the index case (e.g. age, sex, comorbidities), characteristics of the exposure (length and duration of encounter) to the index case or characteristics of the contact (e.g. age, sex, comorbidities. Characteristics of the location (e.g. office space, prison cell proximity, ventilation characteristics).

3.1.4.3 Comparator(s)/control

Contacts of patients with pulmonary tuberculosis in non-household (congregate) settings without these characteristics who do not develop disease.

3.1.4.4 Outcomes

Disease development of either latent or active Mtb infection. MTB latent infection (defined as positive tuberculin skin test (TST) and/ or interferon-gamma release assay (IGRA)) or active infection; clinical, microbiological and/ or radiological confirmation.

3.1.4.5 Eligibility criteria

Types of study to be included

Studies were eligible for inclusion if they reported on contacts of TB index cases in non-household (congregate) settings and identified risk factors relevant to contacts' acquisition of TB in non-household (congregate) settings. Studies needed to have a comparator group and therefore descriptive studies including single case reports, case series, cross-sectional (non-analytical) surveys and qualitative studies were excluded. Analytical studies to be included:

Case-control studies

Cases were considered to be non-household contacts who are diagnosed with Mtb (latent or active). Mtb latent infection (defined as positive tuberculin skin test (TST) and/ or interferon-gamma release assay (IGRA)) or active infection; clinical, microbiological and/ or radiological confirmation. Controls were considered to be non-household contacts who were not diagnosed with Mtb (no evidence of active or latent infection on screening). Exposures in this context were characteristics of the index case, exposure or contact.

Cohort studies

Exposed individuals were to be non-household contacts with specific characteristics (of index case, exposure or contact) thought to be linked to the risk of transmission. Unexposed individuals were to be non-household contacts without these specific characteristics

The outcome will be the diagnosis of Mtb in non-household contacts (latent or active disease). Mtb latent infection (defined as positive tuberculin skin test (TST) and/ or interferon-gamma release assay (IGRA)) or active infection; clinical, microbiological and/ or radiological confirmation.

Analytical cross-sectional studies

Exposed individuals were to be non-household contacts with specific characteristics (of index case, exposure or contact) thought to be linked to the risk of transmission. Unexposed individuals were to be non-household contacts without these specific characteristics

The outcome was the diagnosis of Mtb in non-household contacts (latent or active disease). Mtb latent infection (defined as positive tuberculin skin test (TST) and/ or interferon-gamma release assay (IGRA)) or active infection; clinical, microbiological and/ or radiological confirmation.

Controlled trials (randomised and non-randomised)

Experimental trials were to be included where the intervention targets non-household contacts and aimed to reduce acquisition of Mtb disease (protective interventions). Or examines susceptibilities in non-household contacts.

Other studies

Other studies of non-household contacts with a control or comparison group where risk of TB is compared between contacts with and without specific characteristics were also planned to be included.

Where studies included both household and non-household contacts, they would only be included if the non-household data could be separated and had a sub analysis. Where non-household and household data was not distinct, studies were excluded. This reflects the confounding that can occur when taking into account household settings and their contribution to subsequent Mtb disease development.

Report characteristics

Only human studies were included. Where animal to human transmission had occurred, these studies were excluded given the differing factors involved. Full papers were to be included. No limits were used for language nor any limits placed on search dates.

The inclusion checklist can be seen (table 1).

Table 1 - Inclusion checklist

	Inclusion	Exclusion
Study design	RCT cohort case control comparator study analytical cross sectional Other – please specify	Case reports Case series Observational trials Non-human studies
Report characteristics	Full article Conference abstract (2010 to date) Thesis Other - please specify	Conference abstracts 2009 and earlier
Participants	Contacts of patients with proven TB disease Human studies Separate analysis for non-household setting where mix of household and non-household settings identified. Low incidence country Index case has confirmed pulmonary TB disease (including; laryngeal, bronchial, pleural, miliary disease)	Non-pulmonary disease (to include laryngeal, bronchial, pleural or miliary disease) in index case OR non-confirmed disease in index case. No contacts Non-humans Only household setting Mix of household and non-household setting with no separate analysis.
Comparator	Comparator group present Contacts of patients with pulmonary tuberculosis (including; laryngeal, bronchial, pleural, miliary disease) Separate analysis for non-household setting where mix of household and non-household settings identified.	No comparator group Only prevalence studies
Exposure	Characteristics of the index case leading to transmission Characteristics of the exposure to the index case leading to transmission Characteristics of the contact leading to transmission Characteristics of the location leading to transmission	No Individual, disease, contact or location characteristic identified No assessment of transmission
Outcome	Measure of disease development in contacts	No assessment of disease transmission

3.1.4.6 Search strategy

The search strategy was designed in collaboration with a medical librarian and through several extensive scoping exercises. The following elements will be focused on;

Tuberculosis, contacts, transmission, risk factors and congregate setting. Alternate wording, phrases and spellings as well as plurality based on the key concepts will be used. The following electronic databases will be searched:

MEDLINE (Ovid)

Cochrane Central Register of Controlled Trials (CENTRAL)

Cochrane Database of Systematic Reviews (CDSR)

EMBASE

Web of Science

PubMed

The search strategy for Medline is included in the appendices. Search strategies for remaining databases were adapted accordingly.

Additional searches

These were conducted on the references/ bibliography of selected articles including associated systematic reviews. Abstracts of conferences in international meetings focusing on TB meetings and conference proceedings were searched through web of science. Electronic databases of government publications (clinical guidelines and CDC). Ongoing clinical trials (clinical trials.gov) and dissertations/ theses (through ProQuest Dissertation

Thesis Database and thesis.com) were also searched. No limits were used for, language, year of publication or setting.

3.1.4.7 Data extraction (selection and coding)

All study abstracts and citations identified by the above search criteria were screened by myself and independently confirmed by a second investigator (AM). Full texts of the studies meeting the eligibility criteria were to be obtained. None met the criteria. Exclusions and reasons for exclusions are recorded below. This process was facilitated using the referencing software, Endnote.

3.1.4.8 Risk of bias (quality) assessment

A quality assessment was to be carried out developed in accordance with the Cochrane Collaboration recommendations. [203] Therefore a risk of bias judgement was planned to be given instead of a score. Where non-randomised controlled trials were planned to be included, a ROBINS-I would be used, a Newcastle-Ottawa scale was to be used for any cohort or case-control studies identified.

3.1.4.9 Strategy for data synthesis

Background reading and scoping exercises has led to the conclusion that there is substantial heterogeneity in both study type and outcomes. Where quantitative analysis was predicted and data permitting, studies were planned to be combined and a meta-analysis was to be conducted using the Chi Square test for heterogeneity after visual assessment. Statistical analysis was planned to be carried out using R-statistical software.

Given the likely situation that combined statistical analysis is not possible due to the lack of quantitative data or significant heterogeneity exists between studies, a narrative process was to be conducted and presented as an analysis of the evidence quality for specific risk factors identified in the literature. The Economic and Social Research Council (ESRC) guidance report was the framework for a narrative synthesis. [204] Studies were to be summarized in a tabular format. Details to be recorded were to include the study type, sample size, participant characteristics, outcomes and outcome measures. The Center for Reviews and Dissemination (CRD) process was to be used to inform this process. [205]

3.1.4.10 Analysis of subgroups or subsets

It was not possible to specify the subgroups in advance however immunocompetent and immunosuppressed populations may have differential risks and could represent groups for sub analysis. [176, 177] In addition, age has been shown to be a factor in characteristics affecting acquisition of disease and these populations may have differential risks. [170, 171] Further subgroups anticipated involved exposure to index case as seen in proximity from working location in a congregate setting. [206]

3.2 Results

3.2.1 Study selection and study characteristics

We identified 7046 studies. 600 duplicates were excluded and 6249 were excluded after reviewing titles and abstracts leaving 197 for full text review. Of these 163 failed the inclusion criteria, 21 concerned household contacts, 11 were duplicate data and 2 were letters. No papers met the inclusion criteria. No comparative studies were identified examining risk factors in congregate settings in low-incidence areas. The reasons for exclusion are summarised in figure 1. Because no studies were identified we relaxed the exclusion criteria to include studies in high prevalence settings and conducted a narrative review of these studies. Several excluded studies included comparative designs and examined congregate settings. The findings of these are discussed below as well as the reasons for their exclusions.

Prison populations

Ali et al examined TB prevalence in prisons in Ethiopia. [207] This study did not look primarily at risk factors unique to the congregate settings. However, one reported factor affecting acquisition of TB was noted; a prison with a window reduced the risk of acquiring TB disease OR 0.25 ($p<0.0001$). Of note, duration of stay in prison did not correlate with TB acquisition ($p<0.58$) and neither did positive HIV status ($p<0.91$). Both of there would be considered relevant risk factors.

Fuge et al conducted a cross-sectional study examining active and latent TB in 162 prisoners with a prolonged (>1 week) cough in Southern Ethiopian prison population. The lack of visit from family was the only variable identified as a risk factor for pulmonary TB (PTB) p = 0.029. Almost all of the PTB positive cases were rural residents, farmers, males (aged 20-44) and those who shared a cell with TB patients and chronically coughing persons as well as those who stayed in a cell that contained >100 inmates. [208]

Lopez et al also conducted a cross-sectional study. This was in Spanish prisons and included 936 patients. Lopez Looked at the prevalence of LTBI in prisons and limited risk factors to pre-determined/ known factors involved in TB transmission. The authors did not examine the role of congregate settings in the transmission of Mtb. [209]

Owokuhaisa et al conducted a prevalence study of TB in a prison population in Uganda with 248 participants. The authors identified overcrowding and poorly ventilated cells as a risk factor for TB development. However, the authors did not analyse risk factors related to the development of TB over and above a descriptive report. [210]

Health care population

Park et al undertook a cross-sectional study in South Korean health care workers (HCW's). 499 participants were enrolled. The authors found that there was no statistically significant difference between working in TB departments and non-TB departments. On univariate logistic regression, the authors found that age and duration of work as HCW were significant risk factors for the development of LTBI. On multivariate analysis only age remained as a

significant risk factor. This study did not analyse the factors involved in congregate setting spread of disease examining the risk of TB exposure (TB ward work). [211]

Yoon et al conducted a cross-sectional study of HCW's in military hospitals involving 962 participants. In the univariate analysis, increased age, and providing TB patient care as well as co-morbidities and BCG scars were associated with LTBI. However, after the multivariate analysis only working in a TB department for more than a year was associated with LTBI. [212]

Powell et al. The authors examined the risk of TB infection associated with hospital employment in Vietnam. This was a cross-sectional study that examined hospital and local school personnel. 1188 participants were included from the hospital population and 156 from school personnel. Again, working in high risk TB departments was not significantly associated with development of TB. Additionally, the type of setting in hospital in-patient, out-patient, or both, made no difference. Hospital personnel were twice as likely as school personnel to develop TB (OR 2.0). Age was a significant risk factor for development of TB. The study took into account household contacts and did not differentiate exposures from this setting. [213]

Rafiza et al examined the prevalence and risk factors in HCW's in Malaysia involving 954 participants. Male gender, working as a nurse and increased age were all associated risk factors for developing TB disease. This study did not differentiate between congregate and

household settings and household exposure to TB was a statistically significant risk factor for developing TB. [214]

Tudor et al examined the occupational risk factors in South Africa through a case control study. 145 HCW's were included in the study. The authors found that HIV infection was more likely in cases as was time with infected patients (>/= 8 hours). The authors did not limit the study to congregate settings and cases were highly likely to have TB if they had high household crowding which represented a significant factor. HIV infection and time spent with TB patients were determined as independent risk factors for TB disease development. [215]

Other settings

Hwang et al conducted a cross-sectional study examining TB prevalence in a South Korean homeless population. The authors did not examine risk factors unique to homeless populations and the congregate settings they frequent. They did note the higher rate of TB in this population group, 4x the background general population. [216]

3.3 Discussion

3.3.1 Summary

In summary, we did not identify any comparative studies examining risk factors in congregate locations in a low incidence setting. We did note the small number of studies examining this in high incidence settings. From our findings, themes have emerged. Locations studied have broadly grouped into two categories; prison populations and health care settings. While several studies comment on the risk factors inherent in these locations, it is difficult to ascertain the role interactions in these settings play over and above the background (household) rate of TB in higher incidence setting. A factor acknowledged in several studies when examining statistically significant risk factors. [208, 214, 215]

One identified study in a higher incidence setting looked at a homeless population. This group has previously been described as demonstrating a high rate of TB. [217] Furthermore, the risk for on-going infections is greater in this population given the difficulty with service engagement and follow-up. [218, 219]

3.3.2 Literature Comparison

3.3.2.1 Role of low vs high incidence settings

Low incidence settings differ from higher incidence settings in terms of contact tracing approaches. With more resources allocated to intensive examination of contacts in order to detect latent infection and prevent active disease development. [34] Contacts in low

incidence settings have a higher risk than the background population of acquiring disease warranting close and careful examination. This may not be true of higher incidence settings. [34, 220-222]

3.3.2.2 Congregate vs household environment

Examining contacts beginning in a household environment does not take into account the numerous exposures possible in a non-household, congregate setting. In addition, traditional contact tracing methods and screening guidance advise cessation of further screening if earlier screening episodes in a household environment do not demonstrate evidence of infection. This relies heavily on static, well defined social exposures and does not take into account the plurality and nuance of sociocultural norms. [90, 223]

Non-household exposures have been demonstrated to result in infection. [224]Furthermore these social interactions can result in harder to define outbreaks where the congregate setting has not been taken into account. Consequently, more individuals become contacts and TB disease is propagated. Examining outbreaks and taking into account more nuanced social interactions has demonstrated a higher number of at risk contacts and a better-defined outbreak event allowing health care services to effectively intervene to stop further spread of TB. [54, 165, 186, 225, 226]

That TB transmission occurs outside of a household setting is not in doubt. Congregate settings as locations of transmission have been described by numerous authors. They highlight the significant role these particular settings play and the risk and challenge of

contact tracing effectively. [170, 227-231] Often describing the difficulty in attributing links between contacts when congregate settings are not considered. Transmission in congregate settings has a further dimension over and above the index case and their contacts, namely the ventilation characteristics of the location setting. [182]

3.3.2.3 Lack of studies

The paucity of congregate setting studies or studies examining casual contacts in low to medium incidence settings has been highlighted before. Here the authors again demonstrated the lack of evidence on the subject, with no papers identified in this area. [232] In addition, looking at models of transmission in similar settings has also been examined with no findings. [233]

Oddly, there were limited studies examining educational and work environments identified in this review. Whilst there are case reports on educational and work establishments, I did not identify any comparative studies. [234-236]

3.3.2.4 Risk factors

Age was implicated as a risk factor in the high incidence settings I reported on. This has previously been noted with a trend towards adolescent infection occurring out of a household setting. Social relationships tend to vary with age and occupation. These factors introduce variability in social encounters and potential exposure events. [170, 171, 237].

Prolonged exposure to index cases has been shown to result in a higher risk of infection. The minimum amount of time required and nature of the proximity have not been well established. [172]

Prison locations as high risk congregate transmission sites has been highlighted previously [238], as have healthcare settings. [239-242]

One of the correlating risk factors from the reported high incidence studies was household TB where relatives were visiting. It is difficult to conclude from the high incidence environment studies whether the risk is from the congregate setting rather than from relatives. Even non-household, close relatives have been shown to have a high risk of transmission. [243]

3.3.3 Strengths and limitations

We conducted a large systematic review, not limiting for key words that might restrict findings. One of the reasons for this is that there are numerous ways to refer to non-household, congregate settings i.e. non-household, shared social spaces. These include but are not limited to recreational locations, work and educational settings as well as prisons and healthcare settings as described above. They also include transport. Studies are therefore titled and referenced in many different ways. Scoping searches were carried out to find the best combination of terms to allow for studies relevant to the reviews question.

Two authors assessed the findings and applied inclusion criteria to the available search findings, where there was disagreement, these studies were discussed and if a consensus could not be reached, a third author arbitrated.

The limitations of this study could be inherent bias in the search criteria. It is possible that a broader search or one combining different terms might highlight additional papers. That said, previous review articles have highlighted the limited data on this review question. [232]

A further limitation could be limiting the papers reviewed to comparative studies only. There are numerous case reports that highlight the role congregate settings play as described above. While these could be examined, it is unlikely that they will add much other than to draw attention to congregate settings given the general lack of further analysis in some case reports.

3.3.4 Future research/ further work

Our review highlights the lack of data examining congregate setting specific risk factors in low incidence settings. In order to further reduce transmission events, non-household settings need to be examined and any risk factors taken into account to prioritise case finding in a limited resource setting.

Methods that take into account the further dimensions posed by congregate settings could help. One example is to consider ventilation in these environments. Ventilation in congregate settings has been identified as a risk factor in previous studies. [182, 244-246]. A

study by Ali 2015 identified by this review, touched on the role this may play with improved ventilation (a window) reducing transmission risk.

One method to reliably capture ventilation data would be using carbon dioxide capturing technology can utilise the rebreathed fraction as a measure of circulating air and therefore exposure to Mtb circulating. [246]

In addition, understanding the social dynamics of a congregate location could help to prioritise contacts. This would involve identifying contacts that frequent busy social spaces such as common rooms and dining halls. These contacts may not be identified through normal processes. [237]

There are several barriers to examining congregate setting risk factors. These settings tend to be larger than the average household screened and therefore are resource and time intensive. In low incidence settings, identifying more contacts does not necessarily lead to an increased rate of detection of Mtb disease. [186]

There is still a lack of knowledge about the transmission dynamics of Mtb with no clear minimum time of exposure for transmission identified and very short exposure times leading to infection complicating risk assessment. [172]

As molecular examination of latent infection is not possible, it is difficult to attribute links to cases identified to congregate setting transmission. Where latent infection is detected, this may represent background rates of TB in high risk social and ethnic communities.

3.3.4 Conclusion

In summary, there is a paucity of data on congregate specific risk factors for the spread and acquisition of TB disease. Studies identified in high incidence settings focused on prison and healthcare settings. There were no identified studies in low incidence settings. Given the variety of congregate settings where exposure can take place, this represents a significant gap in data. In particular, educational settings are underrepresented. TB services should take note of this discrepancy. This represents an important area for further work.

As we have seen, this chapter demonstrated the limited evidence examining risk factors for TB transmission in congregate settings. The next chapter will examine the evidence for contact tracing strategies and whether there are alternatives to the existing approach.

Chapter 4 - Contact tracing strategies in household and congregate environments to identify cases of tuberculosis in low- and moderate-incidence populations – A Cochrane review

This chapter reviews evidence for strategies to identify cases of TB in low and moderate incidence populations, it has been published as a Cochrane review. [93] Unlike chapter 3, we are not looking at risk factors that are involved in congregate setting transmission events and the acquisition of TB by contacts. Instead, this chapter focuses on the strategies that are available to identify at risk individuals (contacts) and whether there is an evidence base around the currently employed approach or any alternatives.

4.1 Background

4.1.1 Description of the condition

In order to reduce the incidence of tuberculosis, relevant services engage in contact tracing; this is the evaluation of contacts of infected people for tuberculosis disease. While there are a number of defined screening methods to identify latent and active disease, methodologies used to identify those individuals deemed to be contacts are less obvious. The aim of effective contact tracing is to identify those exposed individuals with infection (either latent or active disease) as quickly and efficiently as possible.

Identifying the contacts of people with tuberculosis usually occurs once an index case has been identified. This screening process can identify a substantial group of contacts depending on the index case's home, work, and travel arrangements.

4.1.2 Description of the intervention

Contact tracing in low- (fewer than 30 per 100,000) and moderate-incidence (30 to 100 per 100,000) countries aims to identify individuals with latent or active tuberculosis infection, or both.[32] This approach often differs from areas of high incidence (greater than 100 per 100,000) where the emphasis remains on identifying and treating active cases as a priority. Given the large number of cases in high-incidence settings and limited resources, this is the most cost-effective approach.

The stone-in-pond model relies on assumed, consistent social relationships for all infected people, as a negative screen in a closer contact group results in the cessation of further contact tracing (for example, if household contacts were negative, no further groups would be screened). This consistency in social relationships and presumed proximity for all individuals may not be universally applicable. The utility of this approach compared to alternate approaches has not been evaluated.

Traditional contact tracing methods may not take into account areas of congregation (that is, areas of common social aggregation; for example, occupational, educational, recreational, and transport environments), which have been shown to be potential sources of transmission for tuberculosis, especially for at-risk population groups [224]. In addition, people with tuberculosis may have varying degrees of contact with alternate groups that do not fit with perceived proximity circles (for example, college students spending more time with close friends than family).

Apart from the above stone-in-pond model, the author is not aware of any named, contact tracing methodologies in clinical use. Recent research papers have examined alternate approaches using methods to draw on non-linear social interactions in congregate settings, an approach known as social network analysis. Patients are questioned as to possible contacts but, importantly, also social activities, hobbies, alcohol and drug use, work and recreational activities. This geotemporal data collection provides clinicians with possible locations of interaction where transmission of disease may have occurred thus expanding the pool of potential contacts. When combined with whole genome sequencing, a novel genomic approach to diagnosis and typing of tuberculosis strains, outbreaks can be mapped to a higher resolution and clearer links between contacts made. This method is not in general clinical use.

4.1.3 How the intervention might work

The effectiveness of the current contact tracing methodology (stone-in-pond approach) is called into question by whole-genome sequencing, where genetic links are found between otherwise unscreened/epidemiologically unrelated contacts, demonstrating deficiencies in current methods. Some research groups have demonstrated the utility of alternative contact tracing approaches, including social network analysis. These methodologies could provide a higher yield of case detection whilst not allowing tuberculosis incidents to propagate and provide epidemiological links that more closely model genetic links [247]. Assessing for alternative contact tracing methodologies could provide a much-needed evidence base to current clinical practice or if lacking, demonstrate the need for more evidence.

4.1.4 Why it is important to do this review

The author is not aware of any published reviews that have examined alternative contact tracing approaches for a higher rate of latent and active case detection. In addition, there is a paucity of evidence of the appropriate contact tracing approach in public settings versus household settings. This Cochrane Review may better inform resource allocation.

This review will add to the evidence base for meaningful strategies in the early detection and prevention of tuberculosis targeting low- and moderate-incidence settings and will assist in achieving the World Health Organization (WHO) tuberculosis elimination goal [167].

New developments in genomic diagnostics, such as whole-genome sequencing, and studies that suggest the benefit of social network analysis raise the prospect of alternative approaches that should be evaluated. It is likely that there will not be many randomized controlled trials (RCTs) that compare contact tracing methodologies. However, by highlighting the lack of comparative evidence and raising the prospect of alternatives, this Cochrane Review could form the basis for future research comparing contact tracing strategies.

4.1.5 Objectives

To assess the effectiveness of novel methods of contact tracing versus current standard of care to identify latent and active cases in low- to moderate-incidence settings.

4.2 Methods

4.2.1 Criteria for considering studies for this review

4.2.1.1 Types of studies

RCTs and cluster-RCTs.

4.2.1.2 Participants

People of any age, gender and ethnicity living in low (< 30 per 100,000) and moderate (30 to 100 per 100,000 population) tuberculosis incidence settings.

4.2.1.3 Intervention

Any contact tracing strategy to identify tuberculosis infection cases other than a stone-in-pond screening approach (standard care).

4.2.1.4 Controls

Stone-in-pond contact screening approach (standard care).

The stone-in-pond method describes the contact tracing approach of prioritising contacts by risk-stratifying cases based on assumed proximity. Household contacts therefore have the highest presumed risk and are the closest circle to be screened followed by the next 'ripple' which may be close friends then casual friends, and so on. This set of outwardly expanding concentric circles is similar to the appearance of a stone being dropped in a pond with the ripples generated representing concentric circles of risk proximity.

4.2.2 Types of outcome measures

4.2.2.1 Primary outcomes

Number of people with tuberculosis infection identified through screening strategies.

4.2.2.2 Secondary outcomes

The number of contacts with disease (latent and active tuberculosis versus non-infected contacts) identified between the two screening approaches.

4.2.3 Search methods for identification of studies

The aim is to identify all relevant trials regardless of language or publication status (published, unpublished, in press, and in progress).

4.2.3.1 Electronic searches

The following databases were searched using the search terms and strategy described in Appendix 1: Cochrane Infectious Diseases Group Specialized Register; Central Register of Controlled Trials (CENTRAL), published in the Cochrane Library; MEDLINE (PubMed); Embase (OVID); LILACS (BIREME); CINAHL (EBSCOHost); Science Citation Index-Expanded and Social Sciences Citation Index (Web of Science). A search was conducted of the WHO International

Clinical Trials Registry Platform (ICTRP; www.who.int/ictrp/), ClinicalTrials.gov (clinicaltrials.gov/ct2/home), and the Clinical Trials Unit of the International Union against Tuberculosis and Lung Disease (IUATLD; www.theunion.org), for trials in progress, using ‘tuberculosis”, ‘contact tracing' and ‘contact screening' as search terms.

4.2.3.2 Searching other resources

The following conference proceedings were searched for abstracts of relevant studies: World Congress on TB, World Lung Conferences of the International Union Against Tuberculosis Lung Disease (IUATLD), American Thoracic Society Meetings Proceedings, and the British Society for Antimicrobial Therapy. References of all included studies to identify additional studies were also checked. [248]

4.2.4 Data collection and analysis

As no relevant studies were identified, study authors did not need to be contacted to identify any additional, relevant unpublished data or results.

4.2.4.1 Selection of studies

Two review authors (DBM, BM) independently screened all study abstracts and citations identified by the above search criteria using a study selection form. I planned to obtain full

texts of studies that potentially meet the eligibility criteria. Exclusions and reasons for exclusions are reported in a ‘Characteristics of excluded studies' table.

4.2.4.2 Data extraction and management

The following information was planned to be extracted:

Study details: start and end dates, study location, study design, funding, tuberculosis prevalence (as stated by the study authors), conflict of interests.

Participant details: demographic details (age and sex), geographic location, index case description. Description of contact relationship.

Details of the intervention: how were these individuals identified as contacts, and the methodology used. Outcome as active or latent.

Details of any co-interventions: did the methodology differ depending on the setting (congregate versus household). Where household setting would be the shared living accommodation and congregate setting describes a non-household, area of common social aggregation, for example, occupational, educational, recreational, and transport environment.

Details of the control: standard of care employing stone-in-pond model of contact tracing. We will identify the number of contacts identified from each described outbreak or incident event. The number of contacts identified and screened with tuberculosis infection (active or latent infection - numerator) over the total number of contacts identified and screened (denominator). An outbreak/incident event is declared where multiple contacts are identified in relation to a new index case. The number of contacts varies, and usually a congregate setting is involved for an outbreak/incident event to be identified. Where the unit of analysis in studies is the same, I intended to group studies.

Cluster-RCTs: I planned to record the number, size and method used for clustering. In addition, I had intended to note the clustered measure of effect and variance if this was adjusted for by study authors. If the study authors did not make any adjustment for clustering, I planned to extract the number of participants experiencing the event and the number randomized to each group (for dichotomous outcomes). For continuous outcomes, we will extract the summary effect (mean or median) and the measure of variance (standard deviation or range).

4.2.4.3 Assessment of risk of bias in included studies

The Cochrane approach assesses risk of bias across six domains: sequence generation and allocation concealment (selection bias), blinding of participants and personnel (performance bias), blinding of outcome assessors (detection bias), incomplete outcome data (attrition bias), selective outcome reporting (reporting bias), and other potential biases.

4.2.4.4 Measures of treatment effect

In order to assess the treatment effect, continuous and dichotomous data was planned for separate analysis. For continuous data, effect would have been assessed by mean differences and for dichotomous data, risk ratios.

4.2.4.5 Unit of analysis issues

Where study authors have not adjusted the results of cluster-RCTs for the effect of the cluster design, I planned to adjust the sample sizes using the methods described in Section 16.3.4 or 16.3.6 [249] using an estimate of the intra-cluster correlation coefficient (ICC). Where possible, I planned to derive the ICC from the trial itself, or from a similar trial. If an appropriate ICC value was unavailable, I planned to conduct sensitivity analyses to investigate the potential effect of clustering by imputing a range of ICC values.

4.2.4.6 Dealing with missing data

There was not planned to be imputation measures for missing data.

4.2.4.7 Assessment of heterogeneity

Statistical heterogeneity between trials will be assessed by visual inspection of the forest plots to detect overlapping CIs, and applying the Chi^2 test and I^2 statistic. [250]

A Chi² P value less than 0.05 will be considered as statistically significant, and an I^2 statistic value greater than 75% as representing considerable heterogeneity. [251]

4.2.4.8 Assessment of reporting biases

The likelihood of reporting bias was planned to be assessed using funnel plots, provided that there are at least 10 included trials. [252]

4.2.4.9 Data synthesis

Data analysis was to be undertaken using RevMan 5. The primary analysis was planned to be stratified by study design (cluster-RCTs and individual RCTs) and meta-analysis was not planned to be performed across different trial designs.

Outcomes were to be stratified according to number of cases detected at a particular time point per contact tracing strategy. Where appropriate, time points were planned to be grouped and a meta-analysis performed (for example, changing number of contacts over time in a single contact tracing episode).

Results from cluster-RCTs that cannot be adjusted for clustering. A random-effects model was planned to be used in the presence of significant statistical heterogeneity and a fixed-effect model in the absence of heterogeneity.

4.2.4.10 Subgroup analysis and investigation of heterogeneity

Potential causes of heterogeneity were to be investigated by performing subgroup analyses by study setting (congregate versus home), screening test used, risk factors in demography (drug and alcohol use, immunosuppressive states), occupation, age of participants, and tuberculosis prevalence in study area.

4.2.4.11 Sensitivity analysis

Sensitivity analyses were to be performed if a minimum of 10 trials meet the inclusion criteria. Sensitivity analyses were planned to be conducted on the robustness of the results to the 'Risk of bias' components.

4.2.4.12 Certainty of the evidence

The GRADE approach to assess the certainty of the evidence was to be used and a 'Summary of findings' table created and Evidence Profiles if relevant (GRADEpro 2015 - GRADEpro GDT [Computer program]. Version accessed 14 May 2018. Hamilton (ON): GRADE Working Group, McMaster University, 2015.Available at gradepro.org).

4.3 Results

4.3.1 Results of the search

755 articles were identified using the pre-specified search strategy, with an additional 16 clinical trials in progress. Of these, none met the inclusion criteria for this review. A description of excluded studies is included in the appendices. No conference proceedings or abstracts were identified. Several trial authors examined the feasibility and effectiveness of network-based approaches to contact tracing, an alternative method. These studies did not meet the inclusion criteria for the reasons below (Figure 1). They are further explored in the Discussion section.

4.4 Discussion

I did not identify any randomized controlled trials (RCTs) or quasi-RCTs that compared alternative contact tracing methods to the standard approach.

To eliminate tuberculosis in low-incidence settings, high-risk groups must be scrutinized as part of an effective screening process. Contacts of patients who have active, infectious tuberculosis represent a higher-risk group than the general population. [34, 221, 222]

Effective contact tracing processes are therefore key to reducing and eventually eliminating tuberculosis. [34]) Several trial authors have demonstrated the effectiveness of alternative strategies for contact tracing in a variety of settings [54, 165, 169, 182, 185, 186, 226, 253, 254]).

Andrews 2014 undertook a modelling study in a high-incidence area. Using a modified Wells-Riley model, researchers described transmission patterns in an endemic setting. They demonstrated the significance of non-household settings for transmission of infection and for correlation with age (adolescents 15 to 19 responsible for school transmission). They included no control group; this was a modelling study examining congregate settings in a high-incidence environment. [254]

Cook 2007 examined the role of SNA in detecting tuberculosis transmission to prioritize contacts. Using a structured questionnaire and supplementing this with patient records, review authors extracted demographic data and information about social aggregation, patients, and their contacts. This was augmented with molecular genotyping. Trial authors found that patients who were not identified through conventional contact investigations were connected through areas of social aggregation or via mutual contacts. These authors found a positive correlation between contact screening tests (positive tuberculin skin tests) and location in the denser portions of constructed networks ($P < 0.1$). This was not an RCT, and trial authors included no comparison group. [226] These trial findings were supported by another study[165], whose authors again examined the role of network analysis in prioritizing contacts by supplementing epidemiological data collected through hospital records and patient interviews with genotyping isolates. They found that contacts prioritized using a network analysis approach were more likely to have latent infection (LTBI) than non-prioritized contacts (odds ratio (OR) 7.8). This again was not a review article and did not include a comparison group for analysis. A further network-based study was identified. [225]

These trial authors retrospectively reviewed outbreak events in a US low-incidence state, applying social network analytical methods. This involved re-interviewing patients and contacts. Researchers discovered previously unrecorded patterns of drug use propagating infection.

Fox 2012 conducted a systematic review to look at the outcomes of contact investigations for tuberculosis. These review authors established the risk to contacts of developing disease and the lack of evidence to support current contact tracing methods. They performed subgroup analyses according to index and contact characteristics. These review authors established the prevalence of contact outcomes according to setting income; low- and middle-income countries have a latent infection prevalence of 45.9%, with tuberculosis disease in 3.1% of all contacts and 3.6% of household contacts. In high-income countries (correlating with low-incidence settings), tuberculosis disease was 1.0%, and amongst household contacts 3.5%, with 26.3% of latent cases detected. There was no comparison of contact tracing methods, and low-quality observational studies provided the basis for conclusions. [255]

There was no clear evidence for the current contact tracing method used - the 'stone-in-pond' approach. [90, 255]) This approach involves assessment of contacts in prescribed proximity rings to the index case, with the most proximal ring traditionally representing household contacts. Several trial authors have demonstrated the shortcomings of this approach, with transmission zones changing depending on age and social activities [169, 170, 182, 185, 256]). The search conducted for our review revealed no RCTs addressing a

comparison of the standard contact tracing approach - the ‘stone-in-pond' model - versus a new approach such as a network analysis approach - social network analysis (SNA).

4.4.1 Summary of main results

This review did not find any RCTs that compared the standard contact tracing approach (the ‘stone-in-pond’ model) to alternative approaches (SNA).

4.4.1.1 Overall completeness and applicability of evidence

Several identified studies demonstrated the benefit of a network analysis approach; however, these were not RCTs, and they did not include a comparison group. [165, 225, 226])

From contemporary research studies, it appears that a more effective approach could possibly be adopted, and that previous assumptions about transmission zones for tuberculosis need to be revisited, with non-household transmission playing a key role. Transmission of tuberculosis is also influenced by demographic factors such as age and gender; therefore, an approach that takes into account geospatial location, activity spaces, and these demographic factors could better identify contacts and help prioritize tracing opportunities while identifying contacts at highest risk of disease development.

Policies advising contact tracing processes should take into account the growing research evidence in support of a network-based approach to contact tracing, considering factors that influence different transmission zones in low-incidence settings. Improved contact tracing

can help to further reduce incidence rates and prevent ongoing active case development in high-risk groups such as contacts of infectious patients. Furthermore, once identified, these factors can inform control and prevention measures in congregate settings to prevent ongoing occurrence of incidents.

4.4.1.2 Potential biases in the review process

Bias was minimised by using a previously described search criterion. In addition, clinical trials in progress were searched to minimize the risk of missed studies; however, this review process may have missed studies that have not yet been published. A broad search including participants and environments studied was conducted, and it is unlikely that there were any missed, published RCTs. Rather, the lack of studies is a reflection of the lack of evidence on this topic. Furthermore, it may reflect differences in the description of interventions used, as well as less clearly defined use of methods by clinical teams (that is, a degree of overlap in contact tracing approaches).

Studies identified through the search have been discussed above and reflect the difficulty involved in conducting a prospective RCT. Trial authors suggest it may be possible to conduct such a trial via a cross-over method using comparative sites.

Several trial authors have suggested the benefit that could be gained from using a social network approach; however, this was not compared to the standard contact tracing approach [54, 165, 169, 182, 185, 186, 225, 226, 254]). Additionally, trial authors have

previously described deficiencies in the current contact tracing approach, lending support to the need for a comparative analysis [169, 170, 182, 185, 256]).

4.4.1.3 Differences between protocol and review

The Methods section of the published protocol was amended regarding inclusion of studies. The text was changed from "any contact tracing strategy to identify tuberculosis infection cases other than a stone-in-pond screening approach (standard care), e.g. the use of SNA to identify contacts prospectively" to "any contact tracing strategy to identify tuberculosis infection cases versus traditional or alternative approaches such as the stone-in-pond screening approach, e.g. the use of SNA to identify contacts prospectively". This change was made on the basis that we found no trials comparing the ‘stone-in-pond' approach to another intervention, and allowed for an expansion of the search.

The outcomes section was amended to further clarify the differences between primary and secondary outcomes and to provide a clear denominator. The primary outcome has been changed from "the proportion of contacts with tuberculosis infection identified through screening strategies" to "contacts with latent TB infection out of all contacts screened". The secondary outcome has been changed from "the proportion of contacts with disease (latent and active tuberculosis versus non-infected contacts) identified between the two screening approaches" to "contacts diagnosed with proven Mycobacterium tuberculosis infection or active clinical tuberculosis disease out of all contacts screened". Neither of these changes affected the search criteria or the number of studies found or investigated.

4.4.2 Conclusions

4.4.2.1 Implications for practice

The current contact tracing method (‘stone-in-pond' approach) does not appear to be based on a comparative evidence-based approach and instead has been adopted from its early proposal as a pragmatic process.

4.4.2.2 Implications for research

This Cochrane Review has identified the dearth of evidence for contact tracing methods. It highlights the need for further research into optimal contact tracing methods to evaluate the most efficient and effective approach.

Contact tracing methods need to take into account the limited resource allocation provided and should be able to effectively identify those at risk of disease development. Ideally, these methods should balance the accurate identification of exposed contacts, whilst allowing prioritization of those at highest risk of disease development.

Future trials should examine the benefit of network-based approaches to contact tracing using current methods as a benchmark. It is likely that the background incidence of study setting will affect outcomes, and trials should take this into account. Furthermore, designing studies to examine the effectiveness of alternative methods for contact tracing (for example, network approach) could take into account novel molecular techniques (for example, whole genome sequencing). This would allow studies to examine the relative decline in the number

of molecularly linked active cases over a specified time period. Finally, to fully inform future policy decisions, a cost-effectiveness analysis should ideally be included.

This review demonstrates that there is limited high-quality evidence for the current contact tracing method and for consideration of alternative approaches. RCTs are therefore needed to assess the best approach to effective contact tracing methods. Ideally these trials would compare a network analysis approach using standard care (the 'stone-in-pond' approach).

New approaches to molecular diagnostics could assist with a network analysis approach. Whole genome sequencing has been shown to provide higher resolution for investigations in outbreaks of tuberculosis infection [54]. In addition, several trial authors have demonstrated the benefit of assessing ventilation when congregate settings are considered in contact tracing processes; this has so far been done using carbon dioxide (CO2) monitors. [172]; [169, 185]

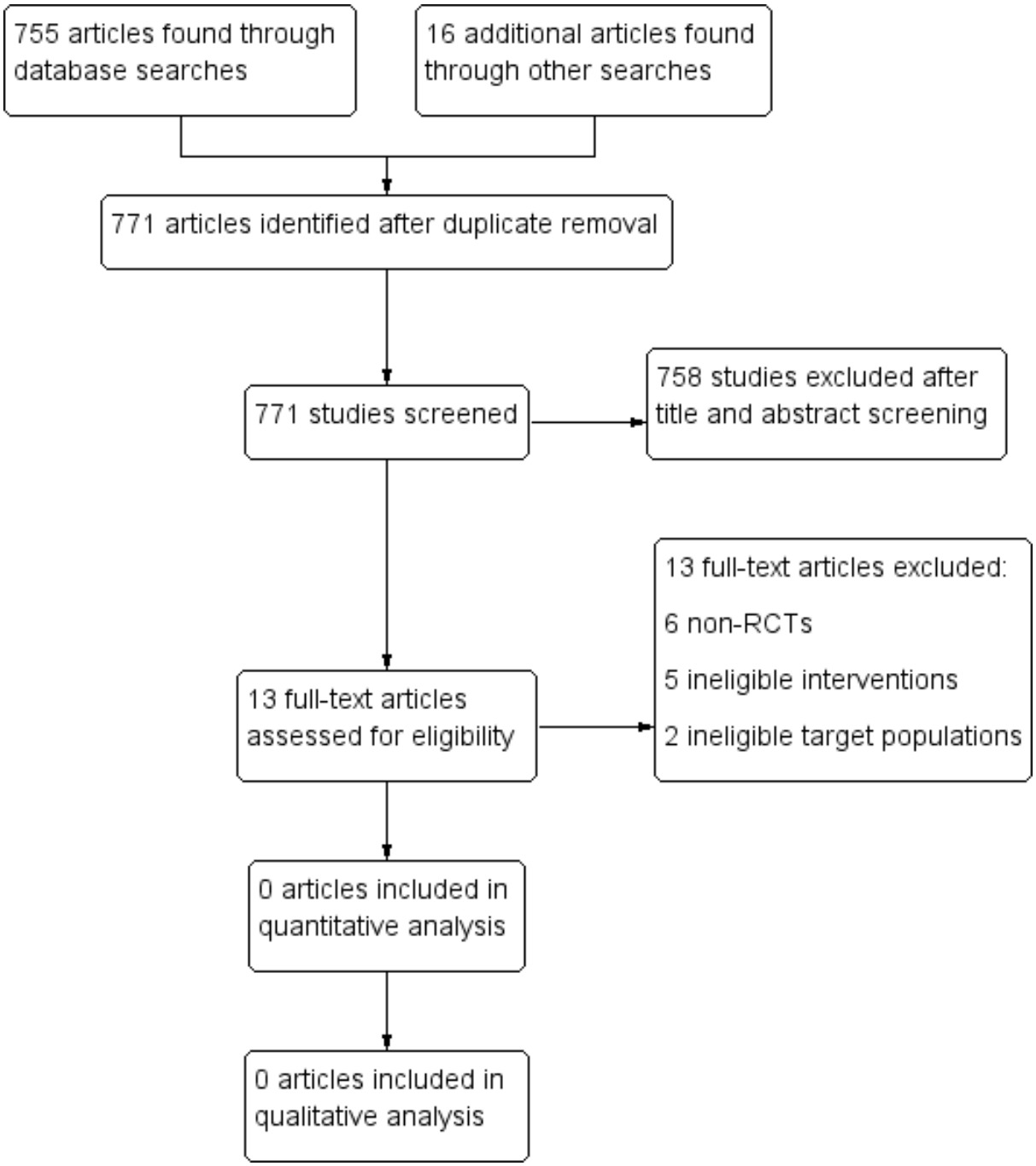

Figure 5 - PRISMA flow diagram - Cochrane review

As we have seen, there is no evidence comparing alternative strategies for contact tracing and no randomised trials to examine the effectiveness of the current contact tracing approach. In the next chapter, I will summarise the data and comparability to national figures for our cohort.

Chapter 5 – A summary of incident data used for analysis and contact tracing outcomes

The aim of this chapter is the describe and summarise the data used for the analysis in this thesis. I will use nationally released public health reports to structure this chapter.

5.1 Introduction

5.1.1 Background

The previous chapters have contextualised the public health efforts to reduce and eliminate TB in the UK. We have the gradual evolution of the contact tracing process and the lack of an evidence base for an alternative approach. This chapter will examine the dataset from which further analysis and description will take place. In addition, it represents a unique presentation in reporting the outcomes of large scale contact tracing investigations not limited to single outbreak or investigations events. This is subsequently compared to annual published public health reports on TB in the UK and is important in understanding the cosmopolitan nature of the dataset as well as its broader generalisability.

Once a case of active TB infection presents to the health services, systematic contact tracing should be conducted in order to identify close contacts who may have latent or active disease. Contact tracing in this setting, to identify active as well as latent cases has higher utility in low incidence settings where the identification and treatment of latent cases

through effective screening programmes is crucial to achieving TB elimination. In areas of high TB incidence, TB prevalence in contacts of up to 5% has been reported. [257]

In the UK, there remains a significant delay in the treatment of pulmonary TB disease with over 25% waiting more than 4 months. Furthermore, though national trends demonstrate a decrease in the total number of cases of TB, this trend is not seen in the native population and in particular certain at risk groups (alcohol/ drug use, prisoners, homeless) who have shown an increase. These groups are also more likely to have pulmonary disease and therefore more likely to transmit TB. In addition to this they also have an increased likelihood of resistant disease and concomitantly poorer outcomes. [164]

Public health England's collaborative strategy aims to achieve a decrease in TB incidence with each year. This strategy is focused through 10 interventional areas of which comprehensive contact tracing and improved earlier diagnosis are key. [258]

Importantly, latent and active disease have previously been seen as two distinct categories of disease. This view has been called into question with subclinical disease and pre-symptomatic transmission of disease possible. [59] Therefore, finding and treating cases of TB in the early latent stages becomes more relevant to TB elimination given the increased risk of active disease development without treatment. [259]

5.1.2 Aims and objectives of paper

We aim to summarise patterns in TB incident episodes in a large UK city, Birmingham, from 2010 to 2017 comparing these to currently published Public Health England (PHE) statistics for England. Birmingham represents a cosmopolitan population with a recognised mix of ethnic groups. Given the structure of TB services, Birmingham represents a large unified data set through a single chest clinic service, making this city ideal for data collection and analysis.

5.2 Methods

5.2.1 Data collection

Cases of TB disease were routinely recorded by public health services in the UK once they were notified by clinical services. These cases were prospectively collated in a database from the Birmingham city TB service including all new cases of pulmonary and laryngeal tuberculosis that presented to services between 2010 and 2017.

Contacts were identified through a multitude of sources. The initial identification coming from the index case who recorded potential contacts in interviews with TB nurses and chest clinic teams. Depending on the congregate setting, local health advisors e.g. occupational health teams or school nurses identified further contacts. In the case of particular congregate settings e.g. schools, classroom colleagues and year group contacts were provided by the school administration where appropriate. All identified contacts were invited to screening as part of a routine TB service and Public Health England contact tracing process. Where data and outcome existed for contacts, these were included in the analysis. Data was grouped based on the primary location of cases and screening where this involved non-household screening e.g. schools/ work place/ activity centre etc.

The interview data was transcribed into the local TB database ‘Dendrite’ by the chest clinic team. Where further information was required and as part of the treatment and follow up process, additional interviews were carried out by the TB nurses working through the chest clinic.

Routine data collected on index cases and contacts is listed in table 2 The contact relationship to index cases was also recorded where available, as reported by the index case. Where this was lacking, chest clinic teams would assign a proximity relationship (e.g. close/ casual). This designated proximity relationship was not based on specific guidelines or definitions and was therefore subject to operator variability. [223]

Table 2 - List of routine variables collected during contact investigations

Location	**Contacts**	**Index**	**Disease**	**Dates**
Category	Date of birth	Date of tuberculosis	Organism	Status of case
	Current age	Date of notification	Sensitivity	Date onset of illness
	Gender	Patient gender	Resistance	Date first presented with symptoms
	Ethnic group	Date of birth	Histology	Date of diagnosis
	Born in the UK	Current age	Variable number tandem repeats (VNTR)	Directly observed therapy
	Year of entry	Ethnicity	ETR	Date start of treatment
	Country of birth	Country birth	Mycobacterial interspersed repetitive units (MIRU)	End of treatment
	Occupation	Year of entry	MIRU Plus	
	Previous TB	Previous TB	Microbiology laboratory notes	
	Previous TB date known	Previous TB date known	Post mortem diagnosis	
	Previous BCG	Previous BCG		
	Date of BCG	Date of BCG		
	Index year	HIV status		
	Contact index close contact	Occupation		
	Contact index relationship to contact	History of drug use		
	Contact index date of notification	Can self-administer?		
	Outcome	Homelessness		
		Incarceration history		
		Alcohol abuse		
		Drug abuse		
		Type of disease		
		Organs		
		Smear positive		
		Culture positive		

5.2.2 Study design

We conducted a descriptive analysis of index and contact cases. We examined patterns within the demographic data collected from our study population and examined how this compared to nationally reported data published by Public Health England (PHE).

5.2.3 Definitions

Active disease cases are categorized as either linked or unlinked. Linkage may be epidemiological, involving acknowledged contact between the index and contact cases, or molecular. The former involving acknowledged contact between the index and contact cases. The latter being demonstrated through molecular techniques, in this case WGS.

Two or more cases of Mtb with known epidemiological links constitute a cluster. This follows the CDC guideline for contact investigation definition of an outbreak; two or more persons exposed to a person with infectious TB. Epidemiological links can be established either through 1 case volunteering the name of a contact when the case was infectious or where both the case and contact shared a social location during the period of infectiousness.

5.2.4 Factors and categorisation

For location events, we grouped congregate settings in discrete categories to enable comparison between events and to analyse for patterns and themes. Types of congregate settings were based primarily on the key function of the location and reason for use of the congregate settings e.g. school/ educational environments, health care settings, work

settings, recreational settings etc. This is in keeping with similar studies examining TB outbreaks in congregate setting. [260, 261]

5.2.5 Ethics

This study was registered as IRAS project ID 221549. Ethical approval for this study was sought from the research ethics committee (REC) based in East Midlands – Derby research ethics committee, reference 17/EM/0253. Approval was granted 20/07/2017. The Health Research Authority (HRA) granted approval on the 21/07/2017. This study was closed on the 01/11/2019.

Anonymised data collection and analysis in tuberculosis is a statutory component of public health policy.

5.3 Results

Over a 7-year period there were 167 index cases with 167 incidents (events generating a community contact tracing investigation). 112 incidents occurred in unique locations, with 45 incidents resulting from recurrent incident events in locations of prior incident events (range from 1-6). These incidents generated 8091 contacts in total. Results are displayed in Table 3.

5.3.1 Locations

5.3.1.1 Types of locations

5 locations categories were used as described above; educational, accommodation, recreational, work and healthcare settings.

5.3.1.2 Sub locations

Each setting was further delineated where appropriate to take into account the different ages represented in different locations as well as the setting itself in order to assess different trends with varying locations. Health care and prison setting for example have previously been described as high risk locations. [230, 239-241]

Table 3 - Cases and contacts in congregate settings

Setting	Cases					Contacts				
	Index cases in setting	Proportion of index cases to all settings (%)	Subcategory of setting	N	Proportion of subcategories to setting (%)	Contacts identified per setting	Proportion of contacts in setting to all contacts (%)	Subcategory of contacts in setting	N	Proportion of subcategory contacts to setting (%)
Educational	89	53.2%	Nursery	1	1.1%	6520	80.6%	Nursery	164	2.5%
			Primary	7	7.9%			Primary	166	2.6%
			Primary/ combined	5	5.6%			Primary/ combined	161	2.5%
			Secondary	40	44.9%			Secondary	3560	54.6%
			College	21	23.6%			College	2171	33.3%
			University	14	15.7%			University	257	3.9%
			Other	1	1.1%			Other	41	0.6%
Work	23	13.8%	Office	15	65.2%	470	5.8%	Office	221	47.0%
			Factory	8	34.8%			Factory	249	53.0%
Recreational	19	11.4%	Recreational centres	9	47.4%	424	5.2%	Recreational centres	173	40.8%
			Restaurants	4	21.1%			Restaurants	150	35.4%
			Service shop	6	31.6%			Service shop	101	23.8%
Accommodation	11	6.6%	Prison	4	36.4%	172	2.2%	Prison	35	20.3%
			Housing	7	63.6%			Housing	137	79.7%
Healthcare	25	15%	Care home	5	20%	505	6.2%	Care home	127	25.2%
			Home care	2	8%			Home care	49	9.7%
			Hospital	13	52%			Hospital	232	45.9%
			GP practice	1	4%			GP practice	16	3.2%
			Outpatients	4	16%			Outpatients	81	16.0%
All	167	100%		167		8091	100%		8091	

5.3.2 Index case characteristics

5.3.2.1 Age

The median age of index cases was 19 with an interquartile range of 17-28. Total range 9-85. Age data was not available for 3 index cases. No index cases were noted to be younger than 9 years of age. The most frequent ages encountered in our data were in the 18-year-old age group and 20-year-old age group (11/167 – 6.6% for both). Under 15 year olds made up 15/167 – 9.0% of all ages presenting as index cases. Females made up 76 (45.5%) of index cases and males made up 91 (54.5%). Males made up the majority of index cases, this correlates with a male predominance in PHE data. The most frequent age period was from 18 and 20 year of age making up 11 out of 167 (6.6%) for both.

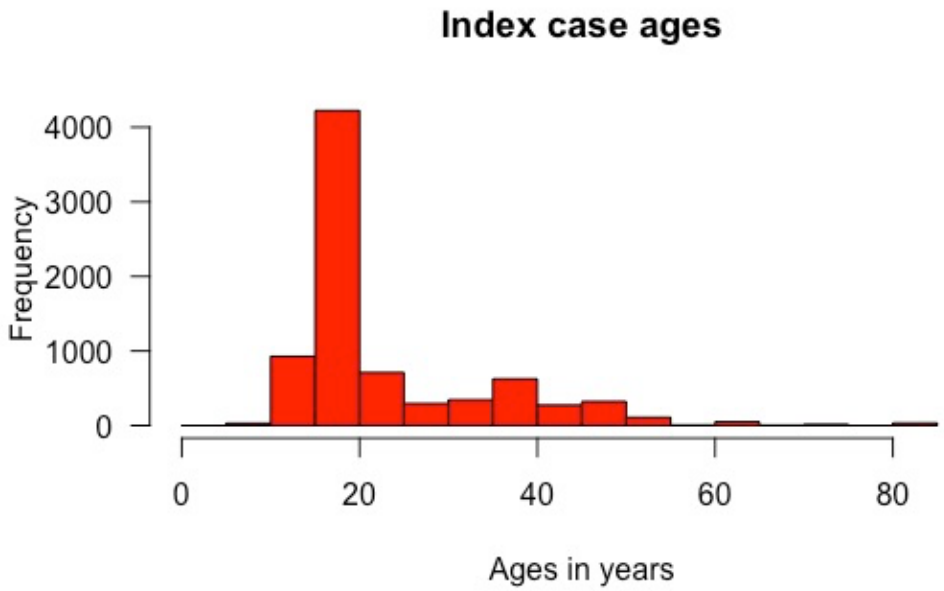

Figure 6 - Frequency of index cases ages

5.3.2.2 Ethnicity

Table 4 - Frequency of ethnicities of index cases

Ethnic group	Frequency of index cases (proportion)
Bangladeshi	5 (3.0%)
Black-African	37 (22.2%)
Black-Caribbean	9 (5.4%)
Black-Other	2 (1.2%)
Chinese	3 (1.8%)
Indian	28 (16.8%)
Mixed/Other	14 (8.4%)
Pakistani	43 (25.8%)
White	26 (15.6%)
Total	167 (100%)

Ethnicity data was available for all the 167 index cases (Table 4). Proportionately, individuals of Pakistani ethnicity represented the highest number of index cases (25.8% compared with 13.5% proportional population in Birmingham from census data) followed by Black-African individuals (22.2% compared with 9% proportional population) and those of Indian ethnicity (16.8% compared with 6% proportional population). Furthermore, it is worth noting that non-British Caucasians made up 4.8% (compared with 2.7% proportional population) of our index cases with British Caucasians making up 10.8% (53.1% proportional population). Health security agency (HSA) data for the region differed in several ways. South Asian ethnicity made up 48% with a majority Indian ethnicity. White ethnic groups made up 23% and Black ethnic groups made up 20%. [262]

5.3.2.3 Immigration

Where immigration data was available, 83 (49.7%) of index cases were from the UK with 84 (50.3%) from abroad. Immigration data was available for 65 out of the 84 (77.4%) index cases who were not born in the UK. Of these, 12 (14.3%) were from Pakistan and 10 (11.9%)

from India. The largest homogenous group of individuals in terms of immigration or native status are British born, 49.7%.

Time from immigration to disease development varied. Median time to disease development was 5 years with an interquartile range of 2-11 years. The total range was 0-52 years. The most frequent time period was 1-2 years since immigration. However, our data demonstrates that cases could occur in this population several decades later with a standard deviation of 10.5.

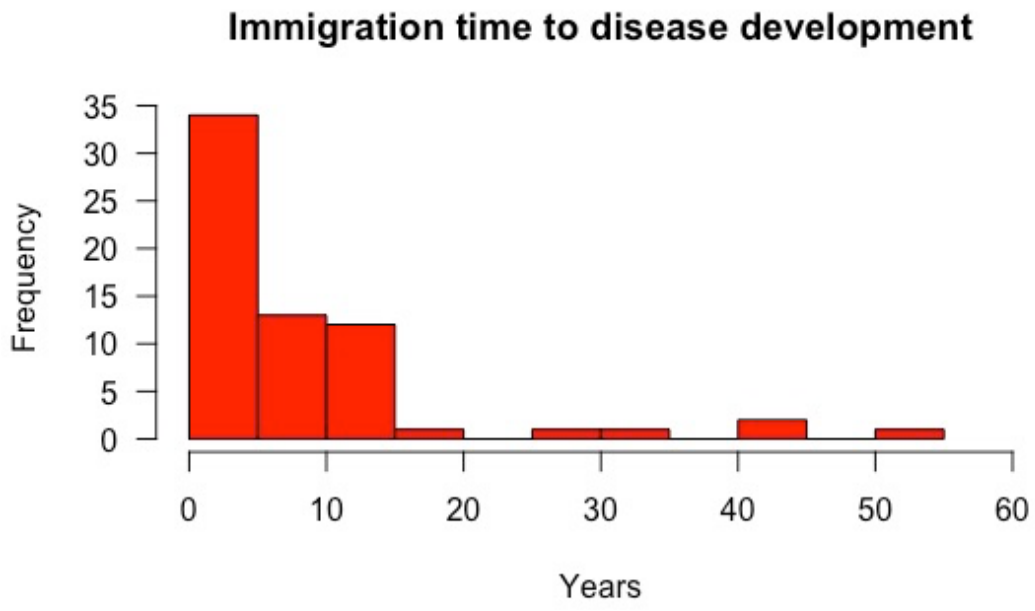

Figure 7 - Time from immigration to disease development in index cases

5.3.3 Lifestyle

Table 5 - Underserved population factors in index cases

	Alcohol history			Illicit drug use						Prison history			Homelessness		
	Missing	Present		Historical			Current			Missing	Present		Missing	Present	
		Yes	No	Missing	Present Yes	No	Missing	Present Yes	No		Yes	No		Yes	No
Index cases	2	4	161	1	12	154	1	0	166	3	6	158	3	3	161
Total	167			167			167			167			167		

Alcohol use information was not available for 2 (1.2%) cases. 161 (96.4%) of index cases did not report any excessive alcohol use. Only 4 (2.4%) were recorded as having heavy alcohol use.

The majority of index cases denied prior drug use, 154 out of 166 (92.2%) index cases. For 1 case, no data was available. 12 (7.2%) cases had a history of historical drug use. Current drug use, during contact investigation and treatment was denied by all index cases.

Incarceration history was noted for 6 out of 164 (3.7%) of contact cases. Data was not available for 3 (1.8%) cases. The majority of cases were data was available, 158 (96.3%) did not have a history of incarceration. Of those with a prison history, 1 case was in prison more than 5 years ago, 2 cases were in prison less than 5 years ago and 3 (1.8%) were currently incarcerated.

Homelessness was recorded in 3 (1.8%) of cases. The majority of our study population 161 (94.4%) had a fixed abode. Data was missing for 3 (1.8%) cases.

5.3.4 Disease characteristics

5.3.4.1 Sputum and culture results

Sputum smear results were available for 143 of 167 cases. 44 (30.8%) index cases had smear negative results on presentation with the majority of 99 (69.2%) cases being smear positive.

Culture results were available for all 167 cases. The majority of cases, 116 (69.5%) were culture positive with 51 (30.5%) being culture negative.

5.3.4.2 Resistant disease

There were a total 12 (7.2%) cases with some level of drug resistant disease. 6 of these cases (3.6% of all index cases) met criteria for multidrug resistant TB (MDR-TB). The varying levels of drug resistance is demonstrated below.

INH: 6 (50.0%)

INH, RMP: 4 (33.3%)

INH, RMP, EMB: 1 (8.3%)

INH, RMP, EMB, ETH: 1 (8.3%)

MDR (all): 6 (50.0%)

5.3.4.3 Site of disease

The majority of cases involved were recorded as having respiratory disease 149 cases (89.2%). Data was not available for 2 cases (1.2%) however given their involvement in a TB incident contact tracing event, it is fair to assume they had respiratory symptoms. 16 cases had bronchial disease (9.6%). Pulmonary disease is assumed to spread TB and while cases may have additional extra pulmonary disease, pulmonary cases would be contact traced.

5.3.5 Contact case characteristics

5.3.5.1 Age

The median age for contacts was 19 years old with an interquartile range of 16-29 (total range 0-97 years of age). School age ranges described the majority of contacts with 12-18 years of age representing 50% of contact cases. The less than 15 years of age group (1189) represented 14.7% of contacts.

5.3.5.2 Gender

Data was not available for 425 (5.3%) of 8091 contacts identified. Where data were available, females made up 4482 (58.5%) with males contributing 3184 (41.5%). There is a female preponderance in contacts even if the missing data is taken into account.

5.3.5.3 Relationship to index

The majority of contacts were casual 4110 (50.8%) with close contacts representing 3423 (42.3%) and household contacts 558 (6.9%). The majority of contacts were identified in a non-household (congregate) setting 7533 (93.1%) with the remainder, 558 (6.9%) occurring in a household setting.

5.3.5.4 Ethnicity

Data was absent for 3370 contacts. Data was available for 4721 contacts:

Table 6 - Frequency of ethnicities in contact cases

Bangladeshi	5 (0.1%)
Black-African	700 (14.8%)
Black-Caribbean	119 (2.5%)
Black-Other	30 (0.6%)
Chinese	83 (1.8%)
Indian	467 (9.9%)
Pakistani	1659 (35.1%)
White	1308 (27.7%)
Mixed/Other	350 (7.4%)

Ethnicity data demonstrates that the largest ethnic group as contacts identified as Pakistani (35.1%). This was followed by Caucasians (27.7%), Black-Africans (14.8%) and Indians (9.9%). This compares to the index cases; Pakistani (25.8%), Black-African (22.2%), Indian (16.8%), White (15.6%).

5.3.6 Contact Investigation outcomes

Of the 8091 contacts screened, 5453 (67.4%) did not have any evidence of disease. A further 1783 (22.0%) did not engage with follow up and investigation and defaulted from contact tracing. A small number of contacts identified received vaccination (BCG) 65 (0.8%).

Of the remaining contacts, 605 (7.5%) were found to have latent infection and 41 (0.5%) had active disease.

Index cases were described as having smear positive or smear negative disease. 7947 smear results were recorded from 8091 contacts. Smear data was absent for 24 index cases. of these outcome data was absent for 120 contacts. The table below compares outcomes for contacts of smear negative and smear positive index cases.

Table 7 - Contact investigation outcomes based on index case smear results

Investigation outcomes for contacts	**Index case smear status (% of total in row)**		
	Smear positive	**Smear negative**	**Total**
Recorded smear data	6616	1331	7947
Missing smear data	76	44	120
Active disease detected	38 (92.7%)	3 (7.3%)	41
Latent infection detected	508 (84.0%)	97 (16.0%)	605
Defaulters	1408 (79.0%)	375 (21.0%)	1783
BCG vaccination conducted	40 (61.5%)	25 (38.5%)	65
Disease free contacts	4622 (84.8%)	831 (15.2%)	5453

Table 8 - Contact investigation outcomes based on contact proximity of relationship to index cases

Contact investigation outcomes	Contact relationship proximity to index case (% of total in row)			Total
	Casual	Close	Household	
Missing data	74	55	15	144
Present data	4036	3368	543	7947
Active disease detected	11 (26.8%)	8 (19.5%)	22 (53.7%)	41
Latent infection detected	182 (30.1%)	314 (51.9%)	109 (18.0%)	605
Defaulters	865 (48.5%)	710 (39.8%)	208 (11.7%)	1783
BCG vaccination conducted	11 (16.9%)	35 (53.9%)	19 (29.2%)	65
Disease free contacts	2967 (54.4%)	2301 (42.2%)	185 (3.4%)	5453

The data demonstrates that with active disease, household contacts represent the largest proportion. The next largest proportion of active disease is in casual not close contacts. When examining active disease development as a proportion of individuals screened, taking into account missing values, the relative proportions remain the same.

Casual contacts and close contacts are not different in the proportion of active cases detected.

This trend is not the same with latent cases detected. Close contacts represent the largest proportion of all latent cases (51.9%) followed by casual and then household contacts. When examined as a proportion of contacts screened in each proximity, the ratio's follow proximity relationships with the highest pick up rate in household followed by close then casual contacts.

There is a large number of defaulting cases. Casual contacts make up the most number of defaulting contacts traced. However, as a proportion of the contacts screened in that setting, both close and casual represent similar results (21%) and household contacts have a higher defaulting rate (38%).

The number of contacts screened who have no disease findings increases progressively with weakening proximity, with household contacts having a lower rate of disease free cases than casual contacts.

5.3.6.1 Contact investigation outcomes by setting

Educational settings represented the largest number of total contacts as well as the largest number of average contacts per location screened. With the exception of healthcare settings, all other settings demonstrated a large number of negative screening results (no disease detected), with educational settings having the most (71.3%) followed by work and recreational settings (53.2% and 51.9% respectively). Accommodation settings demonstrated nearly half of screened candidates having a normal outcome (47.7%). All settings had a significant number of contacts defaulting from screening, at least 20%.

Table 9 - Contact investigation outcomes by congregate setting

Setting	Category	Contacts (% of total)	Subcategory	N	Missing data		Active		Latent		No TB		Defaulters		BCG	
Educational	6520	81%	Nursery	164	10	6%	0	0%	11	7%	138	84%	5	3%	0	0%
			Primary	166	7	4%	0	0%	12	7%	100	60%	44	27%	3	2%
			Primary/ combined	161	3	2%	2	1%	29	18%	91	57%	34	21%	2	1%
			Secondary	3560	30	1%	8	0%	100	3%	1677	47%	214	6%	10	0%
			College	2171	39	2%	12	1%	161	7%	1322	61%	625	29%	12	1%
			University	257	16	6%	0	0%	15	6%	119	46%	105	41%	2	1%
			Other	41	0	0%	0	0%	0	0%	20	49%	19	46%	0	0%
Work	470	6%	Office	221	3	1%	1	0%	17	8%	148	67%	47	21%	5	2%
			Factory	249	5	2%	1	0%	70	28%	102	41%	70	28%	1	0%
Recreational	424	5%	Recreational centres	173	1	1%	0	0%	19	11%	101	58%	51	29%	1	1%
			Restaurants	150	4	3%	1	1%	38	25%	69	46%	31	21%	0	0%
			Service shop	101	2	2%	2	2%	16	16%	50	50%	35	35%	3	3%
Accommodation	172	2%	Prison	35	5	14%	0	0%	1	3%	17	49%	12	34%	0	0%
			Housing	137	3	2%	2	1%	8	6%	65	47%	59	43%	0	0%
Healthcare	505	6%	Care home	127	0	0%	4	3%	10	8%	33	26%	66	52%	0	0%
			Home care	49	0	0%	0	0%	1	2%	42	86%	19	39%	1	2%
			Hospital	232	6	3%	3	1%	26	11%	112	48%	74	32%	11	5%
			GP practice	16	1	6%	0	0%	1	6%	13	81%	1	6%	0	0%
			Outpatients	81	2	2%	0	0%	7	9%	55	68%	17	21%	0	0%
All	8091			8091	137	1.7%	36	0.4%	542	6.7%	4274	52.8%	1528	18.9%	51	0.6%

5.4 Discussion

5.4.1 Summary of findings

Our data demonstrated that 112 congregate settings served as incident locations. However, there were 167 incident events declared and investigated in the study timeframe. This represents recurrent events occurring in the same location. This occurred in 30 locations, with several locations demonstrating more than 1 recurrence.

There may be several explanations for this finding. The number of incidents that occur in the same location could represent new cases with no links to earlier incidents or may represent on-going chains of transmission. This latter link may be due to missed contacts in the initial screening process, or at risk contacts who did not engage with services. In addition, for the contact tracing episodes that occurred earlier on in the study period, failure of the diagnostic tests may also be a cause for missed cases (skin/ Mantoux testing versus the newer IGRA testing).

The largest categories for transmission events are educational settings (53.3%), followed by healthcare and then work settings. Interestingly accommodation settings (included prison populations) is the lowest representative group.

Subcategorising the data allows a more descriptive approach to understanding transmission events at these locations. Additionally, not all broadly grouped congregate settings have the

same social interactions and therefore individually represent a varying risk. A key example here are the sub-categories in educational settings. Here at least 2 factors play a part. The first is age. Children (<15 years of age) are more likely to have acquired TB in a household setting given their age and increased likelihood to have most social interactions with family members. Older students and those at college and university level are less likely than younger children to have acquired TB at home. In addition, their social interactions are more likely to occur with peers and in a variety of locations (non-household). [170] [171] Our data demonstrates this trend with a higher number of educational setting outbreaks occurring in secondary school level (including secondary's with colleges) as well as college level settings. These schools often have pupils aged 11-18 with colleges only including students 16-18. These two settings represent the higher age end of the student spectrum. This trend does not continue to university level, the highest age category. Where the representation of incidents drops back to pre-secondary level. One possible explanation for this finding could be the population being examined. Universities would draw from a larger more disparate population group and therefore represents a more heterogeneous population.

Our data correlates with PHE reported figures for 2017. Nationally there was a male predominance (58.4%), this compares with our finding of 54.5%. The age data for the index cases in our study populations did differ from those reported nationally in particular the most frequent ages and age ranges. Nationally this is reported as between 30 to 34 years (16.1 per 100,000) and 35 to 39 years (16.0 per 100,000). Our findings demonstrated a lower age range for the most frequent presenting ages. [263]

The World Health Organization (WHO) differentiates adults and children in its global report on TB. WHO describes the cut off age as 15. [162] Adults (>15 years of age) made up 149 of the 167 (89.2%) index cases and those </=15 years of age made up 18 out of 167 (10.8%). Under 15 year olds made up 9.0% of all ages presenting as index cases and there were no cases below the age of 9 compared. This compares with low rates of children presenting with TB in national data where those ages 5-9 were the lowest reported category. The most frequent ages encountered in our data were in the 18-year-old age group and 20-year-old age group (11/167 – 6.6% for both). [263]

In terms of ethnic data, proportionately, our data do not correspond with national figures and this divergence likely reflects the demographic make-up of the population in our city mirroring migration patterns and communities with a proportionately larger Pakistani population.

While more than half of the index cases are born abroad and migrate to the UK, subsequently developing active disease, there is a substantial burden of disease occurring in UK born individuals. This is not unexpected, given that the majority of these UK born cases occur in families of migrants, representing second generation migrant groups. Our data also demonstrates the fluctuation over 2 decades where patients could present with active disease. However, the median time to presentation was 5 years and the most frequent time periods were 1-2 years, correlating closely with national data.

Three (1.8%) cases were homeless during screening episodes. Given the national demographic trends, our data may reflect a lower estimate of the affected and at risk group. Contact tracing in this group is difficult given the associated chaotic lifestyles, difficulty engaging with services and complications with outreach interventions given no fixed abode. [218, 264]

Four (2.4%) cases admitted to current problematic alcohol abuse and 12 (7.2%) had a previous history of illicit drug use. Interestingly none of the index cases admitted to current drug use.

Culture positive results were recorded in 69.5% of cases in our study population. This compares favourably to the figure of 62% nationally in 2017. The national figures represent a decrease in culture positive confirmation from 64% in 2016.

Our data demonstrates a significant minority (30.8%) were smear negative. This correlates with the preponderance of smear positivity in presenting index cases; nationally this figure is 52.9% compared with our higher rate of 69.2%. Smear positivity has been associated with a higher infectiousness rate, a negative smear test suggests a lower risk of transmitting disease. [183]Our population demonstrates that smear negative individuals still represent a significant risk in congregate settings and unlike previous papers, suggests the risk is not restricted to the 'first ring' of proximal exposure. From our data 24 sputum results were not recorded, representing 14.4% of the study population, this correlates nationally with a 13% unrecorded sputum smear result. [263]

TB resistant disease is a growing problem globally. [32] National data from PHE demonstrates that resistant disease has not declined and the current proportions of resistance are 1.8%. Our data over a 7-year period demonstrates all levels of resistance at 7.2%. Of these resistant isolates, multidrug resistant strains made up 50% (3.6% of all 167 isolates) with single drug resistance making up the remaining 50%. This is higher than the nationally reported figures. Given our study population represents a higher than average TB prevalence and the representative increase in migrant population and individuals from high risk countries, this may explain the discrepancy. Worth noting is the recognition that associated social risk factors with MDR disease results in an increased mortality rate and loss to follow up. [263]

The average age of contacts was still low at 24 with the most frequent age being 19. Those less than 15 years of age made up just under 15% of all contacts. Interestingly, the secondary and college school age range (11-19 years of age) made up 65% of all contact with the 14-19-year age range representing 53.3% of all contacts. This suggests a higher likelihood of out of house transmission events [170, 171] The majority of contacts were female (58.5%). This is in contrast to the majority of index cases being male. When compared to national testing data, there is a female preponderance (54.9%) however this relates to pre-entrant screening. Despite there being a significant number of absent gender data (5.3%) this would not be enough to explain the gender difference. A much larger proportion of white/ Caucasians occurred as contacts than were represented by index cases.

This finding may represent the populations in congregate settings, itself a reflection of the background population ethnic diversity.

As would be expected, household contacts made up the smallest proportion of all contacts (6.9%). This reflects the relative size of our population households. It is worth noting that household sizes are likely to vary between different ethnic groups, reflecting cultural norms. E.g. those of a South-Asian background potentially have a multi-generational household; involving up to 3 generations of family members. This is a relevant point when assigning relationship status to index and contact cases and may dilute out the relevance of a household vs close contact label.

Casual contacts made up the largest group (50.8%) of all contacts traced. There are several important points to make about this group. Given the nature of the congregate setting, these individuals, in most cases, will not have been named contacts of the index case who would identify close/ household contacts. In addition to this, casual contacts in larger outbreaks and those in certain types of settings such as school will have been screened on the basis of their relative groupings as per national guidance. [223] Screening recommendations use preformed divisions to aid screening e.g. screening classes, year groups or the entire educational establishment. These arbitrary screening levels result in larger screening sizes without a corresponding pick up rate. The screening of these individuals therefore follows the formations they are grouped in and not necessarily their risk of exposure. In addition, the screening of these casual contacts does not take into account the possibility of environmental factors such as shared spaces e.g. lunch locations.

Educational settings had the highest proportion of normal screening outcomes (71.3%) which would fit with this observation. The number of screened but negative findings decreases when the screening size is more tailored. An example of this from our data set would be in a healthcare setting, where staff and patient exposure can be more carefully tracked and appropriate contacts identified. In healthcare settings, those screened who had no findings amounted to 5% compared to the 71% in educational settings.

Screening large groups of people to demonstrate a low pick up rate (8.0%) latent and active disease.

National guidance currently advises a stone in pond model approach. Before large screening episodes of close and casual contacts, more proximal contacts need to demonstrate presence of disease first. [90, 223] Our data demonstrates that this may not reflect the risk of disease acquisition in congregate settings. While household contacts had the highest proportion of active disease of people screened (4.1%), casual contacts had a higher rate of active disease than those identified as close contacts (0.3% and 0.2% respectively).

There is a significant proportion of defaulting individuals, at least 20%. This may represent an additional burden of disease. In addition, those least likely to engage with follow up and investigation may have additional factors or behaviours predisposing to disease acquisition and development.

5.4.2 Strengths and weaknesses

5.4.2.1 Limitations

Secondary schools generally cater to specific catchment areas and therefore mirror the demographics of the pool of neighbourhoods that supply the schools. Ethnic clustering is a possibility and given the increased prevalence of TB in immigrant ethnic groups, could explain the pattern seen.

The number of cases with social risk factors represents a low proportion in the study population. This differs from PHE data and previously reported studies. Drug and alcohol abuse may be underrepresented in our data. This could either be because of the cultural implications and stigma around the use of these substances in some of the larger ethnic groups in our study population. Alternatively, concealment by index cases around questions of illicit activity may also be a factor in underreporting. This is not uncommon in standard contact tracing approaches and has previously been described as a barrier to identifying named contacts. [54, 265] There is also a high rate of non-engagement and defaulters. This population may represent a higher risk group and could potentially change the results.

The assignment of contact relationship to an index case is made by an assessing member of the service delivery team i.e. a TB nurse. The usual labels given are close versus casual contacts with a subcategory of close contacts being household contacts. While this assignment has an element of objective assessment to include proximity to patient and time of exposure, it is usually determined through interrogating the index case. This process is

therefore prone to bias and may not correlate well with practical outcomes. This holds true for index case assignment of proximal relationship also. [265]

In addition, some ethnic groups may have a different relationship to secondary degree relatives. So that even though they do not share a household they may represent close contacts. This again has implications for the assignment of close vs casual contacts. Authors have demonstrated further that such arbitrary contact assignments might not correlate with risk of exposure and subsequent development of disease. [243] Travel in migrant populations may also prove to be a source of acquisition of disease. These journeys to visit friends and relatives in endemic countries would be missed by new entrant screening programs.

A fundamental difficulty in screening in a population with a higher background rate of TB is the difficulty in assigning direction to transmission. Proving an index case has caused latent infection in the population screened can only be confirmed if those latent cases develop active disease and there is a molecular match from samples. While this could certainly affect how epidemiological services describe an incident event, in terms of public health and addressing the burden of latent infection, it is largely irrelevant.

Data completeness from this database remains an issue. Missing data in different variables and categories compromises the data integrity and could impact on the conclusions drawn. There is limited recorded data on social links and relationships in congregate settings. Whilst work/ social relationships are broadly described, the interactions and time spent are lacking.

This limits inferences. Ventilation of a location can have an impact on circulating respiratory droplets and consequently infection rates. No information on congregate settings and ventilation properties or shared spaces are recorded in the dataset.

Over the study time period, the diagnostic technologies used changed. Tuberculin skin test (TST) was replaced by interferon gamma release assay (IGRA). This may have affected earlier detection of cases and resulted in screened cases going on to develop active disease and contributing to recurrent incidence episodes.

Molecular testing also changed during the study period. MIRU-VNTR was the mainstay throughout the study period, however, WGS became increasingly used towards the end of the study period. That said, WGS has been used retrospectively and reinforces the MIRU-VNTR in particular cases.

5.4.2.2 Strengths

This is the largest dataset of its kind to be reported on (in the view of the author). The combination of WGS and completeness of the dataset gives a unique look at TB in a large urban population. Furthermore, despite lacking data and the desire for more information on certain aspects, the dataset collected reflects the statutory public health information requirement and is therefore broadly applicable to other settings.

Given that the study population has a broad ethnic diversity and resides in a high-income country, this increases the generalizability to other low and middle incidence populations.

The study period took place over a 7-year time scale. This would mean that latent to active disease transformation could occur and any molecular links investigated further. In this way, TB in the same locations can be studied and associated cases examined for social and geographic links.

Data was collected contemporaneously by PHE and TB services. Whilst diagnostic and molecular changes occurred during the study period, this transition would have increased the reliability of the data over time and added more confidence in any patterns seen. Furthermore, data collected reflected a variety of locations and settings (schools, work places etc) and a large number of incident events. Both factors that would contribute to reliable patterns drawn.

5.4.3 Literature comparison

TB incidents are defined by the European Centre for Disease Control as:

> *"the exposure to an infectious TB case with the likely occurrence of TB transmission including outbreaks in a specific setting. An outbreak is defined as an incident with two or more epidemiologically linked cases of active TB disease." [266]*

The National Institute for Clinical Excellence (NICE) guidance (UK) also notes TB incidents as an issue and suggests undertaking a risk assessment on a case by case basis. In congregate settings, such an assessment should ideally take into account the risk of transmission (index case infectiousness and duration and proximity of contact and contact case vulnerabilities).

In addition, there is a suggestion that examining the build environment should factor in the risk exposure assessment with the need to assess for ventilation and overcrowding. [267]

Incident episodes are normally declared as part of TB services and Public Health England collaboration and take into account the location for events. Congregate settings naturally pose a higher risk for this reason.

Given the short time periods over which subsequent episodes of recurrence occurred and the molecular links demonstrated with multiple incident episodes, it is likely that missed transmission points are responsible for recurrent incident episodes in the same locations. In some cases, this can be demonstrated by subsequent genetic links where data exists. However, this cannot be proven for all cases. Where recurrent episodes exist and there is no or insufficient molecular evidence, this may represent the prevalence of TB in the population frequenting these locations e.g. schools in an ethnically homogenous area with a higher risk for TB i.e. immigrant population groups.

Where molecular links do exist and there is no clear epidemiological links, certain location specific risk factors should be considered. This could include shared areas of congregation with poor ventilation, thus increasing the risk of TB transmission. [227, 268] This is more of a significant finding where locations have demonstrated a greater number of recurrent incidents (in some cases up to 6 times) and where epidemiological links are lacking.
Given the progressively enlarging numbers of contacts screened for subsequent incidents in the same location, recurrent incidents represent a significant resource investment for public

health and TB services. The ability to identify all the potential at risk individuals early on in a screening episode could prevent needless waste of limited resources, as well as reduce the risk of on-going transmission chains from missed contacts. Consideration of larger social networks should also take into account potentially shared spaces as well as overly stated contacts during the initial screening episode. Furthermore, the need to consider shared social spaces in congregate settings should not serve only as an adjunctive tool for a risk assessment of likelihood of transmission as stated in NICE guidance [267] but should also be used to identify a larger contact pool for tracing.

The location categories represent common themes in the literature on contact tracing [261, 269] they also represent broad social categorisations that would be readily understood by TB services and therefore represent a practical tool for contact tracing in congregate settings [270]

Congregate settings as locations for transmissions have been described by authors as warranting consideration over and above the individual index case investigation i.e. they play an independent role in disease transmission. Several authors have expanded on this theme, demonstrating the key role that locations play in transmission events, particularly when named contacts do not include all possible contacts involved. [227, 265, 271].

Prison populations have often been reported in the literature as sources for TB outbreaks. [218, 238, 272, 273] The reason prison settings as a congregate location for incident events does not feature higher in our data may be due to the demographics of our population. This

may be the over representation of alternative locations such as the relative density of schools and the risk factors inherent in the populations in the catchment locations. In addition, prisons in our setting may be under-represented given the low number of such facilities in our region and the higher awareness in these settings as well as the likely population groups being incarcerated and their relative TB risk compared to the general population in our study group.

Public health data from 2017 demonstrated that the countries representing the highest frequency for the origin of non-UK born active TB cases were India, Pakistan, Romania, Bangladesh and Somalia. [263]

While disease in foreign born individuals most likely represents reactivation of latent infection, there is a risk that TB may be acquired whilst in the UK within the ethnic community as demonstrated by a previous Dutch molecular epidemiological study where 20-30% of TB isolates were shown to have originated within the country of migration and not the foreign-born individual's country of origin. [274]

Nationally, country of birth and time to TB notification demonstrates considerable variability. The highest risk countries (India, Pakistan, Bangladesh and Somalia) all demonstrate time to reporting of 5 years. Other countries such as Romania demonstrate a lower time to presentation of 2 years. [263]

Nationally there is an increasing trend of >/= 11 years since immigration for presentation with TB (55.5% increase since 2010). [263] This correlates with the literature with some studies reporting that the increased risk among foreign-born immigrants could continue for up to 20 years' post migration. Reactivation rates are highest during the first 2–5 years following migration with the children of foreign-born residents having higher active TB rates than the general population. [274, 275]

In most low-incidence countries, the majority of cases of active TB arise among the foreign-born population; this can result in significant transmission within certain foreign-born communities in these countries. [274, 276, 277] However, restriction fragment length polymorphism studies have detected relatively little TB transmission from foreign-born residents to the general population. [278, 279] The estimated proportion of active TB cases among the native-born that can be attributed to transmission from the foreign-born may be as low as 2% or 11%, or as much as 17% In one USA study, foreign-born TB patients were more likely to have acquired TB from USA-born individuals than vice versa. [280] At present, it appears that, in low-incidence countries, the overall public health impact of TB among foreign-born persons is modest. [274, 281]

Recent PHE data suggests a growing problem in native born population groups. Immigrant populations have seen a continuing decline in incidence of TB attributed to new-entrant screening programmes. Native born populations in the UK also tend to have multiple risk factors that impair engagement with TB services. [263]

A significant population group within those affected by TB are the so called underserved population (USP) group. Defined by PHE as:

> *'individuals whose circumstances, language, culture or lifestyle (or those of their parents or carers) make it difficult to recognise the clinical onset of TB; access diagnostic and treatment services; self-administer treatment; or attend regular appointments for clinical follow-up.' [41]*

This demographic block is populated by individuals with one of more risk factors known to propagate disease transmission while impairing service engagement and therapeutic interventions. This includes numerous overlapping individual and contact risk factors including; drug and alcohol use; smoking; unemployment and poverty; co-existing mental health issues; migrant and asylum backgrounds; overcrowding; and poorly ventilated environments.

5.4.4 Clinical implications and further research questions

This chapter demonstrates the first large scale report of outcomes of routine contact tracing over a significant period of time, highlighting the large populations involved in congregate settings and the non-linear relationships with presence of active and latent infection detections in proximity groups. This is at odds with the accepted wisdom and utility of the stone in pond model approach to contact tracing. Previous individual incidents have been reported as distinct events, however entire contact tracing outcomes have not previously

been reported. This represents a significant contribution to services involved in contributing to contact tracing guidelines.

A key limiting factor that reflects clinical practice and outcomes relates to the high rate of 'defaulters'. Defaulters are potentially a large proportion of contacts to be screened and could represent a significant source of infection as an at-risk population. Given the risk of recurrent incident episodes demonstrated, more work is needed to understand and engage with this group. This poses a challenge for TB and public health services largely due to the limitations around compelling compliance.

Smear negative disease in index cases has traditionally been seen as much lower risk. Whilst associated with fewer disease outcomes in contacts (active and latent case detection), our smear negative population still had these associations. This suggests that even smear negative disease poses a risk. What isn't clear is the timeline for this risk and whether the proximity and duration of exposure differs significantly i.e. is there a clear differential risk between smear positive and negative index case exposures enough to inform a different approach to contact tracing and risk assignment to contacts.

Screening in larger population settings (e.g. schools) proceeds on a pre-supposed configuration e.g. class, form, year and school basis. This arbitrarily extends the contact tracing pool without necessarily detecting the at-risk population. A more nuanced approach that takes into account individual movements and exposures might be able to cut down the screened to exposed ration. Given high number of absent findings i.e. no disease detected,

the work to outcome ratios demonstrate poor resource allocation. This is especially true in the most numerous setting, educational locations. The largest number of negative findings occur in casual population groups. A more programmatic approach has been shown to identify more at risk contacts. [54, 186]

The labels used to describe proximity relationships are arbitrary and reliant on patient reported contacts. This has a risk of missing key exposures and is subject to biases from individual awareness and behaviours. In addition, this does not take into account proximity of unrecognised contacts in different location areas nor environmental factors. [265] Correlations between labels of casual, close and household do not appear to be clear cut. There is also a limited basis for which to assign these labels and while proximity is a recognised risk for spread of TB, the duration and distance are not clear. A more objective and reproducible system may need to be investigated. In particular, one based on social distance and less on relationship labels the latter not necessarily reflective of exposure risk.

Latent infection ratios are in keeping with expected proximity changes. This finding is more difficult to explain and may reflect the background rate of TB. Latent infection cannot be molecularly linked to active disease cases and may be unrelated to active case detection. That said, latent cases are still important to track and treat to reduce morbidity and additional costs of active disease. Given the goal to reduce incidence of TB to 1 per 100,000 [258] identifying and treating latent cases early can prevent on-going transmission chains. Ultimately, TB services aim to reduce the risk of recurrent infections (latent to active disease transformation). Barriers to this reflect a limitation in being able to reliably assign a risk to

contacts resulting in large contact screening populations with low yields of disease as well as high rates of defaulters without a clear follow up strategy. This data is comparable to the annual published PHE data and therefore applicable to areas with contact tracing programs. [42]

The next chapter will focus on a subsection of the above data. Understanding the long-term implications of failed contact tracing is important to demonstrate the need for a change. The next chapter will therefore look at outbreak events from initial contact traced incidents.

Chapter 6 – Tuberculosis outbreak investigations in a large UK urban environment

This chapter will look at outbreak events following contact incident investigations where these have been unsuccessful in preventing transmission.

6.1 Background/ Introduction

As we have seen in the previous chapter, uniquely describing the outcomes of large scale congregate contact investigations, there are a large number of contacts generated through the process and there are additional individuals lost to follow up. This chapter will focus more on the investigations where more than one individual with active disease is discovered. The reason for this, is that each active case represents a further node for the propagation of disease over and above the original case. Additional contacts are generated and this can lead to chains of infection spreading.

TB rates increased in the early 1990's reversing a decades long trend of steady decrease. Multiple factors were thought to contribute including the HIV epidemic, global migration and less emphasis on TB control processes. [32, 282] Recently, Mtb rates have decreased in the UK, from a high of 8,280 in 2011 to 4655 in 2018. [283] A number of factors have contributed to this decline including a new-entrant screening program and early and effective treatment of latent TB. [284] TB rates are not the same across the UK, with certain regions (London and the Midlands) and certain population groups (migrants and those with

social risk factors) having higher rates. [285] Despite the decreasing trends, the rate of decrease has slowed (previously 10% year on year, now 1%). In order to continue this decline, the early detection and prevention of Mtb cases is a priority. [34]

TB infection does not result in active disease immediately nor in all exposed individuals. [56] There is therefore a lag between latent infection acquisition and active disease development. Epidemiological investigations can identify latent cases who can then receive preventative treatment. Currently there is no test to predict who will go on to develop disease. Molecular epidemiology techniques like whole genome sequencing (WGS) require active disease to demonstrate links. Given the time gap between epidemiological investigations and subsequent retrospective WGS, cases that were missed by epidemiological investigation could transition from latent to active disease and subsequently be identified by WGS. This would demonstrate missed epidemiological intervention opportunities. By comparing epidemiologically identified active cases with WGS identified active cases, the lack of overlap can demonstrate gaps in current contact tracing approaches.

This chapter will aim to examine outbreak events in the dataset, obtained in a large urban environment, the city of Birmingham in the UK. Data has been recorded in a detailed database and with whole genome sequencing data, providing a unique opportunity to compare this dataset with similar studies in other nations. Birmingham is a large city in the West Midlands of the UK, unique to this area is the larger proportion of migrant population compared with other UK cities with few exceptions. [269]

6.2 Methods

6.2.1 Study design

We used data, collected as part of routine public health investigations in a large UK urban environment between 2010 and 2017 to conduct a descriptive analysis of index and contact cases demonstrating on-going transmission chains. As part of this analysis, we examined occurrences of TB disease in congregate settings and in particular where these occurrences triggered further public health and chest clinic intervention and screening episodes. This follow up period occurred between 2 to 6 years (2019) after the original contact screening episode.

6.2.2 Data sources

Cases of TB disease are routinely recorded by public health services in the UK once they are notified by clinical services. This is a statutory requirement. Where these cases could involve contacts and on-going transmission chains, TB nurses working from local chest clinics investigate.

Contacts are identified through index case interviews as well as local occupational health team cooperation and site visits. The interview data is transcribed into the local TB database. Where further information is required, additional interviews are carried out by TB nurses.

Routine data collected as part of these investigations (table 1) include demographic data on index and contact cases (including age, gender, ethnicity, and migration); Index case risk

factors such as homelessness, substance abuse and incarceration; relevant medical conditions and occupation. The contact relationship to index cases is also recorded, as reported by the index case. Where this is lacking, TB service teams would assign a proximity relationship (e.g. close/ casual). This designation is not based on specific definitions and is therefore subject to operator variability. [223]

WGS has replaced Mycobacterial Interspersed Repetitive Units - Variable Number of Tandem Repeats (MIRU-VNTR) in the West Midlands since December 2016 (with a pilot programme running from Feb 2015). Prior to this VNTR was used and retrospectively supplemented by WGS. It can be used to diagnose Mtb, predict drug resistance and identify genetic links. It therefore has applications in contact tracing. This is possible once a positive sample has been obtained from a patient with active disease. The advantages of WGS have been described previously but are only truly being realised with routine prospective coverage. [54, 286-288]

WGS has demonstrated molecular links between TB cases without recorded epidemiological links. [54, 289] Put differently, molecular evidence demonstrates missed cases of latent Mtb who can subsequently go on to develop active disease.

6.2.3 Definitions

Index patients are the cases of new or recurrent tuberculosis around which a contact investigation is centred. Contacts are individuals who have been exposed to index cases. [257]

Cases are described as having either epidemiological or molecular links. [290] Epidemiological links are acknowledged significant exposures between the index and contact cases either as a named contact or through shared social location during the period of infectiousness. Molecular links are demonstrated through WGS. This is defined as a significant link through the number of single nucleotide polymorphisms (SNP's). where SNP's demonstrate fewer than 5 changes, transmission between cases is presumed to have occurred. Less than 5 SNP variation represents limited genetic variability and therefore correlates with recent transmission. [287, 291]

In accordance with the CDC guideline for contact investigation definition of an outbreak; two or more cases of secondary active Mtb with known epidemiological links constitute a cluster. [290] Latent infection is defined by the WHO as having evidence of TB infection with no clinical, radiological or microbiological evidence of active disease. This definition has been called into question with the possibility of subclinical disease. [59, 68]

6.2.4 Factors and categorisation

In keeping with similar studies of TB outbreaks in congregate settings, shared social locations were categorised based on the key function of the location e.g. school/ educational environments, health care, work etc. [261, 292]

6.2.5 Analysis

We undertook a description of the characteristics of index and secondary cases; a description of the location where outbreaks have occurred; and identify any discrepancies

between epidemiological identified active cases and subsequently linked active cases through retrospective WGS.

6.3 Results

6.3.1 Number of events

Over the study period 2010-2017, there were 167 incident investigations. Of these, 9 incident investigations had ≥3 active cases (i.e. the index case and at least 2 secondary active cases that were linked by typing to the index case) and were eligible for inclusion in the analysis. 20 had 2 active cases (including the index case) and 138 had 1 active case (the index case).

6.3.2 Location of events

The 9 incident investigations included a total of 35 active cases (range 3-6 per outbreak) and occurred in three categories of congregate settings, seven in educational settings, one each in healthcare and accommodation (assisted living environment) settings (Table 10).

Table 10 - Total contacts, active and latent TB cases grouped by setting and outbreak number

Location			Total contacts screened and TB Cases					
Outbreak	Category	Sub-category	Total	Active (%)		Latent (%)		Default* (%)	
1	Educational	Primary/ secondary (4-16 years)	104	2	(1.9)	27	(26.0)	12	(11.5)
2	Educational	Secondary school (>12 years)	270	2	(0.7)	10	(3.7)	56	(20.7)
3	Educational	Secondary school (>12 years)	245	5	(2.0)	21	(8.6)	19	(7.8)
4	Educational	College (16-19 years)	26	2	(7.7)	5	(19.2)	6	(23.1)
5	Educational	College (16-19 years)	1652	3	(0.2)	99	(6.0)	455	(27.5)
6	Educational	College (16-19 years)	33	2	(6.1)	5	(15.2)	2	(6.1)
7	Educational	College (16-19 years)	125	5	(4.0)	14	(11.2)	4	(3.2)
8	Health care	Hospital	12	2	(16.7)	6	(50.0)	1	(8.3)
9	Accommodation	Respite: mental health, substance misuse	44	3	(6.8)	1	(2.3)	30	(68.2)

* *Contacts who did not engage with the screening process either through choice or loss to follow up.*

The first 3 locations (Table 10) were secondary school settings with students aged 12-16, with the exception of 1 school (first educational establishment) which admitted primary aged students as well (4 and above). The following 4 educational settings were colleges with an older population of students (16-19). The remaining 2 facilities represented a district general hospital in a lower socio-economic area near to a prison facility and a congregate setting, which was a hostel for homeless young, mentally ill or substance-misusing individuals.

6.3.3 Active cases

In total, there were 35 active cases; 9 index cases generating 26 active cases in contact tracing. Of the active cases detected; 17/35 (48.6%) were female and 18/35 (51.4%) were male. 29/35 (82.9%) were adults (i.e. >18years of age) and of the remaining cases, 5/35 (14.3%) were 15 years old or younger. Of the implicated index cases, only 1/9 was 15 years old with the remaining cases being greater than 18 years of age at time of transmission.

The index case characteristics are described in table 11. All index cases (9) presented as a result of symptom development. The average time from symptom onset to diagnosis was 81 days (range 3-127 days), median 98 days. The remaining active cases (22/31) were all detected as a result of the contact screening process. 8/9 of the index cases had no history of current or previous illicit drug use, no history of incarceration and there was no history of homelessness or alcohol abuse. The HIV status of all the active cases was negative. None had previous TB disease. 6/9 of the index cases were UK born. Where data was available for

immigration status for all active cases; 26/31 (83.9%) were not born in the UK. The mean time from immigration to active disease was 5 years (range 1-13); median 3 years.

Table 11 - Index case characteristics

Outbreak Number	Gender	Age	Ethnicity	Born in the UK	HIV Positive	Occupation	Social risk factors*
1	Female	18	Pakistani	No	No	Full time Student	0
2	Female	15	Black-African	Yes	No	Child	0
3	Female	20	Pakistani	No	No	Full time Student	0
4	Male	23	Mixed/Other	Yes	No	Full time Student	0
5	Female	19	Pakistani	No	No	Full time Student	0
6	Male	19	Black-African	Yes	No	Full time Student	0
7	Male	19	Pakistani	No	No	Full time Student	0
8	Female	23	Pakistani	No	No	Other	0
9	Male	30	White - British	No	No	Unemployed	4

*Social risk factors include drug use, homelessness, incarceration and alcohol abuse, graded 0-4 depending on absence of or number of risk factors present.

After clinical assessment, all index cases had culture positive pulmonary disease with 1/9 having concomitant lung and disseminated disease. One index case had MDR TB. One index case was smear negative.

6.3.4 Contact cases

In total 2511 contacts were identified from 9 incident events. The mean number of contacts per outbreak event was 279 (range 12 to 1652). Of the congregate setting screened, educational settings had the largest screened populations compared with other settings. 19/2511 contacts had previous TB and of these, 6/19 had latent infection on contact screening. None of the contacts with previous TB had active disease. 957/ 2511 contacts had

previously received BCG of these; 123/957 (12.9%) had latent infection with a further 12/957 (1.3%) having active disease.

6.3.5 Proportion of cases in household vs congregate settings

26 out of 35 active cases were presumed secondary to the index case. Of these, 14 (53.9%) active cases were found from household contact versus 6 (23.1%) from close contact and 6 (23.1%) from casual contact. Household active cases representing 0.56% of all contacts screened and casual and close contacts 0.24% each.

In terms of latent case detection as a proportion of contacts screened and in terms of relationship status, households represented 11.2% (21/187) with close contacts 9.1% (17/187) and casual contacts 79.7% (149/187).

When examined as a proportion of the total number of individuals in that group, that is, the yield per relationship group the ratios differ. Household contacts numbered 83; 21 latent cases (25.3%) and 14 active cases (16.9%). Close contacts amounted to 78; 17 latent cases (21.8%) and 6 active cases (7.7%). Location contacts represented the largest group – 2350; 149 latent cases (6.3%) and 6 active cases (0.3%)

6.3.6 Molecular and epidemiological links

WGS data allows a description of the active cases that have been linked through molecular techniques. As WGS in this population has been conducted retrospectively and cases have occurred over time since the initial investigation, this has allowed time for missed latent cases to develop active disease, present to services and undergo WGS. These molecular links

can be compared to the number of active cases detected through previous routine epidemiological investigation methods. A comparison of these two approaches is shown in table 12 and depicted graphically in figure 8.

Table 12 - Outbreaks and clusters based on congregate setting

Outbreak number	Location category	Location Sub-category	Epidemiologically identified active cases	Molecularly identified active cases in the whole cluster (WGS <5 SNP's)*
1	Educational	Primary/ secondary (4-16)	2	8
2	Educational	Secondary school (>12)	2	8
3	Educational	Secondary school (>12)	5	12
4	Educational	College (16-19)	2	8
5	Educational	College (16-19)	3	0 – not sequenced
6	Educational	College (16-19)	2	4
7	Educational	College (16-19)	5	4 back sequenced, missing people
8	Health care	Hospital	2	20
9	Accommodation	Respite: MH, Substance misuse	3	32 – On-going at time of writing.

** All active cases identified including previous epidemiologically linked cases.*

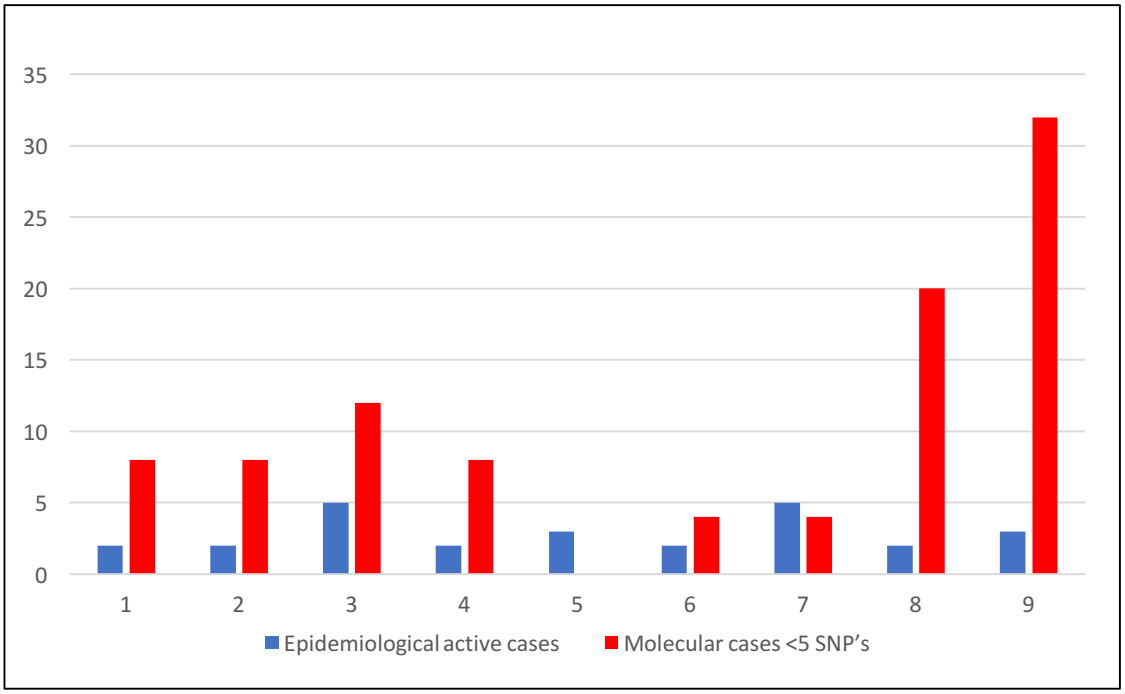

Figure 8 - Comparison of moleculari.e. WGS data (red) and epidemiological (blue) links

WGS cluster investigation (molecular approach) demonstrate transmission links greater than that demonstrated by epidemiological investigations alone. There are some exceptions seen above. Event 5 did not have any sequencing performed due to a satisfaction that there were significant epidemiological links. Additionally, outbreak event 7, whilst some contacts were sequenced, not all the individuals involved were back sequenced, again due to a satisfaction that significant epidemiological links were demonstrated. Outbreak events 8 and 9 represent high-risk populations in terms of risk factors. This is an on-going and expanding cluster from a molecular link point of view. There is a significant difference in epidemiological traced individuals and region wide molecular data.

The important limitations to note are that not all cases have had WGS linkage performed; a proportion were not culture positive and that the prospective observation period for secondary cases is by nature of disease developing, variable.

6.4 Discussion

An outbreak of Mtb represents a cascading spread of infection manifesting as concurrent or sequential infectious presentations with the potential for a chain reaction of subsequent infections. Groupings of infections can be termed outbreaks, which is more an epidemiological term versus clusters, a molecular term. [290]Correlation between the epidemiological and molecular aspects can help services assess contact tracing success.

6.4.1 Summary

6.4.1.1 Index cases

The majority of index cases involved in outbreaks in our study did not demonstrate the traditional risk factors [261, 292, 293] [218, 294] (8/9 cases). Only one outbreak event (outbreak event 9 occurring in an accommodation setting) had associated substance misuse as well as an incarceration history. The index case and contacts in this outbreak were homeless. This may explain the high rate of defaulting contacts (68%) in this setting, as well as the large discrepancy between molecular and epidemiological approaches to linking cases. [256, 264]

6.4.1.2 Contacts

Contact screening populations varied greatly (range 12-1652). Educational facilities demonstrated larger populations than those in other settings (hospital or care facility) without an increase in yield. [295] Furthermore, with the exception of 1 college location, more junior educational establishments had larger populations than more senior facilities

(college vs secondary or primary school facility). Colleges tend to have customised class schedules with different social dynamics. Junior schools tend to group students in class and year groups. Screening along these lines is therefore convenient and reflected in guidance. [223]

6.4.1.3 Defaulters

All outbreak screening events had a number of defaulters with the highest proportion occurring in marginalised populations. Given the subsequent cluster size demonstrated an alternative screening approach may be needed. [256, 296] Surprisingly, several educational establishments also had high levels of defaulters (as many as a quarter in some cases). The reasons for this are unclear.

6.4.1.4 Location

The majority of outbreaks in our population occurred in educational settings. Schools have frequently been reported in case studies as locations for TB transmission [266] and have historically represented locations for screening. [144] Junior educational facilities are described in the literature as having a greater risk. This as a result of transmission occurring in a household setting with an increased likelihood in children under the age of 15 years. [170, 171, 196] Interestingly, our data demonstrates that only one index case was this age, with the remaining index cases being adults. While this may represent household transmission and subsequent active disease development, it could represent alternative sources of infection outside of the household environment. [170] Given the relative sizes of subsequent cluster data, the latter source for infection seems more likely.

6.4.1.5 Proportions of active vs latent cases

All settings had a larger number of latent cases than active cases, this may represent recent transmission given the outbreak nature of the screening event, however it may also represent at risk population groups and background population latent TB infection (LTBI) rate. [297, 298]

6.4.1.6 Yield

A larger proportion of the household contacts screened had active and latent infection when compared with close or casual contacts, correlating with previous findings. [90]However, a significant proportion of "casual" contacts, a perceived lower exposure group, had latent infection detected. Whilst this may be due to the background rates, it still represented an opportunity to screen a higher risk population. [298] Previous studies have implicated incomplete contact tracing approaches to on-going outbreaks. [54] Screening casual contacts, despite the lower yield, may be crucial in preventing additional outbreak events occurring. Furthermore, it may be worth revisiting such labels as casual and close contacts given the lack of consensus on a unified definition and the minimal exposure time for disease acquisition. It is also worth revisiting the "symptomatic period' for contact tracing. The yield from locations differed. Educational settings are the standout example with a large number of contacts screened for no correlating increase in yield.

6.4.1.7 Clusters

We observed (figure 8) that the molecular clusters are greater than the epidemiologically identified active cases in most cases. A key example as discussed above is event 9 –

occurring in an accommodation setting, characterised by risk factors for TB and non-compliance with services. Here, the epidemiological investigation demonstrated 3 active cases. WGS demonstrates a significant cluster with more than 32 cases identified so far as well as context specific index management given the occurrence of this outbreak around a prison population.

6.4.2 Strengths

Our data represents a significant population group in a large urban environment for which there is limited reporting. Furthermore, data for investigations has been collected as per public health guidance and represents a robust data set.

Contact investigations were carried out contemporaneously by experienced clinical staff following a pre-defined method. Trends and outcomes can also be reliably tracked over the time period investigated i.e. a 7-year period, where most active disease from latent transformation would be expected to have occurred. [299]

As epidemiological investigations are essentially contemporaneous and prospective in their approach, the time period investigated allows for an assessment of the utility of the contact tracing approach. This can be measured retrospectively through WGS cluster analysis. If a molecularly linked cluster with close genetic lineage (<5 SNP's) is on-going following an extensive epidemiological investigation, it can be assumed that potential cases of latent or active disease were missed during the original investigation. The availability of WGS data

therefore provides a high-resolution assessment of outbreak investigation long term success.

6.4.3 Limitations

Our population represents a higher than average background rate of TB disease, when compared with other regions in the UK with the exception of London. In addition, there is a higher migrant population impacting the rate of TB. Subgroups of migrant populations are likely to share socio-cultural, familial and geographic links and therefore TB disease may predominate in these groups without impacting on the wider social context. [300] Having said that, it is of note that the majority of outbreak events occurred in educational settings where socio-ethnic mixing occurs. These transmission locations can therefore play a significant role in propagating spread of disease.

The sample size from which we draw conclusions is small and therefore may not be truly representative when compared with other settings.

Data collection places a higher emphasis on index case risk factors. Comparable data are not routinely collected for contacts. It is therefore difficult to assign risk to contacts on a basis other than relationship and proximity status. Furthermore, as multiple active cases exist in outbreaks, the index case may not represent the first infected case, rather the first case to present to services.

Our data demonstrates a high level of defaulters, i.e. individuals identified as contacts not engaging with the screening process. Whilst this is not uncommon in contact investigations, it represents a lack of data for individuals with a greater risk of acquisition and transmission.

WGS data was not present for all individuals identified. Retrospective sequencing was selective with clearly epidemiologically linked cases e.g. families, excluded. Whilst there are acknowledged gaps in the cluster 'footprint' of these outbreak events; namely a variable follow up time period as well as an incomplete transmission picture, it is interesting to note that despite this, the cluster size (molecular footprint) is greater, in most cases, than the outbreak size (epidemiological footprint).

6.4.4 Comparison to existing literature

There are few studies looking at outbreak events on a large scale. The Centre for Disease Control in America has published two series of investigations. These demonstrated several patterns: late presentations and treatment delays propagating infections; and the prevalence of risk factors such as drug and alcohol abuse, incarceration and homelessness. [261, 292] Our findings differ in that the locations are predominately educational facilities with fewer risk factors identified. Case studies do highlight the risk posed for disease acquisition in congregate settings.

6.4.5 Conclusions

Currently guidelines in the UK do not routinely recommend screening casual contacts in a congregate setting [223]. Given our findings and the difficult in assigning responsibility to a household transmission alone, this guidance should be considered on a case by case basis as other countries have done [301] and social network analysis should be considered as a viable alternative. [54, 165, 186]. No comparative evidence exists for the stone-in-pond approach to alternative methods. [93]

Given the lack of close correlation between epidemiological and molecular cases, traditional epidemiological approaches may not be sufficient to quickly identify and curb the spread of latent TB disease and subsequent active disease development. WGS can be used to identify likely transmission events and points of public health intervention. As active TB disease is relatively slow to develop, revisiting contact tracing should be considered in cases where molecularly clustered isolates are subsequently identified. This may be key to stopping TB transmission in the UK. [34]

This chapter demonstrates for the first time the long-term impact of epidemiological contact tracing using the stone in pond model approach and the outcomes of on-going disease transmission as demonstrated by retrospective, linked whole genome sequencing data. Persistent and increasing cluster sizes clearly linked through low SNP (<5) distances demonstrate the ineffectiveness of contact tracing in congregate setting using traditional methods. The next chapter will look at the identifying risk factors through analysis for contacts in congregate exposures.

Chapter 7 – An explanatory model for identifying risk factors in the transmission of tuberculosis in congregate settings

As I have demonstrated, there is a paucity of comparative published data looking at risk factors in congregate settings. Therefore, this chapter will identify risk factors from analysis in order to inform strategies for the risk stratification of contacts.

7.1 Introduction

As we have seen in the preceding chapters, contact tracing remains the primary tool to identify and prevent further propagation of infection with Mtb. There are clearly issues with this approach as demonstrated in the previous chapter with some individuals not identified or not engaging with the contact tracing process going on to spread disease.

Identifying and assessing contacts can represent a significant burden on strained healthcare resources. This is a particular challenge in congregate settings where the contact tracing pool can be significantly larger than household settings. Not all identified contacts have the same risk of acquiring TB disease and therefore may not all require the same level of assessment and follow up. With increasing incubation/ latency periods such as with diseases like TB, reducing delays in contact tracing are not as important as tracing high risk contacts. [302] namely those most likely to go on to develop clinical disease.

The ability to identify individuals at the highest risk of acquiring disease can aid contact tracing services prioritise scarce resource while maximising the impact a contact investigation can have. In addition, for those individuals who do not engage with the process, if they are higher risk, further steps can be taken to engage them, again prioritising resources.

Incubation and latent periods also vary from person to person. The latter being of particular relevance in TB. Incubation periods referring to the time between an infectious exposure and symptom onset in the individual. Latency refers to the time period between infectious exposure and infectiousness. [303] Diseases with a prolonged latent state have the benefit of making contact tracing more effective. [304-306]

Contact tracing does not only have a single impact on disease transmission but can be considered a multifaceted approach. Not only are contacts traced, but secondary case transmission can be further reduced through making potential contacts aware of their status through educational interventions. [302]

Being able to identify these high-risk contacts is the basis of this chapter. The objective is to design a model in order to aid prioritisation of contacts screened as part of a TB exposure in a congregate setting incident. This model would enable services to provide probabilities/ risks dependent on different combinations of predictor values, the latter collected as part of routine screening assessments. This would represent a retrospective analysis in order to establish whether the development of disease is related to collected data in the initial screening phase, prior to the availability of lab data. In order to achieve this, the prediction model will need to assess the relevant variables that suggest a high-risk contact. I will achieve this by identifying the relevant variables and through backwards stepwise logistic regression demonstrate an explanatory model.

7.2 Methods

7.2.1 Modelling design

After a TB index case presents to the healthcare services, routine public health data is collected. If this occurs in a congregate setting, local TB services become involved and establish the need for further intervention based on how many people are potentially exposed.

7.2.1.1 Study population and sampling

Data for modelling have been collected contemporaneously over a 7-year period (2010-2017) this involved the investigation of real time incidents of which there have been 167. These generated 8091 contacts at 112 locations. The study participants for this model would be the contact cases (8091) who were potentially exposed and deemed at risk of developing infection and disease. Outcome data from the screening process is available for these individuals and includes disease detected (latent/ active disease), no disease detected or defaulting (did not engage with screening process/ lost to follow up). This has been extensively described in chapter 5.

The data sampled includes all incident events; the location, index case and contact case characteristics for all active index cases presenting to services. Where required, these data were augmented by interview data collected from index and screened contact cases as well as location data e.g. ages of students attending schools.

7.2.1.2 Outcomes

The primary outcome would be significant predictors for the detection of disease states in contacts i.e. latent or active disease.

7.2.2 Variable Selection

7.2.2.1 Predictors

Data collected reflects 4 domains; Location, contacts, index cases and disease and laboratory characteristics. Further recorded data includes treatment implementation and completion dates (table 21). TB has previously been described in relation to these predictive factors with increasing risk of disease acquisition correlating with factors described. As in chapter 3, there have been no large-scale studies investigating location factors in low incidence congregate settings to inform predictors and locations involved have been included instead. The variable selection and literature related to their justification is discussed in detail in appendix 7.

Data collection occurs over the period from presentation of a symptomatic index case through to laboratory investigations and treatment completion. In real world clinical practice, not all data collected are available at the same time. Demographic and area data can be collected contemporaneously once TB services are alerted to a new index case and incident event and at risk contacts are identified. The latter process involves TB nurses interviewing index cases and assessing locations in order to identify individuals potentially exposed (contacts). Further data such as outcomes, clinical investigations relating to the

extent of the index cases disease severity as well as the microbiological diagnostics can only occur with the evolution of the clinical scenario. Whilst initial clinical investigations such as the extent of disease, organs involved and whether the index case has smear positive disease can be established early on, further data such as culture positive disease, resistance patterns and histology can take several weeks. Contact screening processes therefore rely on limited information at the initial presentation.

Establishing the variables to be analysed must therefore make use of the data available to TB services at the initial screening event. The reason behind the time limitation when conducting the screening event is the increasing difficulty in contact tracing over time as well as the risk that infected contacts may go on to propagate infection chains whilst investigations are pending. Predictor selection will therefore be limited to data that could reliably be available at the early stages of contact tracing. This would remove the disease category (table 13) from predictor selection. Further justification for variable selection is included in the appendices.

Table 13 - Variables included in the analysis

Location	Contacts	Index
Category	Current age	Patient gender
	Gender	Current age
	Ethnic group	Ethnicity
	Year of entry	Country birth
	Country of birth	Year of entry
	Contact index close contact	Previous TB
	Contact index relationship to contact	Previous TB date known
		Previous BCG
	Outcome	Date of BCG
		HIV status
		History of drug use
		Homelessness
		Incarceration history
		Alcohol abuse
		Drug abuse
		Smear positive

7.2.2.2 Missing values

Not all variables have complete observation sets with some variables having missing data (table 22). This could be a result of the data collection process or may reflect data entry issues when recorded.

Several methods have been suggested for accounting for missing data when considering modelling. Imputation methods are the most commonly used and effective. These methods are not universally applicable and depend on sample size and percentage of missing values. With larger sample sizes >50, imputation methods can still result in strong correlations with up to 5% missing values. Correlation with modelling falls significantly when missing values are greater than 10%. [307] Therefore any variables not deemed crucial to the modelling process with >10% missing values have been excluded from the analytic process.

7.2.3 Analysis

For the analysis, the dependent variable will be the outcome i.e. the presence or absence of disease detected. For the purposes of contact tracing, the relevant findings and those required to prevent on going infection chains would include any evidence of disease. Therefore, both latent and active disease have been recorded together. Two further outcomes are possible; these include no disease status and a defaulter status. The latter does not represent missing data, rather it acknowledges the large proportion of contacts who do not engage with the contact tracing process and may represent individuals with disease and risk factors for on-going infection propagation.

7.2.3.1 Model

Analyses were carried out in Stata v 16.0. Multivariate logistic regression was used to develop a prediction model for the development of TB disease (latent or active) in contacts of an index case. Variables informing this model were derived from backward stepwise regression using an elimination alpha-to-remove p-value of 0.25. A received operator curve (ROC) was plotted and goodness of fit was assessed using the Hosmer-Lemeshow chi^2 test (where non-significance indicates a good fit). Statistically significant values were considered with p-values <0.05.

7.3 Results

Using backwards stepwise logistic regression with an alpha-to-remove threshold of 0.25, HIV and drug abuse status were eliminated due to lack of observations with positive results. Contact gender and ethnicity, index ethnicity and relationship status were the only variables that met the pre-specified threshold of 0.25. Location categories were not included in the analysis and were separately analysed given the heterogeneity and likely differing risk factors.

The remaining factors can be ranked according to their fully standardised coefficients with relationship status representing the highest-ranking variable followed by contact gender and ethnicity.

The remaining variables were all included in the model. The results of the backwards stepwise logistic regression as well as the associated robust errors are presented in table 14.

The model generated is seen in the logistic regression equation below. Ln represents the natural logarithm, p is probability of TB disease presence (both latent and active disease have been grouped together), α is the intercept and $\beta_1 - \beta_4$ are the logistic coefficients.

$$\ln\left(\frac{p}{1-p}\right) = \alpha + (\beta_1 \times \text{contact case gender}) + (\beta_2 \times \text{contact case ethnicity}) + (\beta_3 \times \text{index case ethnicity}) + (\beta_4 \times \text{relationship status})$$

4484 observations were included with a probability>chi2 significant at the 0.05 threshold suggesting the null hypothesis can be rejected. A receiver operating characteristic (ROC) curve was generated (figure 9). The R^2 value was 0.0505 with subsequent goodness of fit

testing with the Hosmer-Lemeshow Chi^2 test generated a result of 14.13 which was significant ($p<0.0262$) suggesting a poor model fit. This suggests that the predictive model generated cannot therefore be assumed to be useful for anticipating disease detection in contacts with the variables included.

Table 14 - Variables with an alpha-to-remove threshold of <0.25 - prediction factors based on 4484 observations

Variables	β-Coefficient	Robust standard error	95% Confidence Interval		P value
Contact case gender	-0.39	0.11	-0.61	-0.17	0.001
Contact case ethnicity	-0.34	0.06	-0.46	-0.22	0
Relationship Status	-0.57	0.07	-0.70	-0.43	0
Index case ethnicity	-0.11	0.09	-0.28	0.06	0.194
Hosmer-Lemeshow Chi^2 – 14.13 (Prob > chi^2 - 0.0262), R^2 value – 0.0505					

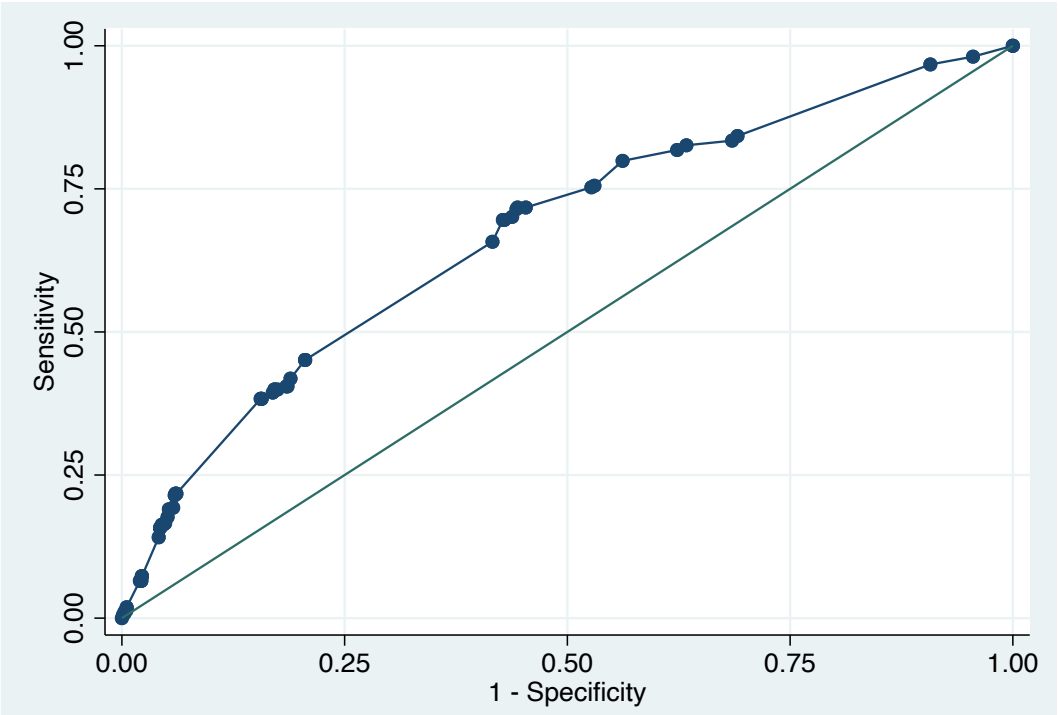

Figure 9 - Receiver operating characteristic (ROC) curve of ability of derived variables in model to predict TB disease based on 6261 observations

7.3.1 Sub analyses

Sub analysis looked at the different location categories assuming they behaved as heterogeneous locations with differing risks and therefore unique variable combinations informing distinct models. Excluding accommodation settings (as these were assumed to be similar to household locations), all 4 remaining categories generated different variable combinations. Healthcare and educational locations produced a model which was significant for the goodness of fit test suggesting poor model fit. However, the remaining categories (recreational and work settings) produced distinct models with insignificance with the Hosmer-Lemeshow test suggesting good model fits and therefore predictive models. All models varied in terms of variables found to have satisfied the alpha-to-remove ratio of 0.25, however, they all contained the relationship status variable. Tables of modelling outcomes and ROC curves are included below.

Given the persistent inclusion of the relationship status variable, a further sub analysis was carried out, constructing a model using close contact and relationship status (table 15), both metrics for the social interactions obtained in the interview process, logistic regression demonstrated a significant chi^2 result of <0.005 refuting the null hypothesis. Additionally, goodness of fit testing using the Hosmer-Lemeshow test was not significant at the 0.05 threshold (0.6578), suggesting a good model fit. Close contact status and relationship status are therefore better predictors alone of outcome status than a combined model with additional variables (as above).

All models were internally validated using the bootstrap technique (300 samples) which confirmed the initial model findings.

Table 15 - Relationship and close contact status modelling outcomes based on 5989 observations

Variables	β-Coefficient	Robust Standard Error	95% Confidence Interval		P value
Relationship Status	-0.46	0.11	-0.67	-0.24	0.000
Close contact status	-0.62	0.07	-0.75	-0.48	0.000
Hosmer-Lemeshow Chi^2 - 0.20 (Prob > chi^2 - 0.6578), R^2 **value – 0.038**					

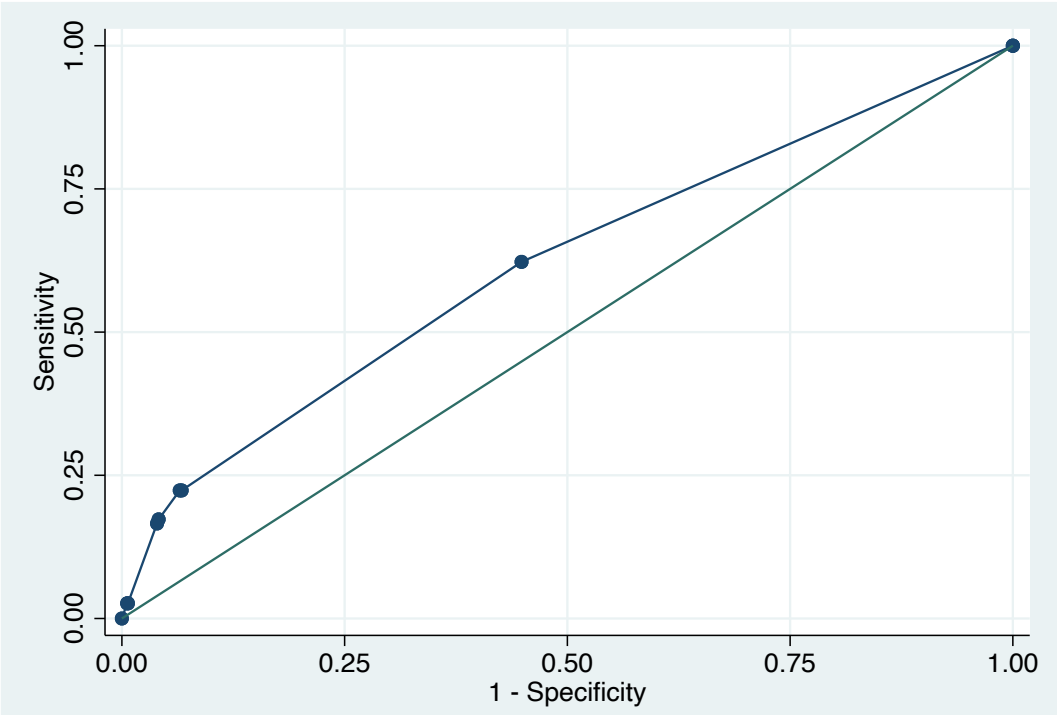

Figure 10 - Receiver operating characteristic (ROC) curve of ability of relationship and close contact status to predict TB disease based on 5989 observations

The coefficients confirm that the closer the relationship status and contact status (i.e. Household/ close contact/ 1st degree relative) the more likely the risk of infection. This correlates with further findings.

7.3.1.1 Educational category

Table 16 - Educational category prediction model based on 5187 observations

Variables	β-Coefficient	Robust Standard Error	95% Confidence Interval		P Value
Contact age	0.01	0.00	0.00	0.02	0.087
Contact gender	-0.21	0.13	-0.45	0.04	0.106
Contact ethnicity	-0.43	0.07	-0.57	-0.30	0.000
Close contact status	-0.22	0.14	-0.49	0.06	0.122
Relationship status	-0.70	0.10	-0.90	-0.51	0.000
Index gender	0.30	0.17	-0.03	0.62	0.075
Index age	0.01	0.01	-0.01	0.03	0.244
Smear positive status	0.25	0.19	-0.13	0.63	0.198
Hosmer-Lemeshow chi^2 – 24.82 (Prob > chi^2 - 0.0017), R^2 **value – 0.0614**					

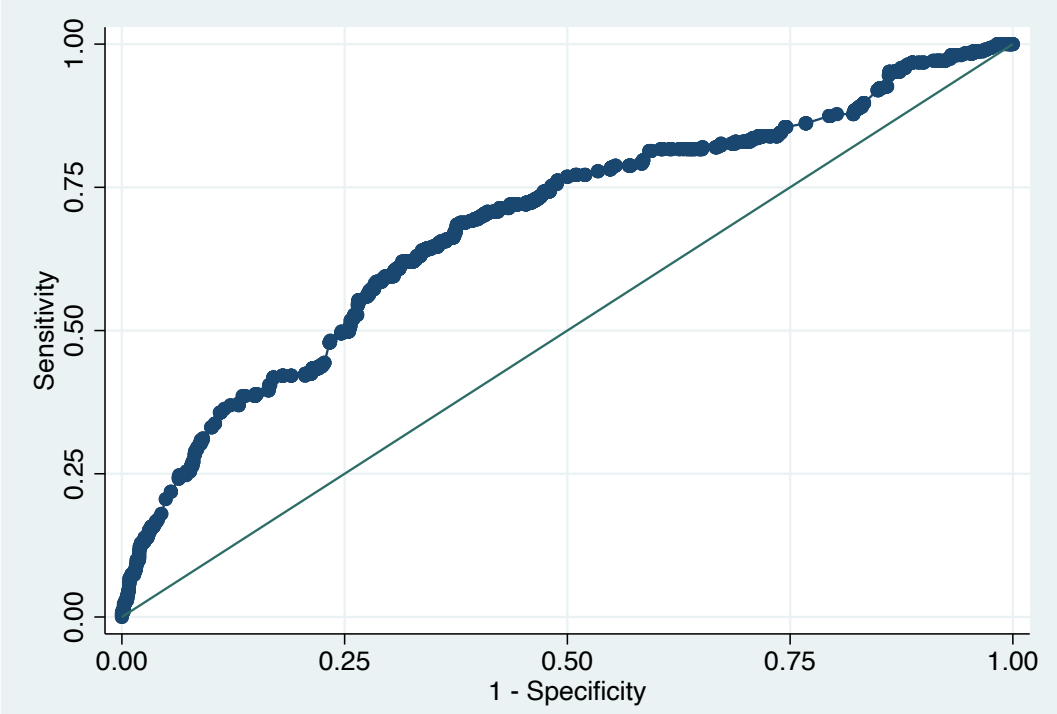

Figure 11 - Receiver operating characteristic (ROC) curve of ability of derived variables from educational settings to predict TB disease based on 5187 observations

The coefficients demonstrate that: younger age in both contacts and index cases; male gender in contacts and female gender in index cases; Asian ethnicity, close contact and

closer relationship status are all factors associated with increasing risk of disease. As is smear positivity.

7.3.1.2 Healthcare category

Table 17 - Healthcare category prediction model based on 310 observations

Variables	β-Coefficient	Robust Standard Error	95% Confidence Interval		P Value
Close contact status	-1.06	0.68	-2.39	0.27	0.119
Relationship status	-0.41	0.20	-0.79	-0.03	0.037
Hosmer-Lemeshow chi^2 – 12.41 (Prob > chi^2 - 0.0146), R^2 **value – 0.0696**					

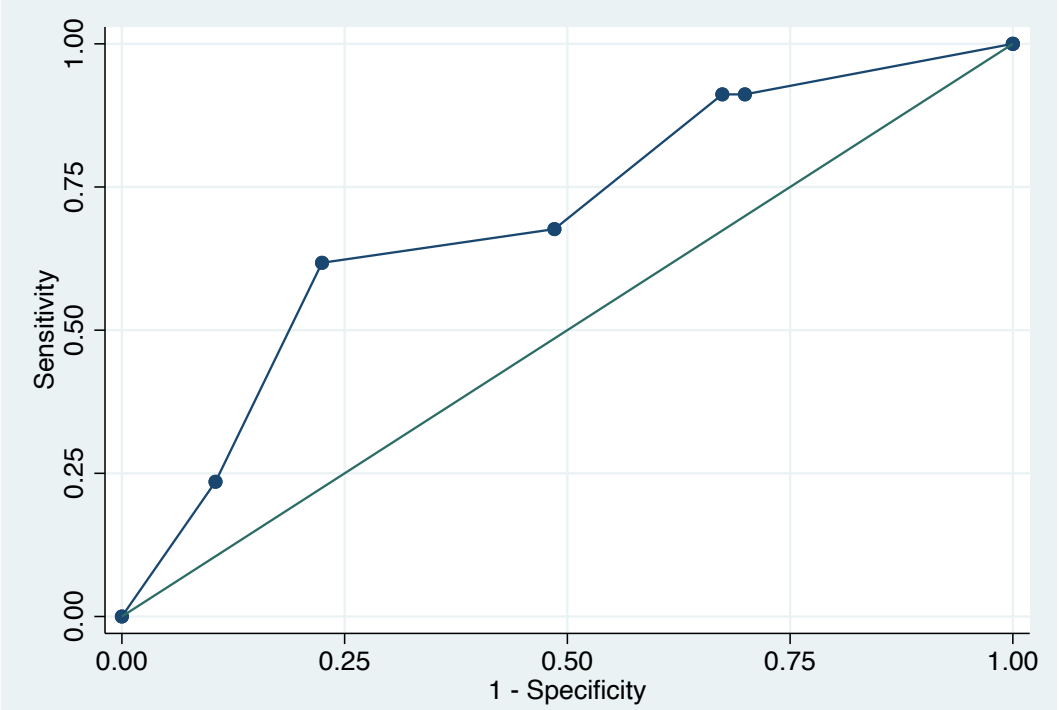

Figure 12 - Receiver operating characteristic (ROC) curve of ability of derived variables from healthcare setting to predict TB disease based on 310 observations

We can see from the coefficients that close contact status (1st degree associations) and close contact with the index case which in this case would represent both healthcare workers and contacts within close contact (nearby patients, staff and visitors) are at increased risk of disease acquisition.

7.3.1.3 Recreational category

Table 18 - Recreational category prediction model based on 333 observations

Variables	β-Coefficient	Robust Standard Error	95% Confidence Interval		P Value
Contact age	-0.02	0.01	-0.04	0.01	0.136
Contact gender	-0.46	0.38	-1.22	0.29	0.225
Contact ethnicity	-0.64	0.18	-1.00	-0.29	0.000
Close contact status	0.69	0.38	-0.04	1.43	0.066
Relationship status	1.18	0.50	0.20	2.16	0.018
Smear positive status	-1.07	0.39	-1.83	-0.31	0.006
Hosmer-Lemeshow chi^2 – 1.93 (Prob > chi^2 - 0.983), R^2 **value – 0.1531**					

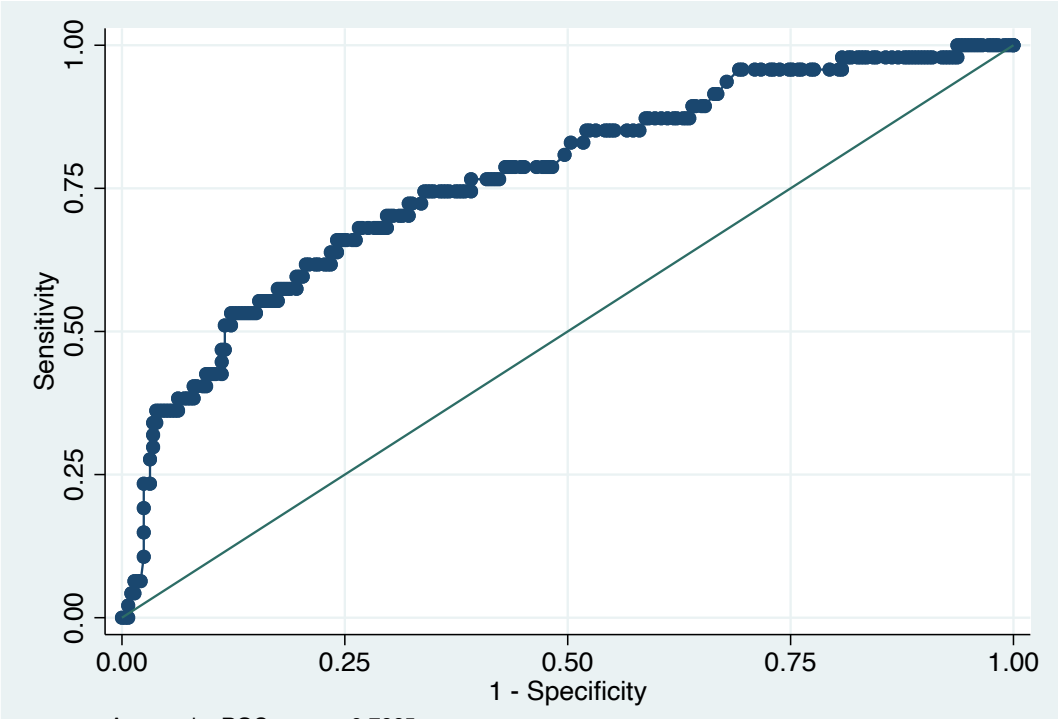

Figure 13 - Receiver operating characteristic (ROC) curve of ability of derived variables from recreational settings to predict TB disease based on 333 observations

Coefficients from this model suggest that younger age in contacts, male gender, and Asian ethnicity are risk factors for the acquisition of disease. In addition, more distant contact status and a close/ casual as opposed to household relationship status are predictive of disease acquisition in a recreational environment. Interestingly, smear negative status is a

risk factor for disease somewhat counterintuitively. These findings are at odds with the previous close relationship status for disease acquisition.

7.3.1.4 Work category

Table 19 - Work category prediction model based on 334 observations

Variables	β-Coefficient	Robust Standard Error	95% Confidence Interval		P Value
Contact age	-0.04	0.02	-0.07	-0.01	0.008
Contact ethnicity	-1.11	0.32	-1.73	-0.49	0.000
Close contact status	1.73	0.65	0.47	3.00	0.007
Relationship status	-1.45	0.33	-2.10	-0.80	0.000
Index gender	-0.95	0.60	-2.13	0.23	0.116
Index drug history	-1.97	1.11	-4.14	0.20	0.075
Hosmer-Lemeshow chi^2 – 6.18 (Prob > chi^2 - 0.6267), R^2 **value – 0.3036**					

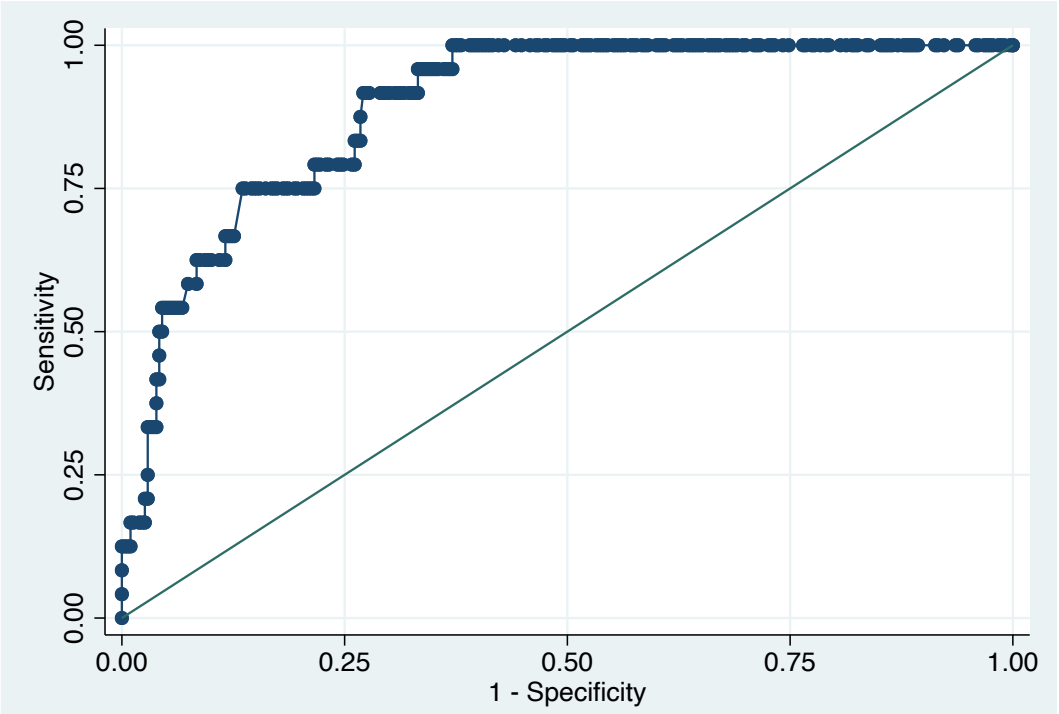

Figure 14 - Receiver operating characteristic (ROC) curve of ability of derived variables from work settings to predict TB disease based on 334 observations

Younger age, Asian ethnicity, and male gender for index cases are predictive of disease as per the coefficients. Contacts that are casual but have a close relationship with the index case are at increased risk of disease. Having no prior drug history is also suggestive of disease presence. Given that this is a working environment the latter findings would correlate.

7.4 Discussion

7.4.1 Summary of results

Backwards stepwise regression produced significant variables coefficients from the data; category of location, contact case gender, contact case ethnicity, close contact status, relationship status, index case ethnicity, homeless status, drug history, alcohol history and prison history. These variables subsequently informed the model derivation, however on goodness of fit testing the model was not found to be a good fit for the data as tested by the Hosmer-Lemeshow chi^2 test and on validation with the bootstrapping method. Given the risk of heterogeneous populations within the data, sub analyses were conducted. Given the heterogeneous nature of locations, a sub analysis was carried out with regressions on individuals in each categorical location. This was due to the similarities of individuals in these categories and the shared risk for disease acquisition. Analysis of categories demonstrated firstly the differing variables informing risk and disease acquisition and secondly the predictive model goodness of fit for the data. In addition, whilst the variable combination for each distinct category were unique – relationship status was a ubiquitous variable suggesting the significance of the interactions between individuals regardless of the location they were located in. Using relationship and close contact status (a similar and related metric) in a global model as the only two variables again produced a model that demonstrated a non-significant goodness of fit test suggesting a good model fit.

Interestingly, historically relevant predictors for infectiousness would be deemed to include HIV status (a positive status conferring higher risk of disease acquisition) and smear positive status of the index case, which has been linked to a higher transmission risk. This did not appear to be the case and may reflect the limited positive values for both this factors in the data.

7.4.2 Existing literature

Clinical prediction models to risk stratify TB cases have been examined previously. Though, importantly there is a lack of prediction modelling using clinically available metrics allowing risk stratification of contacts.

Nanta et al describe a risk prediction score for TB in an HIV infected population group. This was based on 257 HIV infected patients where 66 developed active TB and is not broadly applicable to a non-HIV population given the different levels of risk and exposures. Balcha et al outline the use of a second step clinically based algorithm to further define risk of developing active TB again limited to an HIV infected population in Ethiopia. This study did not take into account extrapulmonary TB disease and was conducted in a high prevalence area (17%) and may therefore not be applicable in a lower prevalence region. In addition, there is no external validation of the scoring system and no assessment of intra-observer variability when recording score relevant variables. [308, 309]

The use of the Kenneth Jones scoring chart is described by Mehnaz et al to determine its applicability in early detection of TB in children. An adult contact with active TB, exaggerated BCG vaccine response and correlating clinical and radiological signs were all important suggestive markers used in the chart. This case control study incorporated 50 children and cases and describes higher scoring figures (5-7) in the case arm only. This study does not take into account different locations of exposure and children less than 5 are at higher risk as are those in household settings. This would not be applicable to congregate transmission areas. [310]

A further clinical prediction model is described by Wisnivesky et al to aid in the decision to isolate cases of suspected pulmonary TB in inpatients. The authors utilised contemporaneously collected data and with univariate logistic regression on a cohort of 50 patients derived a risk prediction tool including the variables: Chronic symptoms (weight loss, malaise, weakness and night sweats), positive PPD, temperature, and upper lobe consolidation. This limits presentation to respiratory illnesses with a high probability of active disease and effectively excludes latent cases. This would not be helpful to identify at risk contacts. [311]

Scoring systems that take into account congregate settings have also been investigated. Using prospective questionnaires to derive a contact scoring to compare LTBI detection by skin test and Elispot, Shams et al derived a contact score based on a sample size of 416 contacts of 72 cases (culture positive pulmonary TB). They identified three key variables: relationship to the patient, infectivity of the patient, and extent of exposure to the patient,

taking into account location settings. Importantly, the weighting of scoring system was based on clinical intuition not statistical analyses with exception of smear positivity where weighting linked to proportional risk. [312]

Public health data has also been used to investigate prediction and risk. Using contemporaneously collected public health data, Chan et al described the development and validation of a risk prediction score for contact of patients exposed to tuberculosis in children under 12. 2 cohorts were used, a derivation cohort 2008-2009 (9411 contacts of 4511 index cases) and a further validation cohort 2005 (2405 contacts of 1130 index cases). Contacts were defined as exposure of >8hrs within 1 day or >40hrs cumulative exposure. Conducted in Taiwan, an area of intermediate/medium burden of TB prevalence (70 per 100,000), though some townships had rates closer to 277/100,000. They used a Cox proportional hazards model to determine risk factors leading to the development of TB. This was then divided by the gender of the index case. Cut offs for scoring and their derivation were not explained. In addition, they did not include clinical features of the children under investigation and relied on skin tests, not IGRA's, the former having important limitations and false positive results. IGRA being more specific than skin tests for populations with wider BCG coverage. [313]

7.3.3 Strengths and limitations

7.3.3.1 Strengths

This study benefited from a much larger data set, compared with previous prediction modelling and derived scoring systems identified in the literature. In addition, the data was informed from contemporaneously collected public health data. Given the mandatory nature of the data required, this would be applicable to other areas of the nation as well as other countries who collect similar data. In addition, the study period was over several years and represented a long period of investigation and follow up allowing for more presentations with latent and active disease. [109] In order to limit the introduction of bias, missing values were minimised and significant data lines missing values were excluded with no imputation of missing points.

The variables themselves were based on frequently demonstrated significant findings and represent a robust set of variables related to TB transmission.

7.3.3.2 Limitations

While this was a large dataset that utilised public health methodology in use nationally, this did represent a single region. Given the population mix, that is a larger migrant population that is more ethnically diverse than significant parts of the UK and with a higher TB burden this data may not be applicable to other regions.

A further limitation was the low number of observed positive outcomes (namely disease detection), given this low value, it would be difficult to make definitive statements on the prediction model generated.

The methodology aimed to limit the level of bias inherent in the modelling process. However, given the nature of deriving variables and coding data, bias may have been introduced. Additionally, there was no independent validation of the model, instead an internal bootstrap method was used. Given the modest to insignificant area under the curves (AUC's), these would not support a predictive model.

A number of variables were excluded during the analytical process, described above. These variables have been previously described as playing a role in TB acquisition risk and transmission risk and as such, could impact on modelling outcomes and subsequent results.

Previous TB and BCG status are key examples, with the latter having been shown to lower the risk of acquisition and the former implying risk of exposure and reactivation. These factors were inconsistently recorded and therefore it is difficult to include them in the final analysis.

An individual's occupation might serve as a useful metric of who they may have exposed and been exposed to. This variable was excluded for reasons described in the analysis section. In terms of a useful predictive tool, occupation may have a role as a proxy for location and exposure.

Subcategories were derived based on information on the location available. Educational establishments in particular are difficult to categorise without running the risk of multiple, nuanced categories and losing the ability to distinguish clearly between any relevant results.

This is largely because of the age ranges of students attending the schools and combined age groups attending a single establishment. An example would be primary and secondary aged students or primary, secondary and college age students in the same establishment. Given the differing risks inherent with increasing ages of children and students, these multi-category locations can impact on meaningful results.

Categorising variables may also have disguised relevant results. Location of birth was categorised as in the UK or not in the UK. This largely removes the risk associated with birth in a high-risk country or exposure to individuals from that risk country in familial and cultural social networks. This is offset to some degree by including an ethnicity grouping for those born in the UK.

In terms of coding, certain variables presented challenges that may have resulted in bias. Coding ethnicity resulted in grouping certain at risk otherwise further delineated populations e.g. Afrocaribbeans combines black African, black Caribbean, black other etc. The original data did not specify where this derivation came from and in particular the public health data had codes that were difficult to interpret e.g. – oriental (coded in Asian other). Given the variable risk of TB disease depending on country of origin, the coding and recoding of ethnic groups could conceal different risks.

Coded in terms of infection, where latent and active disease were combined to demonstrate significant findings on contact tracing i.e. those that would require drug therapy/ intervention. One outcome noted were defaulters, a group who disengaged with therapy

and whose outcome is unknown. Disengaged groups and marginalised populations are the least likely to engage with public health interventions and contact tracing in its current form and therefore represent a high-risk group. The lack of outcome data therefore could potentially result in diluting the outcome data.

Relationship descriptions from the public health data may have introduced bias. It is not clear what underpinned these arbitrary labels, especially as there is no clear guidance on how to apply the labels. Furthermore, there may be significant inter-operator variability. In particular, it is difficult to use the same labels for different cultures and different social groups e.g. Asian families may have multigenerational households and should all be coded as household contacts as opposed to non-nuclear families and separated families with step-parents. Also in schooling, it is difficult to appreciate the degree to which interaction takes place. Work colleagues and school colleagues may have closer interactions depending on the environment in which they interact.

Finally, assumptions from linear regression analysis were made, namely that all the relationships from the data are linear which may not be the case and that there were no other interacting variables.

7.3.4 Future work

A key challenge to model design is ensuring sufficient reflective data to make adequate disease characteristic assumptions. Designing a model with as many parameters as possible

would be ideal. Key to this are the stages involved in epidemic/ outbreak spread; the disease natural history and transmission characteristics, the practical interventions undertaken and control stages. Mathematical models examining viral outbreak transmission events make the assumption of homogeneity in social contact transmission. [303]These assumptions may not be realistic as previous studies have demonstrated both a significant heterogeneity in social contact interactions [314, 315]and more so when looking at age-specific social mixing. [315]

Social network analysis would therefore highlight the nuances in interactions that could result in transmission and has previously demonstrated additional contacts over traditional tracing methods. [54, 186] The positive correlation with relationship status being present in all fitted models is supportive of this notion.

As the basis of the thesis is that the current contact tracing approach misses targets and the data was collected using this supposed flawed approach, the data collected has to be assumed to be inherently flawed. In so much as the relationships we seek to identify are potentially the ones that are not included in the data as they are missed. This may result in better model fitting with a more nuanced description of relationship status.

Significant spread of disease only really occurs when R>1. [316] Cluster sizes are related to reproductive numbers (C = 1/1-R). There are several inherent biases of making assumptions on this basis. Larger clusters are more likely to be recognised and missing transmission events can skew cluster size estimates. Super-spreader events can also generate significant heterogeneity in transmission dynamics. [316]

Utilising prospective questioning of patients with regards to social network questions at the time of contact tracing assessment could add this additional resolution in relationships as well as reduce defaulting populations and underrepresented groups. [54]

Chapter 8 - Conclusions

8.1 Introduction

In this chapter, I seek to answer the aims laid out at the beginning of the thesis. In order to do this, I will summarise the key findings of the above chapters and how they answer the aims specifically as well as where the results fall short of what was expected and the novel information gained from the process.

Following this, I will review the main limitations that have arisen in the results chapters as well as the limitations inherent in the data, that have been alluded to in the above chapters. I will touch on the generalisability of the data and implications for broader applicability.

Finally, I will cover the further work needed to fully realise the implications of the results and on-going research questions as well as the current work in other, related fields that would have an impact on the findings. In the shadow of the COVID pandemic, there is an additional post-script describing the implications this infectious disease has had on contact tracing and implications for disease identification and prevention moving forward.

8.2 Aim 1 – Identify how we arrived at our current contact tracing approach and whether there any alternatives.

TB used to be highly prevalent in the UK. Multiple social factors and preventative measures resulted in sharply decreasing trends. The advent of new treatments allowed for effective therapy and therefore supported early diagnosis. In order to facilitate this, screening of large blocks of the population helped target treatment (e.g. school groups), over time these yielded fewer positive results and with declining utility, this approach was largely phased out. In addition, vaccination of children reduced population level rates as did treating animal hosts and infected milk

The 1980's saw a resurgence in TB rates, this has been described in the context of the emerging HIV epidemic and increasing migration from high incidence countries. The historic epidemiology of TB in the UK therefore changed from a high incidence in the native population to TB retreating to at risk demographic pockets. The approach to screening TB cases and in particular contacts has not adapted to this changing demographic.

Limiting current epidemiological approaches to traditional/ nuclear relationships does not recognise the irregular/ opportunistic transmission event that can and do occur as a result of changing norms and societal practices/ interaction. In addition, whilst societal norms and interactions have changed over the decades, contact tracing approaches have not, being based on initial observations in the forerunner of chest clinics. Furthermore, the lower prevalence of TB disease was not uniform and this is reflected in the changing demographics

of affected individuals, particularly the underserved population groups. The nature of these new demographic pockets further highlighted the difficulty in traditional contact tracing methods.

This latter point can be seen again in the summary data (chapter 5) – where we see the large number of defaulting individuals (those lost to follow up or who did not engage with the screening process). This lost group of individuals clearly has implications for long term TB control and it is from these initial screening episodes that we can see the outbreak events that occur as a result of not identifying or un-screened infected individuals (chapter 6) with the resulting outcome of on-going, linked chains of transmission demonstrated by genomic analysis.

The comparative evidence base behind the stone in pond approach is lacking. A conclusion demonstrated through a Cochrane review process (chapter 4). Several authors have suggested alternative contact tracing approaches which take into account the complexity of social interactions that are a hallmark of modern society. Looking at the evidence base between prior and proposed contact tracing methods has demonstrated a paucity of evidence in terms of controlled trials in support of either approach, but certainly none in defence of existing methods. Given the clear lack of evidence examining the effectiveness of current contact tracing methods, there is a need for alternative approaches to be explored and compared with current methods.

8.3 Aim 2 - Explore whether the current contact tracing approach is effective in congregate settings.

Risk factors for both the acquisition and spread of TB disease have been described in detail previously. Whilst these factors look at an individual risk, they focus on narrow interactions and are generally limited to household settings or a combination of household and congregate settings. Understanding risk factors in congregate settings is key to developing a coherent strategy to prevent spread of infection in this key social space. In addition, by understanding both transmission and acquisition risk factors, contact tracing services can prioritise contacts that have been exposed. This would allow more appropriate use of limited resources and help to target at risk individuals who fail to engage with therapy after initial screening. Congregate setting specific risk factors can also help to inform public space design in order to better protect service users/ customers.

Chapters 3 and 4 demonstrate that there is limited evidence examining risk factors particular to congregate settings. Given the likelihood and public health implications of congregate setting outbreaks, more data on risk factors specific to these settings is needed.

In terms of locations for incident events, these occurred at the highest frequency in educational establishments (schools) having implications for public health services. In addition, the type of educational establishment varied – given the changing nature of childhood exposure to TB, this has implications for social interactions resulting in infection and raises the possibility of non-household acquisition of disease.

Persistent disease in demographic pockets is worthy of examination: Homeless, indigent, prison populations, migrant groups. All of these groups have poor uptake of contact tracing and compliance with treatment and would be less likely to report contacts. This nullifies the impact of conventional contact tracing approaches and larger outbreaks are more likely and less detectable at initial screening episodes.

Chapter 6 builds on this theme and examines the long-term outcomes associated with conventional contact tracing processes. In this chapter I clearly demonstrate the impact of unrecognised and treated individuals and with whole genome sequencing these cases can be linked to original presentations and result in large outbreak clusters over time.

Contact tracing approaches aim to identify all contacts of an exposure in order to screen and treat as appropriate and prevent further propagation of infection. Measures of success are often limited given the lack of real time diagnostics linking latent cases of disease together. Being able to link cases is only possible with active disease cases, where genomic techniques (whole genome sequencing) can identify single nuclear polymorphisms (SNP's) which can if <5 be highly suggestive of a link. Retrospective studies can therefore provide an idea of how 'successful' contact tracing approaches are. By comparing epidemiological investigations (numbers of contacts identified) with subsequent linked cases (WGS), success can be measured. The greater the discrepancy, the higher the likelihood of missed epidemiological cases and ongoing infection chains.

Given the large discrepancy noted in our group of outbreak cases, this points to a lack of coverage with epidemiological screening with contacts not being identified and treated at the early latent stage. Whilst there is clearly a contribution of identified contacts that do not engage with screening and treatment (defaulters). There remains the issue of not identifying and prioritising those at the highest risk of disease acquisition and on-going transmission of disease. Given the significant and persistent difference in epidemiological and genomic interventions in nearly all outbreak scenarios examined, a more effective contact tracing approach is required.

Conventional contact tracing approaches do not examine social interactions and relationships in these groups and do not take into account the atypical nature of interactions as well as the clandestine aspect that would prevent compliance with contact tracing officers. Once this has been established through social networks, the resolution of the outbreak becomes clear. Other authors have demonstrated alternative approaches that have a higher rate of unidentified contact detection such as Gardy et al – social network analysis. [54] Current contact tracing approaches do not allow for the resolution in exploring social interactions, an aspect that underpins TB transmission and given the long-term outcomes of missed contacts, current approaches do not eliminate TB transmission in congregate settings.

8.4 Aim 3 – Assess whether we can be predictive about risk of TB acquisition to contacts in congregate settings

Improving contact tracing approaches to identify more at risk/ exposed individuals is only one part of a more effective contact tracing process. The second aspect for TB services is to generate a more efficient approach to minimise those individuals who do not engage with follow up and tracing as well as reduce the number of screened individuals who have negative results. Finding patients that are at risk of TB disease is key.

Proximity and duration of exposure are recognised as significant predictive factors as demonstrated in earlier chapters. A metric for these factors is the relationship status, despite the caveats inherent in this label.

Issues with being predictive and assigning risk to contacts of index cases are that latent cases are difficult to attribute to active cases (no genomic possibility prior to active disease development. High rates of defaulters could sway the issue of links and bias the data. In addition, it is difficult to establish super spreader events based on type of transmission and latency period i.e. not everyone exposed/ contacted goes on to develop infection, not all who develop infection (latent) progress on to active disease and until active disease develops unable to effectively link to index case using genomic techniques

Within these limitations, chapter 7 examines the possibility of a predictive model with the variables available. The global model, excluding household contacts and therefore focusing

on congregate settings was not a good fit. Despite this, sub analyses demonstrated interesting findings, mainly that a model with a good fit was possible looking at relationship status and closeness of relationship. In addition, looking at the individual settings (congregate settings), bespoke modelling demonstrated good fits identifying different combinations of variables and therefore risk factors. In all these models, relationship status was relevant and included in the model.

While a homogenous model for all congregate settings did not appear to work, individual models based on location type do demonstrate good model fits. Inherently this demonstrates that predictive statements about risk are possible. In addition, it is clear that congregate settings do not represent a homogenous group but rather a heterogeneous group of settings linked by virtue of being non-household settings e.g. educational settings have different infrastructure and functions as well as social interactions and at risk populations when compared with occupational environments or recreational settings such as shops.

Refining a predictive model is likely to need further setting based data to augment contact tracing approaches such as areas of congregation and ventilation status – risk factors that were highlighted in chapter 3 in terms of congregate settings inputs. This data is not routinely collected as part of current contact tracing approaches. Being predictive allows the TB service to target resources efficiently to reduce ongoing disease propagation.

8.5 Limitations

8.5.1 Generalisability of data

The data demonstrates the prevalence of TB in our population. This compares well with annually published public health data. Of note is the higher proportion of migrants in our population group and the lower proportion of at risk groups e.g. incarcerated, HIV, drug taking behaviour, alcohol abuse and homelessness. As such not all of the observations are necessarily applicable to other areas of the UK. Certain incidence areas in the UK fulfil WHO criteria for TB elimination <0.1/100,000 new cases. The lower rate of HIV in the population is also worth noting as this is at odds with international observations. HIV being a significant risk factor for TB globally. The majority of the index cases are migrants. Given the new entry screening programme, this demonstrates re-activation of TB.

8.5.2 Flaws in data

The inclusive model demonstrated in chapter 7 was not a good fit. The reasons for this have been examined in the chapter but further to this, is the issue of the data itself. The collection of data/ database may be inherently biased. Given that the premise of this thesis is that the conventional contact tracing approach misses opportunities and cases in tracing (demonstrated by chapter 6) the data collected is more reflective of the deficient conventional contact tracing approach and as such an appropriate model fit would not be expected and would be highly derivative therefore not an accurate reflection of real world contact transmission events.

A further issue with the data and drawing conclusions is the high rate of contacts lost to follow up, so called defaulters. Given that high risk individuals tend to belong to the underserved population groups and that these groups tend to be lost to follow up more frequently, it follows that the defaulter group likely has undetected disease states (demonstrated by chapter 6) and that this would bias modelling data as well as weakening conclusions drawn from outcomes. That said, acknowledging that this is real world data strengthens the argument for a different approach to contact tracing and by targeting the highest risk contacts (chapter 7) we can increase the yield of contact tracing investigations whilst limiting the contact tracing pool without missing infected cases.

8.6 Further work

Moving forward, TB services are increasingly incorporating whole genome sequencing which is soon to become a national standard. Coupled with a more sensitive approach to contact tracing would allow better mapping of the state of TB disease and transmission in the UK and bring us closer to the goal of TB elimination.

In order to inform this ideal, more robust data collection is needed and ideally prospectively collected data with a view to completeness and avoidance of missing fields. Augmenting public health mandatory data with the use of social network questionnaires would allow relationship and close contact status to be better understood and decrease the subjective nature of current labelling. By objectively describing relationship status, social networks could be constructed and analysed, better informing models given the clear signal seen in chapter 7.

In addition to being able to objectively describe areas of social aggregation and density of social networks, there is a need to establish geographic location data, the latter might prove difficult given the potentially clandestine nature of certain populations in underserved groups. This is broadly being done by TB services in several regions making comparison with traditional approaches more difficult. The latter point is still relevant as demonstrated by the recent Cochrane review (chapter 4). Given the previously demonstrated success with social network questionnaires and the ability to better describe underserved population transmission networks, this could help address the high rates of defaulters.

8.7 Future considerations

8.7.1 Alternative ways to establish social networks

In addition to social network questionnaires, utilising currently active social network platforms with existing metrics to describe links could also have utility, such established platforms include analysing phone contacts, the online media platform, Facebook, new emerging and established platforms such as Instagram etc. Often these platforms and associated software not only link contacts know to the individual but also geotag location and time data, crucial in contact tracing investigations. Modern phones and wearable technology such as smart watches have geolocation software allowing the same ability to map locations travelled by an index case. As the cost and ubiquity of these technologies improve, with individuals informed consent, potential areas of transmission can be better mapped and combined with risk analysis of contacts through improved modelling, contacts can be prioritised and informed and treated appropriately and in a timely manner.

8.7.2 Alternative ways to further derive and describe risk

TB transmission risk does not simply rest with individuals, the location at which transmission events occur can play a role. Congregate settings possess different spaces of congregation and have properties influencing circulation of expectorated pathogen, namely ventilation.

In order to inform risk assessments, an assessment of the ventilation properties of locations could be undertaken by investigatory teams. Whilst historically this has been seen to be a subjective assessment (chapter 2), a more objective approach can be considered. An

established metric for airflow assessments in public locations is CO^2 levels. Monitors capturing CO^2 levels are a sufficient predictor of high footfall and therefore would demonstrate likely nodes of transmission. This could further help investigation teams by limiting the scope of risk locations and therefore contacts to trace.

Ultimately, developing a way to risk stratify contacts could help beleaguered Tb services in a resource restricted setting. Combining index transmission risks; contact acquisition risk and location exposure factors to derive a risk score could inform such a triage and prioritisation process. This would have implications for follow up of defaulters if deemed at a high enough risk. The issues around this approach would be in defining a risk/ action threshold e.g. letter alerting of symptoms vs, active contact tracing approaches vs, public health legislation to compel compliance with screening and treatment.

8.7.3 Unknown aspects

Dogma still exists around the timing of exposure and resultant infection outcome, the 8-hour exposure duration may not be a clear rule and hyper acute transmission events may complicate contact tracing investigations. This change in accepted wisdom is increasingly being challenged by objective outcomes from whole genome sequencing data. Another unknown variable is the role played by so called super spreaders and being able to obtain dispersion values with areas of transmission may help define these prolific index cases further supported by describing social networks.

8.7.4 Molecular developments

As technology progresses, the historical approach to latent vs active disease is increasingly being challenged and what was once thought of as a binary process is increasingly viewed as a spectrum between quiescent and uncontrolled infection determined by immune state and function. Thus far, our ability to distinguish between these two states determines treatment approaches, the ability to detect 'active' disease earlier in the course of clinical assessment could be aided by developments in blood transcriptomics allowing us to monitor the development of subclinical and clinical disease in exposed contacts. [317] It may therefore be possible to differentiate between latent and active disease at an earlier stage. [318-320] This would place even greater importance on identifying contacts early.

8.7.5 Education

All the above future developments rely heavily on the development and incorporation of novel technologies as well as changing an entrenched public attitude to privacy, all of which are likely to encounter resistance and take years to implement. A more readily acceptable approach to improving engagement with contact tracing services is through education. Improving education may help improve compliance in underserved populations. [321] A further approach that has shown some success is the use of monetary incentives and peer advisers also improve adherence. [322]

8.8 Concluding remarks

In conclusion, this thesis has extensively examined the contemporary relevance and suitability of conventional contact tracing methods for TB in congregate settings. In addition to demonstrating the lack of evidence and support for current contact tracing approaches, the gradual evolution of contact tracing has been studied as well as how the shift in TB prevalence has limited its effectiveness.

This study contributes outcome data for contact tracing approaches which is lacking in the literature and demonstrates the longer-term outcomes of incident contact tracing episodes. Subsequent outbreaks that occur as a result of missed contacts at high risk are demonstrated using whole genome sequencing to link cases, further demonstrating the significant growth of infection clusters.

Furthermore, using a large database, predictive risk in contacts in congregate settings is undertaken which demonstrates the heterogeneous nature of social links and infrastructure and how risk factors vary in these settings.

This has implications for policy advisors in order to examine further contact tracing approaches as well as TB services in terms of assigning risk to contacts and following up higher risk groups. Data from this thesis should inform future guidelines on contact tracing in non-household settings.

8.9 Post-script

Since conceiving and writing this conclusion, the world has suffered the effects of a novel coronavirus infection resulting in a pandemic. Unprecedented though this may seem, it has served a purpose from a public health perspective in understanding infectious disease transmission in congregate settings in a dramatic way.

8.9.1 Caveats limiting comparability to TB

Whilst it would be tempting to directly compare airborne infectious diseases, there are some important differences to note between coronaviruses and tuberculosis.

Firstly, the latency period inherent in TB pathogenesis introduces novel locations for transmission distant from initial infection point in a spatio-temporal way.

Prolonged infection from the index case can result in multiple contacts over a longer period of time again separated in time and location. Additionally, unlike limited infectious and shedding periods typical of self-limited viral infections, TB index cases can have persistent infections before the host succumbs with no decrease in infectiousness resulting in a higher number of contacts. This biological difference is further augmented by an aberrant immune control resulting in no systematic defence, unlike a profound viral response which results in death or recovery of the host in a short time space. A further difference is in relation to the fomite component (physical deposition of a virus on a surface allowing further transfer and infection) and unlike novel infections with a high rate of exposure: infection ratio, not all TB exposures result in infection, indeed a minority of exposed individuals develop latent infection and 10% of these ‘reactivate’. Despite these limitations in comparing the two types

of infectious diseases, they both represent epidemics in the UK and preventative measures have some similarities, namely, the effectiveness of identification, treatment, isolation, quarantine and the need to rapidly identify and isolate contacts.

Future suggestions above have therefore in some form or another been tested out as part of the effort to curb and control the COVID-19 infection.

In particular, the recognition social networks play in infection transmission and the ability to identify and provide public health alerts to this process.

Firstly, the recognition that areas of social aggregation result in the spread of infection has resulted in the raft of measures known as lockdown measures the purpose of which is to prevent social contacts and congregations resulting in large infection incidents and subsequent outbreak events. We can see this in measures to close schools, work places, worship places, limited shopping entry/ exit and wearing masks in essential areas of congregate aggregation.

On a more fundamental level, it clearly highlights the key role congregate locations play over and above the household environment in terms of nodes of disease transmission in our modern society and the resultant impact on the population at large despite not being to identify contacts due to lack of personal acquaintance. Contemporaneous relaxation of rules also touch on several suggestions above namely the role ventilation in settings plays in terms of circulating pathogens and increased infection risk. The effectiveness of measures in congregate settings resulted in decreased social contacts and a subsequent reduction in the

number of new cases. This concept was tested with the relaxing of lock down and subsequent breaks in measures resulting in ‘blips’ in terms of rising case numbers.

8.9.2 Identifying and isolating social contacts

Several nations have demonstrated the effectiveness of phone apps, predominately the mandatory health and monitoring systems in South Korea. This is contingent on a recognition that the majority of people carry mobile phones recording geolocation data. Modern phones also have Bluetooth capacity and transmit/ receive local signals in relation to other contacts phones. This has resulted in a public health ability to message people once an index case has been identified and based on time/ extent of exposure, they can be alerted to a potential contact with appropriate public health advice. This has parallels with modern sexual health service confidential alerts and contact informing practices.

Privacy issues notwithstanding, this ability represents a novel and potentially highly effective method in identifying social contacts, particularly as it would identify unrelated and unknown contacts. A significant improvement to an index case having to name and identify contacts from memory with the obvious drawbacks. This would potentially make the social network questionnaire redundant as data recorded on a mobile would supply public health officials with both contact and locational data. The advantages over traditional methods and even over social network questionnaires would be particularly useful in underserved population groups. Combined with WGS, it could allow for rapid identification and curbing of incidents and prevent outbreaks.

8.9.3 Triage and treatment

The data collated would inform more accurate models for risk stratification. Knowing time, location and duration of exposure, contacts could be prioritised to receive either a notification of symptoms to be aware of and when to seek medical attention. Alternatively, higher risk individuals could be offered chest clinic appointments. Higher risk individuals that defaulted could also be prioritised by limited public health resources providing a more responsive and targeted public health approach and effort. The nationwide rollout of a 'track and trace app' despite the limitations has demonstrate the feasibility of this approach as well as the contemporaneous nature of contact identification and alerting systems. Overall, this approach would allow the earlier and more effective treatment of latent cases, decreasing the total burden of TB disease as well as tackling the larger pool of latent infection.

8.9.4 TB trends due to COVID pandemic

COVID-19 public health measures aimed to reduce the transmission of airborne/ droplet spread infectious diseases of which TB is one. Significant advances have been made in terms of public health awareness, contact tracing and direct interruption of spread through distancing and mask wearing. With all this, the initial predicted impact on TB should have been a decrease in transmission and numbers of cases.

That is not the case. Recent figures from public health services, no doubt an underestimate given the ability to record active cases only, have demonstrated an increase in case numbers comparable to 2019 levels. This represents a reversal in the nearly decade long decline in TB

trends and highlights the importance and relevance of TB contact tracing services more than pre-pandemic states. [42]

We are likely to be living with TB for decades to come. In the wake of the pandemic and renewed awareness of infectious threats, we may now have a unique opportunity to advance public health measures in order to develop a robust and adaptable system that not only deals with the TB epidemic once and for all but safeguards public health from large scale congregate transmissible threats into the future.

Appendices

Appendices - Chapter 3

Appendix 3.1: Search strategy in Medline

Tuberculosis
(“tuberculosis”[Title] OR "tuberculosis"[MeSH Terms] OR "mycobacterium tuberculosis"[MeSH Terms] OR "tuberculosis, pulmonary"[MeSH Terms] OR “TB” [Title] OR tuberculosis, latent [MeSH Terms], tuberculosis, miliary [MeSH Terms])

AND

Transmission
(Transmiss*[Title]) OR ("spread*"[Title])

AND

Risk factors
(“risk* factor*”[Title])

Appendix 3.2 – Excluded studies

Table 20 - Comparative papers excluded in systematic review

Number	Study ID	Congregate location	Type of study	Exclusion reason
1	Ali 2015	Prison	Cross-sectional	High incidence setting
2	Fuge 2016	Prison	Cross-sectional	High incidence setting
3	Park 2018	HCW	Cross-sectional	High incidence setting
5	Yoon 2017	HCW	Cross-sectional	High incidence setting
6	Lee 2018	Homeless	Cross-sectional	No risk factors – only prevalence
7	Lopez 2018	Prison	Cross-sectional	Did not focus on congregate settings
8	Owokuhaisa 2014	Prison	Cross-sectional	Did not examine factors in congregate setting
9	Powell 2011	HCW	Cross-sectional	Did not differentiate from household exposure
10	Rafiza 2011	HCW	Cross-sectional	Did not differentiate from household exposure

Appendix 3.3 – PRISMA diagram

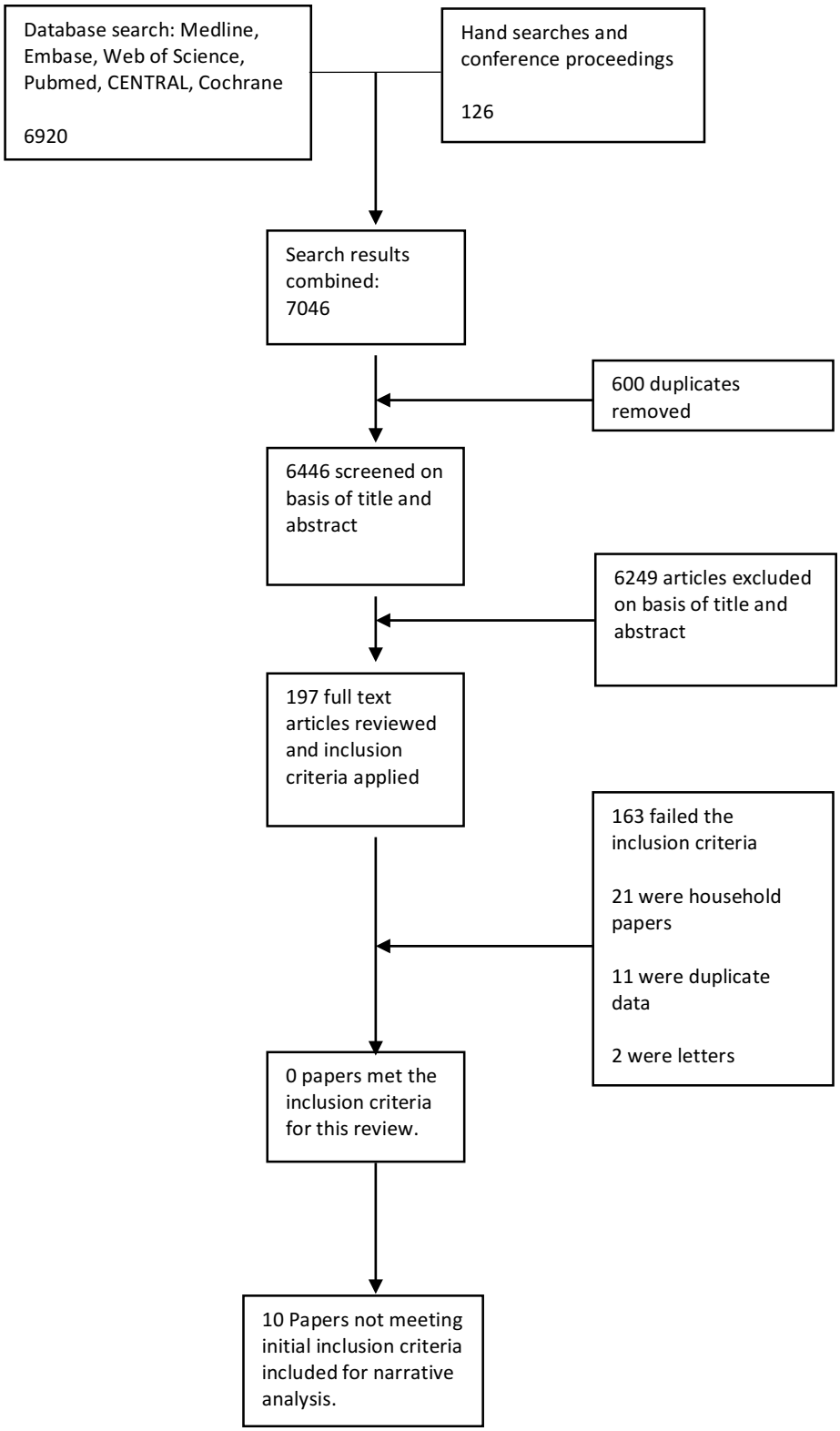

Figure 15 - PRISMA diagram (search conducted 20/06/2019)

Appendices – Chapter 4

Appendix 4.1: Characteristics of excluded studies

Andre 2007

Reason for exclusion	Cluster analysis study. No control group used.

Andrews 2014

Reason for exclusion	No control group. Modelling study examining congregate settings in a high-incidence environment.

Bayona 2003

Reason for exclusion	Cohort study. Investigated multidrug-resistant tuberculosis. Different interventions; study authors examined treatment options and secondary contact cases.

Cook 2007

Reason for exclusion	No control group. Cluster analysis.

de Vries 2006

Reason for exclusion	Cluster analysis study. No control group. Narrow study population.

Driver 2003

Reason for exclusion	Prospective observational study. Ineligible comparison group. Different assessments and outcomes.

Fatima 2016

Reason for exclusion	Different interventions. Examined active and passive case finding. Retrospective study.

Fox 2012

Reason for exclusion	Systematic review and meta-analysis looking at the overall effectiveness of contact tracing. Did not examine alternate methods of contact tracing. Looked at the prevalence of latent tuberculosis infection.

Fox 2013

Reason for exclusion	Different intervention. Examined active case finding. Looked at high incidence setting. Examined only household contacts.

Fox 2018

Reason for exclusion	Different intervention. Examined active case finding. Looked at high incidence setting. Examined only household contacts.

Jensen 2016

Reason for exclusion	Retrospective study. Different intervention. Examined active case finding.

McElroy 2003

Reason for exclusion	Retrospective study. No control group. Outbreak, cluster analysis.

Ribeiro 2015

Reason for exclusion	Spatial and genotypic cluster analysis of tuberculosis transmission events. High incidence setting.

Appendix 4.2 - Excluded studies

Andre 2007 [165]
(CRSSTD: 11976279)
Andre M, Ijaz K, Tillinghast JD, Krebs VE, Diem LA, Metchock B, et al. Transmission network analysis to complement routine tuberculosis contact investigations. American Journal of Public Health 2007;97(3):470-7. [CRSREF: 11976280]

Andrews 2014[254]
(CRSSTD: 11976281)
Andrews JR, Morrow C, Walensky RP, Wood R. Integrating social contact and environmental data in evaluating tuberculosis transmission in a South African township. Journal of Infectious Diseases 2003;7(12):S486-93. [CRSREF: 11976282]

Bayona 2003[323]
(CRSSTD: 11976283)
Bayona J, Chavez-Pachas AM, Palacios E, Llaro K, Sapag R, Becerra MC. Contact investigations as a means of detection and timely treatment of persons with infectious multidrug-resistant tuberculosis. International Journal of Tuberculosis and Lung Disease 2003;7(12):S501-9. [CRSREF: 11976284]

Cook 2007[226]
(CRSSTD: 11976285)
Cook VJ, Sun SJ, Tapia J, Muth SQ, Arguello DF, Lewis BL, et al. Transmission network analysis in tuberculosis contact investigations. Journal of Infectious Diseases 2007;196(10):1517-27. [CRSREF: 11976286]

de Vries 2006[324]
(CRSSTD: 11976287)
de Vries G, van Hest RA. From contact investigation to tuberculosis screening of drug addicts and homeless persons in Rotterdam. European Journal of Public Health 2005;16(2):133-6. [CRSREF: 11976288]

Driver 2003[325]
(CRSSTD: 11976289)
Driver CR, Balcewicz-Sablinska MK, Kim Z, Scholten J, Munsiff SS. Contact investigations in congregate settings, New York city. International Journal of Tuberculosis and Lung Disease 2003;7(12):S432-8. [CRSREF: 11976290]

Fatima 2016[326]
(CRSSTD: 11976291)
Fatima R, Qadeer E, Yaqoob A, Haq MU, Majumdar SS, Shewade HD, et al. Extending ‘contact tracing' into the community within a 50-metre radius of an index tuberculosis patient using Xpert MTB/RIF in urban, Pakistan: did it increase case detection? PLOS ONE 2016;11(11):11. [CRSREF: 11976292]

Fox 2012[232]
(CRSSTD: 11976293)

Fox GJ, Barry S, Marks GB. Outcomes of contact investigation for tuberculosis: a systematic review and meta-analysis. American Journal of Respiratory and Critical Care Medicine 2012;185:140-56. [CRSREF: 11976294]

Fox 2013[327]
(CRSSTD: 11976295)
Fox GJ, Nhung NV, Sy DN, Britton WJ, Marks GB. Cluster randomized controlled trial of household contact investigation in 8 provinces in Vietnam. American Journal of Respiratory and Critical Care Medicine 2013;187(Meetings/ Abstracts):342. [CRSREF: 11976296]

Fox 2018[328]
(CRSSTD: 11976297)
Fox GJ, Nhung NV, Sy DN, Hoa NLP, Anh LTN, Anh NT, et al. Household-contact investigation for detection of tuberculosis in Vietnam. New England Journal of Medicine 2018;378(3):221-9. [CRSREF: 11976298]

Jensen 2016[329]
(CRSSTD: 11976299)
Jensen SG, Lillebaek T, Wilcke T, Pedersen MK, Andersen PH, Olsen NW, et al. Impact of contact investigation and tuberculosis screening among high-risk groups in Denmark. International Journal of Tuberculosis and Lung Disease 2016;20(12):1580-7. [CRSREF: 11976300]

McElroy 2003[225]
(CRSSTD: 11976301)
McElroy PD, Rothenberg RB, Varghese R, Woodruff R, Minns GO, Muth SQ, et al. A network-informed approach to investigating a tuberculosis outbreak: implications for enhancing contact investigations. International Journal of Tuberculosis and Lung Disease 2003;7(12):S486-93. [CRSREF: 11976302]

Ribeiro 2015[330]
(CRSSTD: 11976303)
Ribeiro FK, Pan W, Bertolde A, Vinhas SA, Peres RL, Riley L, et al. Genotypic and spatial analysis of Mycobacterium tuberculosis transmission in a high-incidence urban setting. Clinical Infectious Diseases 2015;61(5):758-66. [CRSREF: 11976304]

Appendix 4.3: Search strategies

Cochrane Central Register of Controlled Trials

#1 tuberculosis:ti,ab,kw (Word variations have been searched)

#2 TB

#3 MeSH descriptor: [Mycobacterium tuberculosis] explode all trees

#4 MeSH descriptor: [Tuberculosis] explode all trees

#5 #1 or #2 or #3 or #4

#6 MeSH descriptor: [Contact Tracing] explode all trees

#7 "contact tracing"

#8 "contact screening" or "contact management"

#9 "contact investigation*"

#10 "transmission dynamics"

#11 referral

#12 "stone in pond"

#13 "household screening"

#14 "social network*"

#15 #7 or #8 or #9 or #10 or #10 or #11 or #12 or #13 or #14

#16 #15 and #5

PubMed (MEDLINE)

Search	Query
#1	tuberculosis [MesH]
#2	Mycobacterium tuberculosis [MesH]
#3	tuberculosis or TB Field: Title/Abstract
#4	((Mycobacterium tuberculosis [MesH]) OR #2) OR #1
#5	"Contact Tracing"[Mesh]
#6	"Contact Tracing" Field: Title/Abstract
#7	"contact screening" or "contact management" Field: Title/Abstract
#8	"contact investigation*" Field: Title/Abstract
#9	"transmission dynamics" Field: Title/Abstract
#10	referral Field: Title/Abstract
#11	"stone in pond" Field: Title/Abstract

#12	"household screening" Field: Title/Abstract
#13	"social network*" Field: Title/Abstract
#14	((((((((#13) OR #12) OR #11) OR #10 OR #9 OR #8 OR #7 OR #6 OR #5
#15	(#15) AND #4
#16	"Randomized Controlled Trial" [Publication Type] OR "Controlled Clinical Trial" [Publication Type]
#17	randomized or placebo Field: Title/Abstract
#18	randomly or trial or groups Field: Title/Abstract
#19	"drug therapy" [Subheading]
#20	(((#19) OR #18) OR #17 OR #16
#21	"Animals"[Mesh]
#22	"Humans"[Mesh]
#23	(#21) NOT #22
#27	(#20) NOT #23
#28	#15 AND #27

Embase

1 (tuberculosis or TB).mp.

2 limit 1 to human

3 Mycobacterium tuberculosis/

4 2 or 3

5 "contact tracing".mp. or contact examination/

6 ("contact screening" or "contact management").mp.

7 "contact investigation".mp.

8 "transmission dynamics".mp.

9 patient referral/

10 "stone in pond".mp.

11 "household screening".mp.

12 social network/ or "social network*".mp.

13 5 or 6 or 7 or 8 or 9 or 10 or 11 or 12

14 4 and 13

15 controlled clinical trial.mp. or Controlled Clinical Trial/

16 randomized controlled trial.mp. or Randomized Controlled Trial/

17 single blind procedure/

18 double blind procedure/

19 crossover procedure/

20 placebo.ti. or placebo.ab.

21 "randomly allocated".mp.

22 (randomized or placebo or double-blind* or single-blind*).mp

23 15 or 16 or 17 or 18 or 19 or 20 or 21 or 22

24 14 and 23

LILACS

Search on : tuberculosis or TB [331] and "contact tracing" or "contact screening" [331] and randomized or trial or groups [331]

CINAHL (EBSCOHost)

Query

S7 S5 AND S6

S6 TX (randomized controlled trial or rct) OR MH controlled clinical trial OR TX ("double blind*" or "single blind" or placebo or crossover)

S5 S1 AND S4

S4 S2 OR S3

S3 MH contact tracing OR TX "transmission dynamics" OR TX referral OR TX "household screening" OR TX "social network"

S2 TX "contact tracing" OR TX "contact screening" OR TX "contact investigation"

S1 TX (tuberculosis or TB) OR MH mycobacterium tuberculosis

Web of Science

5 #4 AND #3

Indexes=SCI-EXPANDED, SSCI, CPCI-S, CPCI-SSH Timespan=All years

4 TOPIC: (randomized trial or clinical trial) OR TOPIC: (double-blind* or single-blind* or placebo)

Indexes=SCI-EXPANDED, SSCI, CPCI-S, CPCI-SSH Timespan=All years

3 #2 AND #1

Indexes=SCI-EXPANDED, SSCI, CPCI-S, CPCI-SSH Timespan=All years

2 TOPIC: ("contact tracing" or "contact screening" or "contact investigation") OR TOPIC: ("household screening" or "transmission dynamics" or "social network*")

Indexes=SCI-EXPANDED, SSCI, CPCI-S, CPCI-SSH Timespan=All years

1 TOPIC: (tuberculosis or tb or "mycobacterium tuberculosis")

Indexes=SCI-EXPANDED, SSCI, CPCI-S, CPCI-SSH Timespan=All years

Appendices - Chapter 7

Appendix 7.1

Justification for variables included

From the summary data chapter, we know that location type plays a role.
There is an increased likelihood of disease acquisition in the following areas; Educational (primary/ combined, college), work (factory) and healthcare (hospital and care home)

Age again plays a role, with increased risk in a younger population group <15 years of age. This risk changes with social interaction which occurs with increasing age. Social contact with infectious diseases has been linked to a variable likelihood of acquisition. Age plays a key role in modelling infectious transmission. Not all age groups have the same number and type of social contact exposure nor are they likely to report multiple locations of exposure. [315] The very young and the very old are likely to have fewer contacts and fewer contact locations than the middle aged. Age has been demonstrated as a key factor in understanding the transmission of respiratory infections through social contacts. [315] The way a particular age group makes, interacts with and reports contacts is not necessarily the same across all age groups. Different age groups may represent different susceptibilities to infection. Social contacts as reported by age groups can be seen as behaviour within the particular age group. This is an important determinant of infection risk. [315, 332] In view of age being used, date of birth of individuals becomes redundant and has been excluded.

Gender has been previously demonstrated as having a role. How much this contributes is unclear but certain gendered behaviours may contribute to disease acquisition in different settings with changing sociocultural norms.

Ethnicity, country of birth and year of entry contribute to TB acquisition. This is due to the inherent risk of imported infection in certain ethnic groups. On average, TB disease can reactivate over a lifetime. The increased risk is in the first 2 years and therefore year of entry plays a role. Country of birth, especially if this is a high-risk location (>100/100,00 cases) can contribute to disease exposure and latent infection state. There are a significant number of missing values noted (41.6%). However, given the relevance of ethnic origin to subsequent risk of Tb disease, this is a crucial factor to be included. Given the inclusion of ethnicity and country of origin, UK birth as a separate factor and year of entry become less important and given the significant number of missing values (>50%), these have been excluded from the analysis.

Previous TB disease if present denotes risk of acquisition, it also raises the possibility of reinfection and reactivation depending on previous treatment compliance. A history of BCG vaccination would be seen as protective and would have an impact on risk of presentation with disease, despite this waning over time. These two variables have been inconsistently recorded for both index and contact cases. This could be due to reporting errors and recall bias from individuals not treated in the UK, or there may be issues with what treatments they may have received. In addition, where recorded, the predominant answer is no

previous BCG/ TB disease. Given the unreliability of recording, these variables have been removed from the analysis process.

The proximity of the contact to the index case and duration of exposure have long been accepted as strongly correlating with disease acquisition. The minimum amount of time needed for transmission to occur is not well established, however closer contact for longer periods would undoubtedly increase risk of disease in contacts.

HIV positive status has been widely seen as a key reason for the resurgence of TB in the early 90's. HIV increases lifetime risk of TB acquisition and can result in more severe disease and therefore higher transmission risk.

The occupation of the index case provides an indicator of social status and exposure risk e.g. teachers are potentially exposed and can expose classes to TB. In addition, occupation can provide some insight in non-traditional social relationships, friends outside work etc. While occupation is a risk factor in terms of how many people could be exposed, given that contacts are included and the proximity of relationship defined, including an occupation variable adds little. Furthermore, the location is included making an occupation category redundant. As the missing values exceed 10% and the categories for occupational status are somewhat arbitrary, this category has been excluded.

An index cases social background can influence presence and transmission of disease, drug and alcohol abuse correlates with disease acquisition and more severe disease. Furthermore, homelessness, incarcerated individuals and those that abuse alcohol and drugs are often harder to reach (services perspective), have erratic lifestyles and unnamed contacts and can result in larger TB contact networks as previously seen in the outbreak chapter. All these factors contribute to disease acquisition and transmission. Contacts of these individuals may be at higher risk by virtue of similar lifestyles. Self-administration ability correlates with an ability to comply with services, thus highlighting less chaotic lifestyles and lower risk of the above social risk factors known to contribute to TB acquisition and transmission. This latter variable is necessary for TB services to understand treatment options, however it does not necessarily impact the initial acquisition or spread of TB and is largely a service assigned level of risk. Therefore, it does not play a part in the assignment or predictability of transmission on initial assessment.

Finally, pulmonary disease and disease affecting the respiratory tract is far more likely to correlate with disease presence in contacts. TB spreads through expectorated respiratory droplets. Purely extra-pulmonary TB disease is less infectious and over and above pulmonary disease is not essential to consider when screening contacts in an incident setting. In addition, smear positive status (the result of lab testing highlighting a high burden of mycobacterium in expectorated sputum) has been shown to result in a poor prognosis in index cases (untreated) and result in a higher rate of transmission to contacts.

Table 21 - Total list of variables and groupings of the data set

Location	**Contacts**	**Index**	**Disease**	**Dates**
Category	Date of birth	Date of tuberculosis	Organism	Status of case
Subcategory	Current age	Date of notification	Sensitivity	Date onset of illness
	Gender	Patient gender	Resistance	Date first presented with symptoms
	Ethnic group	Date of birth	Histology	Date of diagnosis
	Born in the UK	Current age	Variable number tandem repeats (VNTR)	Directly observed therapy
	Year of entry	Ethnicity	ETR	Date start of treatment
	Country of birth	Country birth	Mycobacterial interspersed repetitive units (MIRU)	End of treatment
	Occupation	Year of entry	MIRU Plus	
	Previous TB	Previous TB	Microbiology laboratory notes	
	Previous TB date known	Previous TB date known	Post mortem diagnosis	
	Previous BCG	Previous BCG		
	Date of BCG	Date of BCG		
	Index year	HIV status		
	Contact index close contact	Occupation		
	Contact index relationship to contact	History of drug use		
	Contact index date of notification	Can self-administer?		
	Outcome	Homelessness		
		Incarceration history		
		Alcohol abuse		
		Drug abuse		
		Type of disease		
		Organs		
		Smear positive		
		Culture positive		

Table 22 - Breakdown of the variables and the number of missing values and their relative proportions

Data groups	Variables	Number of missing values	Percentage missing values (%)
Location data	Category (location)	0	0
	Subcategory (setting)	0	0
Contact cases	Date of birth	0	0
	Age	0	0
	Gender	425	5.3
	Ethnicity	3362	41.6
	UK born	4284	52.9
	Year of entry	4767	58.9
	Country of origin	4693	58
	Occupation	1274	15.7
	Previous TB	819	10.1
	Previous TB date	894	11
	Previous BCG	679	8.4
	Previous BCG date	3074	38
	Close contact of index case	0	0
	Relationship to index case	59	0.7
	Date of notification of TB	2477	30.6
	Outcome (disease or not)	144	1.8
Index cases	Date of tuberculosis diagnosis	995	12.3
	Date of notification to services	912	11.3
	Gender	28	0.3
	Date of birth	111	1.4
	Current age	111	1.4
	Ethnicity	0	0
	Country of birth	0	0
	Year of entry to UK	282	3.5
	Has had TB previously	274	3.4
	Previous TB date	474	5.9
	Has had BCG	332	4.1
	Date of BCG	5880	72.7
	HIV status	1212	15
	Occupation	57	0.7
	History of drug use	83	1
	Can self administer treatment	803	9.9
	Current or previous homelessness	59	0.7
	Incarceration	74	0.9
	Current alcohol abuse	30	0.4
	Current drug abuse	8	0.1
Disease characteristics	Extrapulmonary disease	0	0
	Organs involved	9	0.1
	Smear Positive	24	0.3

Table 23 - Data variables and coding for analysis

Variable	Coding			
	0	1	2	3
Gender	Male	Female		
Contact Ethnicity	Asian (dummy variable)	Afrocaribbean (dummy variable)	White/ other (dummy variable)	
Close Contact status	House	Close	Casual	
Relationship Status	1st degree relative or housemate*	2nd degree relative §	Close relationship^	Casual relationship $
Outcome	No disease	Latent or active disease		
Index Gender	male	female		
Index Ethnicity	Asian (dummy variable)	Afrocaribbean (dummy variables)	White/ other (dummy variables)	
HIV status	Negative	Positive		
Drug History	No	Yes		
Homeless	No	Yes		
Prison History	No	Yes		
Alcohol History	No	Yes		
Drug Abuse	No	Yes		
Smear Positive status	Negative	Positive		

* Includes
cellmate and
step-parents
§ Includes cousins, aunts, uncles and grandparents
^ Includes boyfriend, girlfriend and close friends
$ Includes work and school colleagues

Bibliography

1. Lee SH. Tuberculosis Infection and Latent Tuberculosis. Tuberc Respir Dis (Seoul). 2016;79(4):201-6.
2. Lawn SD, Zumla AI. Tuberculosis. Lancet. 2011;378(9785):57-72.
3. Zumla A, Abubakar I, Raviglione M, Hoelscher M, Ditiu L, McHugh TD, et al. Drug-resistant tuberculosis--current dilemmas, unanswered questions, challenges, and priority needs. J Infect Dis. 2012;205 Suppl 2:S228-40.
4. Patterson B, Wood R. Is cough really necessary for TB transmission? Tuberculosis (Edinb). 2019;117:31-5.
5. Turner RD, Chiu C, Churchyard GJ, Esmail H, Lewinsohn DM, Gandhi NR, et al. Tuberculosis Infectiousness and Host Susceptibility. J Infect Dis. 2017;216(suppl_6):S636-s43.
6. Geppert EF, Leff A. The pathogenesis of pulmonary and miliary tuberculosis. Arch Intern Med. 1979;139(12):1381-3.
7. Caws M, Thwaites G, Dunstan S, Hawn TR, Lan NT, Thuong NT, et al. The influence of host and bacterial genotype on the development of disseminated disease with Mycobacterium tuberculosis. PLoS Pathog. 2008;4(3):e1000034.
8. Tiemersma EW, van der Werf MJ, Borgdorff MW, Williams BG, Nagelkerke NJ. Natural history of tuberculosis: duration and fatality of untreated pulmonary tuberculosis in HIV negative patients: a systematic review. PLoS One. 2011;6(4):e17601.
9. Churchyard G, Kim P, Shah NS, Rustomjee R, Gandhi N, Mathema B, et al. What We Know About Tuberculosis Transmission: An Overview. J Infect Dis. 2017;216(suppl_6):S629-s35.
10. Andrews JR, Noubary F, Walensky RP, Cerda R, Losina E, Horsburgh CR. Risk of progression to active tuberculosis following reinfection with Mycobacterium tuberculosis. Clin Infect Dis. 2012;54(6):784-91.
11. Horsburgh CR, Jr. Priorities for the treatment of latent tuberculosis infection in the United States. N Engl J Med. 2004;350(20):2060-7.
12. Comstock GW, Livesay VT, Woolpert SF. The prognosis of a positive tuberculin reaction in childhood and adolescence. Am J Epidemiol. 1974;99(2):131-8.
13. Getahun H, Matteelli A, Abubakar I, Aziz MA, Baddeley A, Barreira D, et al. Management of latent Mycobacterium tuberculosis infection: WHO guidelines for low tuberculosis burden countries. Eur Respir J. 2015;46(6):1563-76.
14. Hayibor KM, Bandoh DA, Asante-Poku A, Kenu E. Predictors of Adverse TB Treatment Outcome among TB/HIV Patients Compared with Non-HIV Patients in the Greater Accra Regional Hospital from 2008 to 2016. Tuberc Res Treat. 2020;2020:1097581.
15. Zingue D, Bouam A, Tian RBD, Drancourt M. Buruli Ulcer, a Prototype for Ecosystem-Related Infection, Caused by Mycobacterium ulcerans. Clin Microbiol Rev. 2018;31(1).
16. Zimmerman MR. Pulmonary and osseous tuberculosis in an Egyptian mummy. Bull N Y Acad Med. 1979;55(6):604-8.
17. Sparacello VS, Roberts CA, Canci A, Moggi-Cecchi J, Marchi D. Insights on the paleoepidemiology of ancient tuberculosis from the structural analysis of postcranial remains from the Ligurian Neolithic (northwestern Italy). Int J Paleopathol. 2016;15:50-64.

18. Posa A, Maixner F, Mende BG, Kohler K, Osztas A, Sola C, et al. Tuberculosis in Late Neolithic-Early Copper Age human skeletal remains from Hungary. Tuberculosis (Edinb). 2015;95 Suppl 1:S18-22.
19. Bos KI, Harkins KM, Herbig A, Coscolla M, Weber N, Comas I, et al. Pre-Columbian mycobacterial genomes reveal seals as a source of New World human tuberculosis. Nature. 2014;514(7523):494-7.
20. Gutierrez MC, Brisse S, Brosch R, Fabre M, Omais B, Marmiesse M, et al. Ancient origin and gene mosaicism of the progenitor of Mycobacterium tuberculosis. PLoS Pathog. 2005;1(1):e5.
21. Murray JF, Rieder HL, Finley-Croswhite A. The King's Evil and the Royal Touch: the medical history of scrofula. Int J Tuberc Lung Dis. 2016;20(6):713-6.
22. Barberis I, Bragazzi NL, Galluzzo L, Martini M. The history of tuberculosis: from the first historical records to the isolation of Koch's bacillus. J Prev Med Hyg. 2017;58:9-12.
23. Furcolow MT, Deuschle KW. A Novel Approach to TB eradication. JAMA. 1965;193(4):3.
24. Citron KM. Home Treatment of Pulmonary Tuberculosis. BMJ. 1966:3.
25. Rodriguez-Campos S, Smith NH, Boniotti MB, Aranaz A. Overview and phylogeny of Mycobacterium tuberculosis complex organisms: implications for diagnostics and legislation of bovine tuberculosis. Res Vet Sci. 2014;97 Suppl:S5-S19.
26. Hunter RL. The Pathogenesis of Tuberculosis: The Early Infiltrate of Post-primary (Adult Pulmonary) Tuberculosis: A Distinct Disease Entity. Front Immunol. 2018;9:2108.
27. Ulrichs T, Kosmiadi GA, Trusov V, Jörg S, Pradl L, Titukhina M, et al. Human tuberculous granulomas induce peripheral lymphoid follicle like structures to orchestrate local host defence in the lung. Journal of Pathogens. 2004(204):217–28.
28. Lienard J, Carlsson F. Murine Mycobacterium marinum Infection as a Model for Tuberculosis. Methods Mol Biol. 2017;1535:301-15.
29. Marakalala MJ, Martinez FO, Pluddemann A, Gordon S. Macrophage Heterogeneity in the Immunopathogenesis of Tuberculosis. Front Microbiol. 2018;9:1028.
30. Hunter RL, Actor JK, Hwang SA, Khan A, Urbanowski ME, Kaushal D, et al. Pathogenesis and Animal Models of Post-Primary (Bronchogenic) Tuberculosis, A Review. Pathogens. 2018;7(1).
31. Orme IM. A new unifying theory of the pathogenesis of tuberculosis. Tuberculosis (Edinb). 2014;94(1):8-14.
32. Organization WH. Global tuberculosis report 2018. 2018.
33. Cohen A, Mathiasen VD, Schon T, Wejse C. The global prevalence of latent tuberculosis: a systematic review and meta-analysis. Eur Respir J. 2019;54(3).
34. Organization WH. Towards TB elimination: An action framework for low incidence countries. 2014.
35. Muhjazi G, Baghdadi S, Dye C, Qusaibaty M. Sustaining a national tuberculosis control programme during civil crisis: 6 years of experience in Syria. Lancet Respir Med. 2018;6(4):255-6.
36. strategy WHOET. Desired decline in TB incidence rates to reach 2035 target. http://wwwwhoint/tb/End_TB_brochurepdf 2015.
37. England PH. Tuberculosis cases in England hit lowest ever levels

PHE PHE; 2019 [updated 22/03/19. Available from: https://www.gov.uk/government/news/tuberculosis-cases-in-england-hit-lowest-ever-levels
38. Reports HP. Provisional TB data for England 2018: HPR; 2019 [updated 08/04/19. Available from: https://www.gov.uk/government/publications/health-protection-report-volume-13-2019/hpr-volume-13-issue-10-news-22-march - provisional-tb-data-england-for-2018
.
39. GOV.UK. Tuberculosis (TB): annual notification and mortality data (1913 onwards). 2019.
40. Agency HP. Health Protection Report volume 13: Issue 10. 2019.
41. O'Moore E, Roy A. Public Health England blog on Tackling tuberculosis in under-served populations 2019 [updated 22/03/19. Available from: https://publichealthmatters.blog.gov.uk/2019/03/22/tackling-tuberculosis-in-under-served-populations/.
42. Agency UHS. Tuberculosis in England: 2020. UK Health Security Agency, London; 2021.
43. Blog PHM. Public Health outcomes. 2019.
44. Mixides G, Shende V, Teeter LD, Awe R, Musser JM, Graviss EA. Number of negative acid-fast smears needed to adequately assess infectivity of patients with pulmonary tuberculosis. Chest. 2005;128(1):108-15.
45. Catanzaro A. What is the appropriate use of the rapid diagnostic tests for tuberculosis? Monaldi Arch Chest Dis. 1997;52(1):27-32.
46. Heym B, Honore N, Truffot-Pernot C, Banerjee A, Schurra C, Jacobs WR, Jr., et al. Implications of multidrug resistance for the future of short-course chemotherapy of tuberculosis: a molecular study. Lancet. 1994;344(8918):293-8.
47. Mazars E, Lesjean S, Banuls AL, Gilbert M, Vincent V, Gicquel B, et al. High-resolution minisatellite-based typing as a portable approach to global analysis of Mycobacterium tuberculosis molecular epidemiology. Proc Natl Acad Sci U S A. 2001;98(4):1901-6.
48. Oelemann MC, Diel R, Vatin V, Haas W, Rusch-Gerdes S, Locht C, et al. Assessment of an optimized mycobacterial interspersed repetitive- unit-variable-number tandem-repeat typing system combined with spoligotyping for population-based molecular epidemiology studies of tuberculosis. J Clin Microbiol. 2007;45(3):691-7.
49. Allix-Beguec C, Fauville-Dufaux M, Supply P. Three-year population-based evaluation of standardized mycobacterial interspersed repetitive-unit-variable-number tandem-repeat typing of Mycobacterium tuberculosis. Journal of Clinical Microbiology. 2008;46(4):1398-406.
50. Didelot X, Bowden R, Wilson DJ, Peto TEA, Crook DW. Transforming clinical microbiology with bacterial genome sequencing. Nat Rev Genet. 2012;13(9):601-12.
51. Schurch AC, van Soolingen D. DNA fingerprinting of Mycobacterium tuberculosis: from phage typing to whole-genome sequencing. Infect Genet Evol. 2012;12(4):602-9.
52. Schurch AC, Kremer K, Kiers A, Daviena O, Boeree MJ, Siezen RJ, et al. The tempo and mode of molecular evolution of Mycobacterium tuberculosis at patient-to-patient scale. Infect Genet Evol. 2010;10(1):108-14.

53. Schurch AC, Kremer K, Daviena O, Kiers A, Boeree MJ, Siezen RJ, et al. High-resolution typing by integration of genome sequencing data in a large tuberculosis cluster. J Clin Microbiol. 2010;48(9):3403-6.
54. 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;364(8):730-9.
55. 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. The Lancet Respiratory Medicine.2(4):285-92.
56. Mack U, Migliori GB, Sester M, Rieder HL, Ehlers S, Goletti D, et al. LTBI: latent tuberculosis infection or lasting immune responses to M. tuberculosis? A TBNET consensus statement. Eur Respir J. 2009;33(5):956-73.
57. Modlin RL, Bloom BR. TB or not TB: that is no longer the question. Sci Transl Med. 2013;5(213):213sr6.
58. Singhania A, Wilkinson RJ, Rodrigue M, Haldar P, O'Garra A. The value of transcriptomics in advancing knowledge of the immune response and diagnosis in tuberculosis. Nat Immunol. 2018;19(11):1159-68.
59. Drain PK, Bajema KL, Dowdy D, Dheda K, Naidoo K, Schumacher SG, et al. Incipient and Subclinical Tuberculosis: a Clinical Review of Early Stages and Progression of Infection. Clin Microbiol Rev. 2018;31(4).
60. Barry CE, 3rd, Boshoff HI, Dartois V, Dick T, Ehrt S, Flynn J, et al. The spectrum of latent tuberculosis: rethinking the biology and intervention strategies. Nat Rev Microbiol. 2009;7(12):845-55.
61. O'Garra A, Redford PS, McNab FW, Bloom CI, Wilkinson RJ, Berry MP. The immune response in tuberculosis. Annu Rev Immunol. 2013;31:475-527.
62. Horsburgh CR, Jr., Rubin EJ. Clinical practice. Latent tuberculosis infection in the United States. N Engl J Med. 2011;364(15):1441-8.
63. Akolo C, Adetifa I, Shepperd S, Volmink J. Treatment of latent tuberculosis infection in HIV infected persons. Cochrane Database Syst Rev. 2010(1):CD000171.
64. Comstock GW. How much isoniazid is needed for prevention of tuberculosis among immunocompetent adults? Int J Tuberc Lung Dis. 1999;3(10):847-50.
65. Comstock GW, Baum C, Snider DE, Jr. Isoniazid prophylaxis among Alaskan Eskimos: a final report of the bethel isoniazid studies. Am Rev Respir Dis. 1979;119(5):827-30.
66. Smieja MJ, Marchetti CA, Cook DJ, Smaill FM. Isoniazid for preventing tuberculosis in non-HIV infected persons. Cochrane Database Syst Rev. 2000(2):CD001363.
67. Menzies NA, Cohen T, Hill AN, Yaesoubi R, Galer K, Wolf E, et al. Prospects for Tuberculosis Elimination in the United States: Results of a Transmission Dynamic Model. Am J Epidemiol. 2018;187(9):2011-20.
68. Organization WH. Latent tuberculosis infection: up- dated and consolidated guidelines for programmatic management. World Health Organization, Geneva, Switzerland; 2018.
69. Swindells S, Ramchandani R, Gupta A, Benson CA, Leon-Cruz J, Mwelase N, et al. One Month of Rifapentine plus Isoniazid to Prevent HIV-Related Tuberculosis. N Engl J Med. 2019;380(11):1001-11.

70. Dye C, Glaziou P, Floyd K, Raviglione M. Prospects for tuberculosis elimination. Annu Rev Public Health. 2013;34:271-86.
71. Warren P. The evolution of the sanatorium: the first half-century, 1854-1904. Can Bull Med Hist. 2006;23(2):457-76.
72. Diacon AH, von Groote-Bidlingmaier F, Donald PR. From magic mountain to table mountain. Swiss Med Wkly. 2012;142:w13665.
73. Kempker RR, Vashakidze S, Solomonia N, Dzidzikashvili N, Blumberg HM. Surgical treatment of drug-resistant tuberculosis. Lancet Infect Dis. 2012;12(2):157-66.
74. Mitchison DA. The diagnosis and therapy of tuberculosis during the past 100 years. Am J Respir Crit Care Med. 2005;171(7):699-706.
75. Citron KM, De Silva DJ. The results of tuberculosis chemotherapy in chest clinic practice. Tubercle. 1971;52(1):31-6.
76. Comstock GW, Woolpert SF, Baum C. Isoniazid prophylaxis among Alaskan Eskimos: a progress report. Am Rev Respir Dis. 1974;110(2):195-7.
77. Parfitt CD. Osler's Influence in the War Against Tuberculosis (The Fifth Osler Oration). Can Med Assoc J. 1942;47(4):293-304.
78. Achkar JM, Jenny-Avital ER. Incipient and subclinical tuberculosis: defining early disease states in the context of host immune response. J Infect Dis. 2011;204 Suppl 4:S1179-86.
79. Corbett EL, Watt CJ, Walker N, Maher D, Williams BG, Raviglione MC, et al. The growing burden of tuberculosis: global trends and interactions with the HIV epidemic. Arch Intern Med. 2003;163(9):1009-21.
80. Shanaube K, Sismanidis C, Ayles H, Beyers N, Schaap A, Lawrence KA, et al. Annual risk of tuberculous infection using different methods in communities with a high prevalence of TB and HIV in Zambia and South Africa. PLoS One. 2009;4(11):e7749.
81. Hoa NB, Cobelens FG, Sy DN, Nhung NV, Borgdorff MW, Tiemersma EW. First national tuberculin survey in Viet Nam: characteristics and association with tuberculosis prevalence. Int J Tuberc Lung Dis. 2013;17(6):738-44.
82. 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;320(9):545-50.
83. Jick SS, Lieberman ES, Rahman MU, Choi HK. Glucocorticoid use, other associated factors, and the risk of tuberculosis. Arthritis Rheum. 2006;55(1):19-26.
84. Sidhu A, Verma G, Humar A, Kumar D. Outcome of latent tuberculosis infection in solid organ transplant recipients over a 10-year period. Transplantation. 2014;98(6):671-5.
85. Chegou NN, Heyckendorf J, Walzl G, Lange C, Ruhwald M. Beyond the IFN-gamma horizon: biomarkers for immunodiagnosis of infection with Mycobacterium tuberculosis. Eur Respir J. 2014;43(5):1472-86.
86. Moss AR, Hahn JA, Tulsky JP, Daley CL, Small PM, Hopewell PC. Tuberculosis in the homeless. A prospective study. Am J Respir Crit Care Med. 2000;162(2 Pt 1):460-4.
87. Cohen T, Dye C, Colijn C, Williams B, Murray M. Mathematical models of the epidemiology and control of drug-resistant TB. Expert Rev Respir Med. 2009;3(1):67-79.
88. Tiberi S, Utjesanovic N, Galvin J, Centis R, D'Ambrosio L, van den Boom M, et al. Drug resistant TB - latest developments in epidemiology, diagnostics and management. Int J Infect Dis. 2022.

89. Dye C, Lonnroth K, Jaramillo E, Williams BG, Raviglione M. Trends in tuberculosis incidence and their determinants in 134 countries. Bulletin of the World Health Organization. 2009;87(9):683-91.
90. Veen J. Microepidemics in tuberculosis: The stone-in-pond principle. Tuber Lung Dis. 1992;73:73-6.
91. Control and prevention of tuberculosis in the United Kingdom: code of practice 2000. Joint Tuberculosis Committee of the British Thoracic Society. Thorax. 2000;55(11):887-901.
92. Abubakar I, Griffiths C, Ormerod P, Guideline Development G. Diagnosis of active and latent tuberculosis: summary of NICE guidance. BMJ. 2012;345:e6828.
93. Braganza Menezes D, Menezes B, Dedicoat M. Contact tracing strategies in household and congregate environments to identify cases of tuberculosis in low- and moderate-incidence populations. Cochrane Database Syst Rev. 2019;8:CD013077.
94. Kasaie P, Andrews JR, Kelton WD, Dowdy DW. Timing of tuberculosis transmission and the impact of household contact tracing. An agent-based simulation model. Am J Respir Crit Care Med. 2014;189(7):845-52.
95. Dodd PJ, Looker C, Plumb ID, Bond V, Schaap A, Shanaube K, et al. Age- and Sex-Specific Social Contact Patterns and Incidence of Mycobacterium tuberculosis Infection. Am J Epidemiol. 2016;183(2):156-66.
96. Wingfield T, Boccia D, Tovar M, Gavino A, Zevallos K, Montoya R, et al. Defining Catastrophic Costs and Comparing Their Importance for Adverse Tuberculosis Outcome with Multi-Drug Resistance: A Prospective Cohort Study, Peru. PLoS Med. 2014;11(7).
97. Siroka A, Law I, Macinko J, Floyd K, Banda RP, Hoa NB, et al. The effect of household poverty on tuberculosis. Int J Tuberc Lung Dis. 2016;20(12):1603-8.
98. Granich R, Akolo C, Gunneberg C, Getahun H, Williams P, Williams B. Prevention of tuberculosis in people living with HIV. Clin Infect Dis. 2010;50 Suppl 3:S215-22.
99. Getahun H, Granich R, Sculier D, Gunneberg C, Blanc L, Nunn P, et al. Implementation of isoniazid preventive therapy for people living with HIV worldwide: barriers and solutions. AIDS. 2010;24 Suppl 5:S57-65.
100. Knight GM, Griffiths UK, Sumner T, Laurence YV, Gheorghe A, Vassall A, et al. Impact and cost-effectiveness of new tuberculosis vaccines in low- and middle-income countries. Proc Natl Acad Sci U S A. 2014;111(43):15520-5.
101. Roy A, Eisenhut M, Harris RJ, Rodrigues LC, Sridhar S, Habermann S, et al. Effect of BCG vaccination against Mycobacterium tuberculosis infection in children: systematic review and meta-analysis. BMJ. 2014;349:g4643.
102. Nemes E, Geldenhuys H, Rozot V, Rutkowski KT, Ratangee F, Bilek N, et al. Prevention of M. tuberculosis Infection with H4:IC31 Vaccine or BCG Revaccination. N Engl J Med. 2018;379(2):138-49.
103. Nemes E, Scriba TJ, Hatherill M. Prospects for a vaccine to prevent HIV-related tuberculosis. Curr Opin HIV AIDS. 2018;13(6):522-7.
104. Agency UHS. TB Action Plan for England, 2021-2026. In: Service NH, editor. Gov.net2021.
105. Santos AP, Silva DR, Mello FCQ. In the time of strategies to end tuberculosis, prevention is better than treatment. J Bras Pneumol. 2020;46(2):e20200017.
106. Daniel TM. The history of tuberculosis. Respir Med. 2006;100(11):1862-70.

107. Houben RM, Dodd PJ. The Global Burden of Latent Tuberculosis Infection: A Re-estimation Using Mathematical Modelling. PLoS Med. 2016;13(10):e1002152.
108. Davies PDO. Tuberculosis in humans and animals: are we a threat to each other? J R Soc Med. 2006;99:539-40.
109. Behr MA, Edelstein PH, Ramakrishnan L. Revisiting the timetable of tuberculosis. BMJ. 2018;362:k2738.
110. Green HT. A Lecture on the Tubercle Bacillus and Phthisis. BMJ. 1883:3.
111. Berry W. An address on the advancement of public health during the last quarter of a century BMJ. 1898:3.
112. Williams CTC. A discussion on the treatment of consumption by climate. Introductory address on classification of climates and comparison of results. . BMJ. 1901:5.
113. Philip RW. Remarks on the universal applicability of open air treatment of pulmonary tuberculosis. BMJ. 1898:4.
114. Hill CG. The Prevention of Tuberculosis: At the tenth annual meeting of the South Durham and North Yorkshire Veterinary Association. BMJ. 1898:1.
115. Maxwell H. British Congress on Tuberculosis: Proceedings of Sections BMJ. 1901:13.
116. Atkins PJ. Milk consumption and tuberculosis in Britain, 1850-1950.', in Order and disorder : the health implications of eating and drinking in the nineteenth and twentieth centuries. 1999:83-95.
117. Ingraham CW. Control of Tuberculosis from a striclty Medico-legal standpoint. JAMA. 1896:3.
118. Thorne R. Sanitary authorities and tuberculosis. BMJ. 1899:2.
119. Carpenter A. Remarks on the First Principles of Sanitary Work. BMJ. 1879:6.
120. Pôrto A. Social representations of tuberculosis: stigma and prejudice. Rev Saúde Pública. 2007;41.
121. Correspondant B. The British Congress on Tuberculosis. BMJ. 1901:1.
122. BMJ GC. The prevention of Tuberculosis: The Ochil Hills Sanatorium. BMJ. 1900:1.
123. Clayson C. Sir Robert Philip and the conquest of Tuberculosis BMJ. 1957:6.
124. Harston W, Wheeler WS. First Tuberculosis Dispensary. BMJ. 1958:2.
125. Dunlop DM. Modern views in the prevention of Tuberculosis BMJ. 1938:4.
126. Sutherland H. The soild and the seed in Tuberculosis. BMJ. 1912:4.
127. Bridge N, Evans WA, Paquin P. Report of the committee on Tuberculosis JAMA. 1900;25(16).
128. Bushnell F. The prevention of Tuberculosis in Leicester. BMJ. 1900:1.
129. Bracken BM. Tuberculosis in State Institutions JAMA. 1902:4.
130. JAMA E. Lessons of the Tuberculosis congress. JAMA. 1908;LI(18):2.
131. Taylor K. The 1913 Dispensary for the care of people with tuberculosis. Wakefield Historical Library.
132. Foster JPC. The State and Tuberculosis. JAMA. 1906:1.
133. Burns J. The relationship between medicine and public health delivered at the seventeenth international congress of medicine. BMJ. 1913:8.
134. Rhodes M. The prevention of Tuberculosis. Sanatoria for the consumptive pauper. BMJ. 1900:1.
135. Philip R. Tuberculin in diagnosis and treatment of tuberculosis. BMJ. 1923:493.

136. Boswell H. Fundamental factors in the control of Tuberculosis with special reference to the Sanatorium. JAMA. 1921;77(7):3.
137. D'Arcy Hart P. Efficacy and Applicability of Mass B.C.G. Vaccination in Tuberculosis Control. BMJ. 1967:6.
138. Smith A. Incidence of Tuberculosis among American medical students JAMA. 1953;152(10):3.
139. CDC. Use of BCG Vaccines in the Control of Tuberculosis: A Joint Statement by the ACIP and the Advisory. MMWR Morb Mortal Wkly Rep. 1988;260(20):9.
140. Editor J. Tuberculosis in families. JAMA. 1936;106(2):3.
141. Opie EL, McPhedran FM. Spread of Tuberculosis within families. JAMA. 1926;87(19):3.
142. Drolet G. World War I and Tuberculosis: A Statistical Summary and Review. 1944;35.
143. Daniels M. Tuberculsis in Europe during and after the second world war BMJ. 1949.
144. Morse LM. Control of Tuberculosis in Wisconsin: The results of four years of mass study in sanitary disctrict No. 5 1936 to 1939. JAMA. 1941;116(19):4.
145. Plunkett RE. Tuberculosis control. JAMA. 1939;113(26):5.
146. Gass RS, Bishop EL. Tuberculosis control in Tennessee JAMA. 1931;97(12):4.
147. Puffer RP. Familial Susceptibility to Tuberculosis: Its Importance as a Public Health Problem. Harvard University Monographs in Medicine and Public Health. 1944(5):106.
148. Hetherington HW. The diagnosis and managment of latent, suspected and early clinical Tuberculosis. JAMA. 1937;109(24):5.
149. Health MP, Association. The Evolution of Tuberculosis as Observed during twenty years at Lymanhurst 1921 to 1941. JAMA. 1944:1.
150. Bliss TL. Tuberculosis case finding: A practical and successful program for non-metropolitan areas. JAMA. 1941;117(23):3.
151. JAMA E. Fiftieth anniversary of the national Tuberculosis association. JAMA. 1954;155(3):2.
152. Mitchell RS. Sanatorium treatment for Tuberculosis JAMA. 1954;154(12):2.
153. Wassersug JD. Tuberculosis and private health JAMA. 1950;143(13):1.
154. BMJ E. Treating Tuberculosis from limited resources. BMJ. 1960:2.
155. Tempel CW. Tuberculosis Control: The Role of the Physician in Private Practice. JAMA. 1965;191(5):7.
156. McConville JH. Tuberculosis Management in the Mid-1970s. JAMA. 1976;235(2):5.
157. Thompson DH, Trachtman L, Greenberg HB, Schilling RP. Orleans Parish Prison Still Needs Program for Tuberculosis Control. JAMA. 1977;237(21):1.
158. Association JTCotBT. Notification of tuberculosis: a code of practice for England and Wales. BMJ. 1982;284:3.
159. Hans L. Rieder M, MPH; George M. Cauthen, ScD; Gloria D. Kelly;, Alan B. Bloch M, MPH; Dixie E. Snider, Jr,. Tuberculosis in the United States. JAMA. 1989;262(3):4.
160. Gostin LO. Controlling the Resurgent Tuberculosis Epidemic: A 50-State Survey of TB Statutes and Proposals for Reform. JAMA. 1993;269(2):6.
161. Bellin E. Failure of Tuberculosis Control: A Prescription for Change. JAMA. 1994;271(9):3.
162. Organisation WH. Global Tuberculosis Report 2016. 2016.

163. England. PH. Tuberculosis in the West Midlands: Annual review (2016 data). Public Health England, Birmingham. 2017.
164. England PH. Tuberculosis in England 2016 Report (2015 data) 2016.
165. Andre M, Ijaz K, Tillinghast JD, Krebs VE, Diem LA, Metchock B, et al. Transmission network analysis to complement routine tuberculosis contact investigations. Am J Public Health. 2007;97(3):470-7.
166. Arnold A, Witney AA, Vergnano S, Roche A, Cosgrove CA, Houston A, et al. XDR-TB transmission in London: Case management and contact tracing investigation assisted by early whole genome sequencing. J Infect. 2016;73(3):210-8.
167. Young KH, Ehman M, Reves R, Peterson Maddox BL, Khan A, Chorba TL, et al. Tuberculosis Contact Investigations--United States, 2003-2012. MMWR Morb Mortal Wkly Rep. 2016;64(50-51):1369-74.
168. Organization WH. Framework towards TB elimination in low incidence countries.
169. Patterson B, Morrow CD, Kohls D, Deignan C, Ginsburg S, Wood R. Mapping sites of high TB transmission risk: Integrating the shared air and social behaviour of TB cases and adolescents in a South African township. Sci Total Environ. 2017;583:97-103.
170. Middelkoop K, Bekker LG, Morrow C, Lee NC, Wood R. Decreasing household contribution to TB transmission with age: a retrospective geographic analysis of young people in a South African township. BMC Infect Dis. 2014;14(221).
171. Middelkoop K, Bekker LG, Morrow C, Zwane E, Wood R. Childhood tuberculosis infection and disease: A spatial and
temporal transmission analysis in a South African township. South African Medical Journal. 2009;99(10):738–43.
172. Jones-Lopez EC, Acuna-Villaorduna C, Ssebidandi M, Gaeddert M, Kubiak RW, Ayakaka I, et al. Cough Aerosols of Mycobacterium tuberculosis in the Prediction of Incident Tuberculosis Disease in Household Contacts. Clin Infect Dis. 2016;63(1):10-20.
173. Sepkowitz KA. How Contagious Is Tuberculosis? Clinical Infectious Diseases November. 1996;23(5):954-62.
174. Faksri K, Reechaipichitkul W, Pimrin W, Bourpoern J, Prompinij S. Transmission and Risk Factors for Latent Tuberculosis Infections among Index Case-Matched Household Contacts. Southeast Asian J Trop Med Public Health. 2015;46(3):486-95.
175. O'Shea MK, Koh GC, Munang M, Smith G, Banerjee A, Dedicoat M. Time-to-detection in culture predicts risk of Mycobacterium tuberculosis transmission: a cohort study. Clin Infect Dis. 2014;59(2):177-85.
176. Keane J, Gershon, S., Wise, R. P., Mirabile-Levens, E., Kasznica, J., Schwieterman, W. D., Siegel, J. N., Braun, M. M. Tuberculosis associated with Infliximab, A Tumour Necrosis Factor alpha neutralising agent. New England Journal of Medicine. 2001;345(15):1098-104.
177. Jick SS, Lieberman, E. S., Rahman, M. U., Choi, H. K. Glucocorticoid use, other associated factors and the risk of tuberculosis. American college of Rheumatology. 2006;55(1):19-26.
178. Dillon ES, Boucot, K. R., Cooper, D. A., Meier, P., Richardson, R. A survey of tuberculosis among diabetics. American diabetes association. 1952.
179. Chia S, Karim, M., Elwood, R. K., FitzGerald, J. M. Risk of tuberculosis in dialysis patients: a population-based study. International Journal of Tuberculosis & Lung Disease. 1998;2(12):989-91.

180. Chiang C, Slama, K., Enarson, D. A. Associations between tobacco and tuberculosis. International Union Against Tuberculosis and Lung Disease. 2007;11(3):258-62.
181. Kaplan MH, Armstrong D, Rosen P. Tuberculosis complicating neoplastic disease: A review of 201 cases. Cancer. 1974;33:850-8.
182. Yates TA, Tanser F, Abubakar I. Plan Beta for tuberculosis: it's time to think seriously about poorly ventilated congregate settings. Int J Tuberc Lung Dis. 2016;20(1):5-10.
183. Sloot R, Schim van der Loeff MF, Kouw PM, Borgdorff MW. Risk of tuberculosis after recent exposure. A 10-year follow-up study of contacts in Amsterdam. Am J Respir Crit Care Med. 2014;190(9):1044-52.
184. Huang L, Li XX, Abe EM, Xu L, Ruan Y, Cao CL, et al. Spatial-temporal analysis of pulmonary tuberculosis in the northeast of the Yunnan province, People's Republic of China. Infect Dis Poverty. 2017;6(1):53.
185. Hella J, Morrow C, Mhimbira F, Ginsberg S, Chitnis N, Gagneux S, et al. Tuberculosis transmission in public locations in Tanzania: A novel approach to studying airborne disease transmission. J Infect. 2017;75(3):191-7.
186. Munang ML, Browne C, Evans JT, Smith EG, Hawkey PM, Welch SB, et al. Programmatic utility of tuberculosis cluster investigation using a social network approach in Birmingham, United Kingdom. Int J Tuberc Lung Dis. 2016;20(10):1300-5.
187. Vynnycky E, Fine PE. Interpreting the decline in tuberculosis: the role of secular trends in effective contact. Int J Epidemiol. 1999;28(2):327-34.
188. Jelip J, Mathew GG, Yusin T, Dony JF, Singh N, Ashaari M, et al. Risk factors of tuberculosis among health care workers in Sabah, Malaysia. Tuberculosis (Edinb). 2004;84(1-2):19-23.
189. Uden L, Barber E, Ford N, Cooke GS. Risk of Tuberculosis Infection and Disease for Health Care Workers: An Updated Meta-Analysis. Open Forum Infect Dis. 2017;4(3):ofx137.
190. Baussano I, Williams BG, Nunn P, Beggiato M, Fedeli U, Scano F. Tuberculosis incidence in prisons: a systematic review. PLoS Med. 2010;7(12):e1000381.
191. Surucuoglu S. [Risk of travel associated tuberculosis]. Mikrobiyol Bul. 2018;52(1):96-107.
192. Kotila SM, Payne Hallstrom L, Jansen N, Helbling P, Abubakar I. Systematic review on tuberculosis transmission on aircraft and update of the European Centre for Disease Prevention and Control risk assessment guidelines for tuberculosis transmitted on aircraft (RAGIDA-TB). Euro Surveill. 2016;21(4).
193. Abubakar I. Tuberculosis and air travel: a systematic review and analysis of policy. Lancet Infect Dis. 2010;10(3):176-83.
194. Edelson PJ, Phypers M. TB transmission on public transportation: a review of published studies and recommendations for contact tracing. Travel Med Infect Dis. 2011;9(1):27-31.
195. Sane Schepisi M, Motta I, Dore S, Costa C, Sotgiu G, Girardi E. Tuberculosis transmission among children and adolescents in schools and other congregate settings: a systematic review. New Microbiol. 2018;41(4).
196. Martinez L, Shen Y, Mupere E, Kizza A, Hill PC, Whalen CC. Transmission of Mycobacterium tuberculosis in Households and the Community: A Systematic Review and Meta-Analysis. American Journal of Epidemiology. 2017;185(12):1327-39.

197. Roberts JR, Mason BW, Paranjothy S, Palmer SR. The transmission of tuberculosis in schools involving children 3 to 11 years of age. Pediatr Infect Dis J. 2012;31(1):82-4.
198. Nava-Aguilera E, Andersson N, Harris E, Mitchell S, Hamel C, Shea B, et al. Risk factors associated with recent transmission of tuberculosis: systematic review and meta-analysis. Int J Tuberc Lung Dis. 2009;13(1):17-26.
199. Fok A, Numata Y, Schulzer M, FitzGerald MJ. Risk factors for clustering of tuberculosis cases: a systematic review of population-based molecular epidemiology studies. Int J Tuberc Lung Dis. 2008;12(5):480-92.
200. Faustini A, Hall AJ, Perucci CA. Risk factors for multidrug resistant tuberculosis in Europe: a systematic review. Thorax. 2006;61(2):158-63.
201. Nardell EA. Tuberculosis in homeless, residential care facilities, prisons, nursing homes, and other close communities. Semin Respir Infect. 1989;4(3):206-15.
202. Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gotzsche PC, Ioannidis JP, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. PLoS Med. 2009;6(7):e1000100.
203. Higgins JPT, Green S, (editors). Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [updated March 2011]. The Cochrane Collaboration Available from http://handbookcochraneorg/. 2011.
204. Popay J, Roberts H, Sowden A, Petticrew M, Arai L, Rodgers M, et al. Guidance on the conduct of narrative synthesis in systematic reviews, ESRC Methods Programme (Unpublished report, University of Lancaster, UK). 2006.
205. Tacconelli E. Systematic reviews: CRD's guidance for undertaking reviews in health care. Lancet Infect Dis 2010;10(226).
206. Shimouchi A, Koda S, Hirota M, Matsumoto K. [Evaluation of tuberculosis contact investigation & examination in congregate settings in Osaka City]. Kekkaku. 2009;84(6):491-7.
207. Ali S, Haileamlak A, Wieser A, Pritsch M, Heinrich N, Loscher T, et al. Prevalence of Pulmonary Tuberculosis among Prison Inmates in Ethiopia, a Cross-Sectional Study. PLoS One. 2015;10(12).
208. Fuge TG, Ayanto SY. Prevalence of smear positive pulmonary tuberculosis and associated risk factors among prisoners in Hadiya Zone prison, Southern Ethiopia. BMC Res Notes. 2016;9.
209. López de Goicoechea-Saiz M, Sternberg F, Portilla-Sogorb J. Prevalence and associated risk factors of latent tuberculosis infection in a Spanish prison. Rev Esp Sanid Penit. 2018;20(1):4-10.
210. Owokuhaisa J, Thokerunga E, Bazira J. Prevalence of Pulmonary Tuberculosis among Prison Inmates at Mbarara Central Prison, South Western Uganda. Adv Res. 2014;2(11):618-25.
211. Park JS. The Prevalence and Risk Factors of Latent Tuberculosis Infection among Health Care Workers Working in a Tertiary Hospital in South Korea. Tuberc Respir Dis (Seoul). 2018;81(4):274-80.
212. Yoon C, Oh SY, Lee JB, Kim MH, Seo Y, Yang J, et al. Occupational Risk of Latent Tuberculosis Infection in Health Workers of 14 Military Hospitals. J Korean Med Sci. 2017;32(8):1251-7.

213. Powell K, Han D, Hung NV, Vu T, Sy DN, Trinh TT, et al. Prevalence and risk factors for tuberculosis infection among personnel in two hospitals in Viet Nam. International Journal of Tuberculosis and Lung Disease. 2011;15(12):1643-8.
214. Rafiza S, Rampal KG, Tahir A. Prevalence and risk factors of latent tuberculosis infection among health care workers in Malaysia. BMC Infect Dis. 2011;11:19.
215. Tudor C, Van der Walt ML, Margot B, Dorman SE, Pan WK, Yenokyan G, et al. Occupational Risk Factors for Tuberculosis Among Healthcare Workers in KwaZulu-Natal, South Africa. Clin Infect Dis. 2016;62(Suppl 3):S255-61.
216. Hwang YW, Lee YJ, Kong SY. Epidemiology and clinical outcomes of tuberculosis among homeless persons visiting emergency department in public hospital. Am J Emerg Med. 2018;36(1):164-6.
217. Aldridge RW, Hayward AC, Hemming S, Yates SK, Ferenando G, Possas L, et al. High prevalence of latent tuberculosis and bloodborne virus infection in a homeless population. Thorax. 2018;73(6):557-64.
218. Story A, Murad S, Roberts W, Verheyen M, Hayward AC, London Tuberculosis Nurses N. Tuberculosis in London: the importance of homelessness, problem drug use and prison. Thorax. 2007;62(8):667-71.
219. Lee CH, Jeong YJ, Heo EY, Park JS, Lee JS, Lee BJ, et al. Active pulmonary tuberculosis and latent tuberculosis infection among homeless people in Seoul, South Korea: a cross-sectional study. BMC Public Health. 2013;13:720.
220. Sloot R, Schim van der Loeff MF, Kouw PM, Borgdorff MW. Yield of tuberculosis contact investigations in Amsterdam: opportunities for improvement. Eur Respir J. 2014;44(3):714-24.
221. Diel R, Nienhaus A, Lange C, Meywald-Walter K, Forssbohm M, Schaberg T. Tuberculosis contact investigation with a new, specific blood test in a low-incidence population containing a high proportion of BCG-vaccinated persons. Respir Res. 2006;7:77.
222. Morrison J, Pai M, Hopewell PC. Tuberculosis and latent tuberculosis infection in close contacts of people with pulmonary tuberculosis in low-income and middle-income countries: a systematic review and meta-analysis. Lancet Infect Dis. 2008;8(6):359-68.
223. Excellence NIfC. Tuberculosis. NICE guideline NG 33. 2016.
224. Barnes PF, El-Hajj H, Preston-Martin S, Cave MD, Jones BE, Otaya M, et al. Transmission of Tuberculosis
Among the Urban Homeless. JAMA. 1996;275(4):305-7.
225. McElroy PD, Rothenberg RB, Varghese R, Woodruff R, Minns GO, Muth SQ, et al. A network-informed approach to investigating a tuberculosis outbreak: implications for enhancing contact investigations. Int J Tuberc Lung Dis. 2003;7(12 Suppl 3):S486-93.
226. Cook VJ, Sun SJ, Tapia J, Muth SQ, Arguello DF, Lewis BL, et al. Transmission network analysis in tuberculosis contact investigations. J Infect Dis. 2007;196(10):1517-27.
227. Kline SE, Hedemark LL, Davies SF. Outbreak of tuberculosis among regular patrons of a neighborhood bar. N Engl J Med. 1995;333(4):222-7.
228. Johnstone-Robertson S, Lawn SD, Welte A, Bekker LG, Wood R. Tuberculosis in a South African prison – a transmission modelling analysis. S Afr Med J. 2011;101(11):809-13.
229. Wood R, Racow K, Bekker LG, Morrow C, Middelkoop K, Mark D, et al. Indoor Social Networks in a South African Township: Potential Contribution of Location to Tuberculosis Transmission. PLoS One. 2012;7(6).

230. Chamie G, Wandera B, Marquez C, Kato-Maeda M, Kamya MR, Havlir DV, et al. Identifying locations of recent TB transmission in rural Uganda: a multidisciplinary approach. Trop Med Int Health. 2015;20(4):537-45.
231. Buu TN, van Soolingen D, Huyen MN, Lan NN, Quy HT, Tiemersma EW, et al. Tuberculosis acquired outside of households, rural Vietnam. Emerg Infect Dis. 2010;16(9):1466-8.
232. Fox GJ, Barry SE, Britton WJ, Marks GB. Contact investigation for tuberculosis: a systematic review and meta-analysis. Eur Respir J. 2013;41(1):140-56.
233. Begun M, Newall AT, Marks GB, Wood JG. Contact tracing of tuberculosis: a systematic review of transmission modelling studies. PLoS One. 2013;8(9):e72470.
234. Gronauer W. Employees and regular customers of restaurants and discos as transmitters of tuberculosis. Gesundheitswesen (Bundesverband der Arzte des Offentlichen Gesundheitsdienstes (Germany)). 1994;56(1):33-6.
235. Franken WP, Koster BF, Bossink AW, Thijsen SF, Bouwman JJ, van Dissel JT, et al. Follow-up study of tuberculosis-exposed supermarket customers with negative tuberculin skin test results in association with positive gamma interferon release assay results. Clin Vaccine Immunol. 2007;14(9):1239-41.
236. Kawatsu L, Izumi K, Uchimura K, Urakawa M, Ohkado A, Takahashi I. Can social network analysis assist in the prioritisation of contacts in a tuberculosis contact investigation? Int J Tuberc Lung Dis. 2015;19(11):1293-9.
237. Johnstone-Robertson SP, Mark D, Morrow C, Middelkoop K, Chiswell M, Aquino LDH, et al. Social Mixing Patterns Within a South African Township Community: Implications for Respiratory Disease Transmission and Control. American Journal of Epidemiology. 2011;174(11):1246-55.
238. Kuhleis D, Ribeiro AW, Costa ER, Cafrune PI, Schmid KB, Costa LL, et al. Tuberculosis in a southern Brazilian prison. Mem Inst Oswaldo Cruz. 2012;107(7):909-15.
239. Silva VM, Cunha AJ, Kritski AL. Tuberculin skin test conversion among medical students at a teaching hospital in Rio de Janeiro, Brazil. Infect Control Hosp Epidemiol. 2002;23(10):591-4.
240. Carbonne A, Poirier C, Antoniotti G, Burnat C, Delacourt C, Orzechowski C, et al. Investigation of patient contacts of heath care workers with infectious tuberculosis: 6 cases in the Paris area. Int J Tuberc Lung Dis. 2005;9(8):848-52.
241. Joshi R, Reingold AL, Menzies D, Pai M. Tuberculosis among health-care workers in low- and middle-income countries: a systematic review. PLoS Med. 2006;3(12):e494.
242. Christopher DJ, Daley P, Armstrong L, James P, Gupta R, Premkumar B, et al. Tuberculosis Infection among Young Nursing Trainees in South India. PLoS One. 2010;5(4).
243. Chheng P, Nsereko M, Malone LL, Okware B, Zalwango S, Joloba M, et al. Tuberculosis case finding in first-degree relative contacts not living with index tuberculosis cases in Kampala, Uganda. Clin Epidemiol. 2015;7:411-9.
244. Matsumoto K, Tatsumi T, Arima K, Koda S, Yoshida H, Kamiya N, et al. An outbreak of tuberculosis in which environmental factors influenced tuberculosis infection. Kekkaku : [Tuberculosis]. 2011;86(5):487-91.
245. Gao X, Wei J, Lei H, Xu P, Cowling BJ, Li Y. Building Ventilation as an Effective Disease Intervention Strategy in a Dense Indoor Contact Network in an Ideal City. PLoS One. 2016;11(9):e0162481.

246. Pankhurst LJ, Anaraki S, Lai KM. Combining environmental assessment and contact investigations to make tuberculosis screening decisions. Int J Tuberc Lung Dis. 2012;16(8):1023-9.
247. Gardy JL, Johnston JC, Sui SJH, Cook VJ, Shah L, Brodkin E, et al. Whole-Genome Sequencing and Social-Network Analysis of a Tuberculosis Outbreak. New England Journal of Medicine. 2011;364(8):730-9.
248. Lefebvre C, Manheimer, E., Glanville, J. Chapter 6. Searching for studies. Higgins J, Green, S., editor. Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 (updated March 2011). 2011.
249. Higgins J, Deeks, JJ., Altman, DG. Chapter 16. Special topics in statistics. Higgins J, Deeks, JJ., Altman, DG, editor. Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 (updated March 2011). The Cochrane Collaboration; 2011.
250. Higgins J, Thompson, SG., Deeks, JJ., Altman, DG. Measuring inconsistency in meta-analyses. . BMJ. 2003;327:557-60.
251. Deeks J, Higgins, JPT., Altman, DG. Chapter 9. Analysing data and undertaking meta-analyses. Higgins J, Churchil,l R., Chandler, J., Cumpston, MS. , editor. Cochrane Handbook for Systematic Reviews of Interventions version 5.2.0 (updated June 2017): Cochrane, 2017; 2017.
252. Sterne J, Egger, M., Moher, D., Boutron, I. . Chapter 10. Addressing reporting biases. Higgins J, Churchill, R., Chandler, J., Cumpston, MS., editor. Cochrane, 20172017.
253. McElroy PD, Rothenberg RB, Varghese R, Woodruff R, Minns GO, Muth SQ, et al. A network-informed approach to investigating a tuberculosis outbreak: Implications for enhancing contact investigations. International Journal of Tuberculosis and Lung Disease. 2003;7(12 SUPPL. 3):S486-S93.
254. Andrews JR, Morrow C, Walensky RP, Wood R. Integrating Social Contact and Environmental Data in Evaluating Tuberculosis Transmission in a South African Township. Journal of Infecious Diseases. 2014;210:597-603.
255. Fox GJ, Nhung NV, Sy DN, Lien LT, Cuong NK, Britton WJ, et al. Contact investigation in households of patients with tuberculosis in Hanoi, Vietnam: a prospective cohort study. PLoS One. 2012;7(11):e49880.
256. Worrell MC, Kramer M, Yamin A, Ray SM, Goswami ND. Use of Activity Space in a Tuberculosis Outbreak: Bringing Homeless Persons Into Spatial Analyses. Open Forum Infect Dis. 2017;4(1):ofw280.
257. Organisation WH. Recommendations for investigating contacts of persons with infectious tuberculosis in low- and middle-income countries. 2012.
258. Abubakar I, Anderson S, Hirani H, Kar-Purkayastha IT, L. , Weil L, Zenner D. Collaborative Tuberculosis Strategy for England 2015 to 2020. In: England PH, editor.: Crown Publishing; 2014.
259. Zenner D, Loutet MG, Harris R, Wilson S, Ormerod LP. Evaluating 17 years of latent tuberculosis infection screening in north-west England: a retrospective cohort study of reactivation. European Respiratory Journal. 2017;50.
260. Mitruka K, Oeltmann JE, Ijaz K, Haddad MB. Tuberculosis Outbreak Investigations in the United States, 2002–2008. Emerg Infect Dis. 2011;17(3):425-31.
261. Mindra G, Wortham JM, Haddad MB, Powell KM. Tuberculosis Outbreaks in the United States, 2009-2015. Public Health Rep. 2017;132(2):157-63.

262. Agency UHS. TB in the West Midlands Annual Review (2020 data). In: UKHSA, editor.: UKHSA; April 2022.
263. England. PH. Tuberculosis in England: 2018. . London: Public Health England; 2018 2018.
264. Yun LW, Reves RR, Reichler MR, Bur S, Thompson V, Mangura B, et al. Outcomes of contact investigation among homeless persons with infectious tuberculosis. Int J Tuberc Lung Dis. 2003;7(12 Suppl 3):S405-11.
265. Asghar RJ, Patlan DE, Miner MC, Rhodes HD, Solages A, Katz DJ, et al. Limited utility of name-based tuberculosis contact investigations among persons using illicit drugs: results of an outbreak investigation. J Urban Health. 2009;86(5):776-80.
266. Control ECfDPa. Investigation and control of tuberculosis incidents affecting children in congregate settings. In: ECDC, editor. Stockholm: ECDC; 2013.
267. (NICE) NIfCE. Costing report: tuberculosis: prevention, diagnosis, management and service organisation Implementing the NICE guideline on tuberculosis (NG33). 2016.
268. Taylor JG, Yates TA, Mthethwa M, Tanser F, Abubakar I, Altamirano H. Measuring ventilation and modelling M. tuberculosis
transmission in indoor congregate settings, rural KwaZulu-Natal. International Union Against Tuberculosis and Lung Disease. 2016;20(9):1155-61.
269. Onorato IM. Tuberculosis outbreaks in the United States. Int J Tuberc Lung Dis. 2000;4(12 Suppl 2):S121-6.
270. Uzicanin A, Gainesm J. CDC Field Epidemiology Manual: Community Congregate Settings CDC Field Epidemiology Manual Oxford University Press; 2018 [updated 13/12/2018. 4th:[Available from: https://www.cdc.gov/eis/field-epi-manual/chapters/community-settings.html.
271. Fitzpatrick LK, Hardacker JA, Heirendt W, Agerton T, Streicher A, Melnyk H, et al. A preventable outbreak of tuberculosis investigated through an intricate social network. Clin Infect Dis. 2001;33(11):1801-6.
272. Aguilera XP, Gonzalez C, Najera-De Ferrari M, Hirmas M, Delgado I, Olea A, et al. Tuberculosis in prisoners and their contacts in Chile: estimating incidence and latent infection. Int J Tuberc Lung Dis. 2016;20(1):63-70.
273. Assefzadeh M, Barghi RG, Shahidi Sh S. Tuberculosis case--finding and treatment in the central prison of Qazvin province, Islamic Republic of Iran. East Mediterr Health J. 2009;15(2):258-63.
274. Dasgupta K, Menzies D. Cost-effectiveness of tuberculosis control strategies among immigrants and refugees. Eur Respir J. 2005;25(6):1107-16.
275. Kambali S, Nantsupawat N, Lee M, Nugent K. A workplace tuberculosis case investigation in the presence of immigrant contacts from high prevalence countries. J Community Health. 2015;40(3):576-80.
276. Martinez-Lirola M, Alonso-Rodriguez N, Sanchez ML, Herranz M, Andres S, Penafiel T, et al. Advanced survey of tuberculosis transmission in a complex socioepidemiologic scenario with a high proportion of cases in immigrants. Clin Infect Dis. 2008;47(1):8-14.
277. Ospina JE, Orcau A, Millet JP, Ros M, Gil S, Cayla JA, et al. Epidemiology of Tuberculosis in Immigrants in a Large City with Large-Scale Immigration (1991-2013). PLoS One. 2016;11(10):e0164736.

278. Bandera A, Gori A, Catozzi L, Esposti AD, Marchetti G, Molteni C, et al. Molecular Epidemiology Study of Exogenous Reinfection in an Area with a Low Incidence of Tuberculosis. J Clin Microbiol. 2001;39(6):2213-8.
279. Chin DP, DeRiemer K, Small PM, de Leon AP, Steinhart R, Schecter GF, et al. Differences in contributing factors to tuberculosis incidence in U.S. -born and foreign-born persons. Am J Respir Crit Care Med. 1998;158(6):1797-803.
280. Bock NN, Mallory JP, Mobley N, DeVoe B, Taylor BB. Outbreak of Tuberculosis Associated with a Floating Card Game in the Rural
South: Lessons for Tuberculosis Contact Investigations. Clinical infectious diseases : an official publication of the Infectious Diseases Society of America. 1998;27:1221-6.
281. Sandgren A, Schepisi MS, Sotgiu G, Huitric E, Migliori GB, Manissero D, et al. Tuberculosis transmission between foreign- and native-born populations in the EU/EEA: a systematic review. Eur Respir J. 2014;43(4):1159-71.
282. Glaziou P, Sismanidis C, Floyd K, Raviglione M. Global Epidemiology of Tuberculosis. Cold Spring Harb Perspect Med. 2015;5:1-17.
283. Tuberculosis Unit NIS, PHE. Tuberculosis in England 2019 report: Executive Summary. PHE, England PH; 2019 July 2019.
284. Thomas LH, Harris RJ, Muzyamba MC, Davidson JA, Lalor MK, Campbell CNJ, et al. Reduction in tuberculosis incidence in the UK from 2011 to 2015: a population-based study. Thorax. 2018;73:769-75.
285. Association PHELG. Tackling tuberculosis. In: England PH, editor. local governments public health role. PHE report: PHE; 2018.
286. Walker TM, Ip CL, 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;13(2):137-46.
287. 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;2(4):285-92.
288. Wyllie DH, Davidson JA, Smith G, Rathod P, Crook DW, Peto TEA, et al. A Quantitative Evaluation of MIRU-VNTR Typing Against Whole-Genome Sequencing for Identifying Mycobacterium tuberculosis Transmission: A Prospective Observational Cohort Study. EBioMedicine. 2018;34:122-30.
289. Munang ML, Browne C, Evans JT, Smith EG, Hawkey PM, Welch SB, et al. Programmatic utility of tuberculosis cluster investigation using a social network approach in Birmingham, United Kingdom. International Journal of Tuberculosis and Lung Disease. 2016;20(10):1300-5.
290. England PH. TB Strain Typing and Cluster Investigation Handbook. London: Public Health England; 2014 February 2014.
291. Xu Y, Cancino-Muñoz I, Torres-Puente M, Villamayor LM, Borras R, Bosque MBM, et al. High-resolution mapping of tuberculosis transmission: Whole genome sequencing and phylogenetic modelling of a cohort from Valencia Region, Spain. PLoS Med. 2019;16(10).
292. Mitruka K, Oeltmann JE, Ijaz K, Haddad MB. Tuberculosis outbreak investigations in the United States, 2002-2008. Emerg Infect Dis. 2011;17(3):425-31.

293. Sterling TR, Thompson D, Stanley RL, McElroy PD, Madison A, Moore K, et al. A multi-state outbreak of tuberculosis among members of a highly mobile social network: implications for tuberculosis elimination. Int J Tuberc Lung Dis. 2000;4(11):1066-73.
294. Mohtashemi M, Kawamura LM. Empirical evidence for synchrony in the evolution of TB cases and HIV+ contacts among the San Francisco homeless. PLoS One. 2010;5(1):e8851.
295. Ward LJ, Hughes SE, Grabau JC. The Evaluation of School-Based Contact Investigations in New York State, Exclusive of New York City, 1997-2001. Journal of Public Health Management & Practice July/August. 2004;10(4):308-15.
296. Li J, Driver CR, Munsiff SS, Fujiwara PI. Finding contacts of homeless tuberculosis patients in New York City. Int J Tuberc Lung Dis. 2003;7(12 Suppl 3):S397-404.
297. Driver CR, Macaraig M, McElroy PD, Clark C, Munsiff SS, Kreiswirth B, et al. Which patients' factors predict the rate of growth of Mycobacterium tuberculosis clusters in an urban community? Am J Epidemiol. 2006;164(1):21-31.
298. Dobler CC, Marks GB. Risk of tuberculosis among contacts in a low-incidence setting. Eur Respir J. 2013;41(6):1459-61.
299. Esmail H, Barry CE, 3rd, Young DB, Wilkinson RJ. The ongoing challenge of latent tuberculosis. Philos Trans R Soc Lond B Biol Sci. 2014;369(1645):20130437.
300. Hayward S, Harding RM, McShane H, Tanner R. Factors influencing the higher incidence of tuberculosis among migrants and ethnic minorities in the UK. F1000Res. 2018;7.
301. Canadian Thoracic S. Canadian Tuberculosis Standards. 7th ed. Canadian Thoracic S, editor. Canada2014 February 2014.
302. Browne C, Gulbudak H, Webb G. Modeling contact tracing in outbreaks with application to Ebola. J Theor Biol. 2015;384:33-49.
303. Kwok KO, Tang A, Wei VWI, Park WH, Yeoh EK, Riley S. Epidemic Models of Contact Tracing: Systematic Review of Transmission Studies of Severe Acute Respiratory Syndrome and Middle East Respiratory Syndrome. Comput Struct Biotechnol J. 2019;17:186-94.
304. Muller J, Koopmann B. The effect of delay on contact tracing. Math Biosci. 2016;282:204-14.
305. Fraser C, Riley S, Anderson RM, Ferguson NM. Factors that make an infectious disease outbreak controllable. Proc Natl Acad Sci U S A. 2004;101(16):6146-51.
306. Kiss IZ, Green DM, Kao RR. Disease contact tracing in random and clustered networks. Proc Biol Sci. 2005;272(1570):1407-14.
307. Cheema J. A Review of Missing Data Handling Methods in Education Research. Review of Educational Research. 2014;84(4):487-508.
308. Nanta S, Kantipong P, Pathipvanich P, Ruengorn C, Tawichasri C, Patumanond J. Screening scheme development for active TB prediction among HIV-infected patients. Southeast Asian J Trop Med Public Health. 2011;42(4):867-75.
309. Balcha TT, Skogmar S, Sturegard E, Schon T, Winqvist N, Reepalu A, et al. A Clinical Scoring Algorithm for Determination of the Risk of Tuberculosis in HIV-Infected Adults: A Cohort Study Performed at Ethiopian Health Centers. Open Forum Infect Dis. 2014;1(3):ofu095.
310. Mehnaz A, Arif F. Applicability of scoring chart in the early detection of tuberculosis in children. J Coll Physicians Surg Pak. 2005;15(9):543-6.

311. Wisnivesky JP, Henschke C, Balentine J, Willner C, Deloire AM, McGinn TG. Prospective validation of a prediction model for isolating inpatients with suspected pulmonary tuberculosis. Arch Intern Med. 2005;165(4):453-7.
312. Shams H, Weis SE, Klucar P, Lalvani A, Moonan PK, Pogoda JM, et al. Enzyme-linked immunospot and tuberculin skin testing to detect latent tuberculosis infection. Am J Respir Crit Care Med. 2005;172(9):1161-8.
313. Chan PC, Shinn-Forng Peng S, Chiou MY, Ling DL, Chang LY, Wang KF, et al. Risk for tuberculosis in child contacts. Development and validation of a predictive score. Am J Respir Crit Care Med. 2014;189(2):203-13.
314. Kwok KO, Cowling B, Wei V, Riley S, Read JM. Temporal variation of human encounters and the number of locations in which they occur: a longitudinal study of Hong Kong residents. J R Soc Interface. 2018;15(138).
315. Kwok KO, Cowling BJ, Wei VW, Wu KM, Read JM, Lessler J, et al. Social contacts and the locations in which they occur as risk factors for influenza infection. Proc Biol Sci. 2014;281(1789):20140709.
316. Cauchemez S, Hoze N, Cousien A, Nikolay B, Ten Bosch Q. How Modelling Can Enhance the Analysis of Imperfect Epidemic Data. Trends Parasitol. 2019.
317. Tabone O, Verma R, Singhania A, Chakravarty P, Branchett WJ, Graham CM, et al. Blood transcriptomics reveal the evolution and resolution of the immune response in tuberculosis. J Exp Med. 2021;218(10).
318. Estevez O, Anibarro L, Garet E, Martinez A, Pena A, Barcia L, et al. Multi-parameter flow cytometry immunophenotyping distinguishes different stages of tuberculosis infection. J Infect. 2020;81(1):57-71.
319. Luo Y, Tang G, Lin Q, Mao L, Xue Y, Yuan X, et al. Combination of mean spot sizes of ESAT-6 spot-forming cells and modified tuberculosis-specific antigen/phytohemagglutinin ratio of T-SPOT.TB assay in distinguishing between active tuberculosis and latent tuberculosis infection. J Infect. 2020;81(1):81-9.
320. Ho J, Bokil NJ, Nguyen PTB, Nguyen TA, Liu MY, Hare N, et al. A transcriptional blood signature distinguishes early tuberculosis disease from latent tuberculosis infection and uninfected individuals in a Vietnamese cohort. J Infect. 2020;81(1):72-80.
321. Peterson Tulsky J, Castle White M, Young JA, Meakin R, Moss AR. Street talk: knowledge and attitudes about tuberculosis and tuberculosis control among homeless adults. Int J Tuberc Lung Dis. 1999;3(6):528-33.
322. Pilote L, Tulsky JP, Zolopa AR, Hahn JA, Schecter GF, Moss AR. Tuberculosis prophylaxis in the homeless. A trial to improve adherence to referral. Arch Intern Med. 1996;156(2):161-5.
323. Bayona J, Chavez-Pachas AM, Palacios E, Llaro K, Sapag R, Becerra MC. Contact investigations as a means of detection and timely treatment of persons with infectious multidrug-resistant tuberculosis. Int J Tuberc Lung Dis. 2003;7(12 Suppl 3):S501-9.
324. de Vries G, van Hest RA. From contact investigation to tuberculosis screening of drug addicts and homeless persons in Rotterdam. Eur J Public Health. 2006;16(2):133-6.
325. Driver CR, Balcewicz-Sablinska MK, Kim Z, Scholten J, Munsiff SS. Contact investigations in congregate settings, New York City. Int J Tuberc Lung Dis. 2003;7(12 Suppl 3):S432-8.

326. Fatima R, Qadeer E, Yaqoob A, Haq MU, Majumdar SS, Shewade HD, et al. Extending 'Contact Tracing' into the Community within a 50-Metre Radius of an Index Tuberculosis Patient Using Xpert MTB/RIF in Urban, Pakistan: Did It Increase Case Detection? PLoS One. 2016;11(11):e0165813.
327. Fox GJ, Nhung NV, Sy DN, Britton WJ, Marks GB. Household contact investigation for tuberculosis in Vietnam: study protocol for a cluster randomized controlled trial. Trials. 2013;14:342.
328. Fox GJ, Nhung NV, Sy DN, Hoa NLP, Anh LTN, Anh NT, et al. Household-Contact Investigation for Detection of Tuberculosis in Vietnam. N Engl J Med. 2018;378(3):221-9.
329. Jensen SG, Lillebaek T, Wilcke T, Pedersen MK, Andersen PH, Olsen NW, et al. Impact of contact investigation and tuberculosis screening among high-risk groups in Denmark. Int J Tuberc Lung Dis. 2016;20(12):1580-7.
330. Ribeiro FK, Pan W, Bertolde A, Vinhas SA, Peres RL, Riley L, et al. Genotypic and Spatial Analysis of Mycobacterium tuberculosis Transmission in a High-Incidence Urban Setting. Clin Infect Dis. 2015;61(5):758-66.
331. Pankhurst LJ, Del Ojo Elias C, Votintseva AA, Walker TM, Cole K, Davies J, et al. Rapid, comprehensive, and affordable mycobacterial diagnosis with whole-genome sequencing: a prospective study. Lancet Respir Med. 2016;4(1):49-58.
332. Mossong JI, Hens N, Jit M, Beutels P, Auranen K, Mikolajczyk R, et al. Social Contacts and Mixing Patterns Relevant to
the Spread of Infectious Diseases. PLoS Med. 2008;3.