Using data analysis and mathematical modelling to study tuberculosis contact tracing in London, with reference to the national strategy and guidance

Sean M Cavany

Thesis submitted in accordance with the requirements for the degree of

Doctor of Philosophy of the
University of London

March 2018

Department of Infectious Disease Epidemiology

Faculty of Epidemiology & Population Health

LONDON SCHOOL OF HYGIENE & TROPICAL MEDICINE

Funded by PHE-LSHTM Infectious Disease Modelling Studentship

Research group affiliations: TB Modelling Group
Centre for the Mathematical Modelling of Infectious Diseases
TB Centre

Declaration

See next page

London School of Hygiene & Tropical Medicine
Keppel Street, London WC1E 7HT
www.lshtm.ac.uk

Registry
T: +44(0)20 7299 4646
F: +44(0)20 7299 4656
E: registry@lshtm.ac.uk

LONDON SCHOOL of HYGIENE &TROPICAL MEDICINE

DECLARATION OF OWN WORK

All students are required to complete the following declaration when submitting their thesis. A shortened version of the School's definition of Plagiarism and Cheating is as follows (*the full definition is given in the Research Degrees Handbook*):

"Plagiarism is the act of presenting the ideas or discoveries of another as one's own. To copy sentences, phrases or even striking expressions without acknowledgement in a manner which may deceive the reader as to the source is plagiarism. Where such copying or close paraphrase has occurred the mere mention of the source in a biography will not be deemed sufficient acknowledgement; in each instance, it must be referred specifically to its source. Verbatim quotations must be directly acknowledged, either in inverted commas or by indenting" (University of Kent).

Plagiarism may include collusion with another student, or the unacknowledged use of a fellow student's work with or without their knowledge and consent. Similarly, the direct copying by students of their own original writings qualifies as plagiarism if the fact that the work has been or is to be presented elsewhere is not clearly stated.

Cheating is similar to plagiarism, but more serious. Cheating means submitting another student's work, knowledge or ideas, while pretending that they are your own, for formal assessment or evaluation.

Supervisors should be consulted if there are any doubts about what is permissible.

DECLARATION BY CANDIDATE

I have read and understood the School's definition of plagiarism and cheating given in the Research Degrees Handbook. I declare that this thesis is my own work, and that I have acknowledged all results and quotations from the published or unpublished work of other people.

I have read and understood the School's definition and policy on the use of third parties (either paid or unpaid) who have contributed to the preparation of this thesis by providing copy editing and, or, proof reading services. I declare that no changes to the intellectual content or substance of this thesis were made as a result of this advice, and, that I have fully acknowledged all such contributions.

I have exercised reasonable care to ensure that the work is original and does not to the best of my knowledge break any UK law or infringe any third party's copyright or other intellectual property right.

To be completed by the candidate

NAME IN FULL (*Block Capitals*): SEAN M. CAVANY

STUDENT ID NO: 1404547

SIGNED: [redacted] DATE: 22/03/18

Registry
Last updated – 04/07/13

The majority of those who profess to be desirous of preventing and curing the disease called consumption must be either hypocrites or fools, for they ridicule the suggestion that it is necessary first to cure and prevent the poverty that compels badly clothed and half-starved human beings to sleep in such dens as this.

- Robert Tressell, *The Ragged Trousered Philanthropists* (1914)

Abstract

Background: In January 2015, Public Health England and NHS England published a collaborative TB strategy for the years 2015-20; this strategy highlighted contact tracing as a key element. In January 2016, the National Institute of Health and Care Excellence (NICE) made changes to the UK TB guidance, including no longer recommending that contacts of non-pulmonary TB cases be screened. This thesis attempts to address several issues arising from these policy documents.

Methods: I utilized a range of quantitative approaches. I undertook a cohort analysis of TB cases in London between 2012-15 (inclusive), including logistic regression, to understand contact tracing outcomes in London, and how these differed between population subgroups. To understand the impact of changes to NICE guidance I carried out an economic analysis using a simple static model. I then utilized a pairwise transmission model to understand how transmission differs between those with primary and reactivation disease.

Results: In London from 2012-15, 91% of pulmonary index cases had at least one contact identified (a median of four per case), and 86% of these identified contacts were evaluated. In this period, 80% of those contacts determined to have TB had an isolate that was indistinguishable from their index case, implying probable transmission. Assuming each contact with PTB infects 1 person/month, screening contacts of ETB cases costs £78000/QALY (95% CI: 39000 to 140000). Pairwise modelling suggests that the number of infections generated by those with a reactivation disease is only slightly greater than those with disease following recent infection.

Conclusions: While contact tracing outcomes in London were good relative to similar countries and previous UK studies, our results highlight several groups for whom outcomes are worse. Our results also show that the impact of contact tracing is not limited to those occasions where transmission between index cases and contact has occurred. Our results also show that screening contacts of non-pulmonary index cases is almost certainly not cost-effective at a £30000/QALY threshold. More work is required if pairwise modelling is to be used effectively to model *M. Tb* transmission.

Contents

Acknowledgements

First and foremost I would like to offer huge thanks to my supervisors Emilia Vynnycky and Tom Sumner. Thank you for your patience when progress has been slow (as it often was), for your encouragement to get into the habit of writing things up, and for your wisdom which has made me a better scientist. It is unlikely I would have finished this thesis or any papers without your support. And hopefully I will never write or say 'data is…' again!

I am very grateful to my advisor Richard White and the rest of the TB modelling group, which always manage to find the perfect balance of work and fun. Our weekly meetings have been invaluable; both in providing constructive criticism of my own work, and in regularly setting amazing examples of how to ensure science has a positive impact on the world. Of course, our monthly visits to the pub were equally invaluable!

I would like to offer great thanks to my other advisors Charlotte Anderson, Lucy Thomas and Helen Maguire, for making sure my work was relevant, providing epidemiological insight and improving my paper-writing no end. I'd also like to thank Charlotte, Helen, Jacqui Carless, Neil MacDonald and the rest of the London field epidemiology team for providing me with access to and helping me get to grips with the London TB Register, not to mention maintenance of this great resource. Many thanks also to all of the TB nurses, doctors and other clinical staff in London, doing the really important work of helping people get better.

I would also like to thank my Public Health England for providing my studentship, and to Andre Charlett and the Statistics Unit: Andre for giving me the chance in the first place and for sporadic football discussions, and the rest of the unit for always welcoming me despite my infrequent appearances at Colindale. I'd also like to thank Ayoub Saei at PHE and Clare Flach at King's College for offering me sage statistics guidance.

Thanks must also go to all inhabitants of room G30a in the last three years, each of whom have made coming into the office a pleasurable experience irrespective of how well my research was going. In particular, thanks to Frank Sandmann, Ketil Tvermosegaard and Matteo Quartagno (honorary G30a alumnus) for helping me put the world to rights, and laugh. Thanks also to Frank for teaching me the rudiments of health economics.

To my family, thanks for providing me with a safe haven in the North. Thank you to my Dad, Debby, Grandad C, Dave, Michelle, Eve, Holly and Beth for getting me out in the countryside to stretch my legs on the hills and my mind with political discussions. Thank you to my Mum, Grandma, Grandad H, Saff and Jade for our occasional pub trips, and for keeping me updated with all the family goings on. Extra thanks to each of my grandparents for their unceasing generosity, and to my Dad for housing me in the final months.

Finally, I reserve my utmost thanks for my wonderful partner Lindsay, whom I met about a week before starting my PhD. Your reading of early drafts of my work (and somehow finding something nice to say), your calming me down when I am stressed, and your celebration of my achievements, however minor, have all been essential to my continued sanity and happiness, and for reaching the point of completing this thesis. Thank you!

List of Abbreviations

ACF	Active case finding
AIDS	Acquired immune deficiency syndrome
ARI	Annual risk of infection
BCG	Bacille Calmette-Guerin
CDC	Centers for Disease Control and Prevention
CI	Contact investigation
CXR	Chest X-ray
EPTB *or* ETB	Extrapulmonary TB
ETS	Enhanced tuberculosis surveillance
GP	General Practitioner
HIV	Human immunodeficiency virus
IBM	Individual based model
ICER	Incremental cost-effectiveness ratio
IGRA	Interferon gamma release assay
IPT	Isoniazid preventive therapy
LSHTM	London School of Hygiene and Tropical Medicine
LTBI	Latent TB infection
LTBR	London TB Register

MDR-TB	Multi-drug-resistant TB
MIRU-VNTR	Mycobacterial interspersed repetitive units, variable number of tandem repeats
NICE	National Institute for Health and Care Excellence
NNS	Number needed to screen
PHE	Public Health England
PTB	Pulmonary TB
QALY	Quality adjusted life year
SEIR	Susceptible – exposed – infectious – recovered epidemic model
SI	Susceptible – infectious epidemic model
SIR	Susceptible – infectious – recovered epidemic model
STI	Sexually transmitted infection
TB	Tuberculosis
TST	Tuberculin skin test
UCL	University College London
US *or* USA	United States of America
WHO	World Health Organization

List of Figures

Note that this list only includes those figures not included in a paper

List of Tables

Outline of thesis

Aim and objectives

The aim of this PhD is to describe TB contact tracing outcomes in London since 2010 and explore ways to improve it, using a combination of data analysis and mathematical modelling.

In order to achieve this aim, the thesis will address four objectives:

1. Quantify the status of contact tracing outcomes in London since 2012, and explore how these differ between population subgroups.
2. Quantify the proportion of cases found through contact tracing that are due to transmission between index case and contact.
3. Use a simple analytic model to evaluate the cost-effectiveness of screening contacts of non-pulmonary, non-laryngeal cases.
4. Explore the relative intensity of transmission from cases found through contact tracing compared to those found through other routes.

Rationale

In recent decades, England has had one of the highest notification rates of tuberculosis in Western Europe, and almost 40% of England's cases occurred in London throughout this period. Within this context, in 2015 NHS England and Public Health England published a 5-year collaborative TB strategy[1], with the overall aim of reducing TB incidence and eliminating TB as a public health problem. The strategy highlighted contact tracing as one of ten key approaches to reducing TB incidence, and focuses on areas in which incidence is highest, such as London. The strategy proposes two indicators (the proportion of cases with pulmonary disease that have at least one contact identified; and the proportion of identified contacts of pulmonary cases that are evaluated) for monitoring progress of contact tracing, but data for these indicators are not collected routinely country-wide; the first objective of this thesis is to estimate these indicators for London. It is also not

known the extent to which contact tracing identifies recent transmission, and estimating this constitutes the second objective.

In 2016, national guidance for tuberculosis care and prevention was changed. One change was to recommend only screening contacts of cases with pulmonary or laryngeal TB, whereas previously contacts of all cases were traced. However, there is evidence from analysis of the contact tracing data in London (see chapter 2) and elsewhere[2] that contacts of non-pulmonary, non-laryngeal cases were more likely to have TB than the general population. The third aim of this thesis is therefore to quantify the cost-effectiveness of screening these contacts to understand the potential impact of this change to guidance. An important unknown parameter determining this cost-effectiveness proved to be the relative number of new infections generated by those found through household contact tracing compared to those found through other routes. Estimation of this quantity forms the final objective of the thesis.

Layout of thesis

This thesis is a 'research paper style' thesis, meaning several of the chapters are publications in peer-reviewed academic journals. I have published one first-author paper, with two more currently in review – these constitute chapters 2-4. Chapter 5 is not currently written as a paper, but may be written up at a later date. There are six chapters in total, comprising the following:

1. Background and literature review: This chapter contains the following subsections: Natural history of tuberculosis; Tuberculosis epidemiology in London and the UK; Tuberculosis care and prevention in the UK; Tuberculosis Modelling.
2. An evaluation of tuberculosis contact investigations against national standards (first paper): This paper estimated the contact tracing indicators from the national strategy, and contact tracing yield, in London in 2012-15. It also estimated which population groups were associated with improved indicators or higher yield.

3. Description of secondary TB cases found through contact tracing in London (second paper): This paper estimated the proportion of contacts that were diagnosed with TB for whom probable transmission had occurred between index cases and contact between 2012-15. It also estimated the average timespan between when the index case was diagnosed and when the contact was diagnosed. It evaluated which population subgroups were associated with greater proportions due to recent transmission, or longer timespans.
4. Cost-effectiveness of screening contacts (third paper): This paper evaluated incremental cost-effectiveness ratios for the screening of contacts, separately for index cases with pulmonary/laryngeal disease and those without disease at these sites.
5. Estimating the rate of transmission from contacts using pairwise equations: The aims of this chapter are two-fold: firstly, to understand whether or not a pairwise modelling approach is feasible for tuberculosis; and secondly, to understand whether cases with reactivation disease generate more transmissions than those with disease following recent infection, due to their contacts being less likely to have been recently infected.
6. Discussion and conclusions.

1. Background and Literature Review

1.1 Natural history

Tuberculosis (TB) is an infectious disease most commonly caused by *Mycobacterium tuberculosis (M. tuberculosis)*. When this bacterium establishes in the lungs, the disease is known as pulmonary TB (PTB), and infection can then be transmitted through the air via exhalation of *M. tuberculosis,* for example, through coughing. TB in other organs is known as extrapulmonary TB (EPTB), and is generally not infectious[3] when not accompanied by PTB, although EPTB may be a sign of subclinical PTB disease. An exception is laryngeal disease, which is often assumed to be infectious, although even this is debated[4,5]. In some countries, including England and Canada, extrapulmonary TB is typically more common in those born in the Indian subcontinent[6–8] compared to other countries of birth, and in Europe, is associated with being aged <15 years[9]. The relationship of site of disease and sex is less clear, with extrapulmonary TB is positively associated with male sex in some studies[10], but in others with female sex[9].

The incubation period for TB is highly variable and is not well-defined. People who have been infected but who have not yet developed disease are sometimes referred to as having "latent *M. tuberculosis* infection" (LTBI). However, recent work has suggested this binary distinction is a false one, and instead describes tuberculosis as a spectrum from infection through subclinical to infectious stages, with patients moving both forwards and backwards along this spectrum during the course of their illness[11]. In the absence of HIV infection or prior preventive therapy, around 10% of those infected will develop disease at some point in their life, with the greatest risk in the first years after infection[12,13]. TB disease is often defined as "primary", "reactivation" or "reinfection"[14]. Primary disease refers to disease soon after initial infection. Reinfection disease is when an infected person becomes infected again subsequent to the previous infection before developing disease. In this case development of disease is thought to occur at a rate much greater than reactivation and lower than following primary infection due to some protection provided by existing infection[14,15]. An exception

to this is if the patient had developed disease prior to reinfection; in that case the risk of disease following reinfection may be higher than the average risk following primary infection[16]. Reactivation is defined as disease many years after infection or reinfection, possibly due to the patient becoming immunocompromised[17], or due to immunosenescence[18], and the rate of disease onset is much lower than for reinfection disease or primary disease[14]. The lifetime risk of developing disease following infection is non-linearly dependent on age: those in the 0-5 and 16-25 year age groups have the greatest risk, whereas those in the 6-15 year age group have the smallest (Figure 1)[19]. The test for infection used in figure 1 is the tuberculin skin test (TST), discussed in more detail in section 1.3. The lower risk in the oldest age-groups is in part due to the fact that those that are TST+ in this age-group will have been infected longer ago on average, and in part that they have less long to live after the skin test, so the lifetime risk is lower.

Pulmonary TB can be diagnosed using smear microscopy, during which the number of bacilli in sputum isolated from the case is counted. Whilst smear-microscopy is not very sensitive, it is highly specific[20] and is also useful as a correlate of infectiousness: smear-positive patients are typically around five times more infectious than smear negative ones[21]. This was determined by grouping tuberculosis cases into clusters whose isolates had indistinguishable DNA, and defining the first case in the cluster as the source case. The probable number of secondary case generated by smear positive source cases was then compared to that for smear negative source cases, after further assuming that for clusters with a smear negative source case, only those cases up-to and including the first smear-positive case were classed as secondary cases.

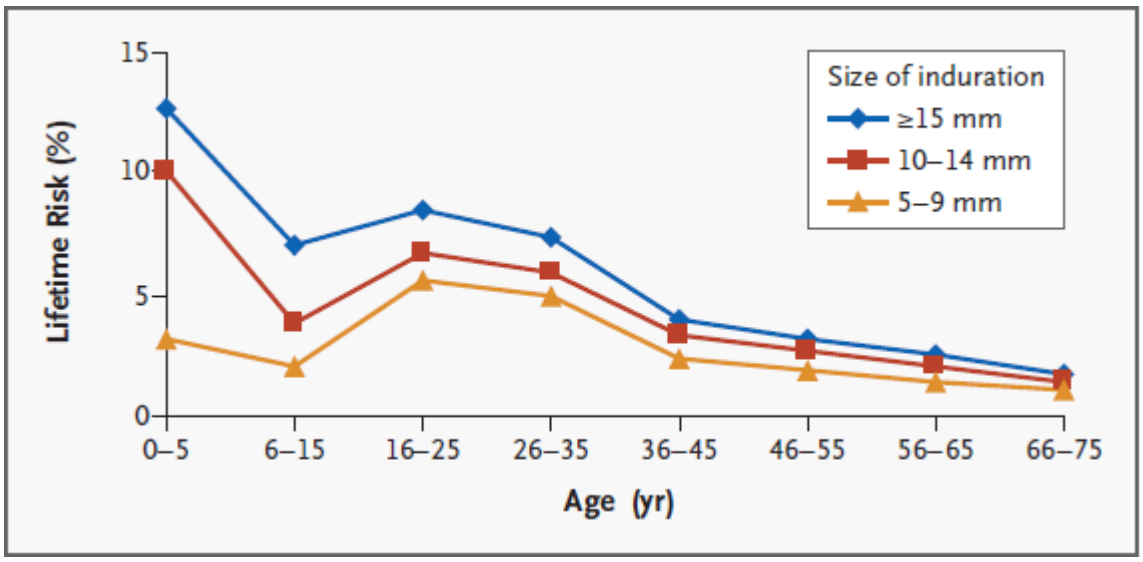


Figure 1: Lifetime risk of active tuberculosis among people with a non-conversion positive tuberculosis skin test (that is, the person did not have a negative TST in the last 2 years, and was not a household contact of someone with TB) (taken from Horsburgh[19])

1.2 TB epidemiology in the UK and London

1.2.1 UK

Excepting the years affected by the World Wars, the notification rate of TB in England and Wales declined almost continually from over 300/100,000/year in 1913 to around 11/100,000/year in 1995 (

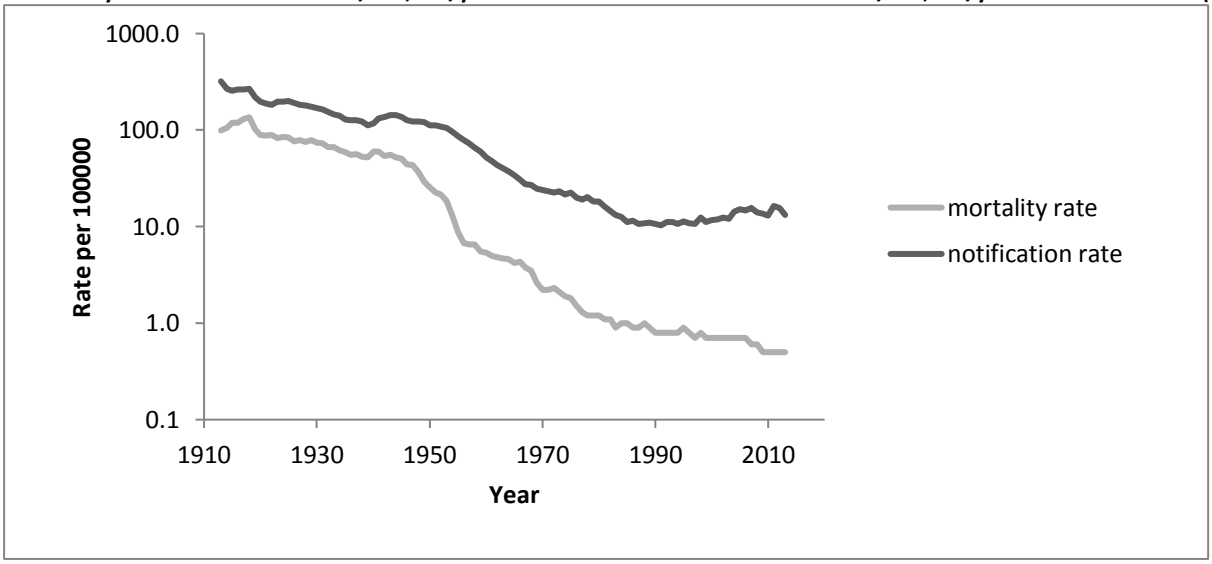


Figure 2). After that there was a slow increase, reaching 15/100,000/year in 2011, and prompting London to be described in national media as Europe's TB capital. In response to this situation, in 2015 PHE and NHS England released a collaborative national TB strategy for the period 2015-2020[1]. Since that peak, there has been a further decline; the case notification rate in England in 2015 was

10.5/100,000, the lowest it has ever been[22]. Over this period the proportion of tuberculosis cases with non-pulmonary disease has increased (Figure 3), perhaps due to an increasing proportion of non-UK born cases (see below).

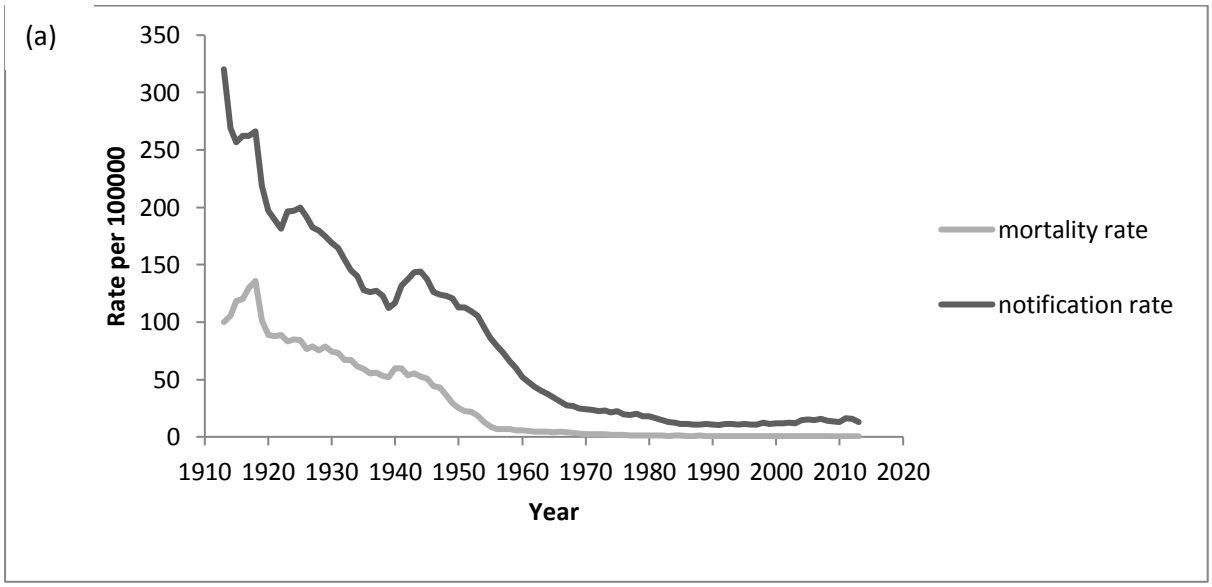
These historical reductions in TB notification rates have been attributed to many factors, including reductions in the number of effective contacts made by infectious patients, resulting from reduced crowding in living conditions, shortened duration of contact due to cases being admitted to sanatoria, improvements in hygiene, the advent of antibacterial drugs for TB and other factors[23].

England has a very heterogeneous geographic distribution of TB, with the disease largely concentrated in large urban centres such as London, Birmingham, Leicester and Manchester (Figure 4)[22]. There are also large inequalities in the demographic distribution of TB in England, and, as in other high-income countries, the incidence is largely influenced by cases in the foreign born population (in 2015, 72.5% of cases in England were non-UK born[22]). In particular, many cases are amongst immigrants from high-TB burden countries, especially the Indian subcontinent (47.4% of cases were amongst people born in India, Bangladesh and Pakistan in 2015) and sub-Saharan Africa, as these have much higher risk of infection in the birth- country than they would in the UK (Figure 5). In 2015, TB notification rates for those born abroad were 18 times higher than for those born in the UK[22]. This could be due to either reactivation of infection acquired abroad, greater local transmission than within the non-UK born group, or a combination of these two[24]. Between 2011 and 2015, the notification rate amongst those born outside the UK fell by over 30% (partly due to a decline in the proportion of migration from high-burden countries, such as India and Pakistan), whereas notification rates have remained lower, but relatively unchanged, amongst the UK-born[22]. This decline may also be linked to institution of pre-entry screening for active disease in the UK in 2012, and of LTBI screening for new entrants (both discussed in the next session).

Those with social risk factors (a history of homelessness, imprisonment or drug use) are also at heightened risk of disease[8], and are typically a group which has more ongoing transmission than the

general population[25,26]. These risk factors are particularly important amongst the UK born population.

Since 2010, the UK has used 24 loci mycobacterial interspersed repetitive units, variable number of tandem repeats (MIRU-VNTR) strain typing to determine genetic relatedness of TB strains. When typed isolates are indistinguishable, the cases are said to be clustered, and if cases are clustered and epidemiological links are present, then it is typically considered that transmission between the cases has likely occurred. In England between 2010-12, 75% of pairs of cases from the same household shared a strain, and hence were due to probable recent transmission[27].



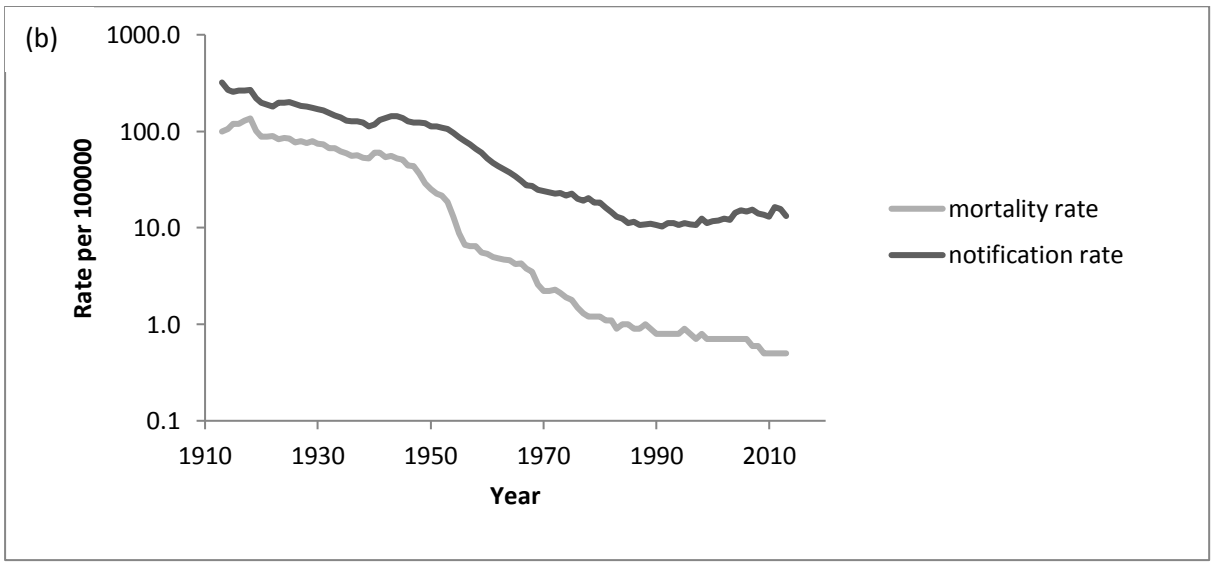


Figure 2: Annual mortality rate and case notification rate of all forms of TB in England and Wales since 1913. Figure (b) shows the same data on a logarithmic scale, highlighting that there have been some continued declines in mortality rates in the late 20th and early 21st centuries, but not in notification rates.[28]

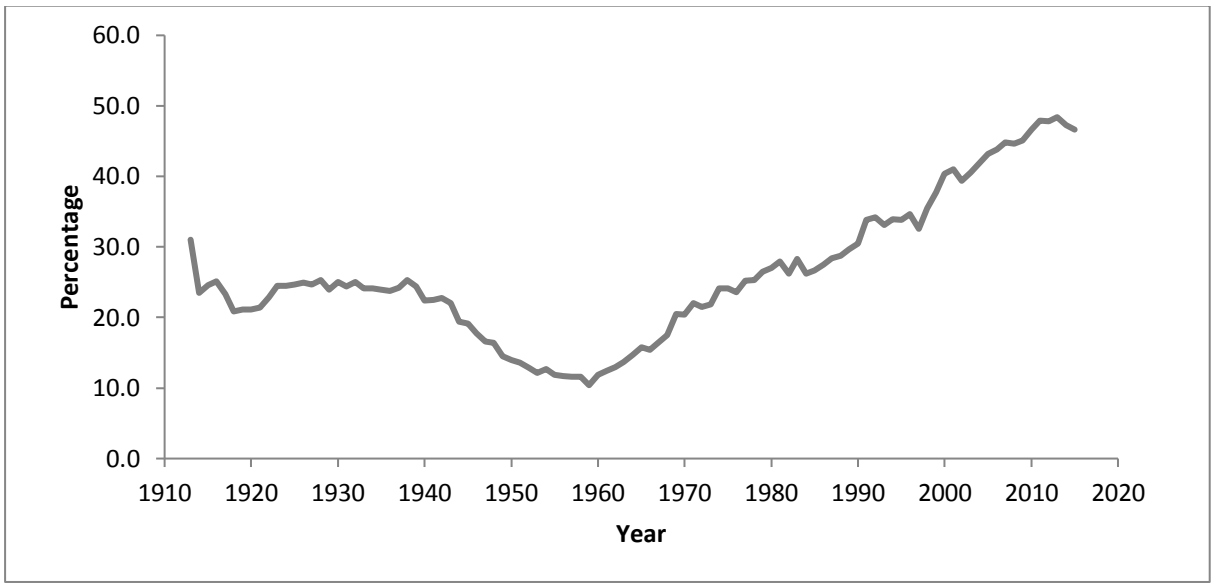


Figure 3: Proportions of cases in England and Wales with non-respiratory disease (1913-1981), or non-pulmonary disease (1982-2015)[28].

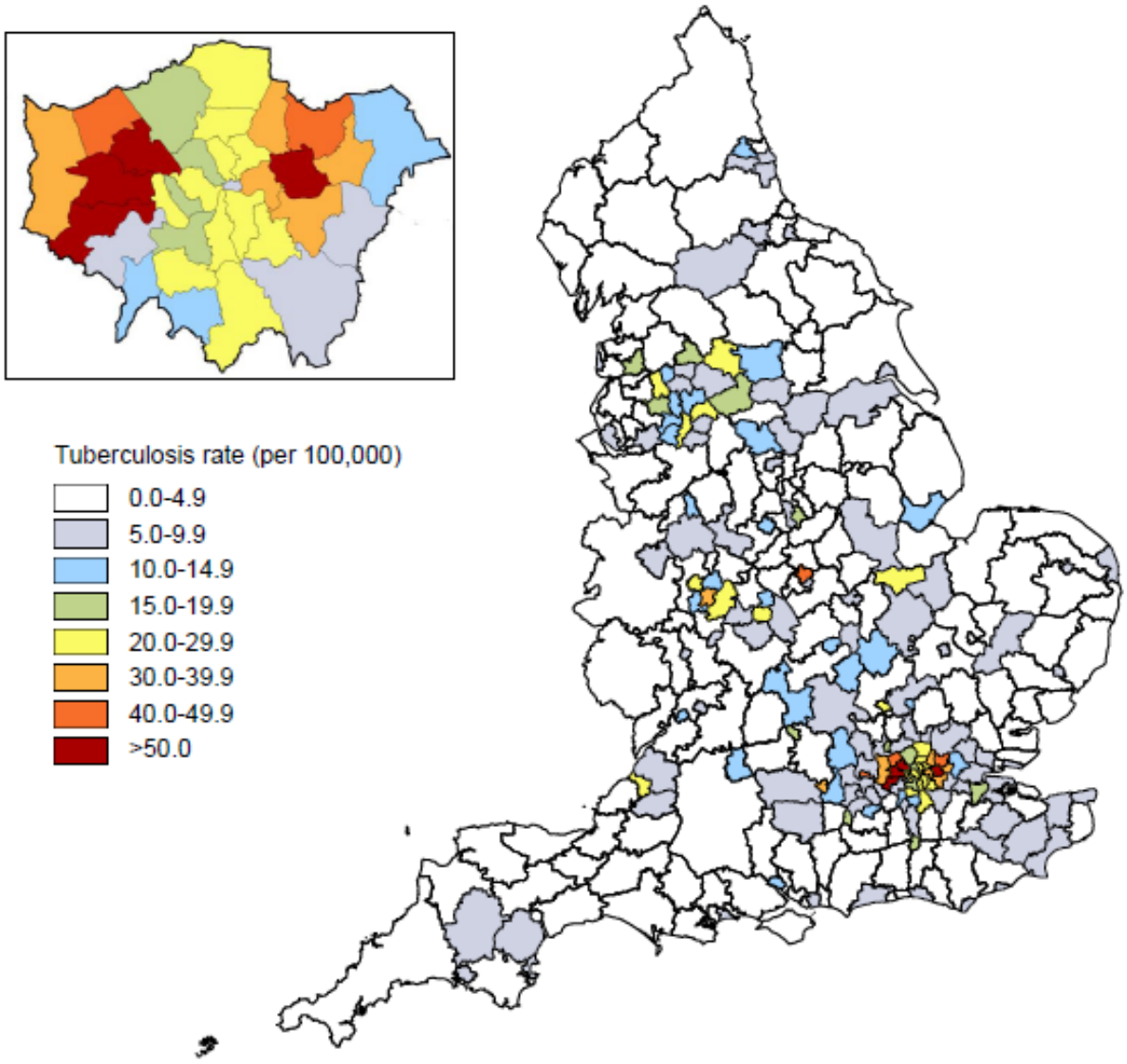


Figure 4: Geographic distribution of TB in England in 2013-2015. The inset shows London, over the same period[1].

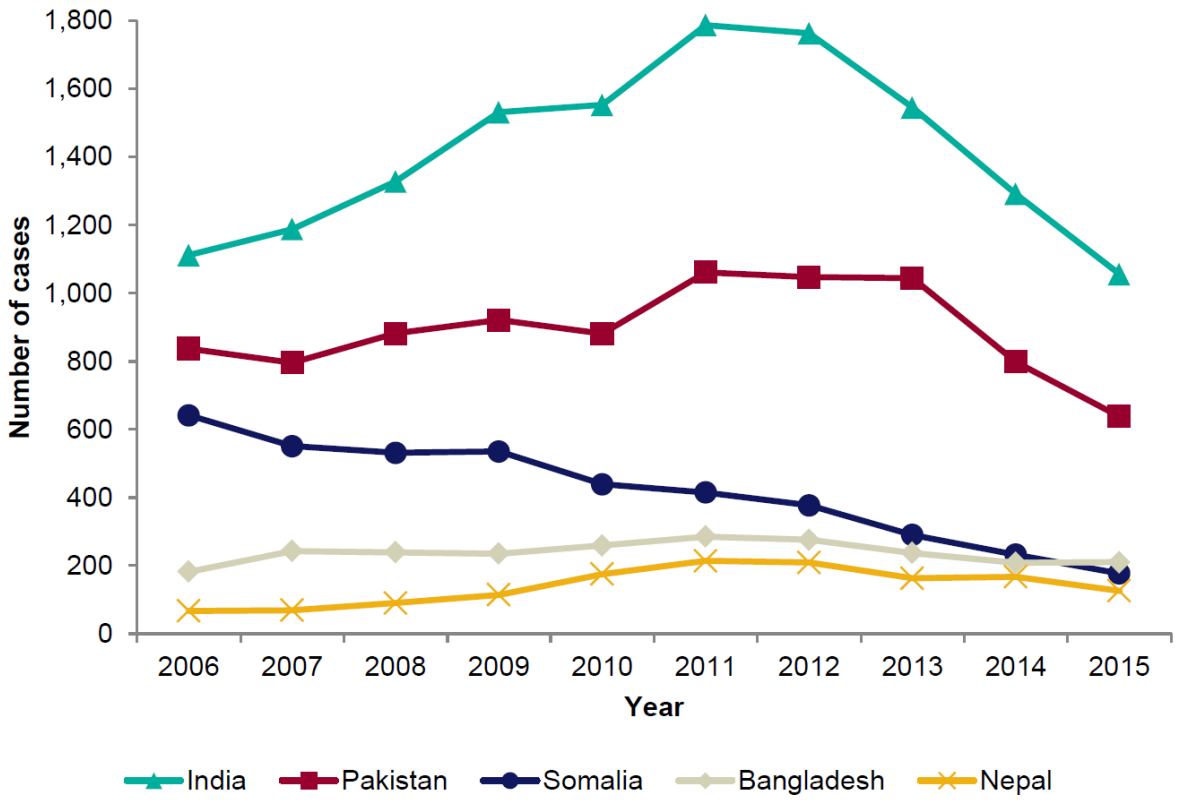


Figure 5: Numbers of TB cases (of all forms) by year in England from the five most common countries of birth, from the Tuberculosis in England, 2015 report[22]. The countries of birth with the highest rates were not provided in the report.

1.2.2 London

The trend in case notification rate in London in recent years has been similar to that seen nationwide: a decline since 2010 (Figure 6). However, the rate is much higher in London compared to the rest of England combined; in 2015 London accounted for 38% of TB cases in England, the notification rate was 26/100000, and one of its boroughs, Newham had a notification rate of 75/100,000 (Figure 4).

The proportion of TB cases with non-pulmonary disease is greater in the foreign born than the UK-born in London, particularly those from the Indian subcontinent (Figure 7), perhaps because of higher rates of reactivation in this group and the association between reactivation and non-pulmonary disease[10]. For example, whilst 36% of the UK born had non-pulmonary TB in this period, of those born in India and Bangladesh 66% and 69% respectively were non-pulmonary.

In London from 2010-2012, 46% of cases where clustered according to typing using 24-loci MIRU-VNTR, which means about 34% of cases are due to likely recent transmission[25]. This second value was calculated using the "n-1" method[29], that is, one case in each cluster is assumed to be a remotely acquired infection or reactivation, and the rest are attributed to recent transmission. Due to the resolution of MIRU-VNTR, matching isolates are only a proxy for recent transmission and ideally would be combined with epidemiological data for a deeper understanding of transmission[30], though this was not done in the London study. These estimates are also sensitive to the sampling fraction (a larger sample will give higher proportion clustered estimates[31]), the timespan of the study, and the variability of circulating strains[32].

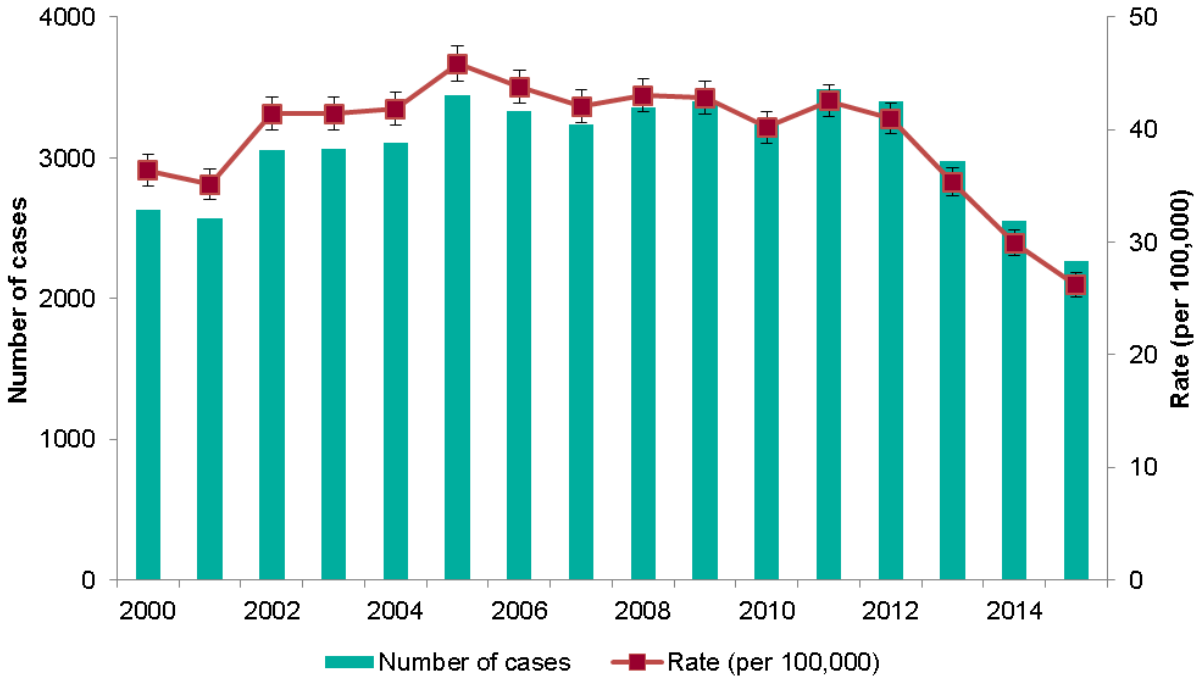


Figure 6: The number of cases of all forms of TB and case notification rate in London from 2000, taken from the London TB report 2015[8]

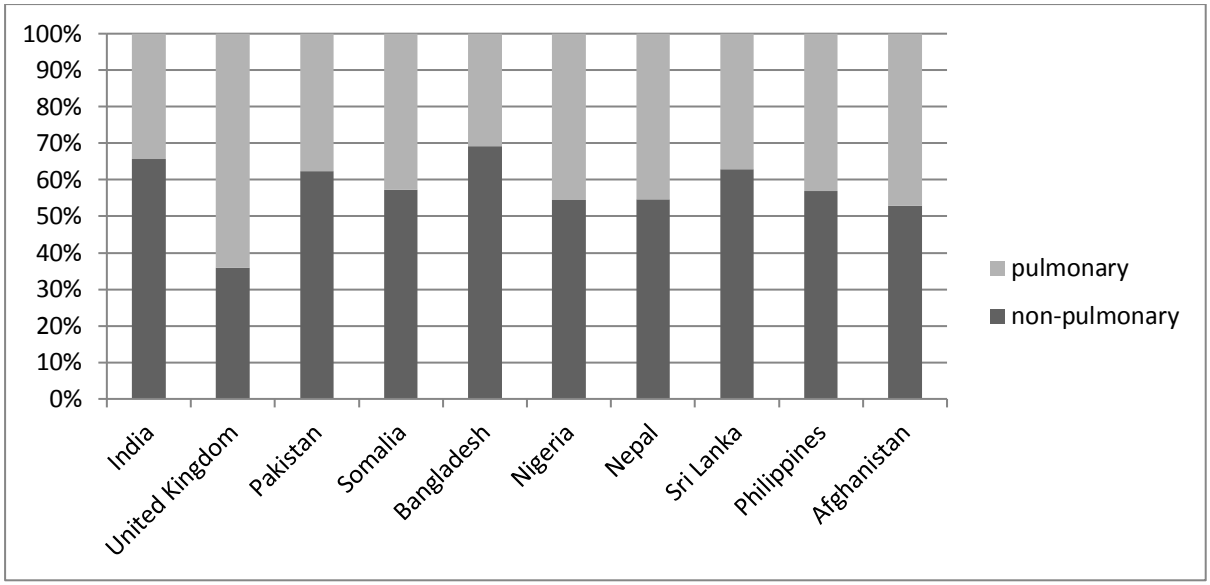


Figure 7: The proportion of cases with pulmonary or non-pulmonary disease by country of birth in London 2010-2015. Only countries with more than 300 cases in this period are included[8].

1.3 TB care and prevention

1.3.1 Vaccine and treatment

From the latter part of the 19th Century until the 1980s, tuberculosis incidence rates declined in developed countries, including the UK. The first part of this decline has been largely attributed to improvements in living conditions and nutrition, and the subsequent decline followed the discovery of anti-TB drugs, starting with Streptomycin in 1943[33]. Prior to this, in 1921, the Bacillus Calmette-Guerin (BCG) vaccine was discovered, and has played a key role in UK TB care and prevention since 1953 when the UK began vaccinating schoolchildren at around age 13 years[34]. The efficacy of BCG varies across the globe[35,36], and was never introduced in certain countries, including the Netherlands and the United States[37]. Reasons for this variation include prior *M. tuberculosis* infection, sensitisation to environmental mycobacteria, and latitude (the efficacy is higher at higher latitudes)[38,39]. Consequently, not all countries have used the vaccine. Perhaps related to the first two of these reasons, the vaccine is also typically more efficacious in infants than older individuals[40]. It has also been observed that protection provided by BCG wanes over time, with protection typically lasting up to 10 years, although some studies have seen protection up to 20 years for school-age

vaccination[40]. This waning is faster in lower latitudes than higher latitudes. A study started in the 1950s showed BCG to have around 80% efficacy in schoolchildren in the UK[41], and it has been shown to be efficacious in preventing severe childhood disease[36]. This led to the policy of vaccinating schoolchildren at age 13 years in the UK, but declining prevalence and incidence have reduced the benefit of BCG amongst schoolchildren[42]. This resulted in the current policy of offering BCG to neonates living in high-incidence (>40/100,000/year) parts of the UK or with relatives from high incidence parts of the world[43].

In the UK, treatment of active disease is with the standard regimen of 2 months of Isoniazid, Rifampicin, Pyrazinamide and Ethambutol, followed by 4 months of Isoniazid and Rifampicin[43]. This period is longer if the patient has central nervous system involvement or drug-resistant TB, particularly if they have Rifampicin-resistant TB. Latent TB infection (LTBI) is treated with 3 months of Isoniazid and Rifampicin or 6 months of Isoniazid, and is offered to contacts of drug-sensitive TB cases aged under 65 years and for whom hepatotoxicity is less of a concern[43].

1.3.2 Contact tracing

Contact tracing (the screening of people exposed to a case of active case of TB for signs of active disease and *M. tuberculosis* infection) has been highlighted as a key element of TB care and prevention in the PHE/NHS England collaborative national TB strategy[1], and is the focus of this thesis. Contact tracing aims to:

1. Reduce morbidity and mortality by finding active TB cases earlier;
2. Arrest further transmission, by finding secondary cases more quickly;
3. Prevent future cases by finding and treating those with latent infection[15].

Contact tracing practice for TB differs between high-income and low/middle-income countries, because of differences in the availability of resources and the prevalence of TB and LTBI. Additionally, high incidence countries may have different aims from the UK (e.g. reduced focus on

finding latent infection). For these reasons, this review focuses on high-income, low-incidence countries[44].

In the UK, a contact investigation (CI) is triggered once a case of laryngeal or pulmonary TB, known as the index case, is diagnosed. At this point the index case is interviewed and asked to name all of their close contacts (defined as those sharing a bedroom, kitchen, bathroom or sitting room, and other people with whom they share contact equivalent to this)[43]. Additionally, sometimes the interview is followed up with a home visit, when a TB nurse visits the home of the index case. The named contacts are invited to screening, which varies by clinic and by age of the contact: contacts aged >65 years are typically given a chest X-ray (to diagnose active disease), and contacts aged <65 years either a TST or an interferon-γ release assay (IGRA), both of which test for infection. This difference between age groups is due to the fact that those >65 years are not eligible for preventive therapy as it is contraindicated in this group. Until recently, the age cut-off for whether infection was tested for was 35 years, but this changed at the start of 2016[43]. Within London, whilst all clinics follow the same guidance, there are some small differences in practice: for instance, some clinics use the T-SPOT.TB test for latent infection, whereas others use QuantiFERON-TB Gold, and some clinics undertake home visits to index cases, whilst others don't.

In many countries, including the UK, variants of the 'stone-in-the-pond' principle is used[45], in which contact investigations are extended if the prevalence of infection is higher than expected amongst close contacts. The exact implementation differs by country. In the UK, household and other close contacts are investigated first and then the investigation is extended to casual (e.g. workplace) contacts if: there is evidence of infection (greater than 10% of contacts infected); the index case is young and the source of infection has not been found; or, there are particularly susceptible casual contacts of a smear positive index case[46].

Policies on screening contacts of non-pulmonary, non-laryngeal cases vary between countries[44,47]. The UK has recently made a change to screen contacts of pulmonary or laryngeal cases only; previously contacts of all cases were screened[43]. An exception to this is if the index is still at school.

There is disagreement amongst UK based epidemiological studies over whether screening contacts of extrapulmonary patients is worthwhile or not, with Rubilar *et al.* finding a lower yield from EPTB index cases (0.029 cases of active disease or contacts with LTBI per index case, in Edinburgh) than Saunders *et al.* (0.096, same units, in Birmingham – see Table 1), and Mandal *et al.* finding higher yield (0.26 same units, also in Edinburgh)[2,48,49]. This is in part due to differences in classifying LTBI between the studies, with Rubilar *et al.* only reporting contacts who began chemoprophylaxis, and the other two studies reporting all those with positive TST or IGRA, without reporting the number of contacts given these tests (in the Rubilar *et al.* study, all contacts were offered Heaf tests, but the number who declined is not stated. The Heaf test is a form of tuberculin skin test used in the UK until 2005, when it was replaced by the Mantoux test).

In 2010 London introduced “cohort review” into its TB program. This process was first implemented in New York City in 1993, where it was considered to be a key driver of huge reductions in TB notification rates[50]. Cohort review involves a quarterly appraisal of both case management and contact investigation of each TB case, in which representatives of TB clinics meet to discuss specific outcomes. In London it was associated with an increase in the proportion of cases who had contacts being identified (77% of cases had at least one contact identified prior to cohort review, versus 86% thereafter) and contacts assessed (74% prior to cohort review versus 81% thereafter), and cohort review has subsequently been implemented across England[51]. As a direct consequence of the recording of cohort review, detailed data on contacts in London has been routinely collected since 2012/13 (the exact date varies by clinic) in the London TB register (LTBR). Cohort review has now been expanded to the rest of the U.K., and so data on contact tracing outcomes is improving, although this data is not routinely reported.

Table 1 (calculated from Saunders et al.[2]) Summary of outcomes and yield (new infections & cases per index case) of contact investigations by site of disease and smear status, from a twenty-one-year retrospective cohort study in Birmingham, 1990-2010. Here LTBI is defined as either a positive TST or IGRA, and no diagnosis of active disease. Number needed to screen (NNS) is the number of contacts screened divided by the number of new cases (equivalently, the reciprocal of the proportion screened positively). Both first- and second-ring contacts are included. It is not clear what the background prevalence of LTBI is. It is also unclear what proportion of contacts were tested for latent infection, although 11% of contacts were screened using IGRA.

	Number of index cases	Total Number of contacts	Number (%) of contacts with LTBI	Number (%) of contacts with TB	Contact tracing yield (contacts with TB or LTBI per index case)	NNS to find one case of TB	NNS to find one case of either TB or LTBI
Smear-positive PTB	1542	16034	1044 (6.51%)	532 (3.32%)	1.02	30	10
Smear-negative PTB	3514	15335	440 (2.87%)	181 (1.18%)	0.177	85	25
EPTB	2309	9875	165 (1.67%)	57 (0.58%)	0.0961	173	44

1.3.3 Other forms of screening and case finding

Since 2012, the UK has also operated a pre-entry screening system, screening immigrants from countries with an incidence of >40/100,000/year as part of visa applications[52]. Prior to this, a combination of pre-, post- and at entry screening had been used[53]. In the period since 2012, the incidence of TB amongst the non-UK born population has declined, although there is not yet evidence that this is caused by the pre-entry screening programme. In addition to the pre-entry programme, the UK has begun to operate a system of latent TB screening for new migrants from very high burden (>150/100000/year) nations[54]. This was introduced after it was shown to be cost-effective in a trial period, costing £20819 per case averted[55]. Additionally, since 2005 in London, an active case finding operation known as Find & Treat has operated, seeking to find cases of active TB amongst the homeless and other marginalised populations of the capital. This operation has also been shown to be highly cost-effective, costing less than £10000/QALY (Quality adjusted life year) gained[56].

1.4 Modelling

1.4.1 *M. tuberculosis* transmission models

Mathematical models of infectious disease have been used for over a century, since Ronald Ross used a simple model to understand Malaria transmission in 1910[57]. In that period, models have been used for many types of questions, which can broadly be grouped into at least three categories (adapted from Knight et al[58]):

1. Projecting impact: both predicting the trajectory of an epidemic[59,60], and predicting the possible impact of control strategies[61,62];
2. Elucidating the natural history and epidemiology of infectious diseases[14,63];
3. Driving empirical research by revealing gaps in data.

In each of these cases mathematical models are useful as they allow exploration of counterfactuals and different scenarios, which may have been expensive, unethical or impossible to do empirically, and synthesis of different data sources.

The first published tuberculosis model may have been that of Waaler et al. in 1962[64]. This paper made forecasts of future prevalence trends in India in order to make the argument for greater use of mathematical models in epidemiology. Since then tuberculosis models have been used to understand TB natural history[14,65], to examine the impact of vaccination[66,67], to explore the interaction between TB and HIV[68–70] and the emergence of drug-resistant TB[71,72], and to understand the impact of contact tracing (see below), amongst other things.

Many different structures have been used to model the transmission dynamics of *M. tuberculosis*, and there are at least two reviews detailing these[73,74]. Many of these are variants of the susceptible (or uninfected) – exposed (or latent) – infectious – recovered (SEIR) model, which have been modified to describe TB's complex natural history. Most compartmental models include at least the following compartments: Uninfected, Latently infected, Infectious, and Recovered[74]. Due to the fact

that a majority of those who will progress to disease will do so in the first 2 years after being infected[15], most models incorporate both an early and a late latent stage. Due to the increased infectiousness of smear positive cases, models also often split the Infectious category into smear positive and smear negative groups[14,75]. Decisions about these and other factors depend on what the model will be used for.

As an example, figures 8, 9 & 10 show model structures of two older tuberculosis modelling papers and one more recent study, those of Vynnycky and Fine[14], Dye *et al*[76] and Lin *et al*[75] respectively. These papers represent a range of different approaches to TB modelling, for a range of different purposes. The first of these was designed with the objective of estimating the age-specific contributions of primary disease, disease resulting from reinfection and disease resulting from reactivation, and hence contains compartments for each of these disease states. The second aimed to investigate the effect of directly observed therapy on TB control, and how this interacts with the HIV/AIDS epidemic. The third model aimed to assess the impact of a new diagnostic tool, and so includes a more detailed sub-model of the diagnostic pathway to elucidate the effect of the new tool on different stages of the pathway.

It should be noted that TB models often do not explicitly include EPTB; of those shown here, only the one by Dye *et al.* does.

1.4.2 Previous TB contact tracing modelling studies

To date, nine modelling papers have studied contact tracing of TB, none of which considered a UK setting (see table 2). The first 5 of these were included in a systematic review[77]; since then, four more papers have been published. Four of these studies focus on a USA setting (two of which have the same authors), three in Canada (two with the same authors), one in Africa and one purely theoretical. Apart from Guzzetta *et al.* (2011)[78] they have each tried to quantify the potential/observed impact of contact tracing or preventive therapy, with varying degrees of detail. Only three of the studies explicitly incorporated extrapulmonary TB: Tuite et al.[79], Guzzetta et al.

(2014)[80] and Aparicio and Hernandez[81]. Tian *et al*[82] considered issues and settings which were most similar to those in this project. The study was based on a region in northern Saskatchewan (Canada), in which 90% of the population are First Nations (an ethnic grouping amongst whom the burden of TB is the greatest in Saskatchewan). This is a high incidence region (notification rate of >100 cases/100,000 per year) in a low incidence country, with notification rates similar to those in the London boroughs of Brent or Newham, though over a much greater area and smaller population. They investigated several different scenarios, for example varying how contacts are prioritized and the speed of the contact tracing process. They did this using an elaborate age- and ethnicity-structured individual-based model (see below). They found that tracing younger contacts (those aged younger than 10 years) first brought significant decreases in incidence, but increasing the speed of contact tracing did not.

The model structures used in these papers varies from a simple SEIR model[83], to a very complex structure[82]. All of the most recent of these studies use individual-based models (IBMs)[78–80,82,84]. The latter type of model explicitly models each individual in the population, by determining stochastically at each time-step whether or not that individual moves to a new disease state. This means that the models are intuitively closer to reality than simple deterministic models, and can give an indication of the uncertainty in model output due to the chance nature of transmission. However, due to their stochastic nature, they must be run multiple times leading to longer run times and results which can be harder to interpret than those of simpler models[85]. Both simpler models and IBMs can represent uncertainty in parameters by running the model with a range of parameter values. The Guzzetta model is the most complex of these models and includes a complex geography in order to explicitly model community and workplace transmission. The Mellor paper uses a discrete-event simulation model, meaning transmission events are generated from a Poisson distribution[86].

1.4.3 Approaches to modelling contact networks

If we imagine *M. tuberculosis* transmission occurring in a network of contacts, and that contact tracing follows the chain of transmission across this network, then the most intuitive way to model contact tracing would be to create a contact structure which mimics typical contact patterns in London. An IBM would then operate on this contact structure, similar to that seen in the Guzzetta et al papers[78]. Modelling of other diseases has also taken this approach, such as Lum *et al.*[87], which models incarceration rates as a disease using a synthetic social network based on US demographic data. One difficulty with this approach arises when trying to create a realistic contact structure, including the level of clustering (in a social network context, taken to mean the proportion of all sets of three contacts which are mutually connected), or equivalently, the proportion of people who share the same contacts. Data from two studies, namely the POLYMOD study and a study by Danon et al, may help with reproducing a realistic model of contact structure[88,89]. The first of these tried to describe contact patterns in a range of European countries, and the other just in the UK. Both studies asked participants to create diaries of their daily contacts. However, these two studies give very different results; for instance, Danon *et al.* reports approximately 27 contacts per day, and the Polymod study 12 per day. Also, as the relationship between the amount of contact and *M. tuberculosis* transmission risk is not well understood, the interpretation of these data in the context of TB is not straightforward. At least one previous TB model was based on a synthetic contact network[90], and whilst it did not explore contact tracing, it is possible the approach could be extended.

Instead of explicitly modelling the network structure, some theoretical studies of diseases, other than TB, have used pairwise equation models[91–93]. These types of models use an equation set which describes how the numbers of pairs of contacts in different disease states changes with time. Eames (2008)[92] adapted this approach in order to incorporate clustering and also used the model to look at contact tracing. Due to TB's complex natural history, the size of the equation set is greater than for

simpler diseases (e.g. measles), due to the larger number of disease states typically used in the model.

Another way to model elements of contact structure using a deterministic or partially stochastic model is through household, or meta-population, modelling. These involve dividing the population into subunits in which the contact rate is different from that in the population overall. These subunits are typically households, but any subunit, for instance workplaces, is possible. Transmission within subunits is then described separately from transmission outside subunits. Meta-population models provide a balance between relevant details, and ease of computation and interpretation[94–96].

1.4.4 Economic evaluations

Cost-effectiveness analyses enable comparison of public health interventions to inform decision making. This is often defined in terms of an incremental cost-effectiveness ratio (ICER), which describes the ratio of change in cost to change in effectiveness ($\mathrm{ICER} = \frac{C_1 - C_0}{E_1 - E_0}$), with effectiveness often defined in terms of cases averted, lives saved or Quality Adjusted Life Years(QALYs) gained/Disability Adjusted Life Years (DALYs) averted. One QALY is equivalent to a year lived in perfect health, and they are calculated by multiplying the time spent in a given health state by the utility of that health state (which is always <1 except for perfect health). DALYs measure the burden of a disease and are defined as the sum of the years of life lost (YLL) and the years live with disability (YLD). The YLL is the number of deaths cause by the disease multiplied by the difference between the maximum expected life-expectancy (i.e. life-expectancy in the country with the longest life-expectancy) and the life-expectancy of someone with the disease. YLD is the product of the number of incident cases, the average time spent with the disease, and a utility weight. DALYs and QALYs are based on similar ideas, but the where the latter is an individualised measure of the effect of an intervention, the former provides a population-level measure of the burden of disease. Another difference is that higher QALY values equate to a healthier individual, whereas for DALYs, smaller values equate to a healthier population. In the UK, the National Institute for Health and Care

Excellence (NICE) produces guidance on clinical practice, much of which is underpinned by economic analyses. In NICE guidance, interventions which cost less than £20000-£30000/QALY gained are taken to be cost-effective[43]. In the UK, the NHS provides a list of reference costs which can be used when considering healthcare costs from the perspective of the NHS[97], and the British National Formulary produces a list of medicine costs[98].

As aforementioned, in recent years two important case-finding strategies in the UK have been informed by cost-effectiveness analysis: Pre-entry screening of migrants[55], and the Find and Treat service in London[56]. The first of these was informed by a multi-centre cohort study and used a simple decision tree model to explore the cost-effectiveness of different incidence thresholds for screening, and expressed results in terms of costs per case averted. The second used a simple compartmental model informed by data collected by the Find and Treat team. Neither included a transmission effect (i.e. they did not calculate the cases prevented, QALYs gained or costs saved by reducing transmission). Some economic evaluations are informed by transmission models where the transmission effect is likely to be an important part of the overall impact[99]. At least one previous study has attempted to evaluate the cost-effectiveness of contact tracing[100]. This study found that close contact screening of pulmonary TB patients in Montreal was actually cost-saving, i.e. costs were lower with contact tracing than without, and so determined contact tracing to be highly cost-effective in that setting. No studies have estimated the cost-effectiveness of contact tracing in the UK as a whole or any parts of the UK.

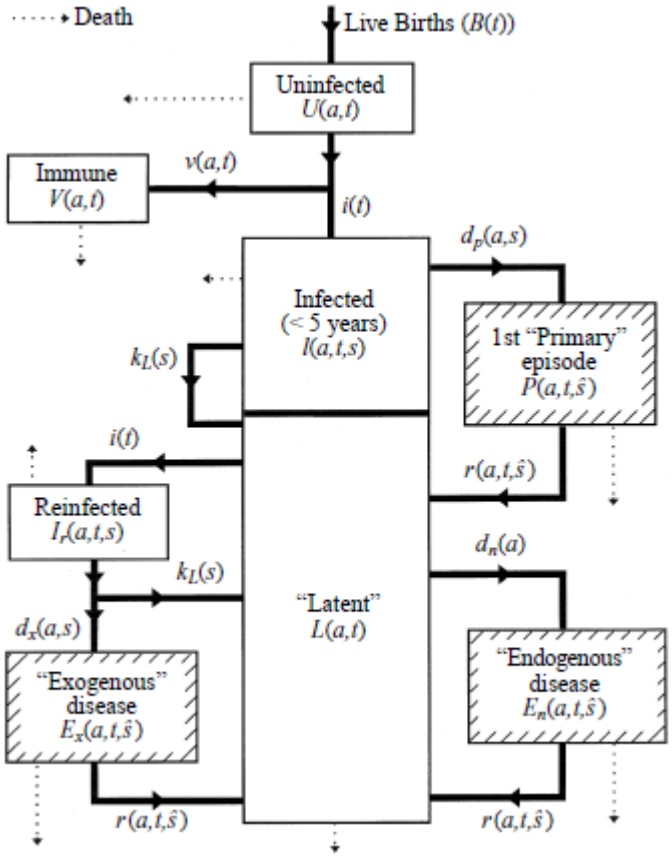


Figure 8: *M tuberculosis* transmission model structure from Vynnycky and Fine[14]. Note that infectious cases are stratified by smear status, although this is not included in the model diagram.

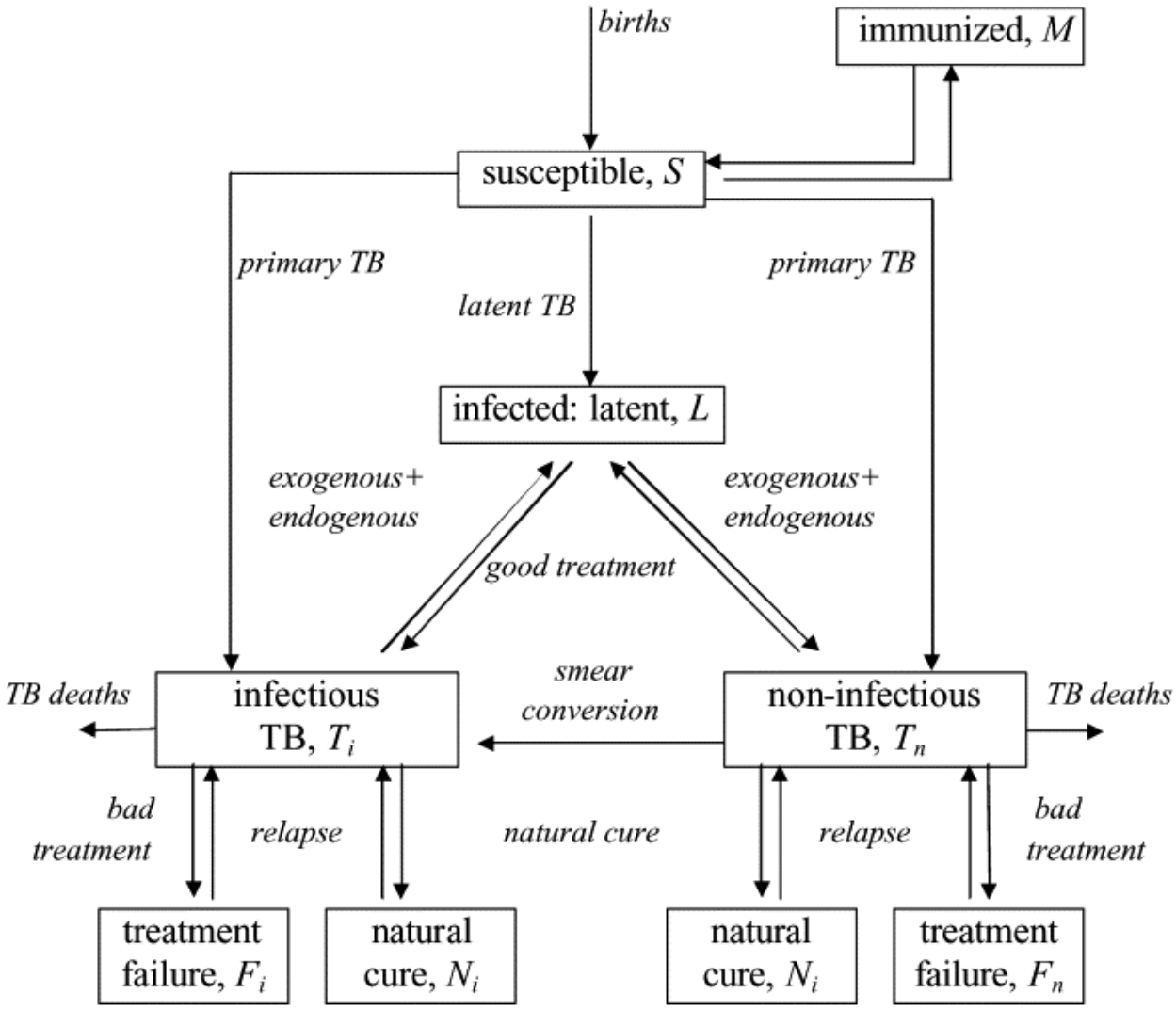


Figure 9: *M tuberculosis* transmission model structure from Dye *et al.*[76]

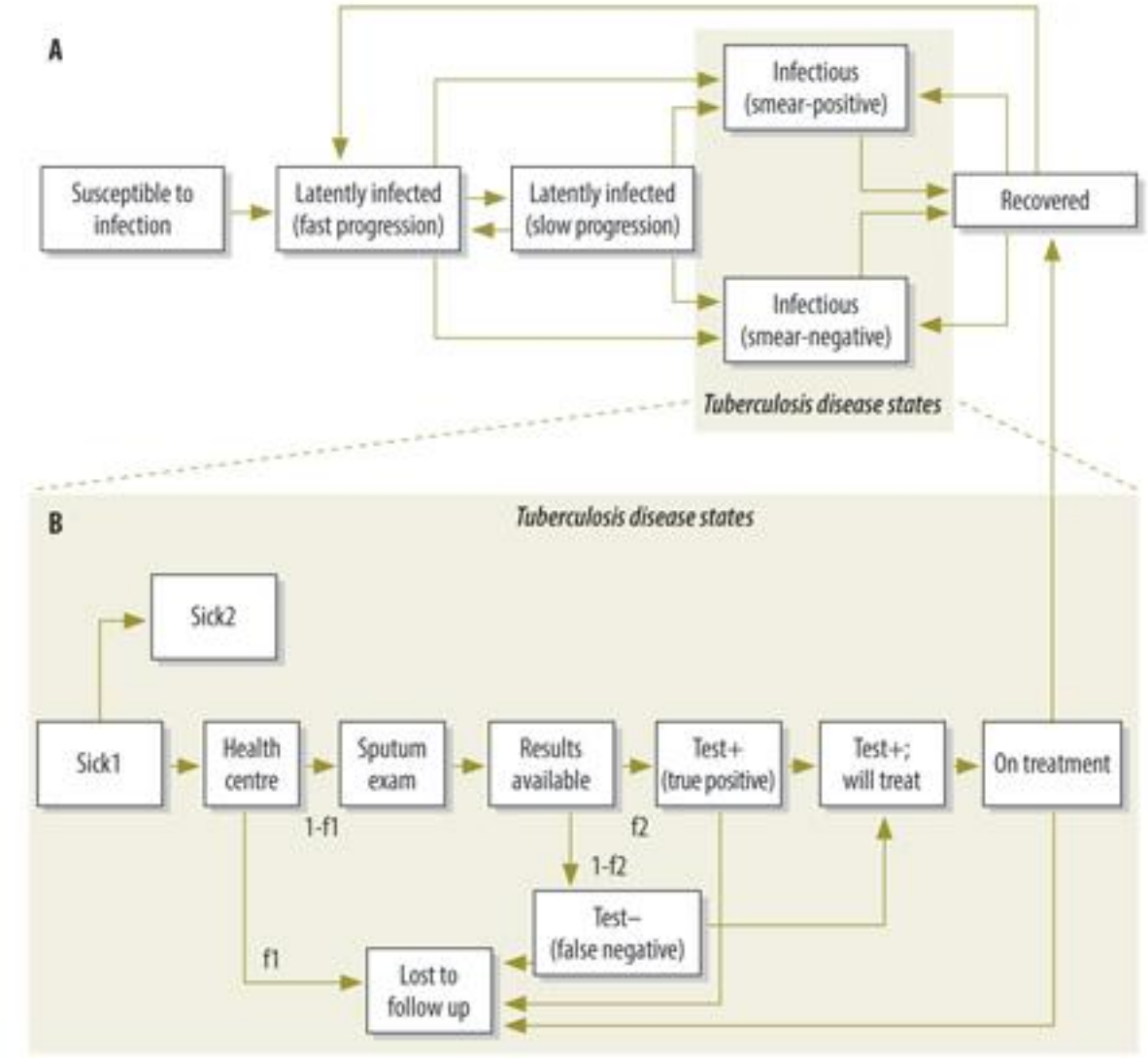


Figure 10: *M. tuberculosis* transmission model structure from Lin *et al.*[75]

Table 2 Summary of some features of all of the previous TB modelling studies that have either incorporated contact tracing, or discussed doing so (adapted from Begun et al.[77], with additional papers identified since review):

Author	Tuite et al.[79]	Guzzetta *et al.*[80]	Kasaie *et al.*[84]	Tian *et al.*[82]	Guzzetta *et al.*[78]	Mellor *et al.*[86]	Tian *et al.*[101]	Aparicio and Hernandez[81]	Ziv *et al.*[83]
Year	2017	2015	2014	2013	2011	2011	2011	2006	2001
Type of model	Individual based	Individual based	Individual based	Individual based	Individual based	Discrete event	Systems Dynamic	Compartmental	Compartmental
Stochastic/ deterministic	Stochastic	Stochastic	Stochastic	Stochastic	Stochastic	Stochastic	Deterministic	Deterministic	Deterministic
Region	Canada	USA	USA	Canada	USA	Africa	Canada	USA	Theoretical
Low/high prevalence?	Medium – High	Low	Medium	Medium	Low	High	Low	Low	N/a
Age-structured	Yes	Yes	No	Yes	Yes	Yes	No	No	No
Explicit extra-pulmonary TB	Yes	Yes	No	No	No	No	No	Yes	No
Population groups considered	Inuit population of the territory of Nunavut	Different work places and education settings incorporated.	Homogeneous, but with household structure incorporated.	90% First Nation, 10% other aboriginal.	Different work places and education settings incorporated.	Age- and Household structure incorporated.	Homogenous population.	N/a	N/a
Background incidence	~200/ 100,000 per year.	~5/ 100,000 per year.	~120/ 100,000 per year.	unknown - but northern Saskatchewan incidence is high ~150/	~5/ 100,000 per year.	~400/100,000 per year at start of experiment,	Unknown.	Decreasing exponentially from 54 /100,000 in 1950 to	250/100,000

				100,000 per year.		declining over time.		4.4/100,000 in 2000.	
Key questions explored	To evaluate which interventions (including contact tracing) would benefit TB control in Nunavut	To use a previously developed model to assess contact tracing success in Arkansas, USA.	To derive maximum contact tracing impact in both household and community scenarios.	Examine a range of modifications to contact tracing protocol: age- and ethnicity-prioritization; increased speed; reduced loss to follow up; greater tracing extent.	To describe an age-structured IBM, with evolving contact structure and spatial map.	To compare CT with other Active Case Finding (ACF) approaches in high-risk groups, in this case HIV+ individuals.	Impact of tracing a. more contacts and b. tracing more quickly.	To develop a model which examines the effect of preventive therapy on TB incidence.	- What is the effect on incidence of treating early latent infection? - How effective does this treatment have to be to eliminate TB?
Contact tracing implementation	Direct simulation. Only household contacts screened. Those with LTBI offered preventive therapy.	Direct simulation. Screening of both smear negative and positive index cases, include non-household contacts.	All household contacts traced, but no others.	Direct simulation. Individuals have on average 60 contacts, and tracing fraction is varied.	No contact tracing implemented.	Direct Simulation.	CT represented as transition rate between un-investigated and investigated states.	CT represented as increased rate of LTBI treatment.	CT represented as increased rate of LTBI treatment.
Key findings	Incidence projections	CT prevents 20% of TB	CT effective if	CT suffers from	The model could	Targeting TB control at	Increasing the number	Significant reductions in	This sort of treatment

	were variable in absence of interventions. Rapid contact tracing and school-based screening had limited effectiveness.	cases and deaths. It's important to trace contacts of smear negatives in this context.	achieves large population coverage. CT effects on incidence lagged by ~3 years.	Diminishing returns. Prioritizing based on ethnicity and age of contact led to particularly strong incidence reductions.	feasibly be extended to include contact tracing; more data required to estimate household transmission accurately.	HIV+ could be cost-effective; more work needed on social network modelling.	of contacts traced had a larger effect than increasing speed but both suffered from diminishing returns.	incidence, even with interventions treating as few as 5% of recent infections.	can have a large effect.

1.5 Summary

The national TB strategy, published by Public Health England in 2015 in response to relatively high incidence rates for a high-income country, forms the focus of the first part of this thesis[1]. The strategy highlighted contact tracing as one of ten key areas for action, and proposed two indicators for monitoring progress on contact tracing. In chapter two we aim to quantify these indicators for London. We focus on London because 40% of TB cases in England are in London[8], and because cohort review was initiated in London before other parts of the country[51], a corollary of which is that there is good data on contact tracing in the capital.

In chapter three, we aim to estimate the proportion of cases found through contact tracing that are due to transmission, and also to estimate the typical time which contact investigations take in London between index case and contact accessing care. A similar figure to the former of these was recently estimated at 75% in the UK[27], though this study was based on postcodes rather than contact investigations. Contact investigation times have not previously been estimated in the UK. As TB incidence in London is largely amongst the non-UK born and those with social risk factors[8], these things are potential important risk factors for differential contact tracing outcomes and heightened transmission, which we examine in chapters two and three.

The third chapter focusses on the question of whether we should screen contacts of non-pulmonary cases, and was precipitated by a change in national guidance to no longer recommend screening these contacts[43]. Whilst we have seen that these contacts are generally not infectious, various studies have found that the proportion of screened contacts of non-pulmonary cases that have TB is quite high relative to the prevalence of tuberculosis in the population, including chapter two of this thesis[2,102,103]. We undertook a cost-effectiveness analysis to understand whether the change in policy was justified. This was the first cost-effectiveness analysis of contact tracing in the UK.

The fifth chapter describes the design of a pairwise TB model. The primary aim of this chapter was to understand the feasibility of this modelling approach for modelling TB, a disease which has not been specifically addressed by a pairwise approach previously. A secondary aim was to quantify whether cases with reactivation disease generate more new infections than those with disease following a recent infection, a question arising out of chapter 4 of the thesis.

2 First paper: An evaluation of tuberculosis contact investigations against national standards

Preamble

As aforementioned, Public Health England and NHS England published a Collaborative TB Strategy in 2015[1], that highlighted contact tracing as an area for action and proposed two indicators to monitor progress in this area. These were:

1. The proportion of pulmonary TB cases that have at least one contact identified;
2. The proportion of identified contacts of pulmonary TB cases that are evaluated.

Currently, data are not collected at a national level to monitor these indicators, but the intention is that ultimately they will be. No target level was set for any of the indicators in the strategy; rather, the aim is that regions should aim to improve upon their own baseline.

Whilst not included as a formal indicator, a statistic that is commonly used in contact investigations is the average yield per contact. This is defined as the proportion of evaluated contacts with active disease or LTBI (note that yield can also be defined per index case, rather than per contact). There are typically large differences between the yields of smear-positive pulmonary TB (PTB), smear-negative PTB and non-pulmonary index cases (EPTB). Smear positive cases are usually the most infectious and so are likely to have highest yields[104], whereas EPTB cases are typically not infectious, so have lower yield[2]. In this case, contact investigations would be undertaken to find the source of infection of the EPTB case[2,105]. One reason for not including yield as a contact tracing indicator, is that it is not clear whether the aim should be for it to increase or decrease; for instance, screening fewer contacts overall may lead to higher yield as a proportion of the number screened (because closer contacts might be screened and/or total TB incidence may rise), but reducing the amount of TB in the population might lead to lower yield.

Whilst the strategy indicators and contact tracing yield are not calculated at national level, some studies have calculated them for parts of the UK[2,48,49,105,106]; for instance, Saunders et al calculated the yield of active TB amongst contacts of PTB cases to be 2.3%[2] in Birmingham (see Table 1 for more details), Underwood et al. calculated it to be 3.5% in Tower Hamlets[105], a borough of London. In low- and middle-income countries, a systematic review in 2008 calculated the yield to be 4.5%[107], whereas a more recent (2013) systematic review calculated 3.1%[108]. Both reviews observed substantial heterogeneity in yield between studies. The latter review also calculated the yield in high-income countries to be 1.4%.

In addition to between-study heterogeneity, some studies have observed within-study heterogeneity in contact tracing screening outcomes. For example, contacts of Indian/Pakistani ethnicity were less likely to complete screening than were white contacts[2] in Birmingham, and those contacts of Caucasian ethnicity less likely to be TST positive than were non-Caucasian ones in south Glamorgan, Wales[106]. Rubilar *et al.* found higher notification rates of EPTB amongst Asian cases than non-Asian in Edinburgh[49].

The aim of the first paper in this thesis was to calculate the strategy monitoring indicators, yield of TB and yield of LTBI in London, and also to find out which population subgroups were associated with improved indicators, or with higher yields. I focussed on London as it is the location of 38% of TB cases in England, and also because it has good data on contact tracing due to cohort review. This paper provided the first quantification of the England collaborative strategy for tuberculosis contact tracing monitoring indicators. In so doing, it provides an important basis for monitoring future trends in tuberculosis contact tracing outcomes in London. The paper also describes groups who may be underserved by current contact tracing practices, thereby highlighting areas of potential improvement or in need of further action.

Cover sheet

See next page

London School of Hygiene & Tropical Medicine
Keppel Street, London WC1E 7HT
www.lshtm.ac.uk

Registry
T: +44(0)20 7299 4646
F: +44(0)20 7299 4656
E: registry@lshtm.ac.uk

LONDON SCHOOL of HYGIENE &TROPICAL MEDICINE

RESEARCH PAPER COVER SHEET

PLEASE NOTE THAT A COVER SHEET MUST BE COMPLETED FOR EACH RESEARCH PAPER INCLUDED IN A THESIS.

SECTION A – Student Details

Student	Sean M. Cavany
Principal Supervisor	Emilia Vynnycky
Thesis Title	

If the Research Paper has previously been published please complete Section B, if not please move to Section C

SECTION B – Paper already published

Where was the work published?	Thorax		
When was the work published?	07/04/17		
If the work was published prior to registration for your research degree, give a brief rationale for its inclusion	Using data analysis and mathematical modelling to study tuberculosis contact tracing in London, with reference to the national strategy and guidance		
Have you retained the copyright for the work?*	**No**	Was the work subject to academic peer review?	**Yes**

**If yes, please attach evidence of retention. If no, or if the work is being included in its published format, please attach evidence of permission from the copyright holder (publisher or other author) to include this work.*

SECTION C – Prepared for publication, but not yet published

Where is the work intended to be published?	
Please list the paper's authors in the intended authorship order:	
Stage of publication	Choose an item.

SECTION D – Multi-authored work

For multi-authored work, give full details of your role in the research included in the paper and in the preparation of the paper. (Attach a further sheet if necessary)	I conceived, deisgned and carried out the work, and wrote the first draft of the paper.

Student Signature: [redacted] **Date:** 28/03/18

Supervisor Signature: [redacted] **Date:** 29.03.18

Paper

See next page

OPEN ACCESS

ORIGINAL ARTICLE

An evaluation of tuberculosis contact investigations against national standards

Sean M Cavany,[1,2] Tom Sumner,[1] Emilia Vynnycky,[1,2] Clare Flach,[3] Richard G White,[1] H Lucy Thomas,[4] Helen Maguire,[5,6] Charlotte Anderson[5]

► Additional material is published online only. To view please visit the journal online (http://dx.doi.org/10.1136/thoraxjnl-2016-209677).

For numbered affiliations see end of article.

Correspondence to
Sean M Cavany, TB Modelling Group, TB Centre and CMMID, Faculty of Epidemiology and Population Health, London School of Hygiene and Tropical Medicine, Keppel Street, London WC1E 7HT, UK;
sean.cavany@lshtm.ac.uk

Received 3 November 2016
Revised 28 February 2017
Accepted 8 March 2017
Published Online First
7 April 2017

ABSTRACT

Background Contact tracing is a key element in England's 2015 collaborative TB strategy, although proposed indicators of successful contact tracing remain undescribed.

Methods We conducted descriptive and multivariable analyses of contact tracing of TB cases in London between 1 July 2012 and 31 December 2015 using cohort review data from London's TB Register, identifying characteristics associated with improved indicators and yield.

Results Of the pulmonary TB cases notified, 60% (2716/4561) had sufficient information for inclusion. Of these, 91% (2481/2716) had at least 1 contact (median: 4/case (IQR: 2–6)) identified, with 86% (10 251/11 981) of these contacts evaluated. 4.1% (177/4328), 1.3% (45/3421) and 0.70% (51/7264) of evaluated contacts of pulmonary smear-positive, pulmonary smear-negative and non-pulmonary cases, respectively, had active disease. Cases who were former prisoners or male were less likely to have at least one contact identified than those never imprisoned or female, respectively. Cases diagnosed at clinics with more directly observed therapy or social workers were more likely to have one or more contacts identified. Contacts screened at a different clinic to their index case or of male index cases were less likely to be evaluated than those screened at the same clinic or of women, respectively; yield of active disease was similar by sex. 10% (490/4850) of evaluated child contacts had latent TB infection.

Conclusions These are the first London-wide estimates of TB contact tracing indicators which are important for monitoring the strategy's success and informing risk assessment of index cases. Understanding why differences in indicators occur between groups could improve contact tracing outcomes.

Key messages

What is the key question?

- In London, what is the baseline level of contact tracing indicators for the Public Health England and National Health Service England collaborative TB strategy?

What is the bottom line?

- In London 91% of pulmonary index cases have at least one contact identified and 86% of those identified are evaluated for signs of TB and latent infection; there were significant differences in these indicators between cases, including when grouped by the sex of the case, whether they have social risk factors, and the staffing levels of the clinic.

Why read on?

- These results provide an important baseline for monitoring progress of England's national TB strategy and highlight areas in which improvements can be made, particularly those which show improved indicators for contacts screened at the same clinic as their index case or clinics with a greater number of directly observed therapy (DOT) or social workers.

► http://dx.doi.org/10.1136/thoraxjnl-2017-210110

To cite: Cavany SM, Sumner T, Vynnycky E, *et al*. *Thorax* 2017;**72**:736–745.

INTRODUCTION

In 2015, there were 10.4 million incident TB cases and 1.4 million deaths worldwide, making TB the largest cause of death by a single infectious agent.[1] While England is a low burden country, with 5758 notified cases in 2015, 39% of these occurred in London,[2–4] where overall incidence was 26/100 000/year; in 2014, it reached 79/100 000/year in one London borough.[4] While incidence remains highest among the foreign-born, and many cases are likely due to reactivation of infection acquired abroad, molecular epidemiological analyses attribute up to a third of new cases in London to recent transmission.[5] UK-born patients with TB and those of white ethnicity, 25% of whom have social risk factors, were frequently 'clustered' and also more likely to have delays exceeding 4 months from symptom onset to starting treatment.[4] The *Collaborative Tuberculosis Strategy for England*,[3] published in 2015, highlights contact tracing as an important tool for improving early TB diagnosis and reducing transmission. Contact tracing, which seeks to identify and diagnose contacts of infectious cases, has been used for decades in high-income countries where the heightened risk of disease among contacts relative to the general population makes it effective.[6–8] The strategy proposes two indicators of improved contact tracing: the proportion of pulmonary TB cases with close contacts identified, and the proportion of identified close contacts of pulmonary TB cases that are evaluated.

TB contact tracing in England broadly follows the stone-in-the-pond principle.[9] Clinics evaluate

close contacts (those with contact similar to that in a household) first, followed by casual contacts (eg, workplace or school contacts) if first round investigations suggest transmission to close contacts has occurred. In England, until 2016, guidance recommended identification and evaluation of household contacts of all index cases, after which tracing just household contacts of index cases with pulmonary or laryngeal disease was recommended.[7] In 2010, cohort review, an approach to case management and contact investigation appraisal that was shown to improve case management outcomes, was first introduced to London, and occurs quarterly.[10 11]

Several studies have found differences in both the proportion of TB contacts evaluated and the yield (the proportion of evaluated contacts diagnosed with TB) between ethnic groups and different disease types in the UK[12–14] and elsewhere.[15–17] The findings and understanding gained from such studies are not readily transferable to the large and ethnically diverse metropolis of London. Using data on contact tracing collected in London through cohort review since mid-2012, the main aim of this study is the presentation of baseline levels of the strategy contact tracing indicators, and of the proportion of contacts who are secondary cases or have latent TB infection (LTBI) (the 'yield'). A secondary aim is the identification of demographic and clinical characteristics associated with different indicators and yield estimates.

METHODS

Definition of terms

During the study period (1 July 2012–31 December 2015, see below), immediately after the diagnosis of the index case, the nurse asked the case for a list of close contacts. These contacts were then requested to attend for screening as soon as possible. Screening begins with symptom-screen; for asymptomatic contacts this is followed by a tuberculin skin test (TST) or interferon γ release assay (IGRA) in those under 35 years and consideration of a chest X-ray (CXR) in those 35 years and over.[7] Those with a positive symptom-screening, TST/IGRA result or CXR are evaluated for signs of active TB. LTBI is defined as a positive result on either a TST or IGRA, which both have variable sensitivity,[6] and the absence of active disease. Those with LTBI are considered for preventive therapy and/or BCG vaccination. The numbers of contacts with TB is recorded in the London TB Register (LTBR) at the initial contact investigation, so all new cases of TB among contacts can be considered prevalent.[18 19]

Data set and inclusion criteria

The primary data source was the LTBR, a web-based register containing demographic and clinical data on all TB cases notified in London since 2002. Clinical and demographic information on patients is entered directly to the LTBR by TB clinic staff. We restricted analyses to data collected after the introduction of cohort review (1 July 2012), after which the database includes data on the aggregate number of contacts per case that were: identified, evaluated, found to have LTBI and/or active TB. Contact data were aggregated by whether the contact was a child or adult (15 years old or above) and whether they were evaluated at the same clinic as the index case or elsewhere. As only contacts aged less than 35 years were evaluated for LTBI, data on LTBI status of contacts was only used for child contacts. No other demographic data on contacts were collected. Individual level data were not available for whether a home visit was undertaken; instead, the clinic's policy was used to provide clinic-level data. Only data from household and other close contacts are included in the analysis.

London is divided into five sectors; in some of these, a non-random selection of cases was reported at cohort review and recorded in the LTBR. To avoid bias, we only included cases in the analysis if the sector in which they were notified reported 80% or more of their cases at cohort review in a given quarter (defined as at least one of the cohort review fields completed) (see online supplementary appendix for a description of included quarters by sector). Selection of quarters was done separately for analyses including all cases, or just pulmonary cases (see online supplementary appendix). We removed further cases if: their line listing contained inconsistencies (eg, more contacts evaluated than were identified); included contacts were probably casual contacts (an incident was declared and more than 20 close contacts were identified); a patient was not reported to cohort review; or the index was detected through a previous contact investigation. In multivariable analyses, we also removed cases if data on an included exposure were missing. If the field describing number of contacts was missing, we assumed it was zero if other cohort review fields were complete and this assumption did not create inconsistencies.

Indicators

We calculated the following four indicators:

1. The proportion of pulmonary index cases who had at least one contact identified;
2. The proportion of identified contacts of pulmonary index cases who were evaluated;
3. The proportion of all evaluated contacts that had active TB ('Yield of active disease per contact');
4. The proportion of all evaluated child contacts that had LTBI ('Yield of LTBI per child contact').

Indicators 1 and 2 were proposed in the National Health Service England and Public Health England collaborative strategy[3] and are recommended in the US Centers for Disease Control and Prevention guidelines,[20] while indicators 2, 3 and 4 were used in a systematic review of European contact investigations.[15] As WHO and ECDC guidelines refer to low-income/middle-income countries and multidrug-resistant cases, respectively, we did not focus on these indicators here.[21 22] Indicators 1 and 2 include only pulmonary index cases, whereas indicators 3 and 4 include all index cases, in order to estimate the yield from screening contacts of non-pulmonary cases. While LTBI treatment is an important outcome of contact tracing in England, it is not covered in this article.

Statistical analysis

We calculated the four indicators London-wide and for subsets of index cases (based on disease site, smear status, age of contact and ethnic group). We used multivariable logistic regression to assess the association of clinical and demographic factors of the index cases, contacts, clinics and local authorities with whether or not the index case (indicator 1) or contact (indicators 2–4) satisfies each of the four indicators (see online supplementary table S1 for details of included variables). As this was an exploratory study all clinical and demographic factors with a plausible direct or confounding impact on the outcome were included in the model (see online supplementary appendix for the exposures included). We assessed all variables for multicollinearity. For indicators 2–4, as each index case may have several contacts, index case exposures tend to cluster; adjustments were made to the p values and CIs to account for this, using the between-cluster variance estimator in Stata.[23] For indicator 4, we excluded adult contacts, as we did not know which of them were tested for LTBI.

Note that there may be discrepancies between numbers presented for the levels of the indicators and for the multivariable analysis, due to cases missing data on variables included in the multivariable analysis. All data were analysed using Microsoft Excel V.14.0 and Stata V.13.1.

RESULTS

Comparison of included and excluded cases

From 1 July 2012 to 31 December 2015 inclusive, 9821 cases were reported in the LTBR of which 4561 were pulmonary. After excluding cases, 5491 cases of all forms of TB and 2716 pulmonary cases remained (figure 1). When considering all cases, there were 971 (18%) pulmonary smear-positive cases, 1095 (20%) pulmonary smear-negative cases, 478 (8.7%) pulmonary cases with unknown smear, four laryngeal cases without pulmonary involvement and 2943 (54%) non-pulmonary, non-laryngeal cases. In general, included and excluded cases shared similar clinical and demographic factors (table 1).

Indicator 1. Proportion of pulmonary index cases who had at least one contact identified

Of 2716 index cases with pulmonary disease, 2481 (91%) had at least one contact identified (table 2). This led to 12 248 contacts identified, a mean and median of 4.5 (95% CI 4.36 to 4.66) and 4 (IQR: 2–6) per index case, respectively. Table 3 shows the predictors of having at least one contact identified; the multivariable analysis is restricted to those with complete data on the variables included, limiting the sample from 2716 to 2327 of which 2168 (93%) cases had at least one contact identified. Male cases were less likely than female cases to have at least one contact identified (adjusted OR (aOR): 0.46 (0.30 to 0.68,)), as were those with a history of imprisonment compared with those without a history of imprisonment (aOR: 0.27 (0.14 to 0.52)), those of black African ethnicity compared with those of Indian ethnicity (aOR: 0.47 (0.27 to 0.83)) and recent migrants compared with long-term migrants (aOR: 0.54 (0.35 to 0.84)). Smear-positive index cases were more likely to have at least one contact identified compared with smear-negative index cases (aOR: 1.84 (1.20 to 2.82)). There was a significant positive linear association between the number of social care or directly observed therapy (DOT) workers (aOR per staff per 100 cases: 1.36 (1.07 to 1.72)) at the clinic and whether a case had at least one contact identified. These associations were similar for the proportion of index cases with three or more contacts named, or with five or more, except that index cases with a history of drug use were more likely to have three or more contacts named than those without (data not shown).

Indicator 2. Proportion of contacts of pulmonary index cases that were evaluated

Of the 12 248 contacts identified, 11 981 had data on age and screening location and were included in this analysis. Of these,

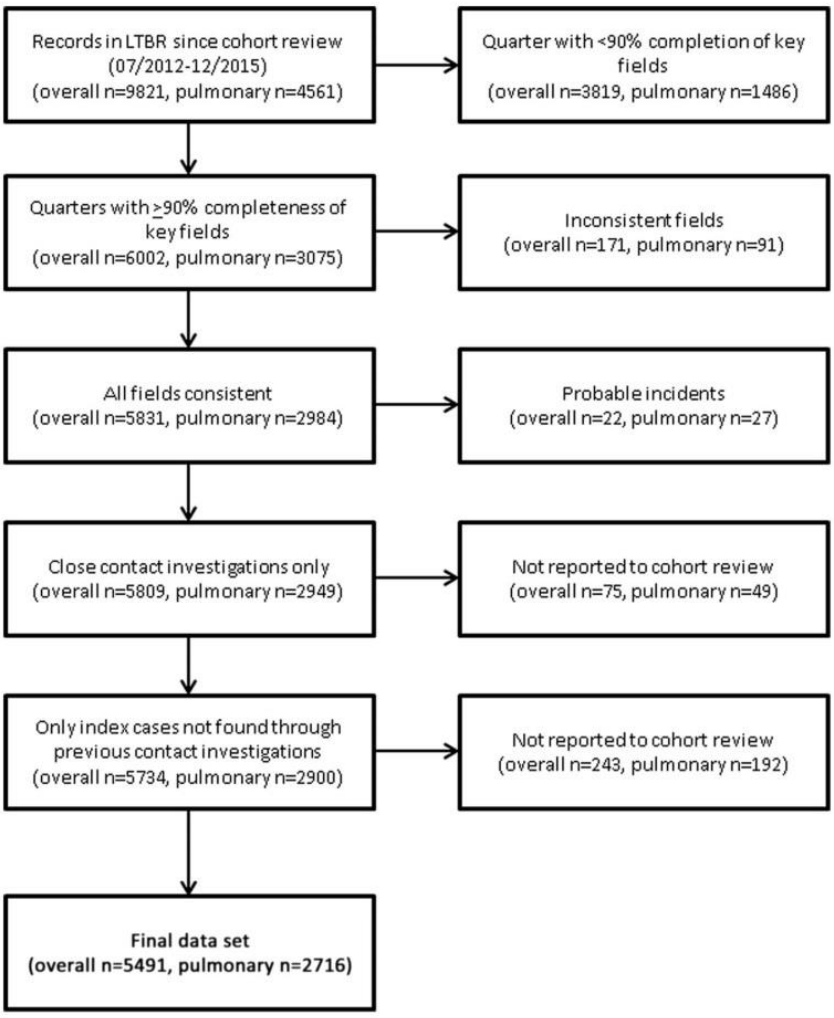


Figure 1 Description of included and excluded cases. LTBR, London TB Register.

Table 1 Comparison of characteristics of cases that were included and excluded in the analyses. Percentages are within-group column percentages with the exception of the total row, where row percentages are given. Within-group totals may be discrepant to the stated total due to cases with missing data

	Index cases with all forms of TB		Index cases with pulmonary TB	
Factor	Number included (%)	Number excluded (%)	Number included (%)	Number excluded (%)
Total	5491 (56%)	4330 (44%)	2716 (60%)	1845 (40%)
UK-born?				
Yes	830 (19%)	943 (17%)	622 (23%)	500 (27%)
No	2880 (67%)	3862 (70%)	1744 (64%)	1093 (60%)
No, recent migrant (<2 years)	578 (13%)	674 (12%)	344 (13%)	233 (13%)
Ethnicity				
Bangladeshi	257 (5%)	317 (7%)	105 (4%)	85 (5%)
Black-African	1278 (23%)	726 (17%)	608 (22%)	361 (20%)
Black-Caribbean	208 (4%)	118 (3%)	116 (4%)	66 (4%)
Black-Other	82 (2%)	58 (1%)	40 (2%)	29 (2%)
Chinese	81 (1%)	37 (1%)	52 (2%)	19 (1%)
Indian	1330 (24%)	1434 (33%)	510 (19%)	461 (25%)
Pakistani	976 (18%)	607 (14%)	496 (18%)	290 (16%)
White	498 (9%)	525 (12%)	201 (7%)	189 (10%)
Other	758 (14%)	476 (11%)	578 (21%)	330 (18%)
Sex				
Male	3287 (60%)	2537 (59%)	1704 (63%)	1122 (61%)
Female	2204 (40%)	1793 (41%)	1012 (37%)	723 (39%)
Site of disease				
Pulmonary or laryngeal	2548 (46%)	2018 (47%)	N/a	N/a
Non-pulmonary and non-laryngeal	2943 (54%)	2312 (53%)	N/a	N/a
Social risk factor				
History of homelessness	230 (4%)	130 (3%)	187 (7%)	90 (5%)
History of imprisonment	127 (2%)	76 (2%)	105 (4%)	50 (3%)
History of drug use	229 (4%)	141 (3%)	179 (7%)	101 (6%)
BCG vaccinated				
Yes	3136 (58%)	2627 (61%)	1541 (58%)	1115 (61%)
No	1117 (21%)	848 (19%)	568 (21%)	354 (19%)
Unknown	1129 (21%)	800 (19%)	563 (21%)	351 (19%)
Age				
15 years old or over	5367 (98%)	4046 (93%)	2641 (97%)	1658 (90%)
Under 15 years old	124 (2%)	284 (7%)	75 (3%)	187 (10%)
Home visits policy				
Yes	1179 (32%)	369 (9%)	907 (33%)	230 (13%)
No	3712 (68%)	3927 (91%)	1801 (67%)	1593 (87%)

10 251 (86%) were evaluated (table 2), with a mean and median of 3.9 (95% CI 3.72 to 4.00) and 3 (IQR: 1–5) per index case, respectively. The multivariable analysis of predictors of a contact being evaluated in table 4 is restricted to those with complete data on the variables included, limiting the sample of contacts from 11 981 to 10 476 of which 8986 (86%) were evaluated. Identified contacts of male index cases were less likely to be evaluated than those of female index cases (aOR: 0.68 (0.54 to 0.85)), as were contacts aged over 14 years (aOR: 0.30 (0.24 to 0.39)) when compared with those aged under 15 years. Contacts screened at the clinic of their index case (aOR: 1.65 (1.26 to 2.16)) were more likely to be evaluated than those screened elsewhere. Identified contacts of index cases notified in boroughs with higher notification rates (aOR per 10 cases per 100 000: 0.93 (0.87 to 0.99)), and contacts of index cases of white ethnicity (compared with Indian index cases, aOR: 0.61 (0.42 to 0.90)) were less likely to be evaluated, though these two associations were weaker.

Indicator 3. Proportion of evaluated contacts diagnosed with active TB

Of the 16 495 contacts (of index cases with disease at any site) evaluated, and no longer under investigation for signs of TB at the time of the study, 294 (1.8%, 95% CI 1.6% to 2.0%) were diagnosed with active TB (table 2). This figure rises to 2.6% (243/9213) (95% CI 2.3% to 3.0%) and 4.1% (177/4328) (95% CI 3.5% to 4.7%) for the contacts of pulmonary index cases and sputum smear-positive pulmonary index cases, respectively. This figure drops to 1.3% (45/3421) (95% CI 0.98% to 1.8%) and 0.70% (51/7264) (95% CI 0.51% to 0.89%) for the contacts of smear-negative pulmonary cases index cases and index cases without pulmonary or laryngeal involvement, respectively. Considering just index cases within the black African ethnic group (the ethnic group with the highest yield of active disease per contact) 2.8% (107/3817) (95% CI 2.3% to 3.3%) of evaluated contacts had active disease. For index cases aged under 15 years, 5.5% (27/491) (95% CI 3.5% to 7.5%) of their evaluated contacts had active

Table 2 Levels of indicators and outcomes over study period

Indicator	Number positive/total	Percentage (95% CI)
1. The proportion of pulmonary index cases who have at least one contact identified	2481/2716	91% (90% to 92%)
The proportion of pulmonary index cases who have at least three contacts identified	1810/2716	67% (65% to 68%)
The proportion of pulmonary index cases who have at least five contacts identified	1093/2716	40% (38% to 42%)
2. The proportion of identified contacts of pulmonary index cases who are evaluated	10 251/11 981	86% (85% to 86%)
3. The proportion of evaluated contacts that have active TB	294/16 495	1.8% (1.6% to 2.0%)
The proportion of evaluated contacts of pulmonary smear-positive index cases that have active TB	177/4328	4.1% (3.5% to 4.7%)
The proportion of evaluated contacts of pulmonary smear-negative index cases that have active TB	45/3421	1.3% (0.98% to 1.8%)
The proportion of evaluated contacts of non-pulmonary and non-laryngeal index cases that have active TB	51/7264	0.70% (0.51% to 0.89%)
4. The proportion of evaluated child contacts that have LTBI	490/4850	10% (9.3% to 11%)
The proportion of evaluated child contacts of pulmonary smear-positive index cases that have LTBI	277/1046	26% (24% to 29%)
The proportion of evaluated child contacts of pulmonary smear-negative index cases that have LTBI	93/980	9.5% (7.8% to 11%)
The proportion of evaluated child contacts of non-pulmonary and non-laryngeal index cases that have LTBI	78/2386	3.3% (2.6% to 4.0%)

LTBI, latent TB infection.

disease. The multivariable analysis of predictors of an evaluated contact being diagnosed with TB in table 5 is restricted to those with complete data on the variables included, limiting the sample from 16 495 to 14 614 contacts of which 263 (1.8%) were diagnosed with TB. Adult contacts were associated with lower yields of active TB per contact (aOR: 0.55 (0.40 to 0.75)) (compared with child contacts). Additionally, index cases aged below 15 years, of black African ethnicity or with pulmonary or laryngeal disease (especially those who are smear-positive) were associated with contacts having an increased risk of active disease. None of the assessed social risk factors (history of homelessness, drug use or imprisonment) for the index case were associated with the contact having active TB.

Indicator 4. Proportion of evaluated child contacts diagnosed with LTBI

Of the 4850 child contacts evaluated and no longer under investigation for signs of TB at the time of the study, 490 (10%, 95% CI 9.3% to 11%) were diagnosed with LTBI (table 2). This figure rises to 26% (277/1046) (95% CI 24% to 29%) for contacts of sputum smear-positive pulmonary index cases only, and drops to 9.5% (93/980) (95% CI 7.8% to 11%) and 3.3% (78/2386) (95% CI 2.6% to 4.0%) for the contacts of smear-negative pulmonary index cases and cases without pulmonary or laryngeal involvement, respectively. For index cases aged under 15 years, 17% (32/184) (95% CI 12% to 23%) of their evaluated child contacts had LTBI. The multivariable analysis of predictors of a child contact being diagnosed with TB in table 6 is restricted to those with complete data on the variables included, limiting the sample from 4850 to 4305 child contacts of which 440 (10%) were found to have LTBI. Pulmonary smear-negative (aOR: 2.92 (1.96 to 4.35)) and pulmonary smear-positive index cases (aOR: 8.39 (5.76 to 12.2)) (both relative to non-pulmonary and non-laryngeal index cases) were positively associated with child contacts having LTBI (table 6). Conversely, culture-negative index cases (aOR relative to culture-positive index cases: 0.51 (0.34 to 0.76)) were negatively associated with child contacts having LTBI. None of the assessed social risk factors (a history of homelessness, drug use or imprisonment) for the index case were associated with child contacts having LTBI, though the numbers were very small.

DISCUSSION

Our analyses provide the first estimates of indicators that will be used to monitor TB contact tracing in England, as part of the Collaborative Tuberculosis Strategy for England. We found that 91% of pulmonary TB cases in London had at least one contact identified and 86% of those contacts were investigated. For all index cases, 1.8% of evaluated contacts had active TB disease and 10% of child contacts had LTBI. The proportion of contacts diagnosed with TB or LTBI was almost fivefold greater for contacts of pulmonary or laryngeal index cases than for index cases with other disease types. Compared with female index cases, male index cases had fewer contacts identified, and fewer of those identified were evaluated, but sex had no significant effect on whether a case's evaluated contacts had LTBI or TB disease. Perhaps surprisingly, social risk factors were generally not significantly associated with either identifying contacts, evaluating those contacts, or for the resultant TB or LTBI yield per contact, the exception being that those with a prison history were less likely to have contacts identified compared with those without a prison history. The study may however have been underpowered to discern these relationships, as only around 5% of included index cases had each social risk factor. Contacts of children were more likely than those of adults to be diagnosed with TB; in the former circumstance, it is more likely that one of the contacts diagnosed with TB was the source case for that child.

By using data elicited through cohort review, we could include more information than that available from routine surveillance. One limitation was that data from several clinics for some periods had to be excluded as cohort review was done selectively (see online supplementary appendix). We mitigated this potential source of bias by only using data from clinics that reported 80% or more of their contacts, although some differences between included and excluded cases remained (table 1) and it reduced the power of the study. Also, cases removed from the multivariable analysis due to missing variables were more likely to be culture-positive, white, UK-born or homeless, and less likely to be screened at the clinic, than those included. It is difficult to predict the impact of these exclusions on the results. However, the proportion of contacts evaluated (86%) may be an overestimate, given that both more contacts of white index cases and fewer contacts screened at the clinic are excluded than included. Further, as more contacts of children were excluded than included (table 1), it is possible that the estimates of yield of TB and LTBI among contacts are underestimates. A second limitation is that little data were collected on individual contacts, potentially masking important determinants of contact outcomes, including the length and intensity of exposure, and risk

Table 3 Associations with pulmonary index cases having at least one contact named (indicator 1)

	At least one contact identified			
Characteristics of the index case	**Yes (row %)***	**No***	**aOR (95% CI)**	**p Value**
Total	2168 (93%)	159	N/a	N/a
UK born				
Yes	506 (94%)	31	1.16 (0.71 to 1.91)	0.01
No, long-term migrant	1390 (94%)	94	1	
No, recent migrant	272 (89%)	34	0.54 (0.35 to 0.84)	
History of drug use				
Yes	133 (92%)	11	1.57 (0.73 to 1.37)	0.25
No	2035 (93%)	148	1	
History of homelessness				
Yes	123 (85%)	21	0.56 (0.31 to 1.01)	0.06
No	2045 (94%)	138	1	
Former prisoner				
Yes	76 (81%)	18	0.27 (0.14 to 0.52)	<0.01
No	2092 (94%)	141	1	
Sex				
Male	1320 (91%)	124	0.46 (0.30 to 0.68)	<0.01
Female	848 (96%)	35	1	
Ethnicity				
Indian	417 (95%)	20	1	<0.01
Black-African	468 (89%)	59	0.47 (0.27 to 0.83)	
White	431 (93%)	34	0.73 (0.38 to 1.41)	
Other	852 (95%)	46	1.06 (0.61 to 1.85)	
Culture				
Positive	1684 (94%)	104	1	0.02
Negative	419 (90%)	46	0.59 (0.40 to 0.89)	
Not done	65 (88%)	9	0.52 (0.23 to 1.16)	
Smear				
Negative	945 (92%)	79	1	<0.01
Not done	367 (90%)	40	0.69 (0.44 to 1.06)	
Positive	856 (96%)	40	1.84 (1.20 to 2.82)	
Clinic case count (linear, 100 cases/year)	$\bar{x}= 1.2$	s=0.58	0.72 (0.46 to 1.12)	0.15
Clinic policy of home visits*				
Yes	714 (92%)	64	0.66 (0.44 to 0.99)	0.04
No	1454 (94%)	95	1	
Age of index case				
0–14 years	63 (97%)	2	1	0.19
15 years and over	2105 (93%)	157	0.36 (0.08 to 1.62)	
Notification rate*† (linear, 10 cases/100 000 population/year)	$\bar{x}= 3.7$	s=0.58	1.10 (0.98 to 1.23)	0.10
Nurses* (linear, nurses/100 cases/year)	$\bar{x}= 3.5$	s=0.87	0.90 (0.63 to 1.28)	0.54
Admin staff* (linear, staff/100 cases/year)	$\bar{x}= 1.0$	s=0.63	0.81 (0.50 to 1.31)	0.39
Social workers/DOT staff* (linear, staff/100 cases/year)	$\bar{x}= 0.88$	s=0.87	1.36 (1.07 to 1.72)	0.01

Adjusted OR (aOR) has been adjusted for everything in the table.
*Mean and SD presented for continuous variables.
†Notification rate in the year and local authority in which the index case was notified. N=2327.
DOT, directly observed therapy.

factors increasing the chance of the contact progressing to disease. A third limitation is that the available data come from a relatively short period (42 months), so temporal trends were indiscernible. Unfortunately, data on the HIV status of index cases are not recorded in the LTBR. Finally, the number of patients with multidrug-resistant and extensively drug-resistant TB was too small to do a subgroup analysis of this group.

The proportion of pulmonary index cases with at least one contact identified (91%) compares favourably with previous figures for one sector of London in 2012, namely 78% and 88% before and after cohort review, respectively.[11] It is also comparable to recent findings from the USA where the corresponding figures were 94% and 86% for smear-positive and smear-negative, culture-positive index cases, respectively,[24] and higher than the figure in Piedmont, Italy (77%).[17] Similarly, the proportion of identified contacts who were evaluated (88%) was higher than the north central London (74% precohort review, 82% postcohort review) and US figures (82% for smear-positives, 81% for smear-negatives). These figures suggest London TB clinics undertake high quality contact tracing, although further improvements in certain groups may be feasible. Note that the higher proportion of smear-positive index cases with at least one contact identified than smear-negative index cases, found in this study and others, is likely due to

Table 4 Associations with contacts of pulmonary index cases being evaluated (indicator 2)

	Contact evaluated			
Characteristics of the index case	**Yes (row %)***	**No***	**aOR (95%CI)**	**p Value**
Total	8986 (86%)	1490	N/a	N/a
UK born				
Yes	2240 (86%)	907	1.22 (0.92 to 1.61)	0.13
No, long-term migrant	5692 (86%)	357	1	
No, recent migrant	1054 (82%)	226	0.82 (0.59 to 1.13)	
History of drug use				
Yes	564 (80%)	137	0.99 (0.60 to 1.62)	0.97
No	8422 (86%)	1353	1	
History of homelessness				
Yes	480 (77%)	147	0.66 (0.40 to 1.06)	0.09
No	8506 (86%)	1343	1	
Former prisoner				
Yes	8708 (86%)	1400	0.63 (0.35 to 1.14)	0.13
No	278 (76%)	90	1	
Sex				
Male	5106 (83%)	1024	0.68 (0.54 to 0.85)	<0.01
Female	3880 (89%)	466	1	
Ethnicity				
Indian	1690 (87%)	263	1	0.05
Black African	1985 (88%)	275	0.88 (0.62 to 1.25)	
White	1739 (82%)	378	0.61 (0.42 to 0.90)	
Other	3572 (86%)	574	0.93 (0.69 to 1.25)	
Culture				
Positive	7221 (86%)	1174	1	0.46
Negative	1507 (84%)	280	0.84 (0.63 to 1.11)	
Not done	258 (88%)	36	1.03 (0.48 to 2.21)	
Smear				
Negative	3391 (84%)	625	1	0.14
Not done	1285 (85%)	220	1.02 (0.74 to 1.41)	
Positive	4310 (87%)	645	1.27 (1.00 to 1.61)	
Clinic case count* (linear, 100 cases/year)	$\bar{x}= 1.2$	s=0.58	0.81 (0.64 to 1.04)	
Clinic policy of home visits				
Yes	3078 (87%)	480	1.09 (0.82 to 1.43)	0.56
No	5908 (85%)	1010	1	
Age of index case				
0–14 years	292 (90%)	34	1	0.93
15 years and over	8694 (86%)	1456	0.97 (0.53 to 1.80)	
Notification rate*† (linear, 10 cases/100 000 population/year)	$\bar{x}= 3.8$	s=2.1	0.93 (0.87 to 0.99)	0.02
Nurses* (linear, nurses/100 cases/year)	$\bar{x}= 3.4$	s=0.87	0.89 (0.72 to 1.09)	0.26
Admin staff* (linear, staff/100 cases/year)	$\bar{x}= 0.99$	s=0.60	1.17 (0.86 to 1.60)	0.32
Social workers/DOT staff* (linear, staff/100 cases/year)	$\bar{x}= 0.89$	s=0.89	0.95 (0.81 to 1.10)	0.47
Age of contact				
15 years and over	6569 (83%)	1344	0.30 (0.24 to 0.39)	<0.01
0–14 years	2417 (94%)	146	1	
Contact screened at clinic				
Yes	7629 (87%)	1147	1.65 (1.26 to 2.16)	<0.01
No	1357 (80%)	343	1	

*Mean and SD presented for continuous variables.
†Notification rate in the year and local authority in which the index case was notified. N=10 476.
aOR, adjusted OR; DOT, directly observed therapy.

clinics' need to focus limited resources on those index cases for whom the yield is likely to be highest. The yield per contact of all forms of TB in this study was higher than from a similar study in Birmingham, another high incidence inner city region of the UK, in 1990–2010.[14] That study found a yield per contact of active TB of 3.3% (compared with 4.1% in this study) and 0.58% (versus 0.70% in this study) for smear-positive pulmonary and non-pulmonary index cases, respectively. The yield per contact of active TB among contacts of pulmonary index cases presented here (2.6%) is high relative to other studies in high-income settings. A systematic review of contact tracing outcomes found a grouped yield per contact of TB in high-income countries of 1.4%,[8] a similar study to ours in Italy found a yield per contact of active disease of 0.71%,[17]

Table 5 Associations with evaluated contacts (of index cases with TB at any site) with active TB

Characteristics of the index case	Contact diagnosed with active TB			
	Yes (row %)*	No*	aOR (95%CI)	p Value
Total	263 (1.8%)	14 351	N/a	N/a
UK born				
Yes	71 (2.3%)	2973	1.31 (0.86 to 1.99)	0.41
No, long-term migrant	159 (1.6%)	9834	1	
No, recent migrant	33 (2.1%)	1544	1.23 (0.71 to 2.15)	
History of drug use				
Yes	18 (2.9%)	612	1.06 (0.56 to 2.01)	0.85
No	245 (1.8%)	13 739	1	
History of homelessness				
Yes	13 (2.6%)	485	0.77 (0.36 to 1.66)	0.51
No	250 (1.8%)	13 866	1	
Former prisoner				
Yes	9 (2.8%)	312	1.02 (0.37 to 2.84)	0.96
No	254 (1.8%)	14 039	1	
Sex				
Male	164 (2.0%)	7884	1.35 (0.95 to 1.91)	0.09
Female	99 (1.5%)	6467	1	
Ethnicity				
Indian	39 (1.2%)	3130	1	<0.01
Black African	103 (3.0%)	3324	1.95 (1.17 to 3.26)	
White	32 (1.7%)	1910	0.81 (0.44 to 1.50)	
Other	89 (1.5%)	5987	1.03 (0.63 to 1.67)	
Culture				
Positive	225 (2.4%)	9316	1	0.12
Negative	71 (0.8%)	4033	0.70 (0.43 to 1.15)	
Not done	7 (0.7%)	1002	0.50 (0.23 to 1.09)	
Disease type				
Non-pulmonary and non-laryngeal	41 (0.6%)	6391	1	<0.01
Pulmonary smear-negative	44 (1.4%)	3107	2.10 (1.20 to 3.69)	
Pulmonary smear-positive	166 (4.3%)	3696	6.71 (4.11 to 11.0)	
Pulmonary, smear unknown or laryngeal	12 (1.0%)	1157	1.39 (0.73 to 2.64)	
BCG				
Not vaccinated	51 (1.6%)	3222	0.78 (0.51 to 1.18)	0.28
Vaccinated	177 (2.0%)	8502	1	
Unknown	35 (1.3%)	2627	0.73 (0.44 to 1.19)	
Clinic case count* (linear, 100 cases/year)	$\bar{x}= 1.1$	s=0.52	0.97 (0.68 to 1.40)	0.89
Age of index case				
0–14 years	25 (5.7%)	417	1	<0.01
15 years and over	238 (1.7%)	13 934	0.25 (0.13 to 0.48)	
Notification rate*† (linear, 10 cases/100 000 population/year)	$\bar{x}= 3.7$	s=1.9	1.05 (0.95 to 1.17)	0.32
Age of contact				
15 years and over	151 (1.5%)	10 133	0.55 (0.40 to 0.75)	<0.01
0–14 years	112 (2.6%)	4218	1	
Contact screened at clinic				
Yes	233 (1.9%)	12 162	1.38 (0.81 to 2.37)	0.24
No	30 (1.4%)	2189	1	

*Mean and SD presented for continuous variables.
†Notification rate in the year and local authority in which the index case was notified. N=14 614.
aOR, adjusted OR.

and a study in Amsterdam found 0.79% of contacts had prevalent TB and 0.39% incident TB,[18] although these comparisons do not account for differences in smear-positivity prevalence. There are no comparable estimates of the yield per contact of LTBI among child contacts. The overall prevalence among contacts (1.8%) compares to 0.50% of contacts with prevalent TB and 0.53% with incident TB in a European study.[19]

Overall, TB contact tracing in London is performing well across the suite of indicators when compared with previous studies in London or elsewhere. Contacts of cases with pulmonary or laryngeal TB are more likely to have TB or LTBI, compared with contacts of cases with non-pulmonary and non-laryngeal TB. However, prevalence of active TB among contacts of non-pulmonary patients is still very high (0.70%) relative to background TB prevalence, (0.027% in 2006[25]),

Table 6 Associations with evaluated child contacts (of index cases with TB at any site) being diagnosed with LTBI

	Child contact diagnosed with LTBI			
Characteristic of the index case	**Yes (row %)***	**No***	**aOR (95% CI)**	**p Value**
Total	440 (10%)	3865	N/a	N/a
UK born				
Yes	113 (16%)	611	1.23 (0.85 to 1.78)	0.44
No, long-term migrant	285 (9%)	2854	1	
No, recent migrant	42 (10%)	400	0.87 (0.51 to 1.48)	
History of drug use				
Yes	29 (18%)	136	1.35 (0.62 to 2.96)	0.45
No	411 (10%)	3729	1	
History of homelessness				
Yes	17 (14%)	108	0.95 (0.47 to 1.94)	0.90
No	423 (10%)	3757	1	
Former prisoner				
Yes	8 (10%)	74	0.53 (0.21 to 1.36)	0.19
No	432 (10%)	3791	1	
Sex				
Male	218 (10%)	1995	0.80 (0.59 to 1.09)	0.16
Female	222 (11%)	1870	1	
Ethnicity				
Indian	62 (8%)	700	1	0.41
Black African	134 (10%)	1232	1.07 (0.69 to 1.64)	
White	72 (18%)	322	1.54 (0.90 to 2.62)	
Other	172 (10%)	1611	1.06 (0.71. 1.58)	
Culture				
Positive	375 (14%)	2283	1	<0.01
Negative	49 (4%)	1260	0.51 (0.34 to 0.76)	
Not done	16 (5%)	322	0.60 (0.33 to 1.11)	
Disease type				
Non-pulmonary and non-laryngeal	67 (3%)	2041	1	<0.01
Pulmonary smear-negative	91 (10%)	823	2.92 (1.96 to 4.35)	
Pulmonary smear-positive	252 (27%)	692	8.39 (5.76 to 12.2)	
Pulmonary, smear unknown or laryngeal	30 (9%)	309	2.35 (1.35 to 4.10)	
BCG				
Not vaccinated	87 (10%)	822	0.83 (0.56 to 1.21)	0.54
Vaccinated	280 (11%)	2343	1	
Unknown	73 (9%)	700	1.08 (0.72 to 1.60)	
Clinic case count* (linear, 100 cases/year)	$\bar{x}= 1.1$	s=0.52	1.22 (0.85 to 1.76)	0.28
Age of index case				
0–14 years	30 (18%)	134	1	0.01
15 years and over	410 (10%)	3731	0.40 (0.20 to 0.77)	
Notification rate*† (linear, 10 cases/100 000 population/year)	$\bar{x}= 3.7$	s=1.8	1.04 (0.93 to 1.16)	0.45
Contact screened at clinic				
Yes	388 (11%)	3306	1.29 (0.83 to 2.01)	0.27
No	52 (8.5%)	559	1	

*Mean and SD presented for continuous variables.
†Notification rate in the year and local authority in which the index case was notified. N=4305.
aOR, adjusted OR; LTBI, latent TB infection.

which has implications for the recent change to the National Institute for Health and Care Excellence (NICE) TB guidelines, limiting screening to contacts of pulmonary and laryngeal patients.[7]

We found that contacts screened at a different clinic to the index case (accounting for 16% of all contacts) are less likely to be evaluated than those screened at the same clinic, suggesting gains can be made by improving cross-clinic contact tracing. Clinics were increasingly likely to successfully evaluate their contacts with increasing numbers of DOT and social workers per index case, indicating the important role played by these staff in building relationships with patients. Future continuation of these relationships may be affected by recent policy recommendations, for example, cessation of screening of contacts of non-pulmonary non-laryngeal patients and screening for LTBI in all contacts aged under 65 years (previously just under 35 years old).

Further work elucidating how performance against contact tracing indicators affects transmission and diagnostic delay would be important, as would the impact of recent changes in procedure to comply with latest national guidance. The recently changed NICE guidelines recommend limiting screening to

pulmonary and laryngeal cases' contacts (bringing the UK into line with many other countries, as well as testing all those aged under 65 years with TST/IGRA and administering preventive therapy to this group (rather than just those aged under 35 years, but the implications for cost-effectiveness are unclear. Cost-effectiveness modelling using data from this study may help. Completion of cohort review fields by all clinics has improved recently; continued high levels of completion of these fields would greatly benefit future contact tracing studies. Further work to understand the full impact of home visits at an individual level would be useful. This intervention has been found to aid identification and evaluation of contacts in the USA[26] and Portugal,[27] and to improve preventive therapy outcomes in the USA and Canada,[28] but our analysis was unable to show a relationship between home visits and improved contact identification and evaluation. Finally, qualitative research into improving engagement with men and those with prison history could potentially improve the proportion of contacts successfully evaluated, as these groups are less likely (compared with women and those without a prison history, respectively) to have contacts identified and evaluated (for men) but do not necessarily have lower yields per contact, suggesting room for improvement.

Author affiliations
[1]TB Modelling Group, TB Centre and CMMID, Faculty of Epidemiology and Population Health, London School of Hygiene and Tropical Medicine, London, UK
[2]Statistics, Modelling and Economics Department, Public Health England, London, UK
[3]Faculty of Epidemiology and Population Health, London School of Hygiene and Tropical Medicine, London , UK
[4]Respiratory Diseases Department, Public Health England, London, UK
[5]Field Epidemiology Services, Public Health England, London, UK
[6]Department of Infection and Population Health, University College London, London, UK

Acknowledgements The authors thank Jacqui Carless, Lamya Kanfoudi and others at Public Health England (PHE) field Epidemiology London for maintaining the LTBR; all nurses and other healthcare workers involved with collecting the data and presenting it at cohort review; Johan van Wijgerden for providing data on London clinic staffing levels; Ayoub Saei for statistical advice; the TB modelling group at LSHTM for advice on presentation and approach; and Lindsay Serene for comments on the manuscript.

Contributors SMC conceived and designed the work with input from TS, EV, RGW, HLT, HM and CA. CA and HM are responsible for the acquisition and maintenance of the data. SMC undertook the analysis with advice from all other authors. All authors contributed to the interpretation of the data. SMC wrote the first draft of the paper and all authors contributed to subsequent drafts. All authors approve the work for publication and agree to be accountable for the work.

Funding SMC is funded by a joint PHE-LSHTM studentship in Infectious Disease Modelling. RGW is funded by the UK Medical Research Council (MRC) and the UK Department for International Development (DFID) (under the MRC/DFID Concordat agreement that is also part of the EDCTP2 programme supported by the European Union (MR/J005088/1)), the Bill and Melinda Gates Foundation (TB Modelling and Analysis Consortium: OPP1084276, and SA Modelling for Policy: #OPP1110334) and UNITAID (4214-LSHTM-Sept15; PO #8477-0-600). CA, EV, HM and HLT are employed by PHE, a government agency, and received no other source of funding. TS is funded by the Bill and Melinda Gates Foundation (SA Modelling for Policy: #OPP1110334).

Competing interests None declared.

Provenance and peer review Not commissioned; externally peer reviewed.

REFERENCES

1 Global Tuberculosis Report 2016. Geneva: World Health Organization, 2016.
2 Tuberculosis in the UK: 2016 report. London: Public Health England, 2016.
3 *Collaborative Tuberculosis Strategy for England: 2015 to 2020*. London: Public Health England, NHS England, 2015.
4 *Tuberculosis in London: Annual review (2014 data)*. London: Public Health England, 2015.
5 Hamblion EL, Le Menach A, Anderson LF, *et al*. Recent TB transmission, clustering and predictors of large clusters in London, 2010–2012: results from first 3 years of universal MIRU-VNTR strain typing. *Thorax* 2016;71:749–56.
6 Erkens CG, Kamphorst M, Abubakar I, *et al*. Tuberculosis contact investigation in low prevalence countries: a European consensus. *Eur Respir J* 2010;36:925–49.
7 *NICE Tuberculosis Guidance*. http://www.nice.org.uk/guidance/CG117
8 Fox GJ, Barry SE, Britton WJ, *et al*. Contact investigation for tuberculosis: a systematic review and meta-analysis. *Eur Respir J* 2013;41:140–56.
9 Veen J. Microepidemics of tuberculosis: the stone-in-the-pond principle. *Tuber Lung Dis* 1992;73:73–6.
10 Munsiff SS, Ahuja SD, King L, *et al*. Ensuring accountability: the contribution of the cohort review method to tuberculosis control in New York City. *Int J Tuberc Lung Dis* 2006;10:1133–9.
11 Anderson C, White J, Abubakar I, *et al*. Raising standards in UK TB control: introducing cohort review. *Thorax* 2014;69:187–9.
12 Mandal P, Craxton R, Chalmers JD, *et al*. Contact tracing in pulmonary and non-pulmonary tuberculosis. *QJM* 2012;105:741–7.
13 Rubilar M, Brochwicz-Lewinski MJ, Anderson M, *et al*. The outcome of contact procedures for tuberculosis in Edinburgh, Scotland 1982–1991. *Respir Med* 1995;89:113–20.
14 Saunders MJ, Koh GC, Small AD, *et al*. Predictors of contact tracing completion and outcomes in tuberculosis: a 21-year retrospective cohort study. *Int J Tuberc Lung Dis* 2014;18:640–6.
15 Mulder C, Klinkenberg E, Manissero D. Effectiveness of tuberculosis contact tracing among migrants and the foreign-born population. *Euro Surveill* 2009;14:19153.
16 MacIntyre CR, Plant AJ. Impact of policy and practice on the effectiveness of contact screening for tuberculosis. *Prev Med* 1998;27:830–7.
17 Borraccino A, Migliore E, Piccioni P, *et al*. Yield of tuberculosis contact investigation in a low-incidence country. *J Infect* 2014;68:448–54.
18 Sloot R, Schim van der Loeff MF, Kouw PM, *et al*. Risk of tuberculosis after recent exposure. A 10-year follow-up study of contacts in Amsterdam. *Am J Respir Crit Care Med* 2014;190:1044–52.
19 Zellweger JP, Sotgiu G, Block M, *et al*. Risk assessment of tuberculosis in contacts by IFN-gamma release assays. A tuberculosis network European trials group study. *Am J Respir Crit Care Med* 2015;191:1176–84.
20 National Tuberculosis Controllers Association; Centers for Disease Control and Prevention (CDC). Guidelines for the investigation of contacts of persons with infectious tuberculosis. Recommendations from The National Tuberculosis Controllers Association and CDC. *MMWR Recomm Rep* 2005;54(RR15):1–47.
21 *Recommendations for investigating contacts of persons with infectious tuberculosis in low- and middle-income countries*. Geneva: World Health Organization, 2012.
22 Sandgren AMI. *Management of contacts of MDR TB and XDR TB patients*. Stockholm: European Centre for Disease Prevention and Control, 2012.
23 Williams RL. A note on robust variance estimation for cluster-correlated data. *Biometrics* 2000;56:645–6.
24 Young KH, Ehman M, Reves R, *et al*. Tuberculosis Contact Investigations—United States, 2003–2012. *MMWR Morb Mortal Wkly Rep* 2016;64:1369–74.
25 Story A, Murad S, Roberts W, *et al*. Tuberculosis in London: the importance of homelessness, problem drug use and prison. *Thorax* 2007;62:667–71.
26 Marks SM, Taylor Z, Qualls NL, *et al*. Outcomes of contact investigations of infectious tuberculosis patients. *Am J Respir Crit Care Med* 2000;162:2033–8.
27 Duarte R., Neto M, Carvalho A, *et al*. Improving tuberculosis contact tracing: the role of evaluations in the home and workplace. *Int J Tuberc Lung Dis* 2012;16:55–9.
28 Fiske CT, Yan FX, Hirsch-Moverman Y, *et al*. Risk factors for treatment default in close contacts with latent tuberculous infection. *Int J Tuberc Lung Dis* 2014;18:421–7.

THORAX

An evaluation of tuberculosis contact investigations against national standards

Sean M Cavany, Tom Sumner, Emilia Vynnycky, Clare Flach, Richard G White, H Lucy Thomas, Helen Maguire and Charlotte Anderson

Thorax 2017 72: 736-745 originally published online April 7, 2017
doi: 10.1136/thoraxjnl-2016-209677

Addendum

Clarifications

- On page 54 in the first paragraph of the results, we state that there were 2716 pulmonary cases, and also that there were 971 pulmonary smear positive cases, 1095 pulmonary smear negative cases and 478 pulmonary cases with unknown smear, which totals 2548 pulmonary cases. The reason for this discrepancy is that we applied the exclusion criteria separately for pulmonary cases and for all cases combined. The former figure (2716) refers to the number of pulmonary cases included in the pulmonary-only dataset, whereas the latter (2548) refers to the number of pulmonary cases included in the all-cases dataset.
- Similarly, in figure one more cases are excluded from the pulmonary-only dataset for being probable incidents than were excluded in the all-cases dataset for that reason. This is again a consequence of applying the exclusion criteria separately, meaning that these two figures represent slightly different time periods.
- The reason for the separate pulmonary-only and all-cases datasets was so that indicators 1 and 2, which refer only to pulmonary index cases, and indicators 3 and 4, which refer to all cases, could be estimated using the maximum amount of available data. Several other apparent discrepancies between the text and the tables occur for similar reasons.

Corrections

- There is a discrepancy between the exclusion criteria shown in figure 1 and the methods: the figure states that sectors are included if they reported 90% or more of their cases, whereas the text states 80%. The figure given in the text (80%) is correct.
- In the fourth row of figure 4, the entry in the second column should read “index cases found through previous contact investigations” rather than “not report to cohort review”.

3. Second paper: Transmission events revealed in tuberculosis contact investigations in London

Preamble

The first paper of this thesis showed that at a practical level (i.e. the things within clinics' control), contact tracing in London is being undertaken well when compared to other countries, as measured by both indicator 1 (the proportion of pulmonary cases with at least one contact identified) and indicator 2 (the proportion of identified contacts of pulmonary cases that are evaluated), as both of these indicators are typically as high or higher than values found in other locations. It also showed that the proportion of cases found to have either active TB or LTBI during contact investigations in London is high relative to other high-income, low-incidence nations or regions of the UK. For instance, 2.9% of all evaluated contacts of pulmonary cases had active TB, more than double the average yield of contact investigations amongst high-income countries (1.4%[108]). However, this says little about the impact of contact tracing as in the UK approximately 95% of cases are notified, so cases would probably be found anyway if they had not been found through contact tracing[109]. Additionally, part of the potential impact of contact tracing is derived from preventing cases, i.e. those which would have been infected by contacts found to have active TB had they been infectious for longer. In order to get a better understanding of the impact, it would help to know how many of the cases found are involved in chains of ongoing transmission, and also whether contact tracing found them sooner and by how much. This is what the second paper of this thesis sought to do for London. This paper is in review at Scientific Reports.

To do this I used the same dataset as that used in the first paper (i.e. the LTBR, a web-based register of cases diagnosed in London which includes some contact tracing outcomes), but focus on pairs of cases and contacts diagnosed with TB, and supplement the dataset with mycobacterial-interspersed-repetitive-units variable-number-tandem-repeats (MIRU-VNTR) typing data to look at transmission. When two cases have indistinguishable isolates and are known to have epidemiological links (e.g.

they know each other, or work in the same place), then transmission between them is usually described as confirmed[27]. In our study, all case-contact pairs had epidemiological links by definition, and so when they shared isolates this was probably due to recent transmission.

Two studies in the US (2002, 2004) calculated the proportion of cases found through contact tracing which had an indistinguishable isolate from their index[110,111], and a recent study in England (2017) attempted to estimate the proportion of pairs which had indistinguishable isolates and shared an address[27]; all three of these studies are discussed in the paper. They calculated that 71%, 70% and 75% of pairs shared isolates respectively. Additionally, a study in Cape Town (2004) calculated that in households with more than one case, 46% of cases had an isolate that was indistinguishable from that of at least one other member of their household[112]. This implies that less than or equal to 46% of household case-contact pairs would have indistinguishable isolates in this setting. We would expect the proportion to be lower in high incidence settings than low incidence settings, as this study found, because the risk of community transmission is likely to be increased as the proportion of those encountered in the community that have TB will be greater. A study in Poland (2012) found an intermediate figure of 63% of cases due to intra-household transmission[113]. However, to our knowledge, no study has attempted to estimate the proportion of those cases found through contact tracing that are due to recent transmission in a UK setting, nor has any study attempted to estimate how long contact investigations typically take in a UK setting. This study provides the first estimates of both of these things in a UK setting, and describes groups for whom contact investigations are longer and for whom cases found through contact investigations are more likely to be due to transmission. The results contain the important finding that a large minority (20%) of cases found through contact tracing had an isolate which did not match their index case, implying that cases of TB are found through contact tracing irrespective of whether transmission has taken place.

Cover sheet

See next page

London School of Hygiene & Tropical Medicine
Keppel Street, London WC1E 7HT
www.lshtm.ac.uk

Registry
T: +44(0)20 7299 4646
F: +44(0)20 7299 4656
E: registry@lshtm.ac.uk

LONDON SCHOOL of HYGIENE &TROPICAL MEDICINE

RESEARCH PAPER COVER SHEET

PLEASE NOTE THAT A COVER SHEET MUST BE COMPLETED FOR EACH RESEARCH PAPER INCLUDED IN A THESIS.

SECTION A – Student Details

Student	Sean M. Cavany
Principal Supervisor	Emilia Vynnycky
Thesis Title	Using data analysis and mathematical modelling to study tuberculosis contact tracing in London, with reference to the national strategy and guidance

If the Research Paper has previously been published please complete Section B, if not please move to Section C

SECTION B – Paper already published

Where was the work published?			
When was the work published?			
If the work was published prior to registration for your research degree, give a brief rationale for its inclusion			
Have you retained the copyright for the work?*	Choose an item.	Was the work subject to academic peer review?	Choose an item.

**If yes, please attach evidence of retention. If no, or if the work is being included in its published format, please attach evidence of permission from the copyright holder (publisher or other author) to include this work.*

SECTION C – Prepared for publication, but not yet published

Where is the work intended to be published?	Scientific Reports
Please list the paper's authors in the intended authorship order:	Sean M. Cavany, Emilia Vynnycky, Tom Sumner, Neil Macdonald, H. Lucy Thomas, Jacqui White, Richard G. White, Helen Maguire, Charlotte Anderson
Stage of publication	**Undergoing revision**

SECTION D – Multi-authored work

For multi-authored work, give full details of your role in the research included in the paper and in the preparation of the paper. (Attach a further sheet if necessary)	I conceived, deisgned and carried out the work, and wrote the first draft of the paper.

Student Signature: [redacted] **Date:** 28/03/18

Supervisor Signature:  **Date:** 29.03.18

Paper

See next page

Transmission events revealed in tuberculosis contact investigations in London

Word count: 3163

Authors: Sean M. Cavany*[1,2], Emilia Vynnycky[1,2], Tom Sumner[1], Neil Macdonald[4], H. Lucy Thomas[3], Jacqui White[5], Richard G. White[1], Helen Maguire[4,6], Charlotte Anderson[4]

Author affiliations:

1. TB Modelling Group, TB Centre and CMMID, Faculty of Epidemiology and Population Health, London School of Hygiene and Tropical Medicine, London.
2. Statistics, Modelling and Economics Department, Public Health England, London
3. Respiratory Diseases Department, Public Health England, London
4. Field Epidemiology Service, Public Health England, London
5. North Central London TB Service, Whittington Health NHS Foundation Trust, London
6. Institute for Global Health, University College London

***Corresponding author:** Sean M. Cavany, email: sean.cavany@lshtm.ac.uk

Abstract

Contact tracing is a key part of tuberculosis prevention and care, aiming to hasten diagnosis and prevent transmission. The proportion of case-contact pairs for which recent transmission occurred and the typical timespans between the index case and their contact accessing care are not known; we aimed to calculate these. We analysed individual-level TB contact tracing data, collected in London from 20/01/2011-31/12/2015, linked to tuberculosis surveillance and MIRU-VNTR 24-locus strain-typing information. Of pairs of index cases and contacts diagnosed with active tuberculosis, 85/314 (27%) had strain typing data available for both. Of these pairs, 79% (67/85) shared indistinguishable isolates, implying probable recent transmission. Of pairs in which both contact and the index case had a social risk factor, 11/11 (100%) shared indistinguishable isolates, compared to 55/75 (75%) of pairs in which neither had a social risk factor (*P*=0.06). The median time interval between the index case and their contact accessing care was 42 days (IQR: 16, 96). As over 20% of pairs did probably not involve recent transmission between index case and contact, the effectiveness of contact tracing is not necessarily limited to those circumstances where the index case has transmitted disease to their close contacts.

Key words: tuberculosis, contact tracing, London, United Kingdom, MIRU-VNTR, transmission.

Introduction

Contact tracing, the systematic screening of contacts of tuberculosis (TB) cases, is an important part of TB control in the UK and other high-income, low-incidence countries, and is highlighted as a key element of the Public Health England (PHE)/National Health Service (NHS) England collaborative tuberculosis strategy[1]. It aims to reduce transmission from and morbidity in contacts with active TB, and find contacts with latent *M. tuberculosis* infection that are eligible for preventive therapy[2]. London accounts for nearly 40% of England's TB cases[3], and during 20/01/2011-31/12/2015, 13 692 TB cases were notified in London, with a median of four contacts identified per case[3,4].

If a contact is diagnosed with active TB, there may have been recent transmission between index case and contact, or the contact may have been infected, recently or historically, by another source. Understanding the extent to which cases found through contact tracing are due to recent transmission shows the value of contact tracing in interrupting ongoing transmission. Where the proportion of cases identified due to recent transmission is high, this is more likely to reflect active and ongoing transmission. Reducing the time between the index case and their contact accessing care means earlier diagnosis of an active case, benefitting the individual and reducing the risk of onward transmission from an infectious source. Whilst around 5% of cases in London are found through contact tracing [unpublished data], less is known if those cases are due to recent transmission, nor of the typical timescales (such as the median time between the index case and their contact accessing care) involved in contact investigations.

Since 2010 in the UK, isolates from culture confirmed TB cases have been routinely strain-typed using 24-loci mycobacterial-interspersed-repetitive-units – variable-number-tandem-repeats (MIRU-VNTR) strain-typing[5]. Using strain-typing data, transmission between cases is considered not to have occurred if their isolates are distinct and probable if the isolates are indistinguishable and supported by evidence of contact between cases. Previous studies using strain-typing data in the United States found that around 70% of contacts with TB may have been infected by or have infected their index

case[6-8], but no studies have estimated this in the UK. Whilst a recent study in France estimated the time between index case notification and the contact being screened to be 48 days[9], to our knowledge no studies have directly estimated the time interval between index case notification and diagnosis of TB in a contact in the UK.

We aimed to describe the extent of transmission that was identified through contact tracing, and the time taken from index case identification to finding the active case among their contacts. This was in order to provide TB services with evidence for the value of contact investigations, and where efforts might be targeted or strengthened in order to give the biggest benefit.

Our first objective was to estimate the proportion of index case-contact pairs for whom probable recent transmission had occurred, and determine factors associated with differences in this proportion. The second objective was to estimate the time interval between the index case and the contact accessing care, as a proxy measure for contact investigation length, and determine which factors are associated with longer or shorter intervals. An additional aim of the study was to understand whether the patient characteristics of those contact tracing pairs found to be due to recent transmission were also common among the pairs tracing for whom investigations are longer.

Methods

Dataset & inclusion criteria. The primary data source was the London TB Register (LTBR; a web-based register containing demographic and clinical data on all TB cases notified in London since 2002). TB cases are notified in England either if they are culture confirmed, or based on the clinician's decision to treat with a full course of anti-TB therapy. From 2012, the LTBR has incorporated data on contact tracing from 'cohort review'; this is a quarterly case management and contact tracing appraisal conducted by clinical staff for TB cases, introduced incrementally across London from 2010[10]. This paper utilizes contact tracing data collected as part of cohort review, linked to surveillance data from the LTBR and strain typing data held by PHE.

Contacts identified during contact investigations and diagnosed with active TB are linked in the LTBR to the index case of the investigation. For the period of the study (20/01/2011-31/12/2015) contact tracing was conducted according to national guidance CG117[11], which recommended screening household and other close contacts of all cases. During the study period, contact investigations began immediately after the diagnosis of the index case, whereupon the nurse asked the case for a list of close contacts. Screening begins with symptom-screen; for asymptomatic contacts this is followed by a tuberculin skin test (TST) or interferon-gamma release assay (IGRA) in those aged under 35 years and consideration of a chest X-ray (CXR) in those aged 35 years and over. Those with a positive symptom-screening, TST/IGRA result or CXR are evaluated for signs of active TB. Those with LTBI are considered for preventive therapy.

The study population included any pair for which the index case was notified in the study period and had at least one contact diagnosed with active TB. Pairs were excluded if the linked contact began their current episode of care prior to their index case. For the first objective, analysis was further limited to pairs for which both contact and index case had strain typing results.

In London, isolates of culture-confirmed TB cases are typed at the PHE Mycobacterium Reference Laboratory, with results matched with surveillance data using the Enhanced Matching System

81

(EMS)[12]. Isolates are defined as indistinguishable if at least one in the cluster had 24 loci typed and all others had 23 or 24 loci typed and matched in all typed loci.

Within the LTBR, the episode of care start date is the date when a patient was first seen by the clinic at which they were notified, and is a mandatory field recorded for all patients.

Analysis. We took the following approach for the two objectives:

1. To determine the proportion of strain typed pairs for whom index and contact had indistinguishable isolates. This was stratified by country of birth, site of disease & smear status, social risk factors (current or history of: imprisonment, drug or alcohol misuse and/or homelessness), age and sex, based on attributes of the index, contact or shared by both. We evaluated the sensitivity of these results to excluding contacts not recorded as presenting through contact screening, and to including pairs for whom the contact accessed care first. There were four instances when the same contact with TB was named by two different index cases; we evaluated the sensitivity of results to the exclusion of these links so that each contact only appeared in one pair.
2. To estimate the median and distribution of the time interval between episode of care start dates of index and contact(s). We measured the interval in days between episode of care start date of the index case and contact, and explored whether the factors mentioned in objective one were associated with longer time intervals, in each instance adjusted for the site of disease of the index. Differences in medians were assessed for significance using Mood's median test.

Software: All data were analysed using Microsoft Excel 14.0 and Stata 13.1.

Results

Comparison of excluded and included data. There were 451 cases of TB in the study period recorded as having one or more contacts diagnosed with TB (286 when restricting to strain typed pairs), resulting in 697 potential case-contact pairs (406 when restricting to strain typed pairs) (Figure 1). After applying the inclusion criteria, 85 pairs (21%) were included in the analyses for objective one and 314 pairs (45%) for objective two, corresponding to 81/286 (28%) and 247/451 (55%) of all strain-typed index cases and all index cases, respectively. There were 44 index cases included in more than one case- contact pair overall, and three in more than one strain-typed pair.

In the analyses for each objective, a varying number of pairs were also removed where there were insufficient data on the demographic trait or clinical characteristic of interest. Those included in analysis for objective one had a different ethnic profile to those excluded (Table 1). Index cases included for objective two were more likely to be male, have pulmonary disease or be 15 years old or above, than those excluded (Table 1). As children are less likely to be culture positive than adults, adults were over-represented in both included and excluded pairs for analysis of objective one relative to objective two.

Pairs of index cases and contacts with indistinguishable isolates. Overall, 67/85 (79%) of contacts who were diagnosed with TB had indistinguishable isolates from their index case. This was similar across a range of clinical and demographic factors relating to both the index case and the contact (Table 2). For pairs in which both case and contact had a social risk factor, 11/11 (100%) pairs had indistinguishable isolates (compared to 55/70 (75%) where either or both had no social risk factors, P=0.07). For all other factors the p-value for the association with indistinguishable isolates was above 0.1.

If we remove pairs for whom the contact was listed as presenting through a route other than contact tracing, the proportion who have indistinguishable isolates was similar at 48/61 (79%). If we include pairs for whom the contact accessed care prior to the index, the proportion is slightly lower at

85/112 (76%), and the association between a higher proportion of pairs with indistinguishable isolates and the presence of social risk factors is weaker. Removing pairs which contain contacts already named in another pair did not change the results.

Timescales of contact tracing. The median time interval between episode of care start dates for index and contact was 42 days (interquartile range (IQR): 16, 96). The time interval was slightly shorter for pulmonary index cases (41 days, IQR: 16, 96) compared to non-pulmonary index cases (56 days, IQR: 15, 103), and shorter for smear positive pulmonary index cases (37 days, IQR 14, 91) compared to smear negative pulmonary index cases (47 days, IQR 20, 96); the significance level of these differences was *P*=0.12 and *P*=0.57 respectively. There was also no evidence of a difference when comparing smear positive pulmonary, smear negative pulmonary and non-pulmonary index cases (*P*=0.25). The median time interval was 42 days (IQR: 14, 96.5) among contacts diagnosed with pulmonary TB, and 47 days (IQR: 19.3, 94.5) for non-pulmonary contacts (*P*=0.69).

This time interval between accessing care had a positively skewed distribution (Figure 2) with most contacts accessing care within six weeks (52% for pulmonary index cases, and 33% for non-pulmonary index cases) and 39% (122/314) of all contacts first accessing care within one month of their index.

Contacts that were UK-born or recent migrants (entered within two years; numbers were small in this group) (aOR: 2.0 [1.2, 3.4] and 3.9 [1.3, 12] respectively) were more likely to be identified and access care within six weeks of their index case (Table 3), compared to longer-term migrants. Adult contacts (aged greater than 14 years) (aOR: 0.38 [0.23, 0.65]) were less likely to have a short investigation compared to children.

Discussion

Our analysis estimates almost 80% of contacts diagnosed with TB and strain-typed in London are part of a recent transmission event involving the index case and the contact. This implies that 20% of contact investigations that find new cases do so even though no transmission has occurred between the index case and their contact. When both the case and the contact had one or more social risk factors, recent transmission was more likely to have occurred. The median time-interval between index and contact starting care was six weeks (42 days). Contacts who were adults (compared to children) or non-UK born migrants who entered >2 years previously (compared to UK born) were more likely to have an interval longer than six weeks. Contacts with social risk factors were not associated with delayed intervals of longer than six weeks.

A limitation of this study was the small number of pairs (85) where cases and contacts had strain typing results (21% of 406 pairs with strain-typed index cases; 12% of all 697 pairs). As a result, our analysis may have lacked power to discern all associations. In addition, findings may not be generalizable to other TB patients as for the included pairs the index cases were almost all pulmonary (95%), were more often of white or other ethnicity, more often UK born (35%), male (73%) and were almost all (98%) adults. We may have over-estimated the contribution of social risk factors as a result of this inclusion bias.

We may have overestimated the proportion of indistinguishable isolates compared to a higher resolution method such as whole genome sequencing (WGS)[13]. However, the combination of microbiological and epidemiological links is good evidence that patients with indistinguishable isolates represent recent transmission.

We were also not able to include 176 pairs where cases were identified through contact tracing, but the contact was not linked in the LTBR.

The proportion of index cases and contacts in London with indistinguishable isolates (79%) was higher than estimates of 70-71% in three previous studies in the United States of America[6-8]. The first of these studies found that pairs with unconfirmed transmission were more likely to include smear-negative source cases and a foreign-born secondary case, and less likely to include secondary cases under the age of 15 years, compared to pairs with confirmed transmission; the significance level of these relationships in our study was 0.75, 0.19 and 0.29 respectively. A recent UK-wide study found that 75% of pairs of cases with the same address had indistinguishable isolates[14], supporting results presented here. Other recent studies based on MIRU VNTR typing data estimated the proportion of all cases due to recent transmission to be 34% in London and 10% in North-West England[15,16]. While these results are not directly comparable with ours, both findings support the notion that contacts of new cases of TB are more likely to have been recently exposed to tuberculosis than other TB cases, and so contact tracing is an essential tool to identify individuals at increased risk of disease.

A recent study in France found the mean period from index notification to completion of contact screening was 48 days[9]. A previous modelling study used a value of a quarter of a year (in rural Saskatchewan, Canada: a setting with greater barriers to screening than London)[17], more than twice the median of 42 days found in London. A study in the Netherlands estimated the incubation period of TB cases (with any site of disease) to be around 1.3 years (95% CI: 1.1-1.4), with 30% of cases having an incubation period of less than six months[18]. This suggests that, for pairs with pulmonary index cases where there has been recent transmission from index to contact, for every week that a contact investigation is shortened, it may be possible to find and prevent disease in ~1% $\left(=\frac{30\%}{26\ \text{weels}}\right)$ of infected contacts. However, a modelling study found that shorter contact investigations had little population-level impact[17].

That case-contact pairs with social risk factors were more likely to have indistinguishable isolates than those without, coupled with the higher prevalence of disease seen in those with social risk

factors[19], suggests ongoing transmission occurring in this group in London, supporting previous studies[15,20]. This supports the emphasis given to this group in national guidelines, in particular recommendations for a programme of active case-finding amongst homeless and drug-users using a mobile X-ray unit , and to coordinate contact investigations around those with social risk factors in locations frequented by the index[11]. The higher proportion of indistinguishable isolates amongst these pairs may be because the index cases were more likely to be infectious [19,21] and for longer, or because of the contact's increased susceptibility.

Previous analysis of contact tracing outcomes in London found the prevalence of active disease amongst contacts of pulmonary patients to be 2.6%[4]. Assuming that, as found here, 80% of contacts of pulmonary index cases that develop active TB do so following recent transmission from their index case, this suggests that 0.52% of the contacts of pulmonary cases developed disease without transmission having occurred between the case-contact pair. This proportion can be seen as the risk of TB disease in contacts of TB patients that comes from sources other than the known index, perhaps due to shared risk factors and/or community contacts. A study in London in 2006 found the prevalence of TB amongst the homeless population to be 0.79%, amongst problem drug users to be 0.35%, and amongst prisoners to be 0.21%. Our study suggests that, even after removing the effect of transmission from the known index case, the risk of TB in a close contact of a TB case is high when compared with other risk groups in London[19]. However, contacts may self-present more quickly than the aforementioned risk groups.

In London, some clinics aim to screen contacts of smear positive patients within two weeks of contact identification, and contacts of non-smear positive cases within six weeks. While some contacts will be identified subsequent to diagnosis of the index case after the building of a relationship between case manager and patient, it is possible that these targets are not met for some contacts: 12 pairs had a time interval of more than six months between index case and contact accessing care.

WGS of TB isolates has recently been rolled out across the UK[22]. This will enable studies looking at transmission and clustered cases to link cases more accurately, which will in turn enable greater understanding of transmission networks and better target the allocation of resources.

We only found two pairs with an index case with non-pulmonary TB in which probable recent transmission occurred, but there were only four typed pairs with non-pulmonary index cases in total. Further research to understand the impact of 2016 changes to guidance, which no longer recommends screening contacts of non-pulmonary, non-laryngeal index cases[11], would be useful. The results presented here could be utilized in modelling studies to assess the impact of contact tracing in different groups. Improved linkage of contacts would enable future research to have sufficient power to find risk factors with greater significance for the types of contacts which are most likely to have been part of a transmission event.

Whilst those pairs with social risk factors are more likely to involve recent transmission, these contacts may also be harder to identify and reach[4]. This highlights the importance of services such as Find & Treat in identifying these patients in London[23]. While we have quantified the typical times between index cases and their contacts accessing care, our study was not able to estimate the impact of shortening contact investigations, and when this is greatest. Further work to quantify this would be useful, perhaps incorporating mathematical modelling as well as data on the infectiousness of contacts. Finally, our results show that on the whole contact tracing in London happens in a timely manner thanks to the great effort of healthcare staff.

Figures and tables

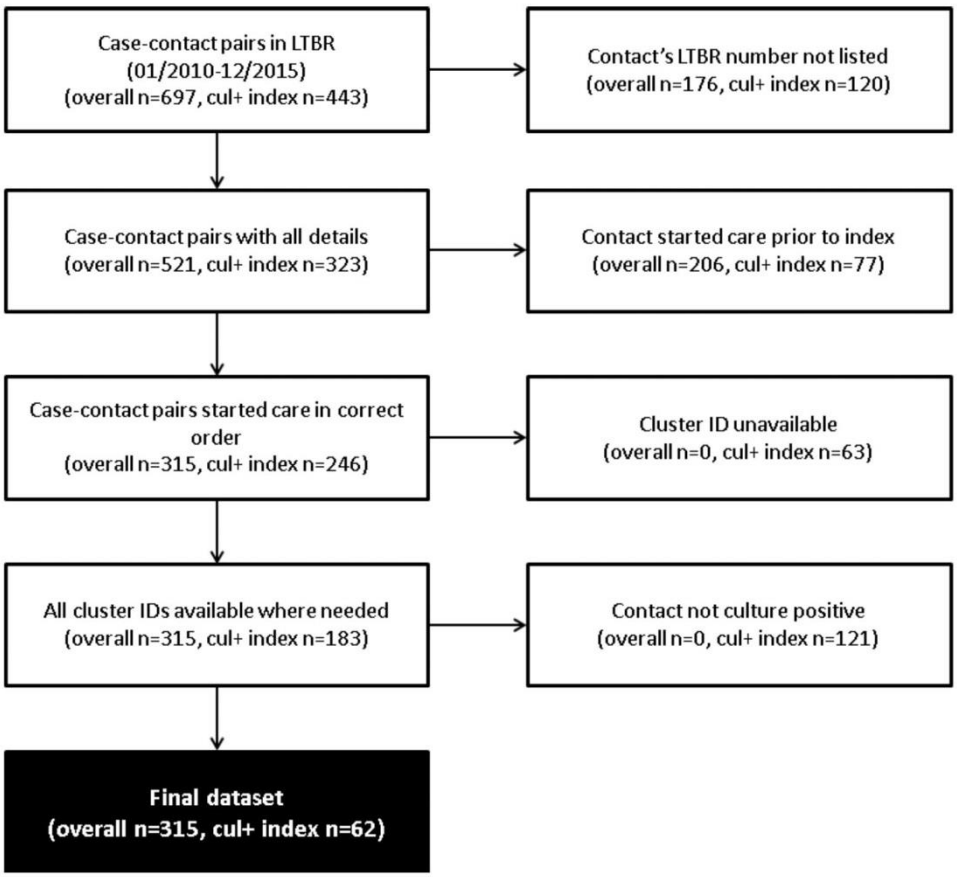


Figure 1: Flowchart of included and excluded case-contact pairs, for objectives one ('typed index') and objective two ('overall'). Note that boxes three and four on the right were not exclusion criteria for objective two.

Table 1: comparison of characteristics of index cases included in the analyses and those with one or more contacts who were diagnosed with TB but were excluded from the analyses. The analyses involving strain typing data (objective 1) only included index cases that had a strain typed isolate and at least one contact who also had a strain typed isolate. Percentages are column percentages except for the total row. P-value is chi-squared p-values for differences between groups

Factor		Analyses of indistinguishable isolates of case-contact-pairs (objective 1)			Analyses of time between episode of care start date of index case and contact (objective 2)		
		Number included (%)	Number excluded (%)	p-value	Number included (%)	Number excluded (%)	p-value
Total		81 (28%)	205 (72%)	N/a	247 (55%)	204 (45%)	N/a
UK-born?	Yes	28 (35%)	50 (24%)	0.22	82 (33%)	70 (34%)	0.68
	No	44 (54%)	127 (62%)		135 (55%)	103 (50%)	
	No, recent migrant (<2 years)	9 (11%)	28 (14%)		30 (12%)	29 (14%)	
Ethnicity	Bangladeshi	1 (1.2%)	6 (2.9%)	0.02	4 (1.6%)	5 (2.5%)	0.42
	Black-African	15 (19%)	62 (30%)		65 (26%)	63 (31%)	
	Black-Caribbean	2 (2.5%)	9 (4.4%)		10 (4.1%)	6 (2.9%)	
	Black-Other	5 (6.2%)	3 (1.5%)		11 (4.5%)	5 (2.5%)	
	Chinese	0 (0%)	3 (1.5%)		2 (0.81%)	4 (2.0%)	
	Indian	21 (26%)	50 (25%)		53 (22%)	54 (26%)	
	Pakistani	1 (1.2%)	13 (6.4%)		15 (6.1%)	12 (5.9%)	
	White	17 (21%)	25 (12%)		34 (14%)	26 (13%)	
	Other	19 (23%)	33 (16%)		52 (21%)	29 (14%)	
Sex	Male	59 (73%)	126 (61%)	0.07	164 (66%)	106 (52%)	<0.01
	Female	22 (27%)	79 (39%)		83 (34%)	98 (48%)	
Site of disease	Pulmonary	77 (95%)	179 (87%)	0.05	211 (85%)	150 (74%)	<0.01
	Non-pulmonary	4 (4.9%)	26 (13%)		36 (15%)	54 (26%)	
Social Risk Factor	History of homelessness	7 (8.6%)	14 (6.9%)	0.61	15 (6.2%)	9 (4.5%)	0.45
	History of imprisonment	5 (6.3%)	6 (3.0%)	0.20	10 (4.1%)	4 (2.0%)	0.20
	History of drug use	11 (14%)	19 (9.4%)	0.30	21 (8.6%)	12 (6.0%)	0.29
	Alcohol misuse	7 (8.9%)	11 (5.5%)	0.30	15 (6.3%)	7 (3.5%)	0.19
Age	15 years old or over	79 (98%)	192 (94%)	0.19	216 (87%)	148 (73%)	<0.01
	Under 15 years old	2 (2.5%)	13 (6.3%)		31 (13%)	56 (27%)	

Table 2: The proportion of contacts diagnosed with TB who share a strain with their index case. Only index cases and contacts who both had typed isolates are included. The denominator is the number of case-contact pairs for whom that risk factor applies and the numerator is the number of these that have indistinguishable isolates.

Risk factors	(Number pairs with indistinguishable isolates) / (total number of pairs) (%)	p-value
Overall	67/85 (79%)	N/a
Index case pulmonary	65/81 (80%)	0.15
Index case extrapulmonary	2/4 (50%)	
Index case smear positive pulmonary	54/68 (79%)	0.75
Index case smear negative pulmonary	10/12 (83%)	
Index case UK born	26/30 (87%)	0.19
Index case non-UK born	41/55 (75%)	
Contact UK born	30/35 (86%)	0.19
Contact non-UK born	37/50 (74%)	
Both index and contact UK born	20/23 (87%)	0.26
Either index or contact or both non-UK born	47/62 (76%)	
Both index and contact non-UK born	31/43 (72%)	0.12
Either index, contact or both UK born	36/42 (86%)	
Both index and contact have one or more social risk factors	11/11 (100%)	0.06
Either index, contact or both have no social risk factors	55/70 (75%)	
Child contact	10/11 (91%)	0.29
Adult contact	57/74 (77%)	
Index case female	17/23 (74%)	0.50
Index case male	50/62 (81%)	
Contact female	30/38 (79%)	0.98
Contact male	37/47 (79%)	
Both index and contact female	7/10 (70%)	0.76
Both index and contact male	27/34 (79%)	
Index and contact different sex	33/41 (80%)	

Table 3: Comparison of characteristics of contacts whose episode of care start date is six weeks or less after their index case with those of contacts whose episode of care start date is more than six weeks after that of their index case. Percentages are within-group column percentages except for the total row. All odds ratios are adjusted for the site of disease of the index except for site of the disease of the contact, where it is omitted due to collinearity.

Factor		Number (%) of contacts with short or long time intervals between index case and contact accessing care		p-value	Adjusted odds ratio for investigation being short (95%confidence interval)
		six weeks or less	More than six weeks		
Total		142 (51%)	138 (49%)	N/a	N/a
UK-born?	Yes	78 (55%)	55 (40%)	<0.01	2.03 [1.24, 3.36]
	No	51 (36%)	77 (56%)		1
	No, recent migrant (<2 years)	13 (9.2%)	5 (3.7%)		3.88 [1.29, 11.7]
Ethnicity	Indian	25 (18%)	33 (24%)	0.06	1
	Black-African	52 (37%)	33 (24%)		1.95 [0.98, 3.89]
	White	24 (17%)	17 (12%)		1.59 [0.69, 3.62]
	Other	41 (29%)	55 (40%)		0.90 [0.46, 1.77]
Sex	Male	78 (55%)	73 (53%)	0.57	1.15 [0.71, 1.85]
	Female	64 (45%)	65 (47%)		1
Site of disease	Pulmonary	98 (69%)	87 (63%)	0.98	1.28 [0.78, 2.12]
	Non-pulmonary	44 (31%)	51 (37%)		1
Social Risk Factor	History of homelessness	3 (2.1%)	4 (3.0%)	0.66	0.71 [0.15, 3.28]
	History of imprisonment	2 (1.4%)	6 (4.4%)	0.20	0.34 [0.07, 1.74]
	History of drug use	7 (5.0%)	5 (3.7%)	0.60	1.44 [0.44, 4.75]
	Alcohol mis use	3 (2.1%)	4 (3.0%)	0.56	0.63 [0.14, 2.89]
	Any social risk factor	11 (7.9%)	11 (8.4%)	0.83	0.91 [0.38, 2.19]
Age	15 years old or over	75 (53%)	99 (72%)	<0.01	0.46 [0.28, 0.76]
	Under 15 years old	67 (47%)	39 (28%)		1

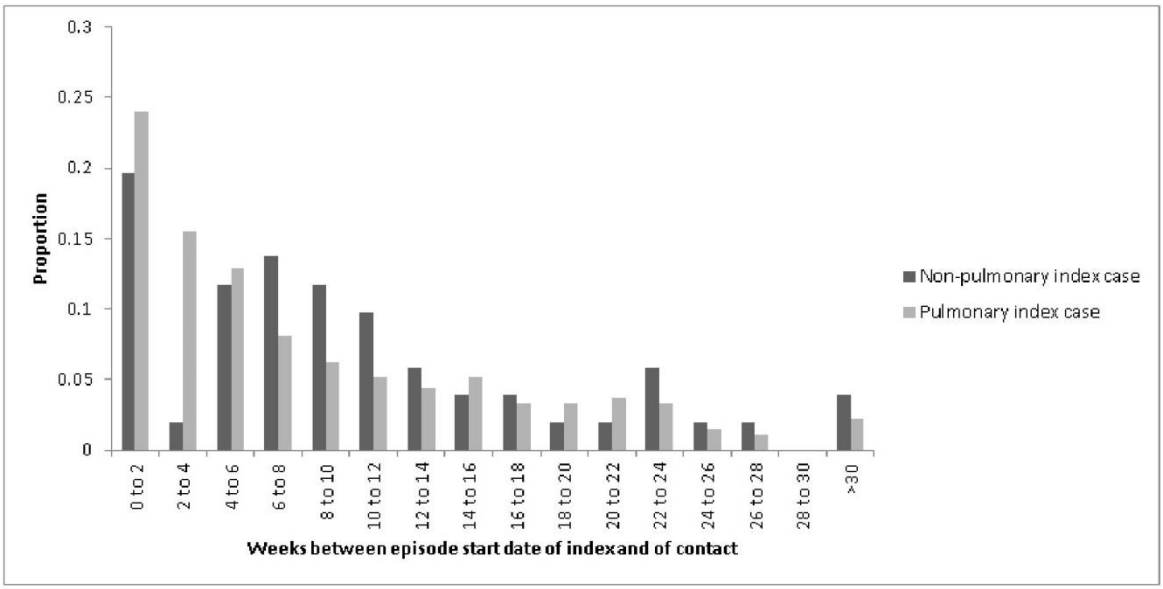


Figure 2: Distribution of the time in weeks between the episode of care start date of an index case and that of the contact, delineated by the site of disease of the index. Ranges include the upper bound, and exclude the lower bound, after the first bar.

Acknowledgements

The authors would like to thank Jacqui Carless, Lamya Kanfoudi and others at PHE field Epidemiology London for maintaining the LTBR; all nurses and other healthcare workers involved with collecting the data and presenting it at cohort review; the TB modelling group at LSHTM for advice on presentation and approach; and Lindsay Serene for comments on the manuscript.

Statements

Competing interests: We declare that the authors have no competing interests as defined by Nature Research, or other interests that might be perceived to influence the results and/or discussion reported in this paper.

Funding sources: This work was supported by a joint Public Health England-London School of Hygiene and Tropical Medicine studentship in Infectious Disease Modelling to SMC; the UK Medical

Research Council (MRC) and the UK Department for International Development (DFID) , under the MRC/DFID Concordat agreement that is also part of the EDCTP2 programme supported by the European Union [MR/J005088/1] to RGW; the Bill and Melinda Gates Foundation [TB Modelling and Analysis Consortium: OPP1084276 to RGW, and SA Modelling for Policy: OPP1110334 to RGW and TS], and UNITAID [4214-LSHTM-Sept15; PO #8477-0-600] to RGW. CA, EV, HM, NM and LT are employed by PHE, a government agency, and received no other source of funding.

Author Contributions: SMC conceived and designed the work with input from TS, EV, RGW, HLT, HM & CA. CA, HM, JW and NM are responsible for the acquisition and maintenance of the data. SMC undertook the analysis with advice from all other authors. All authors contributed to the interpretation of the data. SMC wrote the first draft of the paper and all authors contributed to subsequent drafts. All authors approve the work for publication and agree to be accountable for the work.

Data availability: Aggregate data that support the findings of this study are available on reasonable request from the corresponding author (SMC). The individual level data generated and analysed during the current study are not publicly available as the data were collected in adherence with the legal framework governing use of confidential personally identifiable information.

Ethics: Ethical approval was not required. The data analysed were routinely collected surveillance data held by PHE under Section 251 of the NHS Act 2006. All records were anonymised before analysis.

References

1 Collaborative Tuberculosis Strategy for England: 2015 to 2020. (Public Health England, NHS England, London, 2015).

2 Erkens, C. G. *et al.* Tuberculosis contact investigation in low prevalence countries: a European consensus. *The European respiratory journal* **36**, 925-949, doi:10.1183/09031936.00201609 (2010).

3 Tuberculosis in London: Annual review (2015 data). (Public Health England, London, 2016).

4 Cavany, S. M. *et al.* An evaluation of tuberculosis contact investigations against national standards. *Thorax*, doi:10.1136/thoraxjnl-2016-209677 (2017).

5 Makki, S. L., M; Hamblion, E; Burkitt, A; Anderson, S; Thomas, L; Anderson, L; Maguire, H. (ed Board Public Health England Strain Typing Project) (Public Health England, London, 2014).

6 Bennett, D. E. *et al.* DNA fingerprinting of Mycobacterium tuberculosis isolates from epidemiologically linked case pairs. *Emerging infectious diseases* **8**, 1224-1229, doi:10.3201/eid0811.020420 (2002).

7 McNabb, S. J. *et al.* Added epidemiologic value to tuberculosis prevention and control of the investigation of clustered genotypes of Mycobacterium tuberculosis isolates. *American journal of epidemiology* **160**, 589-597, doi:10.1093/aje/kwh253 (2004).

8 Behr, M. A. *et al.* Transmission of Mycobacterium tuberculosis from patients smear-negative for acid-fast bacilli. *Lancet* **353**, 444-449 (1999).

9 Yanogo, P. K. *et al.* Factors associated with delayed screening of contacts of tuberculosis cases in the Somme, France. *Revue d'epidemiologie et de sante publique* **64**, 247-253, doi:10.1016/j.respe.2016.03.162 (2016).

10 Anderson, C. *et al.* Raising standards in UK TB control: introducing cohort review. *Thorax* **69**, 187-189, doi:10.1136/thoraxjnl-2013-203751 (2014).

11 Tuberculosis: clinical diagnosis and management of tuberculosis, and measures for its prevention and control. Guidance and guidelines. NICE [CG117]. 2011. *http://www.nice.org.uk/guidance/cg117*. Accessed 02 Feb 2018.

12 Aldridge, R. W., Shaji, K., Hayward, A. C. & Abubakar, I. Accuracy of Probabilistic Linkage Using the Enhanced Matching System for Public Health and Epidemiological Studies. *PloS one* **10**, e0136179, doi:10.1371/journal.pone.0136179 (2015).

13 Gardy, J. L. *et al.* Whole-genome sequencing and social-network analysis of a tuberculosis outbreak. *The New England journal of medicine* **364**, 730-739, doi:10.1056/NEJMoa1003176 (2011).

14 Lalor, M. K. *et al.* Recent household transmission of tuberculosis in England, 2010-2012: retrospective national cohort study combining epidemiological and molecular strain typing data. *BMC medicine* **15**, 105, doi:10.1186/s12916-017-0864-y (2017).

15 Hamblion, E. L. *et al.* Recent TB transmission, clustering and predictors of large clusters in London, 2010-2012: results from first 3 years of universal MIRU-VNTR strain typing. *Thorax* **71**, 749-756, doi:10.1136/thoraxjnl-2014-206608 (2016).

16 Saavedra-Campos, M. *et al.* Identifying areas and risk groups with localised Mycobacterium tuberculosis transmission in northern England from 2010 to 2012: spatiotemporal analysis incorporating highly discriminatory genotyping data. *Thorax*, doi:10.1136/thoraxjnl-2014-206416 (2015).

17 Tian, Y., Osgood, N. D., Al-Azem, A. & Hoeppner, V. H. Evaluating the effectiveness of contact tracing on tuberculosis outcomes in Saskatchewan using individual-based modeling. *Health education & behavior : the official publication of the Society for Public Health Education* **40**, 98S-110S, doi:10.1177/1090198113493910 (2013).

18 Borgdorff, M. W. *et al.* The incubation period distribution of tuberculosis estimated with a molecular epidemiological approach. *International journal of epidemiology* **40**, 964-970, doi:10.1093/ije/dyr058 (2011).

19 Story, A. *et al.* Tuberculosis in London: the importance of homelessness, problem drug use and prison. *Thorax* **62**, 667-671, doi:10.1136/thx.2006.065409 (2007).

20 Love, J. *et al.* Molecular epidemiology of tuberculosis in England, 1998. *The international journal of tuberculosis and lung disease : the official journal of the International Union against Tuberculosis and Lung Disease* **13**, 201-207 (2009).

95

21 Anderson, C., Story, A., Brown, T., Drobniewski, F. & Abubakar, I. Tuberculosis in UK prisoners: a challenge for control. *Journal of epidemiology and community health* **64**, 373-376, doi:10.1136/jech.2009.094375 (2010).

22 Walker, T. M. *et al.* Tuberculosis is changing. *The Lancet. Infectious diseases* **17**, 359-361, doi:10.1016/S1473-3099(17)30123-8 (2017).

23 Jit, M. *et al.* Dedicated outreach service for hard to reach patients with tuberculosis in London: observational study and economic evaluation. *Bmj* **343**, d5376, doi:10.1136/bmj.d5376 (2011).

Addendum

Clarifications

- When we refer to ‘contact investigation length’, we mean the typical timescale associated with contact investigation. Ideally this would be the time between when the contact is elicited and when they are screened, though in this paper we use the time between index case and contact diagnosis as a proxy for this timespan.

Corrections

- There are some discrepancies between the figure 1 and the first paragraph of the results; the figures given in the text are correct.

4. Third paper: Should NICE reconsider the 2016 UK guidelines on tuberculosis contact tracing? A cost-effectiveness analysis of contact investigations in London

Preamble

As mentioned in the discussion for both papers one and two, in 2016 the NICE guidelines changed to no longer recommend screening of contacts of non-pulmonary, non-laryngeal (ETB) index cases (unless the index was under 15 years old)[43]. These cases are typically not infectious[3], and this is the reason given for this change to guidance[43]. However, it is possible these ETB cases have infectious contacts, and that they were infected recently. Analysis of the yield of contact tracing (paper one[102], chapter 2) found that the proportion of contacts of non-pulmonary cases that had active disease was quite high (0.7%), relative to the prevalence of TB in the population (0.027%), and comparable to other risk groups (e.g. homeless, 0.79%)[114]. In addition, analysis of strain typing data (paper two, chapter 3) found that 20% of contacts who had active disease were infected by a source other than the index case that triggered the contact investigation. Both of these findings imply that there may be some benefit to screening contacts of ETB cases, even though they're not infectious, as they show that the risk of TB amongst contacts is high irrespective of whether transmission from the index case has occurred. It is also possible that ETB cases form an outer node of a transmission cluster, and in the contact tracing may uncover, and potentially avert, transmission related to this cluster. Furthermore, while the WHO does not recommend the screening of ETB cases in low- and middle-income countries[115], a number of countries in Europe did screen contacts of ETB cases in 2008[44]; these ranged from very low incidence countries like Cyprus and Israel (4.4/100000 and 6.0/100000 respectively at the time of the survey) to high incidence countries like Kazakhstan (210/100000 at the time of the survey).

Cost-effectiveness analyses have become an important part of informing tuberculosis care and prevention policy, and are used by NICE to inform their guidelines[43,116]. They have been used in the UK to evaluate the screening of migrants[55] and the homeless[56]; both of these studies are discussed in more detail in the following paper. Over the years at least nine studies have evaluated the cost-effectiveness of screening for LTBI amongst migrants[117], typically finding it to be effective. However, while a number of studies have evaluated the effectiveness of giving preventive therapy to contacts[118–120], or the effectiveness of case finding[121,122], only one study has attempted to evaluate the cost-effectiveness of contact tracing as a whole[100]. That study, by Dasgupta et al. (2000), was based on just six cases found, did not evaluate contact tracing as its main aim (which was migrant screening), and did not include the effect of reduced transmission from contacts by finding them sooner. As aforementioned, no studies have previously estimated the cost-effectiveness of contact tracing in the UK or London.

The third paper, which is in review at Thorax, evaluates the cost-effectiveness of contact investigations in London in the period 2012-15, and in so doing, improves understanding of the impact of the aforementioned change to NICE guidance. This paper is the first cost-effectiveness analysis of contact tracing to incorporate all of the potential impacts of contact tracing (reduced morbidity in contacts, prevented cases, reduced transmission from contacts, and reduced mortality), the second overall, and the first in the UK. It addresses an area of direct relevance to national policy guidelines, following recent, potentially controversial, changes to guidance[43]. It also provides a novel framework for delineating and quantifying the effect of contact tracing.

Cover sheet

See next page

London School of Hygiene & Tropical Medicine
Keppel Street, London WC1E 7HT
www.lshtm.ac.uk

Registry
T: +44(0)20 7299 4646
F: +44(0)20 7299 4656
E: registry@lshtm.ac.uk

LONDON SCHOOL of HYGIENE &TROPICAL MEDICINE

RESEARCH PAPER COVER SHEET

PLEASE NOTE THAT A COVER SHEET MUST BE COMPLETED FOR EACH RESEARCH PAPER INCLUDED IN A THESIS.

SECTION A – Student Details

Student	Sean M. Cavany
Principal Supervisor	Emilia Vynnycky
Thesis Title	Using data analysis and mathematical modelling to study tuberculosis contact tracing in London, with reference to the national strategy and guidance

If the Research Paper has previously been published please complete Section B, if not please move to Section C

SECTION B – Paper already published

Where was the work published?			
When was the work published?			
If the work was published prior to registration for your research degree, give a brief rationale for its inclusion			
Have you retained the copyright for the work?*	Choose an item.	Was the work subject to academic peer review?	Choose an item.

**If yes, please attach evidence of retention. If no, or if the work is being included in its published format, please attach evidence of permission from the copyright holder (publisher or other author) to include this work.*

SECTION C – Prepared for publication, but not yet published

Where is the work intended to be published?	Thorax
Please list the paper's authors in the intended authorship order:	Sean M. Cavany, Emilia Vynnycky, Charlotte Anderson, Helen Maguire, Frank Sandmann, H. Lucy Thomas, Richard G. White, Tom Sumner
Stage of publication	**Undergoing revision**

SECTION D – Multi-authored work

For multi-authored work, give full details of your role in the research included in the paper and in the preparation of the paper. (Attach a further sheet if necessary)	I conceived, deisgned and carried out the work, and wrote the first draft of the paper.

Student Signature: [redacted] **Date:** 28/03/18

Supervisor Signature: [redacted] **Date:** 29.03.18

Paper

See next page

Should NICE reconsider the 2016 UK guidelines on tuberculosis contact tracing? A cost-effectiveness analysis of contact investigations in London

Authors: Sean M. Cavany[1,2], Emilia Vynnycky[1,2], Charlotte Anderson[4], Helen Maguire[4,5], Frank Sandmann[2, 6], H. Lucy Thomas[3], Richard G. White[1], Tom Sumner[1]

Affiliations and positions:

1. TB Modelling Group, TB Centre and CMMID, Faculty of Epidemiology and Population Health, London School of Hygiene and Tropical Medicine, London.
2. Statistics, Modelling and Economics Department, Public Health England, London
3. Respiratory Diseases Department, Public Health England, London
4. Field Epidemiology Service, Public Health England, London
5. Institute for Global Health, University College London
6. CMMID, Faculty of Epidemiology and Population Health, London School of Hygiene and Tropical Medicine, London.

*__Corresponding author__: SM Cavany, email: sean.cavany@lshtm.ac.uk, phone: +447861234239

Word Count: 3575

Abstract (222)

Background – In January 2016, clinical TB guidance in the UK changed to no longer recommend screening contacts of non-pulmonary, non-laryngeal (ETB) index cases. However, no new evidence was cited for this change, and there is evidence that screening these contacts may be worthwhile. The objective of this study was to estimate the cost-effectiveness of screening contacts of adult ETB cases and adult pulmonary or laryngeal tuberculosis (PTB) cases in London, UK.

Methods – We carried out a cross-sectional analysis of data collected on tuberculosis index cases and contacts in the London tuberculosis register, and an economic evaluation using a static model describing contact tracing outcomes. All adult TB cases (≥15 years old) in London from 2012-15, and their contacts, were eligible (2465/5084 PTB and 2559/6090 ETB index cases were included).

Results – Assuming each contact with PTB infects 1 person/month, screening contacts of ETB cases costs £78000/QALY (95% CI: 39000 to 140000) and screening contacts of PTB cases costs £30000/QALY (95% CI: 18000 to 50000). Screening contacts of ETB cases costs £30000/QALY if each contact with PTB infects 3.4 people/month. Limitations of this study include the use of self-reported symptomatic periods, and lack of knowledge about onward transmission from PTB contacts.

Conclusions –Screening contacts of ETB cases in London was almost certainly not cost-effective at a willingness-to-pay threshold of £30000/QALY, supporting recent changes to NICE national guidelines.

Key Messages

What is the key question? Was NICE correct to change its tuberculosis clinical guidelines to no longer recommend screening contacts of non-pulmonary TB cases?

What is the bottom line? It is almost certainly not cost-effective to screen contacts of non-pulmonary TB cases in London at a willingness-to-pay-threshold of £30000/QALY, providing strong evidence that the decision to cease recommending screening contacts of non-pulmonary cases was the correct one.

Why read on? In addition to helping an answer an important policy question that has been questioned by several recent papers, this article provides the first cost-effectiveness analysis of contact tracing in the UK and the first to incorporate non-pulmonary cases, and proposes a novel way to evaluate contact tracing effectiveness.

Introduction (440)

Contact tracing, the systematic screening of contacts of tuberculosis (TB) cases, is a fundamental part of TB control in high-income countries, and is highlighted as a key element of the Public Health England (PHE)/National Health Service (NHS) England collaborative tuberculosis strategy 2015-2020[1]. The aim of contact tracing is threefold: to reduce morbidity and mortality in contacts with TB by finding them sooner; to reduce transmission from those contacts with active TB; and to find contacts with latent *M. tuberculosis* infection (LTBI) who are eligible for preventive therapy (PT)[2].

In January 2016, the UK National Institute for Health & Care Excellence (NICE) TB guidelines changed from recommending screening contacts of all cases, to only screening contacts of pulmonary or laryngeal TB (PTB) cases. No new evidence was cited to justify this change[3]. Although the guidance on whether contacts of non-pulmonary cases are screened differs between countries[4,5], most advocate not screening.

Whilst non-pulmonary, non-laryngeal TB (ETB) cases are typically not infectious, there is evidence that their contacts are more likely to have TB than the general population. Between 2012-15, the prevalence of active TB amongst contacts of ETB index cases in London was 0.7%[6], compared to 0.027% in the general population[7]. Similar patterns are observed in Birmingham[8,9], and in both cities the prevalence of disease amongst contacts of ETB cases was higher than the prevalence of disease amongst migrants eligible for pre-entry screening[10] ,and more than 10 times higher than the NICE threshold for new entrant screening[9]. Additionally, studies have shown only 25% of pairs of cases sharing an address in the UK[11], and 20% of case-contact pairs in London[12] had different *M. tuberculosis* isolates, implying the risk of disease in household contacts is high irrespective of whether transmission has occurred. This suggests that the fact that ETB cases are not infectious may not be a valid justification for not screening their contacts.

In light of this evidence, key stakeholders have questioned the change in guidance and a cost-effectiveness analysis has been called for[9]. To our knowledge, only one previous study has

attempted to evaluate the cost-effectiveness of contact tracing[13], and no studies have done so in the UK or London, nor have any studies attempted to evaluate the cost-effectiveness of contact tracing delineated by site of disease of the index case. In this study we aim to evaluate the effectiveness and cost-effectiveness of contact tracing, for ETB and PTB index cases, in London. We first estimate symptomatic periods and the number of contacts found with active disease or LTBI per index case. We then use these values alongside previously published data to develop a simple static model to calculate the cost-effectiveness.

Methods (1016)

Data analysis: We used data on adult and adolescent (≥ 15 years old) TB cases notified to the London TB register (LTBR) during 2012-2015. The LTBR is a web-based register containing demographic and clinical data on all TB cases notified in London since 2002[6]. We excluded index cases that were notified in a region and year where the completeness was less than 80%, or were children (≤ 14 years old) (because contacts of children with ETB will still be screened under new guidelines)[3]. When estimating yield we excluded index cases who first accessed health-care through contact investigation, as the number of contacts is not recorded consistently[6] . Costs were calculated based on national accounting expenditures and current treatment guidance for England[3,14] (see Appendix part 1 for details).

Other data sources: Estimates of utility scores were taken from Jit et al[15]. The life-time risk of developing disease following infection was taken from Sloot et al.[16] and the efficacy of PT from Smieja et al.[17] and Ayieko et al.[18] See Table 1 for details of data sources.

Effectiveness: We quantified the effectiveness of contact tracing with four outcomes:

1. Morbidity: the reduction in time contacts with TB are symptomatic if they are found earlier due to contact tracing.
2. Prevention: the number of contacts with LTBI prevented from developing active TB following PT.
3. Transmission: the number of cases prevented by reducing transmission from: a) contacts with prevalent TB found earlier through contact tracing; b) cases prevented from occurring due to PT.
4. Mortality: the number of TB deaths prevented by contact tracing.

Model description: We developed a simple static model to estimate the cost-effectiveness of screening contacts of ETB and PTB cases in London during the period 2012-2015. The model was

109

103 used to calculate the four measures of effectiveness and estimate the quality-adjusted life-years
104 (QALYs) gained by contact tracing using the following equations (see Table 1 and Table 2 for
105 definitions of symbols).

106 In all equations σ is either P or E, and represents the site of disease of the index cases under
107 analysis. The number of PTB index cases is given by $N_P = (1 - f_E)N$ and the number of ETB index
108 cases is given by $N_E = f_E N$, where f_E is the fraction of all adult cases that have ETB.

109 The reduction in morbidity was calculated using the number of contacts with TB per index case (Y_σ),
110 the proportion of contacts with TB that have ETB (ϵ_σ) and the difference in symptomatic period of
111 cases found through contact tracing and those found through other routes:

$$t_{\text{morbidity},\,\sigma} = N_\sigma \left(Y_\sigma (1 - \epsilon_\sigma) \left(\frac{S_{P,\,\text{passive}} - S_{P,\,\text{traced}}}{365.25} \right) + Y_\sigma \epsilon_\sigma \left(\frac{S_{E,\,\text{passive}} - S_{E,\,\text{traced}}}{365.25} \right) \right)$$

112 The number of cases of TB prevented by PT, assuming contacts with LTBI are recently infected is:

$$N_{\text{prevention},\,\sigma} = N_\sigma \left(\sum_{j=a,c} y_\sigma \phi_{\sigma,j}\, \theta_{j,\sigma,B} \theta_{j,\sigma,C} \tau_j \right) P$$

113 where $\phi_{\sigma,c} = 1 - \phi_{\sigma,a}$. As the efficacy of PT is different in children (c) and adults (a), and children
114 are more likely to begin preventive therapy than are adults (Appendix part2, Table G), we calculated
115 the effectiveness of PT separately for these two groups.

116 The number of cases of TB prevented by reducing transmission from contacts with PTB by finding
117 them sooner is:

$$N_{\text{transmission},\,\sigma} = N_\sigma Y_\sigma (1 - \epsilon_\sigma) \left(\frac{S_{P,\,\text{passive}} - S_{P,\,\text{traced}}}{365.25} \right) rP$$

118 The prevention of subsequent generations of TB cases which would have occurred in the absence of
119 contact tracing is given by $N_{\text{later generations},\,\sigma}$ (see Appendix part 3).

The number of TB-related deaths prevented by screening contacts is calculated as follows:

$$N_{\text{mortality},\,\sigma} = \left(\frac{365.25\,t_{\text{morbidity},\,\sigma}}{S_{\text{overall}}} + N_{\text{transmission},\,\sigma} + N_{\text{prevention},\,\sigma} + N_{\text{later generations},\,\sigma}\right)\mu$$

where μ is the case fatality ratio. The first term in this equation describes the reduction in mortality among prevalent cases in contacts identified sooner via contact tracing.

To calculate the amount of onward transmission from prevented cases, we assumed a range of values for the number of new infections per PTB case per month infectious, r, and explored the dependence of results on this parameter. This parameter, r, can be related to the updated Styblo rules developed by Trunz *et al.* and van Leth *et al.*[19,20]; these studies calculated that each case of smear positive TB would lead to approximately 3 to 6 new infections, equating to a value of r between 0.5 and 1 (see Appendix part 4).

Cost-effectiveness: Costs were calculated from a health system perspective. We excluded diagnostic and treatment costs of contacts with TB, as we assumed these contacts would be treated later regardless of whether the contact investigation took place. However, we subtracted the costs of diagnosis and treatment of cases that are prevented. We assumed latently infected contacts are given a 3 month course of rifampicin and isoniazid (with pyridoxine)[3], and assumed this has the same efficacy as 6 months of isoniazid[21].

We calculated the resulting incremental cost-effectiveness ratio (ICER) for contact tracing of both PTB and ETB index cases, using no screening as the baseline comparator for both. Equations for these calculations are given in the Appendix part 3. Following NICE recommendations, we assumed a ratio of £20000-30000/QALY as being cost-effective[22]. We included secondary cases which occurred at any time after infection, but assumed most occur in the first year[16]. Consequently, most costs and QALY gains occurred in the first year, and so no discounting was included in the main analysis (see Appendix part 5 for a discussion of discounting).

111

142 **Uncertainty and sensitivity**: 95% confidence intervals were calculated by randomly selecting 10000
143 parameter sets from the distributions shown in Table 1 and Table 2. Correlation coefficients were
144 calculated between the distribution of each parameter and distribution of the ICER.

145 We explored the sensitivity to the symptomatic period by doubling each of these periods, and to
146 assumptions about risk of disease following infection and preventive therapy by using estimates of
147 these from Erkens *et al.*[23] instead of the estimates from Sloot *et al.*[16].

148 We explored sensitivity to utility scores by using values from Mears et al.[15]. These were derived from
149 the same source [24] as those of Jit et al.[25] used in our primary analysis, but differ as the Jit et al values
150 were based on London specific data.

151 **Role of finding source**: The funding sources played no part in the study design, data analysis, writing
152 of the manuscript or decision to submit for publication.

112

Results (795)

Mean symptomatic periods. During the period 2012-2015 in London, there were 5084 PTB cases, of whom 2465 met the inclusion criteria and had data on symptomatic period. Of these, 82 were found through contact tracing, and were symptomatic for a mean period of 76.6 days (95% CI: 58.5, 94.7). Those who accessed care through other routes were symptomatic for a longer mean period of 110 days (95% CI: 103, 117 days) (p=0.0016) (Table 2)

During the same period there were 6090 ETB cases, of whom 2559 were included and had data on symptomatic period. Of these, 26 were found through contact tracing and had a mean symptomatic period of 152 days (95% CI: 15.0, 289 days). Those who accessed care through other routes had a mean symptomatic period of 180 days (95% CI: 165, 195 days) (p=0.36). See Table E in Appendix part 2 for further details.

Preventive therapy. Of 1497 contacts with LTBI identified in the study period, 1165 (77.8% (95% CI: 74.9%, 80.7%) started PT and 918 of those that started (78.6% (95% CI: 75.4%, 81.8%) completed PT (Table 2). See Table G in Appendix part 2 for further details; of note is that children are much more likely than adults to start PT and, for contacts of PTB cases, to complete PT.

Effectiveness

Reduction in morbidity of contacts: On average, in a single year, not screening contacts of adult ETB cases would have led to those contacts with TB being undiagnosed for a combined additional 2.58 years (95% CI: 0.660 to 8.59) (Table 3). For contacts of PTB cases this would be 10.5 years (95% CI: 4.02 to 26.4).

Cases prevented by preventive therapy: By giving PT to contacts of ETB cases we would expect to prevent 5.45 (95% CI: 3.71 to 7.59) cases. This value would be 18.9 (95% CI: 13.1 to 25.8) cases prevented by giving PT to contacts of PTB index cases.

113

176 *Cases prevented by reduced transmission from contacts:* Finding contacts of ETB index cases with TB
177 sooner via contact tracing, thereby reducing onward transmission, could prevent 1.71 cases (95% CI:
178 0.584 to 3.33) when $r = 1$ new infections per PTB case per month infectious. The corresponding
179 value for PTB index cases is 8.76 (95% CI: 3.56 to 14.9). This reduction in cases is directly
180 proportional to the assumed value of r.

181 *Prevention of subsequent generations of cases:* Preventing cases from occurring amongst contacts of
182 contacts of ETB cases could avert 1.62 cases (95% CI: 0.772 to 3.11) when $r = 1$, and 5.19 cases (95%
183 CI: 2.08 to 12.2) when $r = 2$. The corresponding figures for PTB index cases are 8.63 (95% CI: 4.77 to
184 14.7) and 33.1 (95% CI: 16.1 to 66.7).

185 *Reduction in mortality:* When $r = 1$, screening contacts of ETB cases could prevent 0.551 deaths
186 (95% CI: 0.303 to 1.14) and screening contacts of PTB cases 2.27 deaths (95% CI: 1.36 to 3.94).

187 **Cost-effectiveness:** The cost per QALY of screening the contacts of ETB cases is £101000/QALY (95%
188 CI: 46200 to 178000) when transmission is not included ($r = 0$), £77700/QALY (95% CI: 38800 to
189 139000) for $r = 1$ new infection per PTB case per month infectious and £56400/QALY (95% CI:
190 29300 to 102000) for $r = 2$ (Table 3, Figure 1a). The equivalent values for PTB cases are
191 £43700/QALY (95% CI: 23700 to 70100), £30300/QALY (95% CI: 17700 to 50100) and £18700/QALY
192 (95% CI: 10500 to 32700) respectively (Figure 1b). Screening contacts of ETB cases becomes cost-
193 effective at a £30000/QALY threshold when $r = 3.40$. If $r = 1$, the yield of ETB index cases would
194 need to be 0.0959 (an almost 5-fold increase above the observed yield, and greater than current
195 PTB yield) in order for screening contacts of ETB cases to become cost-effective at £30000/QALY.

196 **Sensitivity:** Cost-effectiveness results are most sensitive to the symptomatic period of those found
197 through contact tracing (Table 4) (especially of contacts of ETB index cases), the probability of
198 developing disease, and the yield of ETB index cases. At low levels of transmission from PTB
199 contacts, the symptomatic period of contacts with ETB explains most of the variation in the ICER. As

the amount of transmission from contacts is increased, the results become more sensitive to the probability of developing disease and the symptomatic period of PTB index cases, and less sensitive to the symptomatic period of ETB index cases. Increasing each symptomatic period by a factor of 2 (Figure 1(c) and (d)), then for $r \geq 1.60$ the mean cost-effectiveness of screening contacts of ETB cases is below the £30000/QALY threshold. Calculating the probability of developing disease from Erkens et al. rather than Sloot et al. does not qualitatively change the cost-effectiveness results (not shown). Using utility scores used by Mears et al.[15] instead of those used by Jit et al.[25] leads to a slight decrease in cost-effectiveness (Appendix part 6).

115

208 # Discussion (1334)

209 **Principal findings:**

210 On average, we estimate that in a single year, screening contacts of ETB would save a total of 2.58
211 years of morbidity in contacts with prevalent TB, and prevent at least 5.45 cases through reduced
212 transmission and PT. However, screening ETB contacts was very unlikely to be cost-effective at a
213 threshold of £30000/QALY, even with the assumption of high levels of transmission from contacts.
214 Hence, the results presented here support recent changes to the NICE guidelines to remove
215 screening of contacts of ETB cases from their guidance. In contrast, screening contacts of PTB cases
216 was probably cost effective at a £30000/QALY threshold, especially when assuming high levels of
217 transmission from contacts. Neither was likely to be cost-effective at a £20000/QALY threshold at
218 plausible levels of transmission.

219 **Strengths and limitations:**

220 This study used high quality data on contact tracing yield in London to answer an important question
221 for TB care and prevention, which has implications for TB policy in the UK. The approach used
222 proposes a novel way of quantifying the effectiveness of contact tracing across four potential
223 impacts (reduced morbidity, preventive therapy, reduced transmission and reduced mortality). The
224 main limitation of the study is the large uncertainty in several parameters. However, we explored
225 this first by varying the number of infections generated by each case (r), and by carrying out a
226 probabilistic sensitivity analysis of all other parameters. A related limitation is the treatment of
227 transmission. It is difficult to know the rate at which infectious contacts would infect further
228 contacts, so we explored a range of assumptions. We did not characterise the indirect effect of
229 contact tracing on transmission at a population level, though as only five percent of all cases in
230 London are found through contact tracing, this is probably negligible over short time-scales. The
231 quantitative nature of this approach is unable to assess broader outcomes of contact tracing, such as
232 community engagement and tackling stigma. Finally, we used the self-reported symptomatic period

to estimate the time during which cases are infectious. Due to issues with patient recall and the fact that the ratio of estimated prevalence to incidence in London[7,26] is much greater than the mean self-reported symptomatic period found in this study, it is likely that this value systematically underestimates the true time people are symptomatic. Our sensitivity analysis showed that cost-effectiveness of contact tracing would increase and screening contacts of ETB cases would be possibly cost-effective at a £30000/QALY threshold if the symptomatic period was double that estimated by self-reported symptom onset (Figure 1).

Relation to other studies:

In recent years, studies in the UK have evaluated the cost-effectiveness of screening new migrants[27] and hard to reach populations using a mobile X-ray unit (MXU, known as Find & Treat)[25]. In 2011 Pareek et al.[27] found that screening migrants from countries with an incidence exceeding 150/100000 cost £21000 per case averted. This is cheaper than screening ETB contacts, and similar to screening PTB contacts for $r = 1$ new infections per PTB case per month infectious (Table 3). Jit et al. found that screening hard-to-reach groups in London cost £6400-£10000/QALY gained, so was more cost-effective than screening PTB cases even if $r = 2$. In their study, Jit et al.[25] found that about 80% of QALYs gained were due to improved case-management of these complex cases, and the cost-effectiveness of screening alone was similar to screening contacts of PTB cases. The case management impact would likely be smaller for contact tracing than for the MXU, because the population of contacts is less complex, and case management is not an explicit aim of contact tracing. When Dasgupta et al.[13] compared the cost-effectiveness of screening close contacts to migrant screening in Montreal, they found that close contact investigation was cost saving. This was due to much lower treatment costs of contacts as opposed to cases found through other routes, due largely to much higher rates of hospitalization amongst passively detected cases. However, this assumption was based on only six cases found through contact tracing. We did not explore the impact of decreased hospitalization rates here due to a lack of data. Finally, a 2008 study in British

Columbia, Canada[28] found that giving PT to contacts was cost-effective, though this study focused on infectious index cases. Our results are not directly comparable with this study due to its focus on PT, but both support the continued screening of contacts of PTB cases.

Interpretation of results:

These results support the recent decision to remove screening contacts of adult ETB cases from NICE guidance. In order for screening these contacts to be cost-effective at a £30000/QALY threshold, r would need to be 3.40 new infections per PTB case per month infectious, which would mean each smear positive case would need to generate 21 new infections. This is likely to be high for some settings[19], but may be plausible in crowded environments, such as homeless shelters[29]. Additionally, we found that if the yield per ETB index case was above 0.0959, then the ICER for screening contacts of these cases was below £30000/QALY. In London, ETB cases with a history of homelessness or drug use have a yield greater than this (unpublished data), supporting recommendations for active case-finding amongst this group. Additionally, subgroups for whom the yield is higher, are also those for whom r is likely to be higher, further increasing the impact of screening contacts of those subgroups. It is unlikely that the average yield of ETB cases in other parts of the UK are much higher than those seen in London[8], implying that it would also not be cost-effective to screen contacts of ETB cases nationally.

The impact on the ICER caused by changing the amount of transmission (r) indicates the importance of reducing transmission from contacts as one of the impacts of contact tracing. It is plausible, though, that the number of infections generated by a contact with PTB (i.e. the value of r) will be lower than that suggested by the re-estimated Styblo rule[19,20], as the household contacts of someone themselves found through contact tracing are more likely to have already been infected.

The main reason for the low ICER for ETB index cases was the small difference in symptomatic period of contacts with ETB and cases with ETB found through other routes (Table 4), suggesting that the

impact may be improved by hastening contact tracing for these contacts. The NICE guidelines now recommend PT for anyone aged under 65 years. This may cause a small improvement in cost-effectiveness, as we would now expect a higher yield of LTBI per case, as more contacts will be tested for LTBI, provided it is not accompanied by lower rates of PT enrolment and completion. The introduction in 2017 of whole genome sequencing (WGS) in the UK[30] may also affect our conclusions. Whilst a study of the current strain typing service found no impact on contact tracing[15], it is plausible that faster turnaround times and improved targeting available with WGS may affect contact tracing yields.

Further research:

This work would benefit from an improved understanding of the rate of onward transmission from contacts. Mathematical modelling work incorporating transmission on a network structure may help to understand this. It would also help to have a greater understanding of the proportion of contacts that have pulmonary TB and how this differs across groups. If there are subgroups for whom a greater than average proportion of contacts with TB have PTB, then this would increase the cost-effectiveness in these groups. Whilst we were able to estimate this proportion for the whole population, our small sample meant we could not stratify this estimate. Work to understand how the different screening approaches (migrant, hard-to-reach populations and contacts) interact would help our understanding of the impact of each. Our results were very sensitive to estimates of the symptomatic period of contacts, both due to the uncertainty of these estimates and the fact that they are based on self-reported periods. A more thorough understanding of diagnostic delay amongst both contacts and non-contacts is needed.

Tables

Table 1: Variables and constants from other sources. CI = confidence interval, ETB = non-pulmonary and non-laryngeal tuberculosis, LTBR = London TB register, NICE = National Institute for Health & Care Excellence, PT = preventive therapy, TB = tuberculosis, BNF = British National Formulary, QALY = quality-adjusted life years, UK = United Kingdom. †=this was calculated using the age-specific case-fatality ratios given in Mears et al. and the age-structure of cases calculated from the LTBR. Note that some confidence intervals differ slightly from those in the literature due to the use of beta distributions. Following current treatment guidance (NICE 2016), we used the following references to calculate cost values: NICE 2011, Pareek et al. 2011, Reference costs 2016, Dowdy et al. 2008, Dinnes et al. 2007, BNF 2017; where necessary, we inflated costs according to inflation to the base year 2016. See Appendix parts 1 and 6 for details of cost and utility calculations.

Name of variable (units, if applicable)	Symbol	Value	95% CI, (or *range)	Distribution	Source
Life-time probability of developing disease following infection	P	0.1	(0.08, 0.12)	Beta	Sloot et al.[16]
Efficacy PT in adults	τ_a	0.6	(0.49, 0.70)	Beta	Smieja et al.[17]
Efficacy PT in children	τ_c	0.4	(0.16, 0.57)	Beta	Ayieko et al.[18]
Average number of cases per year	N	2790	N/a	N/a	LTBR
Fraction of all adult cases that have ETB	f_E	0.545	N/a	N/a	LTBR
Fraction of those tested for active TB that have active TB	f_c	0.2	N/a	N/a	Mears et al.[15]
Case fatality ratio	μ	0.0363	N/a	N/a	Mears et al.[15†] and LTBR
Relative average treatment length of non-completed PT	f_i	0.33	N/a	N/a	Assumption
Contact tracing, per contact traced, £	C_0	244	N/a	N/a	See Appendix part 1
Further tests if case is suspected to have	C_1	497	N/a	N/a	See Appendix part 1

active disease, £					
Cost per full course PT (3 month rifampicin and isoniazid, with pyridoxine), £	C_{PT}	852	N/a	N/a	See Appendix part 1
Cost per full course (6 months) of treatment of tuberculosis disease, £	C_{FT}	1694	N/a	N/a	See Appendix part 1
Average utility of a healthy person, given age structure of TB cases in London	U_H	0.876	N/a	N/a	Calculated from Kruijshaar et al *via* Mears et al
Symptom onset to diagnosis	U_0	$0.68U_H$	N/a	N/a	Kruijshaar et al *via* Jit et al
On treatment	U_1	$0.79U_H$	N/a	N/a	Kruijshaar et al *via* Jit et al
Utility preventive therapy	U_{PT}	$0.9992U_H$	N/a	N/a	Kruijshaar et al *via* Mears et al
Average # of QALYs at death for someone living in UK	A_H	72.6	N/a	N/a	Calculated from Mears et al. and LTBR
Average # of QALYs at death for someone living in UK with TB as cause of death	A_{TB}	52.2	N/a	N/a	Calculated from Mears et al. and LTBR

Table 2: Estimates of parameters calculated from the LTBR. All parameters are chosen from a normal distribution. ETB = non-pulmonary and non-laryngeal tuberculosis, PTB=Pulmonary or laryngeal tuberculosis. LTBI = latent M.Tb infection, LTBR = London TB report, PT = preventive therapy, TB = tuberculosis

Name of variable (units, if applicable)	Index case disease type	Symbol	Value	Standard error
Number of contacts screened per index case	ETB	n_E	2.50	0.0464
	PTB	n_P	3.86	0.0729
Number of contacts found with TB per index case	ETB	Y_E	0.0196	0.00395
	PTB	Y_P	0.0938	0.00838
Proportion of contacts with TB that have ETB	ETB	ϵ_E	0.486	0.0802
	PTB	ϵ_P	0.337	0.0301
Number of contacts found with LTBI per index case	ETB	y_E	0.119	0.00767
	PTB	y_P	0.471	0.0219
Proportion of index contact's with LTBI that are children	ETB	$\phi_{E,c}$	0.206	Not varied
	PTB	$\phi_{P,c}$	0.360	Not varied
Proportion of contacts with LTBI that begin PT, adult contact	ETB	$\theta_{a,E,B}$	0.611	0.0515
	PTB	$\theta_{a,P,B}$	0.666	0.0318
Proportion of contacts with LTBI that begin PT, child contact	ETB	$\theta_{c,E,B}$	0.931	0.0477
	PTB	$\theta_{c,P,B}$	0.969	0.0238
Proportion of contacts starting PT that complete PT, adult contact	ETB	$\theta_{a,E,C}$	0.875	0.042
	PTB	$\theta_{a,P,C}$	0.803	0.031
Proportion of contacts starting PT that complete PT, child contact	ETB	$\theta_{c,E,C}$	0.81	0.0876
	PTB	$\theta_{c,P,C}$	0.906	0.0311
Mean symptomatic period of PTB cases not found through contact tracing (days)	N/a	$S_{P,\text{ passive}}$	110	3.66
Mean symptomatic period of PTB cases found through contact tracing (days)	N/a	$S_{P,\text{ traced}}$	76.6	9.26
Mean symptomatic period of PTB cases (days)	N/a	$S_{P,\text{ overall}}$	109	3.55
Mean symptomatic period of ETB cases not found through contact tracing (days)	N/a	$S_{E,\text{ passive}}$	181	7.41
Mean symptomatic period of ETB cases found through contact tracing (days)	N/a	$S_{E,\text{ traced}}$	152	69.9
Mean symptomatic period of all cases (days)	N/a	S_{overall}	147	4.26

Table 3: Summary of the effectiveness measures included, costs incurred, quality adjusted life years(QALYs) gained and resulting incremental cost effectiveness ratio (ICER) for screening contacts of the indicated index cases compared to a baseline of not screening those contacts. Numbers are given for a year with a case-load that is the average caseload of the years 2012-15 (i.e. 2790 cases); note that the case-load does not affect the ICER. No discounting was applied; see Appendix part 5 for a discussion of discounting. Case-equivalents averted refers to both cases averted, and the reduction in the time contacts are symptomatic divided by the mean symptomatic period of TB cases. ETB = non-pulmonary, non-laryngeal; PTB = pulmonary or laryngeal; r = the number of infections generated by a pulmonary contact per month infectious; ICER = incremental cost-effectiveness ratio; PT = preventive therapy (3 months of isoniazid and rifampicin). Numbers in brackets indicate the 95% confidence intervals.

Quantity (units, if applicable)	ETB indexes			PTB indexes		
	$r = 0$	$r = 1$	$r = 2$	$r = 0$	$r = 1$	$r = 2$
Reduction in time contacts are symptomatic (years)	2.58 [0.66, 8.59]	2.58 [0.66, 8.59]	2.58 [0.66, 8.59]	10.5 [4.02, 26.4]	10.5 [4.02, 26.4]	10.5 [4.02, 26.4]
Cases prevented by administering PT (cases)	5.45 [3.71, 7.59]	5.45 [3.71, 7.59]	5.45 [3.71, 7.59]	18.9 [13.1, 25.8]	18.9 [13.1, 25.8]	18.9 [13.1, 25.8]
Transmission reduced by finding contacts sooner (cases)	0.0 [0.0, 0.0]	1.71 [0.584, 3.33]	3.41 [1.17, 6.62]	0.0 [0.0, 0.0]	8.76 [3.56, 14.9]	17.5 [7.02, 29.8]
Transmission reduced from prevented cases (cases)	0.0 [0.0, 0.0]	1.62 [0.772, 3.11]	5.19 [2.08, 12.2]	0.0 [0.0, 0.0]	8.63 [4.77, 14.7]	33.1 [16.1, 66.7]
Reduction in mortality (deaths)	0.431 [0.238, 0.977]	0.551 [0.303, 1.14]	0.743 [0.408, 1.45]	1.64 [0.997, 3.08]	2.27 [1.36, 3.94]	3.47 [2.04, 5.89]
Total case-equivalents averted	11.9 [6.56, 26.9]	15.2 [8.34, 31.4]	20.5 [11.2, 39.9]	45.0 [27.5, 85.0]	62.4 [37.5, 109.0]	95.6 [56.2, 162.0]
Total QALYs gained	10.6 [5.98, 23.4]	13.7 [7.66, 27.6]	18.7 [10.4, 35.6]	39.9 [24.8, 73.9]	56.3 [34.2, 95.9]	87.5 [51.7, 148.0]
Total costs incurred (£ 000 000s)	1.07 [1.03, 1.12]	1.06 [1.02, 1.11]	1.05 [1.01, 1.1]	1.74 [1.67, 1.82]	1.71 [1.63, 1.78]	1.63 [1.52, 1.72]
Incremental cost-effectiveness ratio (£ 000s/QALY)	101.0 [46.2, 178.0]	77.7 [38.8, 139.0]	56.4 [29.3, 102.0]	43.7 [23.7, 70.1]	30.3 [17.7, 50.1]	18.7 [10.5, 32.7]

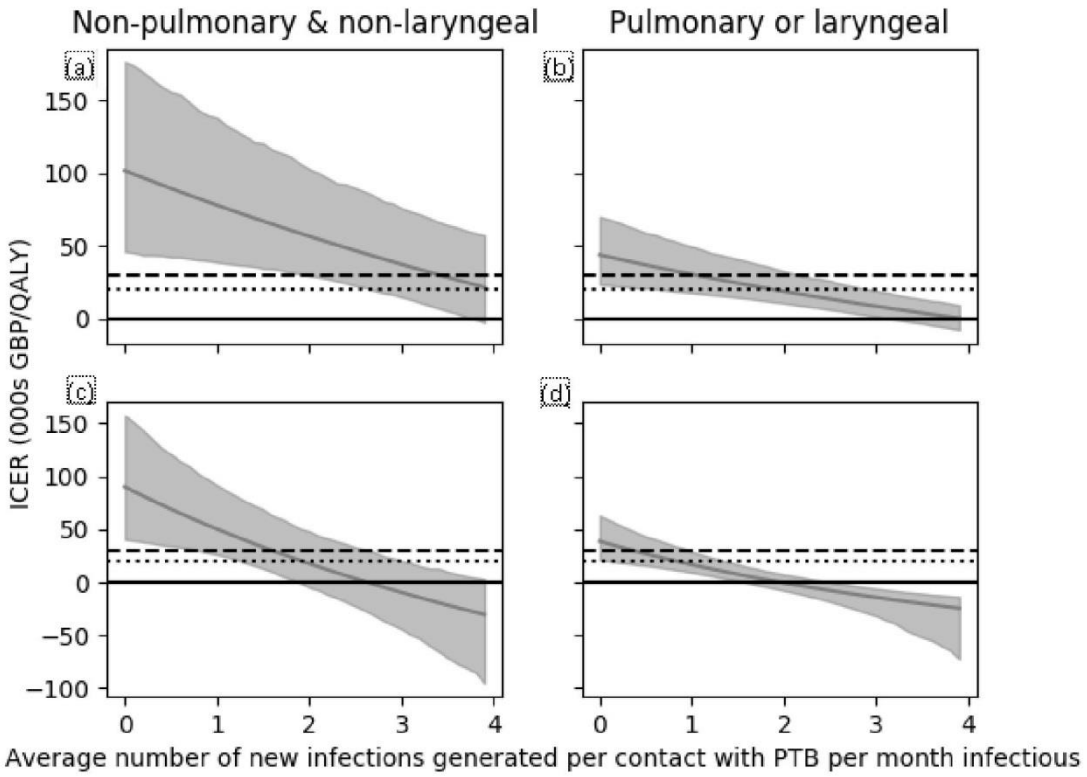
Table 4: Sensitivity of the ICER to each parameter. The numbers presented are the square of the correlations of each parameter value with the ICER across all 10000 parameter sets. A higher number indicates the given parameter explains a greater proportion of the uncertainty in the ICER. The shading indicates the correlation; deeper shading indicates a higher correlation. r = number of new infections per PTB case per month infectious. ETB = non-pulmonary and non-laryngeal tuberculosis, LTBI = latent *M.Tb* infection, PT = preventive therapy, PTB = pulmonary or laryngeal tuberculosis, QALY = quality-adjusted life years, ICER = incremental cost-effectiveness ratio.

	r=0		r=1		r=2	
	ETB indexes	PTB indexes	ETB indexes	PTB indexes	ETB indexes	PTB indexes
Number of contacts screened, ETB index	0.00	0.00	0.00	0.00	0.00	0.00
Number of contacts screened, PTB index	0.00	0.00	0.00	0.00	0.00	0.00
Yield of active tuberculosis, ETB index	0.05	0.00	0.08	0.00	0.09	0.00
Yield of active tuberculosis, PTB index	0.00	0.02	0.00	0.02	0.00	0.02
Proportion of contacts with ETB, ETB index	0.00	0.00	0.03	0.00	0.11	0.00
Proportion of contacts with ETB, PTB index	0.00	0.00	0.00	0.01	0.00	0.02
Yield of LTBI, ETB index	0.00	0.00	0.01	0.00	0.00	0.00
Yield of LTBI, PTB index	0.00	0.00	0.00	0.00	0.00	0.00
Probability of starting PT, ETB index, adult contact	0.00	0.00	0.00	0.00	0.00	0.00
Probability of ending PT, ETB index, adult contact	0.00	0.00	0.00	0.00	0.00	0.00
Probability of starting PT, PTB index, adult contact	0.00	0.00	0.00	0.00	0.00	0.00
Probability of ending PT, PTB index, adult contact	0.00	0.00	0.00	0.00	0.00	0.00
Probability of starting PT, ETB index, adult contact	0.00	0.00	0.00	0.00	0.00	0.00
Probability of ending PT, ETB index, adult contact	0.00	0.00	0.00	0.00	0.00	0.00
Probability of starting PT, PTB index, adult contact	0.00	0.00	0.00	0.00	0.00	0.00
Probability of ending PT, PTB index, adult contact	0.00	0.00	0.00	0.00	0.00	0.00
Efficacy of PT, adult	0.01	0.01	0.01	0.01	0.01	0.01
Efficacy of PT, child	0.02	0.03	0.03	0.04	0.02	0.03
Probability of developing disease ever	0.03	0.03	0.07	0.12	0.13	0.28
Symptomatic period, passively detected PTB cases	0.01	0.03	0.01	0.04	0.02	0.05
Symptomatic period, contact traced PTB cases	0.06	0.16	0.11	0.26	0.14	0.31
Symptomatic period, PTB cases	0.00	0.00	0.00	0.00	0.00	0.00
Symptomatic period, passively detected ETB cases	0.01	0.00	0.00	0.00	0.00	0.00
Symptomatic period, contact traced ETB cases	0.60	0.56	0.48	0.38	0.30	0.15

Symptomatic period, passively detected cases	0.00	0.00	0.00	0.00	0.00	0.00

125

325 # Figure



326

327 **Figure 1: Summary of incremental cost-effectiveness ratios and 95% confidence intervals (shaded region) for different**
328 **levels of transmission from contacts. The comparator is no screening. The dashed horizontal line indicates the**
329 **£30000/QALY cost-effectiveness threshold and the dotted horizontal line the £20000/QALY threshold. The solid**
330 **horizontal line indicates when contact tracing becomes cost-saving. (a) and (b) represent the main results for ETB and**
331 **PTB index cases respectively. (c) and (d) represent results for a symptomatic period which is double the self-reported**
332 **period. GBP = pounds sterling, ETB = non-pulmonary, non-laryngeal tuberculosis, PTB = pulmonary or laryngeal**
333 **tuberculosis, QALY = quality-adjusted life years, ICER = incremental cost-effectiveness ratio.**

Acknowledgements

We would like to thank Jacqui Carless, Lamya Kanfoudi and others at PHE field Epidemiology London for maintaining the LTBR; all nurses and other healthcare workers involved with collecting the data and presenting it at cohort review; the TB modelling group at LSHTM for advice on presentation and approach; and Lindsay Serene for comments on the manuscript.

References

1. Tuberculosis (TB): collaborative strategy for England - GOV.UK [Internet]. [cited 2017 Jun 14]. Available from: https://www.gov.uk/government/publications/collaborative-tuberculosis-strategy-for-england

2. Erkens CGM, Kamphorst M, Abubakar I, Bothamley GH, Chemtob D, Haas W, et al. Tuberculosis contact investigation in low prevalence countries: a European consensus. Eur Respir J. 2010 Oct 1;36(4):925–49.

3. Tuberculosis | Guidance and guidelines | NICE [Internet]. [cited 2017 Jun 13]. Available from: https://www.nice.org.uk/guidance/ng33/chapter/Recommendations

4. Bothamley GH, Ditiu L, Migliori GB, Lange C, Contributors T. Active case finding of tuberculosis in Europe: a Tuberculosis Network European Trials Group (TBNET) survey. Eur Respir J. 2008 Oct 1;32(4):1023–30.

5. Rodriguez CA, Sasse S, Yuengling KA, Azzawi S, Becerra MC, Yuen CM. A systematic review of national policies for the management of persons exposed to tuberculosis. Int J Tuberc Lung Dis Off J Int Union Tuberc Lung Dis. 2017 Aug 1;21(8):935–40.

6. Cavany SM, Sumner T, Vynnycky E, Flach C, White RG, Thomas HL, et al. An evaluation of tuberculosis contact investigations against national standards. Thorax. 2017 Apr 7;thoraxjnl-2016-209677.

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

8. Saunders MJ, Koh GCKW, Small AD, Dedicoat M. Predictors of contact tracing completion and outcomes in tuberculosis: a 21-year retrospective cohort study. Int J Tuberc Lung Dis Off J Int Union Tuberc Lung Dis. 2014 Jun;18(6):640–6.

9. Wingfield T, MacPherson P, Cleary P, Ormerod LP. High prevalence of TB disease in contacts of adults with extrapulmonary TB. Thorax. 2017 Nov 16;

10. Aldridge RW, Zenner D, White PJ, Muzyamba MC, Loutet M, Dhavan P, et al. Prevalence of and risk factors for active tuberculosis in migrants screened before entry to the UK: a population-based cross-sectional study. Lancet Infect Dis. 2016 Aug;16(8):962–70.

11. Lalor MK, Anderson LF, Hamblion EL, Burkitt A, Davidson JA, Maguire H, et al. Recent household transmission of tuberculosis in England, 2010-2012: retrospective national cohort study combining epidemiological and molecular strain typing data. BMC Med. 2017 Jun 13;15(1):105.

12. Cavany SM, Vynnycky E, Sumner T, Macdonald N, Thomas HL, White J, et al. Transmission events revealed in tuberculosis contact investigations in London. Submitted. 2017;

13. Dasgupta K, Schwartzman K, Marchand R, Tennenbaum TN, Brassard P, Menzies D. Comparison of Cost-Effectiveness of Tuberculosis Screening of Close Contacts and Foreign-Born Populations. Am J Respir Crit Care Med. 2000 Dec 1;162(6):2079–86.

14. Committee JF, editor. BNF 73 (British National Formulary) March 2017. 73rd Revised edition edition. London: Pharmaceutical Press; 2017. 1480 p.

15. Mears J, Vynnycky E, Lord J, Borgdorff MW, Cohen T, Crisp D, et al. The prospective evaluation of the TB strain typing service in England: a mixed methods study. Thorax. 2016 Aug 1;71(8):734–41.

16. 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 Sep 29;190(9):1044–52.

17. Smieja M, Marchetti C, Cook D, Smaill FM. Isoniazid for preventing tuberculosis in non-HIV infected persons. In: The Cochrane Collaboration, editor. Cochrane Database of Systematic Reviews [Internet]. Chichester, UK: John Wiley & Sons, Ltd; 1999 [cited 2017 May 2]. Available from: http://doi.wiley.com/10.1002/14651858.CD001363

18. Ayieko J, Abuogi L, Simchowitz B, Bukusi EA, Smith AH, Reingold A. Efficacy of isoniazid prophylactic therapy in prevention of tuberculosis in children: a meta–analysis. BMC Infect Dis. 2014;14(1):1.

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

20. Trunz BB, Fine P, Dye C. Effect of BCG vaccination on childhood tuberculous meningitis and miliary tuberculosis worldwide: a meta-analysis and assessment of cost-effectiveness. Lancet Lond Engl. 2006 Apr 8;367(9517):1173–80.

21. 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 Dec;46(6):1563–76.

22. Guide to the methods of technology appraisal 2013 | Guidance and guidelines | NICE [Internet]. [cited 2017 May 31]. Available from: https://www.nice.org.uk/process/pmg9/chapter/the-appraisal-of-the-evidence-and-structured-decision-making

23. Erkens CGM, Slump E, Verhagen M, Schimmel H, Cobelens F, Hof S van den. Risk of developing tuberculosis disease among persons diagnosed with latent tuberculosis infection in the Netherlands. Eur Respir J. 2016 Oct 6;ERJ-01157-2016.

24. Kruijshaar ME, Lipman M, Essink-Bot M-L, Lozewicz S, Creer D, Dart S, et al. Health status of UK patients with active tuberculosis. Int J Tuberc Lung Dis Off J Int Union Tuberc Lung Dis. 2010 Mar;14(3):296–302.

25. Jit M, Stagg HR, Aldridge RW, White PJ, Abubakar I. Dedicated outreach service for hard to reach patients with tuberculosis in London: observational study and economic evaluation. BMJ. 2011 Sep 13;343:d5376.

26. Tuberculosis (TB): regional and devolved administration reports - GOV.UK [Internet]. [cited 2017 Jun 22]. Available from: https://www.gov.uk/government/publications/tuberculosis-tb-regional-reports

27. Pareek M, Watson JP, Ormerod LP, Kon OM, Woltmann G, White PJ, et al. Screening of immigrants in the UK for imported latent tuberculosis: a multicentre cohort study and cost-effectiveness analysis. Lancet Infect Dis. 2011 Jun 1;11(6):435–44.

28. Tan MC, Marra CA, Sadatsafavi M, Marra F, Morán-Mendoza O, Moadebi S, et al. Cost-effectiveness of LTBI treatment for TB contacts in British Columbia. Value Health J Int Soc Pharmacoeconomics Outcomes Res. 2008 Oct;11(5):842–52.

29. Crisan A, Wong HY, Johnston JC, Tang P, Colijn C, Otterstatter M, et al. Spatio-temporal analysis of tuberculous infection risk among clients of a homeless shelter during an outbreak. Int J Tuberc Lung Dis. 2015 Sep 1;19(9):1033–8.

30. Walker TM, Cruz ALG, Peto TE, Smith EG, Esmail H, Crook DW. Tuberculosis is changing. Lancet Infect Dis. 2017 Apr;17(4):359–61.

Statements

Declaration of interests: The authors have declared that no competing interests exist.

Details of contributors: SMC, EV and TS conceived and designed the work, with input from all other authors. CA and HM are responsible for the acquisition and maintenance of the data. SMC undertook the analysis with advice from all other authors. FS calculated the costs. All authors contributed to the interpretation of the data. SMC wrote the first draft of the paper and all authors contributed to subsequent drafts. All authors approve the work for publication and agree to be accountable for the work.

Funding sources: SMC and FGS are funded separately by a joint PHE-LSHTM studentship in Infectious Disease Modelling. RGW is funded by the UK Medical Research Council (MRC) and the UK Department for International Development (DFID) (under the MRC/DFID Concordat agreement that is also part of the EDCTP2 programme supported by the European Union (MR/J005088/1)), the Bill and Melinda Gates Foundation (TB Modelling and Analysis Consortium: OPP1084276, and SA Modelling for Policy: #OPP1110334) and UNITAID (4214-LSHTM-Sept15; PO #8477-0-600). CA, EV, HM and LT are employed by PHE, a government agency, and received no other source of funding. TS is funded by the Bill and Melinda Gates Foundation (SA Modelling for Policy: #OPP1110334). SMC made the final decision to submit and had access to all data.

Software: All data were analysed using Microsoft Excel 14.0, Stata 13.1 and Python 2.7.9 (including pandas 0.15.2, NumPy 1.13.1 and matplotlib 2.0.2).

Ethics. Ethical approval was not required. The data analysed were routinely collected surveillance data held by PHE under Section 251 of the NHS Act 2006. All records were anonymised before analysis.

Paper appendix

See next page

Appendix

1. Cost Calculations

Table A: Cost of contact tracing, per contact traced (C_0), † indicated costs which were inflated to the base year 2016. HRG = healthcare resource group, IGRA = Interferon Gamma Release Assay, NHS = National Health Service, NICE = National Institute for Health & Care Excellence, TB = tuberculosis.

Item	Quantity	Unit costs (£)	Total costs, including uprating (£)	Reference
TB specialist nurse - non face to face (HRG code N28AN)	1	21.10	21.10	NHS Reference costs 2015-2016[31]
TB specialist nurse - face to face (HRG code N28AF)	2	61.93	123.86	NHS Reference costs 2015-2016[31]
Mantoux skin test†	1	1.22	1.32	NICE[32,33]
IGRA blood test†	0.5	56	30.21	Pareek et al.[33,34]
Outpatient appointment for IGRA positive (HRG code WF02B)	0.25	239.83	59.96	NHS Reference costs 2015-2016[31]
Chest X-ray (to rule out active disease)†	0.25	28	8.04	NICE[33,35]
Per contact traced			**244**	

Table B: Further tests if case is suspected to have active disease, per contact with suspected TB (C_1), † indicated costs which were inflate. HRG = healthcare resource group, NHS = National Health Service, NICE = National Institute for Health & Care Excellence, TB = tuberculosis.

Item	Quantity	Unit costs (£)	Total costs, including uprating (£)	Reference
TB specialist nurse - face to face (HRG code N28AF)	3	61.93	185.78	NHS Reference costs 2015-2016[31]
Outpatient appointment for diagnosis (HRG code WF02B)	1	239.83	239.83	NHS Reference costs 2015-2016[31]
Chest X-ray†	1	28	32.17	NICE[33,35]
Sputum smear microscopy†	1	1.56	1.83	Dowdy[33,36]
Culture and MDR identification†	1	30	36.37	Dinnes[33,37]
Liver function test (HRG code DAPS04)	1	1.18	1.18	NHS Reference costs 2015-2016[31]
Per contact with suspected TB			**497**	

Table C: Cost per full course PT (3 month rifampicin and isoniazid with pyridoxine) (C_{PT}) . BNF = British National Formulary, HRG = healthcare resource group, NHS = National Health Service, PT = preventive therapy.

Item	Quantity (pack-size if applicable)	Unit costs (£)	Total costs (£)	Reference
Follow-up appointments nurse only (HRG code N28AF)	3	91.93	185.78	NHS Reference costs 2015-2016[31]
Follow-up appointments nurse & consultant (HRG code WF02A)	2	188.50	377	NHS Reference costs 2015-2016[31]
Isoniazid 300mg daily, 3 months	10 (28)	19.24	192.40	BNF 2017[11]
Rifampicin 600mg daily, 3 months	4 (60)	21.98	87.92	BNF 2017[11]
B6 pyridoxine 10mg tablets (per month)	1 (500)	8.48	8.48	BNF 2017[11]
Per person completing treatment			**852**	

Table D: Cost per full course of treatment for tuberculosis disease (C_{FT}), * assuming 15% do not complete treatment, after an average of 2 months. BNF = British National Formulary, HRG = healthcare resource group, IGRA = Interferon Gamma Release Assay, NHS = National Health Service, NICE = National Institute for Health & Care Excellence, TB = tuberculosis.

Item	Quantity (pack-size if applicable)	Unit costs (£)	Total costs (£)	Reference
Admission (weighted average of HRG codes DZ14F, DZ14G, DZ14H, DZ14J, DZ51Z)	0.05	3904.16	195.21	NHS Reference costs 2015-2016[31]
Follow-up appointments nurse only (HRG code N28AF)	5	61.93	309.64	NHS Reference costs 2015-2016[31]
Follow-up appointments nurse & consultant (HRG code WF02A)	2	188.50	942.51	NHS Reference costs 2015-2016[31]
Rifater (R,I ,P) 120mg, 50mg, 300mg (Sanofi, 100 tabs) 6 tablets daily for 2 months	4 (100)	26.34	105.36	BNF 2017[11]
Ethambutol hydrochloride (solo) 400mg (non-proprietary, 56 tabs) 3 tabs daily for 2 months	4 (56)	42.74	170.96	BNF 2017[11]
Rifinah (R,I) 300/150 (Sanofi, 56 tabs) 2 tab daily for 4 months	5 (56)	25.22	126.10	BNF 2017[11]
Pyridoxine B6 (solo) 10mg (non-proprietary, 500 tabs) 1 tab daily for 6 months	1 (500)	8.48	8.48	BNF 2017[11]
Per person starting treatment*			**1694**	

Comment on inflation of costs: Cost values taken from past years were inflated to 2016 according to consumer price inflation[33].

Comment on quantities: Unless stated otherwise, estimated quantities for items other than drugs or NHS reference costs were taken from Mears et al.[12], who based the average estimates on standard treatment protocols, informed by expert judgement.

Comment regarding tariff costs: Using NHS tariff costs appeared to give higher costs for hospital admission than did the reference costs, and so the use of these costs was not further explored.

Comment regarding sensitivity and uncertainty of costs: We did not vary costs when calculating uncertainty, or in the sensitivity analyses, because the standard errors in the costs were very small relative to the standard errors in other parameters. We estimated the standard deviation of costs from the mean cost and quartile costs using the approach of Wan et al.[38]. We then calculated the standard error using the sample size in the NHS reference costs. For example, the unit cost for a TB specialist nurse (non-face to face) was £21.10 (see above), the lower and upper quartiles are £15.36 and £28.77, and the sample size is 53742. The approach referred to above calculates the standard deviation as £9.94, and hence the standard error is £0.04, or less than 0.5% of the cost.

2. Data tables

Table E: Number of cases in each category and the respective mean symptomatic period (in days), with information on missing data. Note slightly different exclusion criteria used for each of the final three columns, hence the final two columns do not align exactly eih the first columns. See Cavany et al.[6] for details of exclusion criteria. CT = Contact tracing, s.e. = standard error

	All cases	Pulmonary or Laryngeal index	Non-pulmonary and non-laryngeal index
Total number of cases in study period	11174	5084	6090
Total number included after applying exclusion criteria (% is of all cases in study period)	5509 (49%)	2758 (54%)	2897 (48%)
With data on symptomatic period (% is of all those included)	4906 (89%)	2465 (89%)	2559 (88%)
Found through CT (% is of all those included)	142 (2.6%)	113 (4.1%)	41 (1.4%)
With data on symptomatic period (% is of all those included & found through CT)	102 (72%)	82 (73%)	26 (63%)
Found through other routes (% is of all those included)	5367 (97%)	2645 (96%)	2856 (99%)
With data on symptomatic period (% is of all those included & found through other routes)	4804 (90%)	2383 (90%)	2533 (89%)
Mean symptomatic period , all cases (days)	147 [s.e. = 4.26]	109 [s.e. = 3.55]	180 [s.e. = 7.36]
Mean symptomatic period if found through CT (days)	97.6 [s.e. = 19.3]	76.6 [s.e. = 9.26]	152 [s.e. = 69.9]
Mean symptomatic period if found through other routes (days)	148 [s.e. = 4.33]	110 [s.e. = 3.66]	180 [s.e. = 7.41]

Table F: Break down of the number of LTBI cases starting and completing preventive therapy by disease type of the index. Note that different cases were included for the three categories (as the exclusion criteria were applied independently to each category), hence the lack of agreement. See Cavany et al.[6] for details of exclusion criteria. Cases found through contact tracing are excluded as their contact tracing results are not reported consistently. LTBI = latent *M.Tb* infection, PT = preventive therapy, s.e.=standard error.

	All cases found through routes other than contact tracing
Number of index cases included and with data on PT	5367 (55%)
Pulmonary/Laryngeal included	2645 (55%)
Non-pulmonary & non-laryngeal included	2856 (56%)
Mean number of contacts with LTBI per index	0.279 [s.e. = 0.0114]
Pulmonary/Laryngeal	0.471 [s.e. = 0.0219]

Non-pulmonary & non-laryngeal	0.119 [s.e. = 0.00395]
Mean number of contacts starting PT per index	0.217 [s.e. = 0.00939]
Pulmonary/Laryngeal	0.366 [s.e. = 0.0178]
Non-pulmonary & non-laryngeal	0.0907 [s.e. = 0.00668]
Mean number of contacts completing PT per index	0.171 [s.e. = 0.00819]
Pulmonary/Laryngeal	0.289 [s.e. = 0.0157]
Non-pulmonary & non-laryngeal	0.0683 [s.e. = 0.0219]

Table G: Break down of the proportion of contacts with LTBI starting and completing preventive therapy by disease type of the index and age of the contact. LTBI = latent *M.Tb* infection, PT = preventive therapy, s.e.=standard error. Child refers to ages 0-14 and adult to 15 and above.

	All contacts	**Adult contacts**	**Child contacts**
Proportion of those with LTBI that start PT	0.778 [s.e. = 0.0147]	0.655 [s.e. = 0.0279]	0.962 [s.e. = 0.0231]
Pulmonary/laryngeal	0.776 [s.e. = 0.0168]	0.666 [s.e. = 0.0318]	0.969 [s.e. = 0.0238]
Non-pulmonary & non-laryngeal	0.762 [s.e. = 0.0274]	0.611 [s.e. = 0.0515]	0.931 [s.e. = 0.0477]
Proportion of those who start PT that complete PT	0.786 [s.e. = 0.0164]	0.822 [s.e. = 0.0266]	0.917 [s.e. = 0.0245]
Pulmonary/Laryngeal	0.790 [s.e. = 0.0186]	0.803 [s.e. = 0.0310]	0.906 [s.e. = 0.0311]
Non-pulmonary & non-laryngeal	0.753 [s.e. = 0.0319]	0.875 [s.e. = 0.0420]	0.810 [s.e. = 0.0876]

3. Equations

We use the following equations to quantify the ICER:

1. The prevention of subsequent generations of cases is calculated as follows. We say each prevented case would have generated

$$F = (1 - \epsilon_\sigma) S_{P,\,\text{overall}} rP$$

cases, and each of those prevented cases would have generated F cases etc. Hence, taking the limit of this sum, we find that

$$N_{\text{later generations},\sigma} = \frac{(N_{\text{prevention},\,\sigma} + N_{\text{transmission},\,\sigma})F}{1 - F}$$

2. The costs are calculated in the following ways:

The costs of screening contacts (note that contacts with TB are excluded from the second part of the equation, as we assume they will ultimately be diagnosed anyway):

$$\text{cost}_{\text{screening},\,\sigma} = N_\sigma \left(n_\sigma C_0 + Y_\sigma C_1 \left(\frac{1}{f_c} - 1 \right) \right)$$

The costs of administering preventive therapy to contacts, including those who begin but do not complete preventive therapy, and deducting costs of cases which do not occur. Costs are assumed to be incurred at a constant rate:

$$\text{cost}_{\text{preventive therapy},\,\sigma} = N_\sigma C_{\text{PT}} D \sum_{j=a,c} y_\sigma \phi_{\sigma,j} \theta_{j,\sigma,B} \left(\theta_{j,\sigma,C} + (1 - \theta_{j,\sigma,C}) f_i \right) - N_{\text{prevention},\,\sigma} (C_1 + C_{\text{FT}})$$

The costs saved by preventing transmission from contacts and stopping transmission from prevented cases:

$$\text{cost}_{\text{transmission},\,\sigma} = -(C_1 + C_{\text{FT}}) \left(N_{\text{transmission},\,\sigma} + N_{\text{later generations},\,\sigma} \right)$$

3. The effectiveness measured in QALYs is calculated as follows:

The QALYs gained by finding cases sooner:

$$\text{QALY}_{\text{morbidity},\,\sigma} = (U_H - U_0) t_{\text{morbidity},\,\sigma}$$

The QALYs gained for each case prevented is:

$$\text{QALY}_{\text{extra case}} = (U_H - U_0)\frac{S_{\text{overall}}}{365.25} + \frac{6(U_H - U_1)}{12}$$

The QALYs gained by administering preventive therapy to contacts with latent infection (the number of QALYs gained by preventing cases by preventive therapy), and subtracting QALYs lost from three months of preventive therapy:

$$\text{QALY}_{\text{prevention, }\sigma} = N_{\text{prevention, }\sigma}\text{QALY}_{\text{extra case}} - \frac{3N_\sigma(U_{\text{H}} - U_{\text{PT}})}{12}\sum_{j=a,c} y_{j,\sigma}\theta_{j,\sigma,B}\left(\theta_{j,\sigma,C} + (1 - \theta_{j,\sigma,C})f_i\right)$$

The QALYs gained by reducing transmission from contacts with TB:

$$\text{QALY}_{\text{transmission, }\sigma} = \left(N_{\text{transmission, }\sigma} + N_{\text{later\_generations},\sigma}\right)\text{QALY}_{\text{extra case}}$$

The QALYs gained by reducing the number of deaths:

$$\text{QALY}_{\text{mortality, }\sigma} = \left(N_{\text{mortality, }\sigma} - N_{\text{later generations, }\sigma}\right)\mu(A_H - A_{TB})$$

4. Putting all this together we get an incremental cost effectiveness ratio (ICER) per QALY gained, using no screening as the baseline, of

$$\text{ICER}_\sigma = \frac{\text{cost}_{\text{screening, }\sigma} + \text{cost}_{\text{preventive therapy, }\sigma} + \text{cost}_{\text{transmission, }\sigma}}{\text{QALY}_{\text{morbidity, }\sigma} + \text{QALY}_{\text{prevention, }\sigma} + \text{QALY}_{\text{transmission, }\sigma} + \text{QALY}_{\text{mortality, }\sigma}}$$

Note that the ICER is independent of the number of index cases (N) as this appears once in each term on the numerator and denominator

4. Styblo Rule

Given that in London 45% of PTB cases are smear positive(21), and assuming that smear negative PTB cases are 0.22 times as infectious as smear positive PTB cases(22), if each PTB case is symptomatic for 0.3 years , then a value of $r = 1$ (or each PTB case generates 12 infections per year) means that each smear positive case equates to about 6.3 new infections.

Let r_+ be the number of new infections per PTB smear positive case per month infectious (and recall that r is the number of new infections per PTB case per month infectious), p_+ the proportion of PTB cases that are smear positive, and ρ_- be the relative infectiousness of smear negative cases (estimated as 0.22[19]). Then $r = r_+ p_+ + r_+ \rho_- (1 - p_+)$, or rearranging, $r_+ = \frac{r}{p_+ + \rho_- (1 - p_+)}$.

Then, with $p_+ = 0.45$ and $\rho_- = 0.22$, with $S_{P,\,\text{overall}} = 0.3$ years, we find that $r = 1$ corresponds to each smear positive case generating $r_+ S_{P,\,\text{overall}} = 6.3$ infections. This is broadly in line with the two re-estimations of the Styblo rule[16,17], although it is at the upper end of the ranges found in those studies.

5. Discounting

Whilst discounting was not included in the main analysis, we explored the potential impact of including discounting in a rudimentary fashion, using the following approach. First, we calculate a discounting factor, D, by assuming new cases occur after infection according to the distribution given in Sloot et al.[13] , and assume all those with infection have been recently infected. We then discount cases which occur after the first year at a rate of 3.5% per year and 1.5% per year (as recommended by NICE[21]). Finally we multiply those costs and QALYs which occur in the future by this factor D (i.e. costs and QALYs associated with mortality and cases prevented through reduced transmission and preventive therapy). As most cases develop disease in the first year using this approach, this equated to multiplying the total number of QALYs and costs gained/incurred from these cases by $D = 0.987$ for a rate of 3.5%/year and $D = 0.994$ for a rate of 1.5%/year.

For example, we calculate the discounting factor for a 3.5% per year rate of discounting according to Table H. Column 2 is taken from Sloot et al. and shows the cumulative risk of developing disease by year, and column 3 the yearly risk. The fourth column indicates the factor by which we should discount costs and QALYs occurring this year, calculated as $\frac{1}{(1+0.035)^{\text{year}-1}}$, so that cases in the first year are undiscounted. The final column is then the discounted risk in the given year, i.e. the product of columns three and four. The discounting factor is then the cumulate discounted risk (sum of last column) divided by the cumulative undiscounted risk (sum of third column).

Table H: Calculating the discounting factor, see text for explanation. The discounting factor is the ratio of the sum of the final column to the sum of the third column

Year	Cumulative risk of developing disease	Risk of developing disease this year	Discount cases in this year by a factor of…	Discounted risk this year
1	0.083005	0.083005	1	0.083005
2	0.094874	0.011869	0.985222	0.011694
3	0.096503	0.001629	0.970662	0.001581
4	0.096503	0	0.956317	0
5	0.096503	0	0.942184	0
6	0.096503	0	0.92826	0
7	0.099067	0.002564	0.914542	0.002345
8	0.099067	0	0.901027	0
9	0.099067	0	0.887711	0
10	0.099067	0	0.874592	0
11	0.099067	0	0.861667	0
12+	0.1	0.000933	0.848933	0.000792

The following two tables show results when discounting is set to 3.5%/year, and 1.5%/year. As most cases occur in the first year, which is undiscounted, discounting has little impact on results.

Table I: Summary of the effectiveness measures included, costs incurred, quality adjusted life years(QALYs) gained and resulting incremental cost effectiveness ratio (ICER) for screening contacts of the indicated index cases compared to a baseline of not screening those contacts, with a discounting rate of 3.5%/year. Numbers are given for a year with a case-load that is the average caseload of the years 2012-15 (i.e. 2790 cases); note that the case-load does not affect the ICER. Case-equivalents averted refers to both cases averted, and the reduction in the time contacts are symptomatic divided by the mean symptomatic period of TB cases. ETB = non-pulmonary, non-laryngeal; PTB = pulmonary or laryngeal; r = the number of infections generated by a pulmonary contact per month infectious; ICER = incremental cost-effectiveness ratio; PT = preventive therapy (3 months of isoniazid and rifampicin). Numbers in brackets indicate the 95% confidence intervals. Discounting makes little difference to the ICER in this analysis.

Quantity (units, if applicable)	ETB indexes			PTB indexes		
	$r = 0$	$r = 1$	$r = 2$	$r = 0$	$r = 1$	$r = 2$
Reduction in time contacts are symptomatic (years)		2.58 [0.665, 8.78]			10.5 [4.0, 27.0]	
Cases prevented by administering PT (cases)		5.45 [3.72, 7.59]			18.9 [13.2, 25.5]	
Transmission reduced by finding contacts sooner (cases)	0.0 [0.0, 0.0]	1.71 [0.591, 3.31]	3.41 [1.16, 6.62]	0.0 [0.0, 0.0]	8.76 [3.54, 14.9]	17.5 [7.08, 29.7]
Transmission reduced from prevented cases (cases)	0.0 [0.0, 0.0]	1.62 [0.754, 3.13]	5.19 [2.11, 12.2]	0.0 [0.0, 0.0]	8.63 [4.73, 14.8]	33.1 [16.0, 66.7]
Reduction in mortality (deaths)	0.431 [0.239, 1.0]	0.551 [0.307, 1.15]	0.743 [0.399, 1.44]	1.64 [0.999, 3.15]	2.27 [1.38, 3.89]	3.47 [2.02, 5.89]
Total case-equivalents averted	11.9 [6.58, 27.6]	15.2 [8.44, 31.7]	20.5 [11.0, 39.8]	45.0 [27.5, 86.9]	62.4 [37.9, 107.0]	95.6 [55.6, 162.0]
Total QALYs gained	10.4 [5.93, 23.7]	13.5 [7.65, 27.3]	18.4 [9.96, 35.2]	39.4 [24.5, 74.6]	55.5 [34.0, 94.0]	86.0 [50.2, 146.0]
Total costs incurred (£ 000 000s)	1.07 [1.03, 1.12]	1.06 [1.02, 1.11]	1.05 [1.01, 1.1]	1.75 [1.67, 1.82]	1.71 [1.63, 1.78]	1.64 [1.53, 1.73]
Incremental cost-effectiveness ratio (£ 000s/QALY)	103.0 [45.6, 179.0]	78.8 [39.2, 138.0]	57.4 [29.9, 107.0]	44.2 [23.5, 71.1]	30.8 [18.1, 50.7]	19.0 [10.7, 33.4]

Table J: Summary of the effectiveness measures included, costs incurred, quality adjusted life years(QALYs) gained and resulting incremental cost effectiveness ratio (ICER) for screening contacts of the indicated index cases compared to a baseline of not screening those contacts, with a discounting rate of 1.5%/year. Numbers are given for a year with a case-load that is the average caseload of the years 2012-15 (i.e. 2790 cases); note that the case-load does not affect the ICER. Case-equivalents averted refers to both cases averted, and the reduction in the time contacts are symptomatic divided by the mean symptomatic period of TB cases. ETB = non-pulmonary, non-laryngeal; PTB = pulmonary or laryngeal; r = the number of infections generated by a pulmonary contact per month infectious; ICER = incremental cost-effectiveness ratio; PT = preventive therapy (3 months of isoniazid and rifampicin). Numbers in brackets indicate the 95% confidence intervals. Discounting makes little difference to the ICER in this analysis.

Quantity (units, if applicable)	ETB indexes			PTB indexes		
	$r = 0$	$r = 1$	$r = 2$	$r = 0$	$r = 1$	$r = 2$
Reduction in time contacts are symptomatic (years)		2.58 [0.683, 8.93]			10.5 [4.02, 26.8]	
Cases prevented by administering PT (cases)		5.45 [3.73, 7.61]			18.9 [13.3, 25.8]	
Transmission reduced by finding contacts sooner (cases)	0.0 [0.0, 0.0]	1.71 [0.569, 3.29]	3.41 [1.16, 6.68]	0.0 [0.0, 0.0]	8.76 [3.4, 14.8]	17.5 [7.02, 29.7]
Transmission reduced from prevented cases (cases)	0.0 [0.0, 0.0]	1.62 [0.765, 3.11]	5.19 [2.12, 12.3]	0.0 [0.0, 0.0]	8.63 [4.81, 14.6]	33.1 [15.8, 66.7]
Reduction in mortality (deaths)	0.431 [0.24, 1.0]	0.551 [0.307, 1.14]	0.743 [0.399, 1.43]	1.64 [1.01, 3.13]	2.27 [1.38, 3.88]	3.47 [2.02, 5.82]
Total case-equivalents averted	11.9 [6.61, 27.7]	15.2 [8.46, 31.5]	20.5 [11.0, 39.3]	45.0 [27.9, 86.3]	62.4 [38.1, 107.0]	95.6 [55.7, 160.0]
Total QALYs gained	10.5 [5.98, 23.9]	13.6 [7.68, 27.4]	18.5 [10.1, 35.1]	39.7 [24.9, 74.6]	55.9 [34.5, 94.1]	86.8 [50.9, 145.0]
Total costs incurred (£ 000 000s)	1.07 [1.03, 1.12]	1.06 [1.02, 1.11]	1.05 [1.01, 1.1]	1.75 [1.68, 1.82]	1.71 [1.64, 1.78]	1.64 [1.52, 1.73]
Incremental cost-effectiveness ratio (£ 000s/QALY)	102.0 [45.3, 177.0]	78.2 [39.1, 138.0]	56.9 [29.8, 105.0]	43.9 [23.4, 69.9]	30.5 [18.1, 49.7]	18.8 [10.7, 33.1]

6. Utility Calculations

Table K: This table shows how we calculate the number of quality adjusted life years (QALYs) accrued at death by someone who does not have tuberculosis, and the number accrued by someone who dies prematurely from TB. The case fatality ratio (CFR), life expectancy (LE) and utilities are taken from Mears et al.[12], the population estimates are taken from Office for National Statistics[30], and the caseloads are taken from the London TB register. The final three columns are calculated. The QALYs at this LE is calculated by summing the age-specific healthy utilities up until the gender-specific life-expectancy given in that row. From these columns we can calculate the values of $A_H = 72.4$ and $A_{TB} = 52.1$ given in Error! Reference source not found..

Age	CFR (%)	LE (men)	LE (women)	Utility with no TB	Male population, 2015	Female population, 2015	Cases in London , 2015	QALYs/year up to this age	QALYs at this LE (men)	QALYs at this LE (women)
0-4	0.3	78	82.1	0.94	324,404	308,866	254	0.94	69.83	72.7
5-9	0.2	73.5	77.5	0.94	289,083	276,855	177	0.94	70.18	72.98
10-14	0.2	68.5	72.5	0.94	241,496	231,172	347	0.94	70.18	72.98
15-19	1.2	63.5	67.6	0.94	240,789	227,381	890	0.94	70.18	73.05
20-24	1.2	58.7	62.6	0.94	284,353	290,053	2152	0.94	70.32	73.05
25-29	1.2	53.9	57.7	0.94	410,745	421,885	3040	0.94	70.46	73.12
30-34	1.2	49	52.8	0.91	441,020	420,357	2670	0.938	70.53	73.19
35-39	1.2	44.3	47.9	0.91	378,059	358,584	1995	0.934	70.74	73.26
40-44	1.2	39.6	43.1	0.88	319,017	308,709	1522	0.929	70.95	73.37
45-49	4.8	34.9	38.3	0.86	293,289	296,029	1323	0.923	71.16	73.5
50-54	4.8	30.3	33.6	0.83	261,918	271,143	1019	0.915	71.44	73.695
55-59	4.8	25.9	29.1	0.82	213,163	219,039	840	0.908	71.86	74.02
60-64	4.8	21.7	24.6	0.81	166,398	178,540	659	0.900	72.42	74.345
65-69	17.6	17.7	20.4	0.8	147,623	163,020	527	0.893	73.12	74.865
70-74	17.6	14.1	16.4	0.78	103,809	121,884	481	0.886	74.02	75.515
75-79	17.6	10.8	12.6	0.75	85,856	103,614	382	0.878	75.125	76.295
80-84	17.6	8	9.4	0.7	59,937	79,369	242	0.869	76.555	77.465
85-89	17.6	5.8	6.7	0.65	32,967	52,762	138	0.858	78.375	78.96
90-	17.6	4.2	4.6	0.65	15,586	34,939	43	0.846	80.585	80.845

In order to calculate A_H from this table, we take the product of the "Population" and "QALYs at this LE" columns, divided by the total population, to get the QALYs at death for both men and women. We then take the sum of these, weighted by the population of each group to get A_H. To calculate A_{TB}, we take the product of the "CFR", "Cases in

London" and "QALYS/year up to this age" columns and the midpoint of the age ranges, and divide by the product of the "Cases in London" and "QALYS/year up to this age" columns.

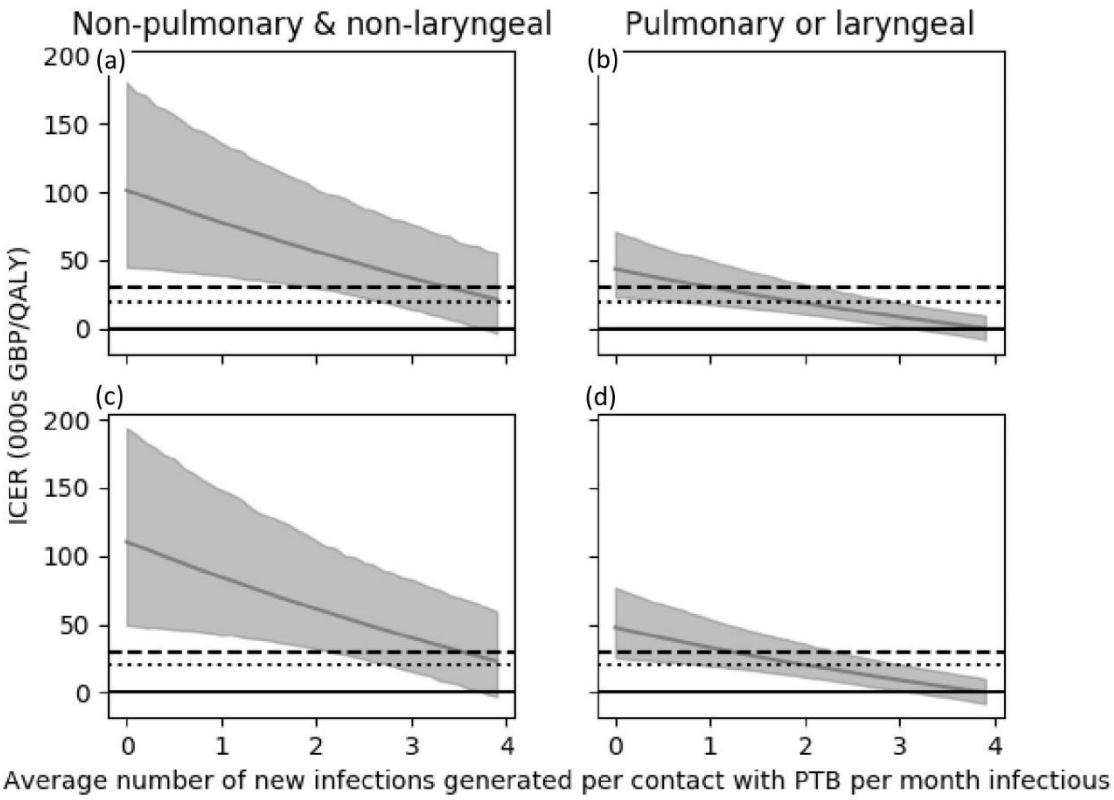


Figure A: Summary of incremental cost-effectiveness ratios and 95% confidence intervals (shaded region) for different levels of transmission from contacts if we use utility scores given in Mears et al. (c and d) as opposed to Jit et al. (a and b) The comparator is no screening. The dashed horizontal line indicates the £30000/QALY cost-effectiveness threshold and the dotted horizontal line the £20000/QALY threshold. The solid horizontal line indicates when contact tracing becomes cost-saving. GBP = pounds sterling, ETB = non-pulmonary, non-laryngeal tuberculosis, PTB = pulmonary or laryngeal tuberculosis, QALY = quality-adjusted life years, ICER = incremental cost-effectiveness ratio

Addendum

Clarifications

In the methods section, we state that 'infected contacts are given a 3 month course of rifampicin and isoniazid'. To clarify: this is the current standard of care in the UK[43].

- In section 3 of the appendix ('Equations'), the equation for $N_{\text{later generations},\,\sigma}$ will only converge if $F \leq 1$, and also neglects overlapping contacts. As $F \ll 1$ with parameters used in the model, both of these assumptions are valid.

Corrections

- The last subsection in the methods section should read 'role of funding source'.

5. Estimating the relative transmission intensity from tuberculosis cases with reactivation disease compared to those with primary disease using a pairwise model

Introduction

The results presented in chapter four suggest that it is very unlikely that screening contacts of non-pulmonary, non-laryngeal cases is cost effective in London, and hence in the UK. However, it was uncertain whether screening contacts of pulmonary or laryngeal cases was cost-effective at either of the cost-effectiveness thresholds used in the UK (i.e. £20000 and £30000). This uncertainty was largely derived from uncertainty in the symptomatic period of cases found through contact tracing, and in the number of infections generated by cases per day (see Table 4 and Figure 1 in chapter 4). It is plausible that recently infected cases (as those found through contact tracing often are – see chapter 3) will generate fewer new infections, on average, than cases with reactivation disease, due to the likelihood of the recently infected cases sharing contacts with their index case (a 'saturation effect'). In this final section, we aim to explore this saturation effect.

In order to do so it is necessary to consider the local network structure of the population around each case, in order to estimate how many of their contacts are exposed. Often, this has been incorporated into TB models by using an individual based model with an explicit network structure[78–80,84]. A significant drawback of this type of model is that it can be difficult to parametrize and computationally expensive[85]. An approach which has been used for a range of other diseases, although never for TB, is that of pairwise equations. As discussed in chapter 1, this approach, rather than having equations which describe the number of individuals at each stage of the natural history of a given disease, instead has equations which describe the number of pairs across which transmission can occur[91–93,123]. This is particularly intuitive for sexually transmitted diseases, for which the pair is the natural unit, and has been used in theoretical contact tracing studies before[93].

Previous studies have extended the pairwise equation system to include a random, 'mass action' transmission term to the equations[92]. If we consider pairs to represent close (e.g. household) contacts, and the random transmission term to represent casual contacts, then a pairwise system may plausibly describe *M. Tb* transmission. However, as a pairwise equation system needs one differential equation for each possible pair of disease states, the complexity of TB's natural history (e.g. compared to gonorrhoea) will cause the resultant system of equations to be correspondingly complex.

Our primary aim in this study was to understand the extent to which a pairwise model is useful for modelling *M. Tb* transmission. Our secondary aim was to use the pairwise model to understand the extent of the saturation effect when considering the amount of transmission from cases that have been recently infected or reinfected (for simplicity, taken to be within a year), relative to those with reactivation disease. This has relevance for contact tracing, as cases found through contact tracing are likely to have been infected in the past year (see chapter 3). We believe this is the first study to quantify whether the number of infections attributable to cases that have themselves been recently infected differs from that from cases experiencing reactivation disease, and the first pairwise model applied specifically to tuberculosis.

Methods

Model structure: As the number of differential equations required for a pairwise model increases quadratically with the number of compartments used to describe the natural history of the disease, it is beneficial to keep the number of compartments as small as possible, at least for this exploratory study. Consequently, we used a simplified model of tuberculosis natural history which doesn't account for age structure (Figure 11), parameterized using values from the literature and from surveillance data for London (Table 3). In the model, everyone is born uninfected. People can be infected either by a close (pairwise) contact (represented by τ in Figure 11 and the equations) or by a casual contact (represented by β). Infected people are stratified into those that are recently infected or reinfected (L_f) and those that have latent infection (L_s) – people stay in the recently (re)infected compartment for one year on average before moving to the latent compartment[12]. People with a recent infection (L_f compartment) progress to disease at a rate, p_f, and those with latent infection (L_s compartment) develop reactivation disease at a rate p_s. As the recently infected develop disease at a faster rate than those with latent infection, $p_s < p_f$.

Those in the L_s compartment can also be reinfected, upon which there is a transition to the L_f compartment, but the chance of reinfection is reduced by a factor σ compared to initial infection, due to a protective effect of previous disease or infection. There are two infectious compartments, but this is merely to record from which infected compartment disease was developed; there are no differences in the recovery or mortality rates, nor in infectiousness. Both infectious compartments recover to the latent compartment, implying a life-long risk of relapse/reinfection. Both infected and both infectious compartments have a constant rate of mortality, with a higher rate in the infectious compartments. To maintain a constant population size, the birth rate (into the susceptible compartment) is equal to the mortality rate.

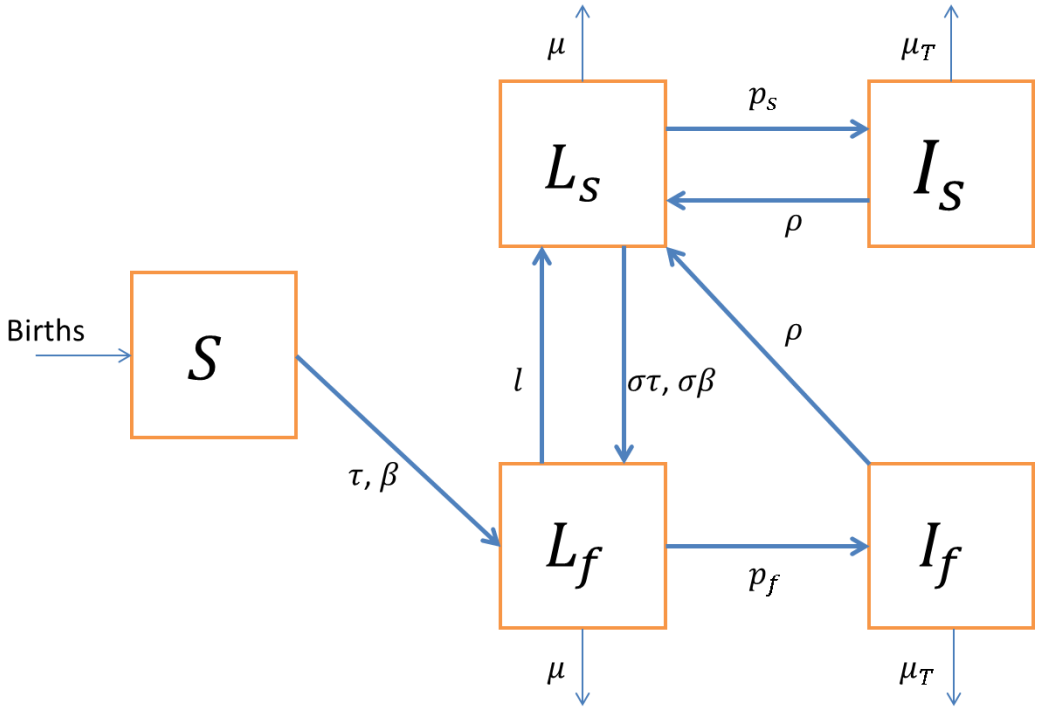


Figure 11: Model diagram. *S* refers to susceptible (previously uninfected) people, L_s refers to latently infected, or recovered people, L_f refers to recently infected people, I_s refers cases who developed reactivation disease, I_f refers to cases who developed disease after a recent infection or reinfection. The two rates of infection, τ and β, represent pairwise and mass-action transmission rates respectively

Table 3: Parameter values, or prior distributions of parameters that were fitted. *The ranges for p_f and p_s were chosen so that their geometric mean value approximately matched the rates of progression to disease following infection given in Sloot et al.[12]

Parameter	Value, or prior distribution	Units	Source
Contacts per person, n	5	Contacts	London TB register
Pairwise contact parameter, τ	Log-Uniform(4×10^{-5}, 4×10^{-1})	Number of effective pairwise contacts made per day	Plausible range
Random contact parameter, β	Uniform(0, 1)	Number of effective casual contacts made per day	Plausible range
Mortality rate, μ	0.01/365.	Deaths per day per capita	To give an average life expectancy of 100 years, in absence of TB.
TB mortality rate, μ_T	0.07/365.	Deaths per day per capita	PHE surveillance database, *via* Mears *et al.*[124]
Recovery rate, ρ	Log-Uniform(5×10^{-4}, 5×10^{-1})	Cases per day per capita	Plausible range. Implies an average duration of infectiousness of between 2 and 2000 days (in the absence of death).
Protection against reinfection by being latently infected, σ	0.79		Andrews et al[125]
Rate recently (re)infected move to latent, l	1/365.	Cases per day per capita	Calculated so that people remain for an average of 1 year in L_f
Rate that latent progress to disease, p_s	Log-Uniform(2×10^{-6}, min(p_f, 2×10^{-3}))	Cases per day per capita	Plausible range*, with $p_s < p_f$.
Rate that recently (re)infected progress to disease, p_f	Log-Uniform(6×10^{-5}, 6×10^{-2})	Cases per day per capita	Plausible range*, geometric mean 1.9 cases per day per 1000 people
Population size, N	8640000	People	Office for National Statistics[126]

Clustering coefficient, ϕ	Uniform(0,1)	The proportion of sets of pairs connected by a central contact that form triples	Uninformative prior

Pairwise equation system: We wish to understand whether cases who have developed TB following recent (re)infection (i.e. the I_f compartment) generate fewer infections than those with reactivation disease due to a saturation effect amongst their contacts. That is, in our pairwise system, do those who have disease following a recent (re)infection have fewer uninfected contacts, and so cause fewer infections than do those with reactivation disease?

The pairwise system is a deterministic set of differential equations for which each equation describes the number of each type of pair. Terms of the form [AB] represent a pair of people, one of whom is in the A compartment, and the other in the B compartment. Similarly, a term of the form [ABC] represents a triple in which A is paired with B and B with C. The transitions shown in Figure 11 can be found in the equations by locating the corresponding parameter. In order to simplify the equations, the number of like pairs (e.g. S-S pairs) are counted twice in each direction, whereas unlike pairs are counted once in each direction. This results in the appearance of a factor of 2 in some of the equations. In this system, each infectious-susceptible pair and each infectious-latent pair results in transmission at a rate τ or $\sigma\tau$ respectively. There is also a random contact term, highlighted in these equations in red, which acts on the susceptible and latent compartments at a rate proportional to the total number in the infectious compartment. These terms can be thought of as representing transmission between close contacts (τ) and casual contacts (β) respectively. The number of close contacts is fixed throughout a model run at 5, and each person in the model has the same number of close contacts. The equations for this system are as follows:

0. $\frac{d[SS]}{dt} = 2\left(\mu([SL_s] + [SL_f]) + \mu_T([SI_s] + [SI_f]) - \tau([SSI_s] + [SSI_f]) - \frac{\beta[SS]([I_s]+[I_f])}{N}\right)$
1. $\frac{d[SL_s]}{dt} = \mu([L_sL_s] + [L_sL_f]) + \mu_T([L_sI_s] + [L_sI_f]) - \tau\left([L_sSI_s] + [L_sSI_f] + \sigma([SL_sI_s] + [SL_sI_f])\right) + l[SL_f] + \rho([SI_s] + [SI_f]) - (p_s + \mu)[SL_s] - \frac{(1+\sigma)\beta[SL_s]([I_s]+[I_f])}{N}$
2. $\frac{d[SL_f]}{dt} = \mu([L_fL_s] + [L_fL_f]) + \mu_T([L_fI_s] + [L_fI_f]) + \tau\left([SSI_s] + [SSI_f] + \sigma([SL_sI_s] + [SL_sI_f]) - ([L_fSI_s] + [L_fSI_f])\right) - (l + p_f + \mu)[SL_f] + \frac{\beta([I_s]+[I_f])([SS]+\sigma[SL_s]-[SL_f])}{N}$

3. $\frac{d[SI_s]}{dt} = \mu([I_sL_s] + [I_sL_f]) + \mu_T([I_sI_s] + [I_sI_f]) - \tau([SI_s] + [I_sSI_s] + [I_sSI_f]) + p_s[SL_s] - (\rho + \mu_T)[SI_s] - \frac{\beta[SI_s]([I_s]+[I_f])}{N}$
4. $\frac{d[SI_f]}{dt} = \mu([I_fL_s] + [I_fL_f]) + \mu_T([I_fI_s] + [I_fI_f]) - \tau([SI_f] + [I_fSI_f] + [I_sSI_f]) + p_f[SL_f] - (\rho + \mu_T)[SI_f] - \frac{\beta[SI_f]([I_s]+[I_f])}{N}$
5. $\frac{d[L_sL_s]}{dt} = 2\left(l[L_sL_f] + \rho([L_sI_s] + [L_sI_f]) - \sigma\tau([L_sL_sI_s] + [L_sL_sI_f]) - (p_s + \mu)[L_sL_s] - \frac{\sigma\beta[L_sL_s]([I_s]+[I_f])}{N}\right)$
6. $\frac{d[L_sL_f]}{dt} = l[L_fL_f] + \rho([L_fI_s] + [L_fI_f]) + \tau\left([L_sSI_s] + [L_sSI_f] + \sigma\left([L_sL_sI_s] + [L_sL_sI_f] - ([L_fL_sI_s] + [L_fL_sI_f])\right)\right) - (l + p_f + 2\mu + p_s)[L_fL_s] + \frac{\beta([I_s]+[I_f])([SL_s]+\sigma[L_sL_s]-\sigma[L_sL_f])}{N}$
7. $\frac{d[L_sI_s]}{dt} = l[I_sL_f] + \rho([I_sI_s] + [I_sI_f]) - \sigma\tau([L_sI_s] + [I_sL_sI_s] + [I_sL_sI_f]) + p_s[L_sL_s] - (\rho + \mu_T + p_s + \mu)[I_sL_s] - \frac{\sigma\beta[L_sI_s]([I_s]+[I_f])}{N}$
8. $\frac{d[L_sI_f]}{dt} = l[I_fL_f] + \rho([I_sI_f] + [I_fI_f]) - \sigma\tau([L_sI_f] + [I_fL_sI_f] + [I_sL_sI_f]) + p_f[L_sL_f] - (p_s + \mu + \rho + \mu_T)[L_sI_f] - \frac{\sigma\beta[L_sI_f]([I_s]+[I_f])}{N}$
9. $\frac{d[L_fL_f]}{dt} = 2\left(\tau\left([L_fSI_s] + [L_fSI_f] + \sigma([L_fL_sI_s] + [L_fL_sI_f])\right) - (l + p_f + \mu)[L_fL_f] + \frac{\beta([L_fS]+\sigma[L_fL_s])([I_s]+[I_f])}{N}\right)$
10. $\frac{d[L_fI_s]}{dt} = \tau\left([SI_s] + [I_sSI_s] + [I_sSI_f] + \sigma([L_sI_s] + [I_sL_sI_s] + [I_sL_sI_f])\right) + p_s[L_fL_s] - (\rho + \mu_T + l + p_f + \mu)[L_fI_s] + \frac{\beta([I_sS]+\sigma[I_sL_s])([I_s]+[I_f])}{N}$
11. $\frac{d[L_fI_f]}{dt} = \tau\left([SI_f] + [I_fSI_f] + [I_sSI_f] + \sigma([L_sI_f] + [I_fL_sI_f] + [I_sL_sI_f])\right) + p_f[L_fL_f] - (l + p_f + \mu + \rho + \mu_T)[L_fI_f] + \frac{\beta([I_fS]+\sigma[I_fL_s])([I_s]+[I_f])}{N}$
12. $\frac{d[I_sI_s]}{dt} = 2(p_s[I_sL_s] - (\rho + \mu_T)[I_sI_s])$
13. $\frac{d[I_sI_f]}{dt} = p_s[I_fL_s] + p_f[I_sL_f] - 2(\rho + \mu_T)[I_sI_f]$
14. $\frac{d[I_fI_f]}{dt} = 2(p_f[I_fL_f] - (\rho + \mu_T)[I_fI_f])$

We can calculate exactly the number of singles from these terms:

15. $[A] = \frac{1}{n}\sum_B [AB]$

As transmission can occur within a pair, or from outside the pair, triple terms appear in the above equations. This is approximated using the closure equation:

16. $[ABC] = \frac{[AB][BC](n-1)}{n[B]}\left[(1 - \phi) + \frac{\phi N[AC]}{n[A][C]}\right]$

Where n is the average number of contacts, and ϕ is the clustering coefficient (the proportion of triples that form triangles, or in other words, for what proportion of triples in which A is a contact of B and B is a contact of C, is A also a contact of C?). When $\phi = 0$, the term in the square brackets is 1 and none of the triples form triangles, i.e. if A is connected to B, and B to C, then is never connected to C. Similarly, if $\phi \ll 1$, a negligible proportion of the triangles will form triples. If $\phi = 1$, then every triangle forms a triple, i.e. if A is connected to B and B to C, then A is always connected to C

In order to separately quantify the total number of transmissions which take place from the $I_s$and I_f compartments, we use the following eight equations. In these equations, T_s terms represent transmission from those who developed disease following reactivation of latent infection, T_f terms represent transmission from those who developed disease following recent (re)infection, the superscript p represents pairwise transmission, r represents transmission from casual contacts, and the even numbered equations each represent transmission to previously infected contacts:

17. $\frac{dT_{s,p}}{dt} = \tau\left([SI_s] + [I_sSI_f] + [I_sSI_s]\right)$
18. $\frac{dT_{s,p,\text{ reinfection}}}{dt} = \sigma\tau\left([L_sI_s] + [I_sL_sI_f] + [I_sL_sI_s]\right)$
19. $\frac{dT_{s,r}}{dt} = \frac{\beta[S][I_s]}{N}$
20. $\frac{dT_{s,r,\text{ reinfection}}}{dt} = \frac{\beta\sigma[L_s][I_s]}{N}$
21. $\frac{dT_{f,p}}{dt} = \tau\left([SI_f] + [I_sSI_f] + [I_fSI_f]\right)$
22. $\frac{dT_{f,p,\text{reinfection}}}{dt} = \sigma\tau\left([L_sI_f] + [I_sL_sI_f] + [I_fL_sI_f]\right)$
23. $\frac{dT_{f,r}}{dt} = \frac{\beta[S][I_f]}{N}$
24. $\frac{dT_{f,r,\text{ reinfection}}}{dt} = \frac{\beta\sigma[L_s][I_f]}{N}$

We also estimate the incidence using the following equations:

25. $\frac{d\text{Inc}_s}{dt} = p_sL_s$
26. $\frac{d\text{Inc}_f}{dt} = p_fL_f$

Model fitting: We aimed to reproduce the incidence and prevalence of pulmonary disease in London, and to then identify values for the ratio of infections generated by cases with reactivation disease to infections generated by cases with disease following recent (re)infection that are

consistent with the observed data. We fitted to an annual incidence of pulmonary disease of 12.3/100000[8] and a prevalence of pulmonary disease of 13.3/100000 persons[114]. This latter figure is over ten years old (from 2003), and is an approximation, as no prevalence survey was undertaken, so it was calculated using a point prevalence of cases who were or should have been on treatment. It is also further approximated, as the data were not stratified by site of disease; to arrive at the value of 13.3/100,000 we assumed 49% of cases had pulmonary disease, which is the proportion of incident cases in the most recent year with pulmonary disease.

We varied: the clustering coefficient, ϕ; the recovery rate, ρ; the progression rate following recent (re)infection, p_f; the reactivation rate, p_s; the pairwise contact parameter, τ, and the casual contact parameter, β. We fit the model using the sampling-importance-resampling algorithm, sometimes known as Bayesian melding. This consists of the following steps:

1. Sample each parameter from its prior distribution. In our case we take M = 1500000 parameter sets from the prior distribution.
2. Obtain the model output for each of these M parameter sets. In our case, the model output is the equilibrium prevalence and incidence of disease. This was obtained by running the model for 1000 simulated years, using a time-step of 1 day..
3. Estimate the likelihood of each of these model outputs with respect to the data point we are fitting to. In our case this means estimating the likelihood of the model predicted equilibrium prevalence or incidence, to the prevalence or incidence (respectively) in London in 2016.
4. Compute the weights of each input parameter set which are proportion to the overall likelihood, calculated by multiplying the estimated likelihoods with respect to each data point. As initial attempts to use the combined likelihood of the incidence and prevalence yielded few posterior parameter sets, possibly due to inaccuracies in the estimate of

prevalence, we separately use either the likelihood with respect to prevalence or with respect to incidence.

5. Generate the posterior distribution of the model output, by resampling from the original output sets with weights proportional to the likelihood. In our case we resample m=12000 sets.

To estimate the likelihood of the prevalence data for a given model run we used the binomial distribution, with $k = \frac{13.3N}{100000}$ (i.e. the number of prevalent cases in the data), $n = N$ (the population size), and p equal to the model predicted prevalence. We calculate the likelihood of the incidence data using the Poisson distribution, with $k = \frac{12.3N}{100000}$ (i.e. the number of incident cases in one year in the data) and λ equal to the model predicted number of incident cases in one year.

For model fitting, we calculate the incidence, prevalence and prevalence of infection, using the following equations, where superscripts represent time-steps and T is the final time-step

25. $\text{Prevalence} = \frac{I_s^{(T)} + I_f^{(T)}}{\text{N}}$
26. $\text{Incidence} = \frac{\text{Inc}_s^{(T)} + \text{Inc}_f^{(T)} - \text{Inc}_s^{(T-1)} - \text{Inc}_f^{(T-1)}}{\text{N}\Delta t}$
27. $\text{Prevalence of infection} = \frac{L_s^{(T)} + L_f^{(T)}}{\text{N}}$

If the region of parameter space which maximises the likelihood is sharply 'ridge-like', it may be difficult for a Bayesian melding algorithm to effectively explore the optimum region of parameter space. For this reason, alongside the fact that the Bayesian melding model fit did not prove satisfactory, we also fitted the model using the Metropolis-Hastings Markov Chain - Monte Carlo (MCMC) algorithm. We used uniform prior distributions for each of the six parameters and calibrated to: the incidence of disease, the prevalence of disease, the prevalence of infection (5%, taken from a US study as no recent estimate exists in the UK), and the proportion of all pairs containing at least one infectious person that contain two infectious people (2.6%, chapter three).

We estimated the likelihood of each model output using the Poisson distribution for the incidence, and the binomial distribution for the other three outputs. We adapted the size and standard deviation of the proposal distribution during the Metropolis-Hastings algorithm and ran the algorithm for 5000 iterations. We selected the initial parameter sets sing Latin Hypercube Sampling in the pyDOE package in Python, implemented the MCMC algorithm in Python, and analysed the results using the fitR and coda packages in R.

Main analysis: We first estimate the distribution of each of the six free parameters (the clustering coefficient, ϕ; the recovery rate, ρ; the progression rate following recent (re)infection, p_f; the reactivation rate, p_s; the pairwise contact parameter, τ, and the casual contact parameter, β) when fitted to prevalence and to incidence, and look at the corresponding estimated incidence or prevalence distributions for plausibility. We use the Kolmogorov-Smirnov test to test for a difference between the parameter distributions when fitted to prevalence compared to when fitted to incidence. We then estimate the ratio of infections generated by those with reactivation disease to those with disease following recent (re)infection, or $\frac{T_s^m - T_s^{m-1}}{T_f^m - T_f^{m-1}}$ (where m represents the final time-step of the model run). We also estimate this per infectious case, or $\left(\frac{T_s^m - T_s^{m-1}}{T_f^m - T_f^{m-1}}\right)\left(\frac{I_f^m + I_f^{m-1}}{I_s^m + I_s^{m-1}}\right)$. Finally, we calculate the correlation of each free parameter with the per case ratio to understand which parameter most strongly determines the outcome.

Additional analyses: We undertook a number of additional analyses to explore the results:

- Stratify the results based on values of β into nine equal width bands, in order to understand the impact of β on the model predicted value of τ and on the ratio.
- Include other things into the model fit: An estimate of infection prevalence (5%, taken from a US study as no recent estimate exists in the UK[127]); and the proportion of all pairs containing at least one infectious person that contain two infectious people (2.6%, taken from chapter 2). Both are calculated using the binomial distribution.

- Stratify the results by values of the per case ratio, to understand how the distributions for parameters change with the output of interest. We stratify the ratio into three bands: <1 (slightly more transmission per case from those with disease following recent (re)infection), between 1 and 1.05 (slightly more transmission per case from those with reactivation disease), and >1.05
- Fit to incidence and prevalence just among the UK-born, 4.5/100000/year and 4.6/100000 respectively[8,114], to account in a simple way for the lack of inclusion of immigration.
- Fit to the incidence from 2003, when the prevalence estimate was made. This value is 20.4/100000[28], assuming the proportion of cases with pulmonary TB has not changed significantly. This analysis was not repeated for the UK born population because, while the overall incidence has changed a lot since 2003, the incidence amongst UK born has been fairly constant, with nearly all of the decline in incidence happening amongst the non-UK born[22]

Results

Parameters

Figure 12 and Figure 13 show the predicted prevalence and incidence when the model is fitted to prevalence alone. Figure 14 and Figure 15 show the predicted prevalence and incidence when the model is fitted to incidence alone. When we fit to prevalence, the predicted incidence is too high (on average), whereas when we fit to incidence the prevalence is too low. When fitting to prevalence of disease, the prevalence of infection (latent and recent) is much too high (Figure 16), but when we fit to incidence, the prevalence of infection is more reasonable, though in this case is too low (Figure 17).

Posterior distributions of each parameter when fitting to prevalence are shown in Figure 18, and when fitting to incidence are shown in Figure 19. In both cases τ and ϕ have uniform posterior distributions which match their prior distributions. Whilst, according to the Kolmogorov-Smirnov test, the parameter distributions are different for each parameter when fitted to incidence as compared to prevalence (Table 4), the most stark difference occurs for β. While when we fit to prevalence the posterior distribution of β is uniform, matching its prior distribution, when we fit to incidence the posterior is positively skewed, and hence the median value is lower (Table 5).

When fitting to prevalence the annual risk of infection (ARI) is 5.0% (95% CI: 0.51%, 16%), whereas when fitting to incidence it is 0.027% (95% CI: 0.0020%, 0.12%). Whilst the ARI is not well known for the UK[128] it is likely to be between these two estimates.

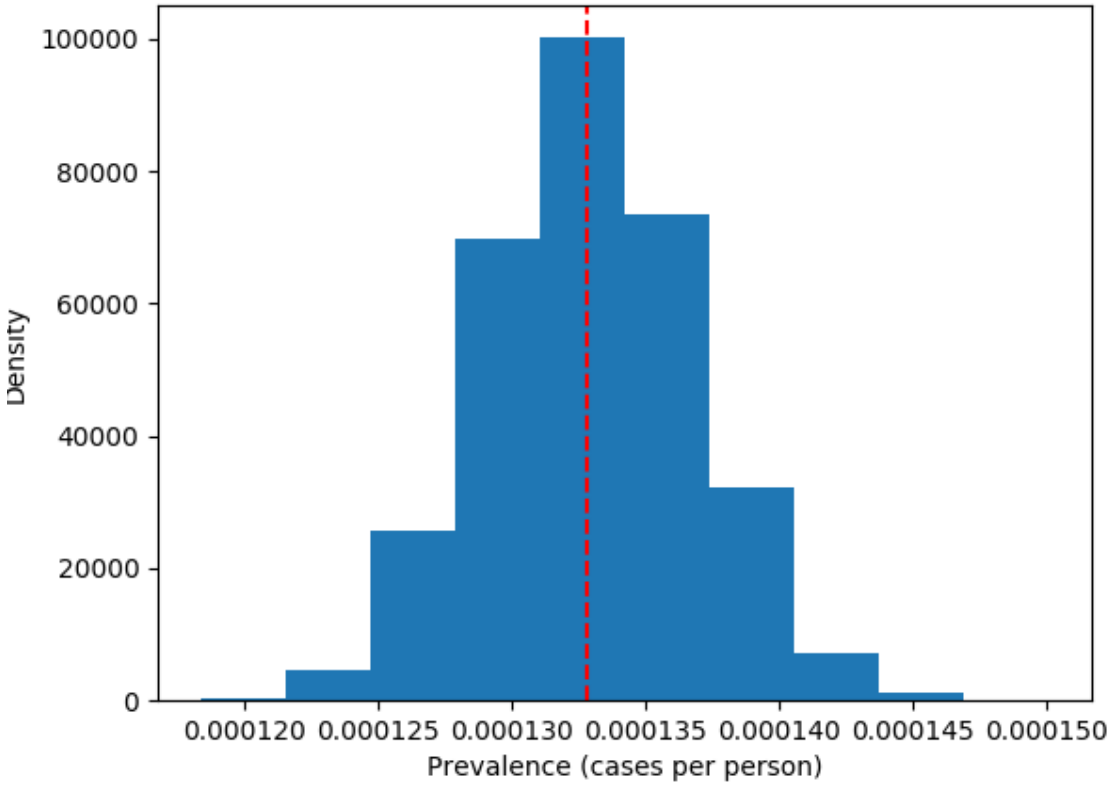


Figure 12: Distribution of prevalence when fitted to prevalence. The vertical dashed line is the target value. There are no values beyond the scale shown here.

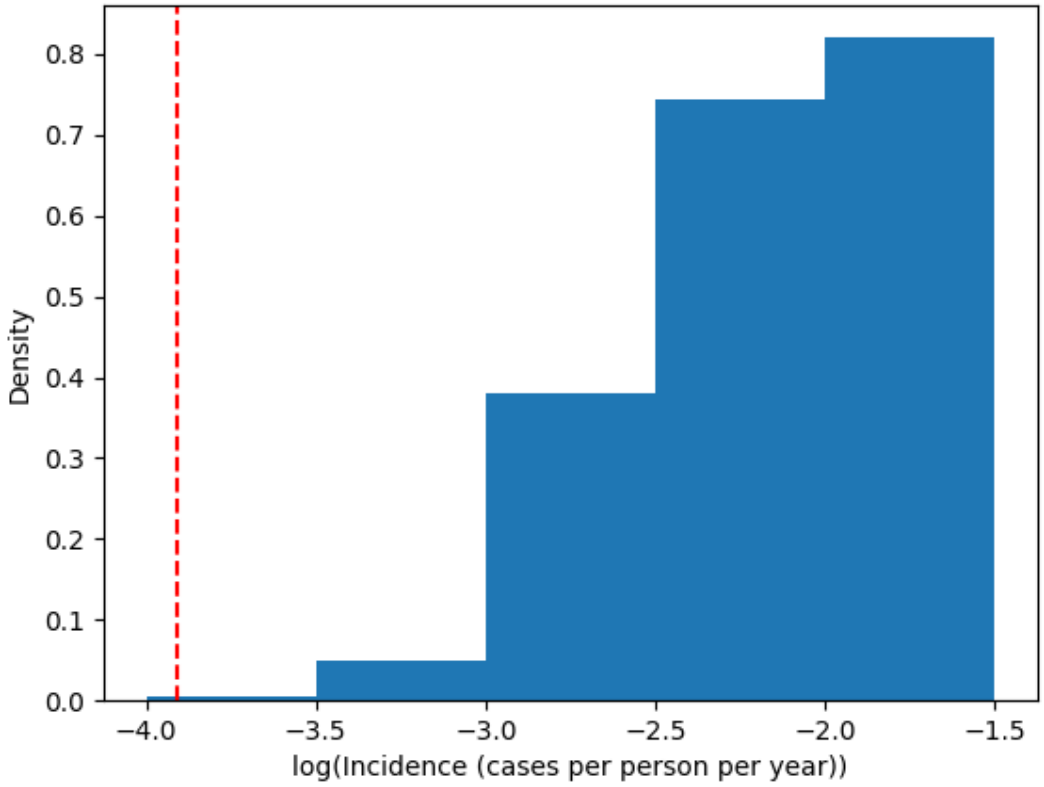


Figure 13: Distribution of incidence when fitted to prevalence. The vertical dashed line is the target value. There are no values beyond the scale shown here.

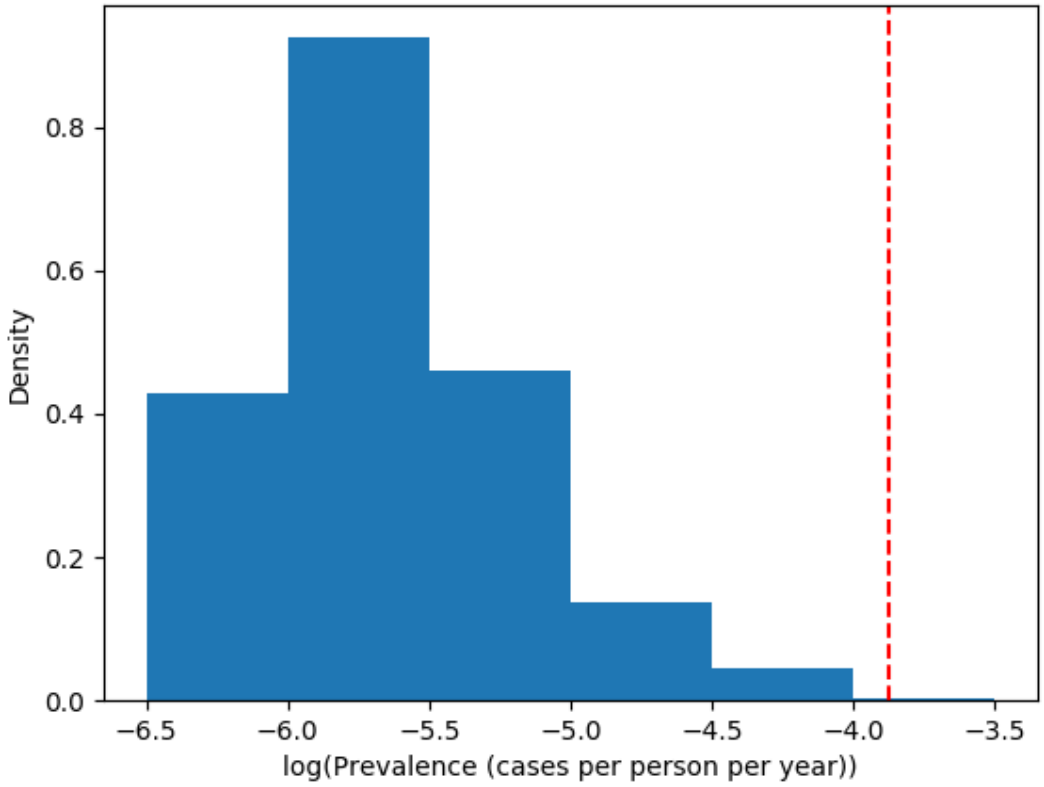


Figure 14: Distribution of prevalence when fitted to incidence. The vertical dashed line is the target value. There are no values beyond the scale shown here.

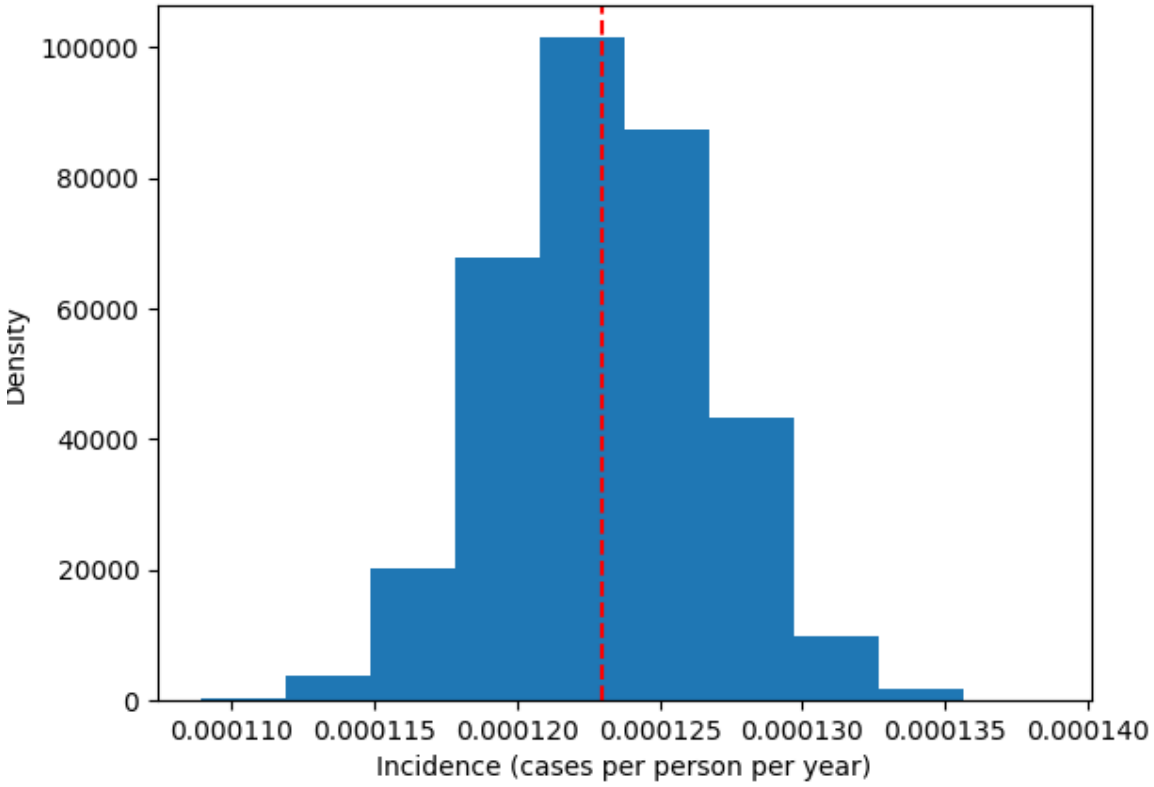


Figure 15: Distribution of incidence when fitted to incidence. The vertical dashed line is the target value. There are no values beyond the scale shown here.

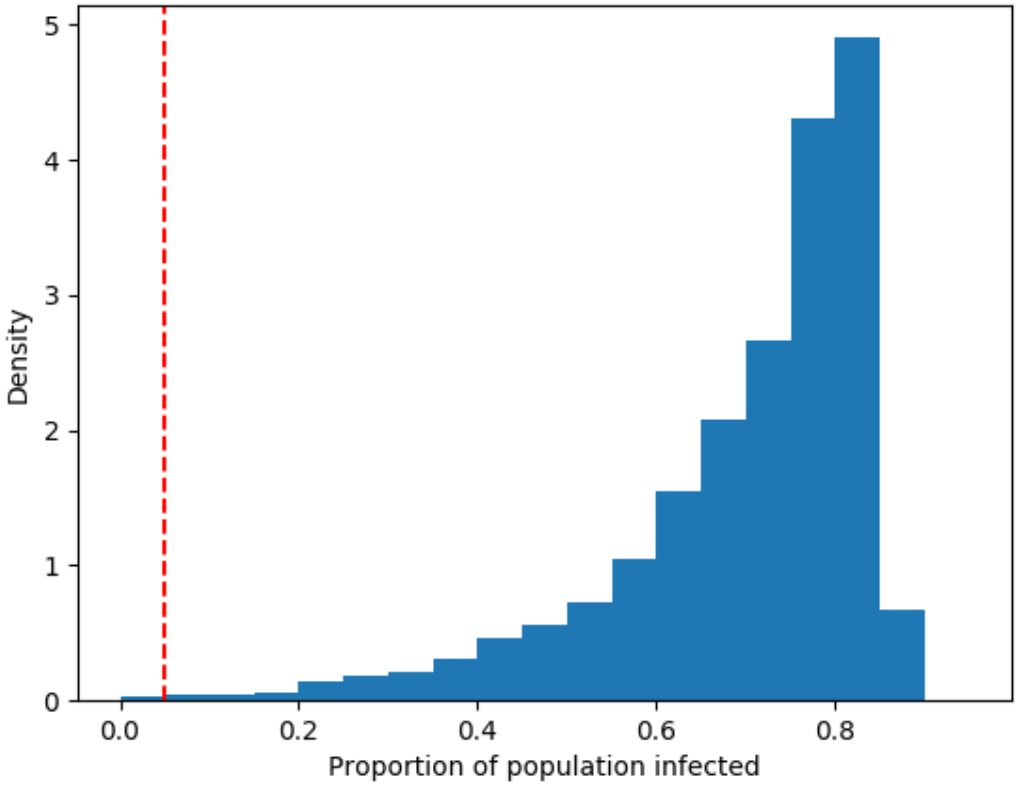


Figure 16: Prevalence of M. Tb infection when model is fitted to prevalence of disease. The red vertical line indicates 5%

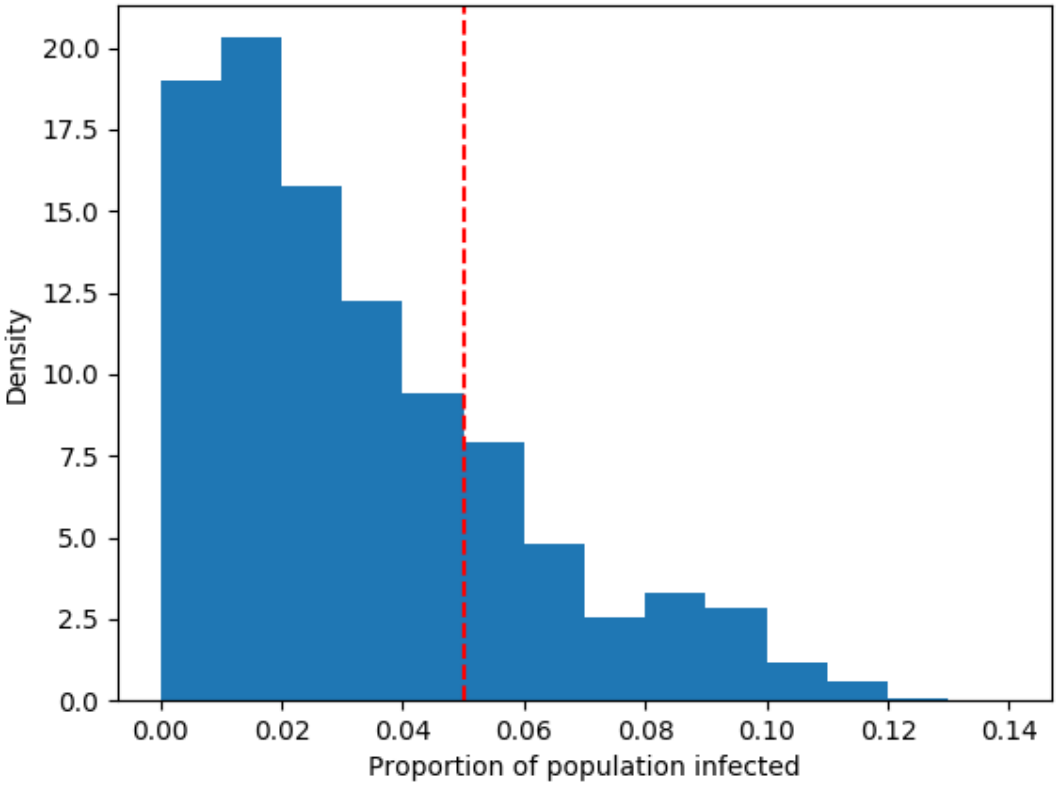


Figure 17: Prevalence of M. Tb infection when model is fitted to incidence of disease. The red vertical line indicates 5%

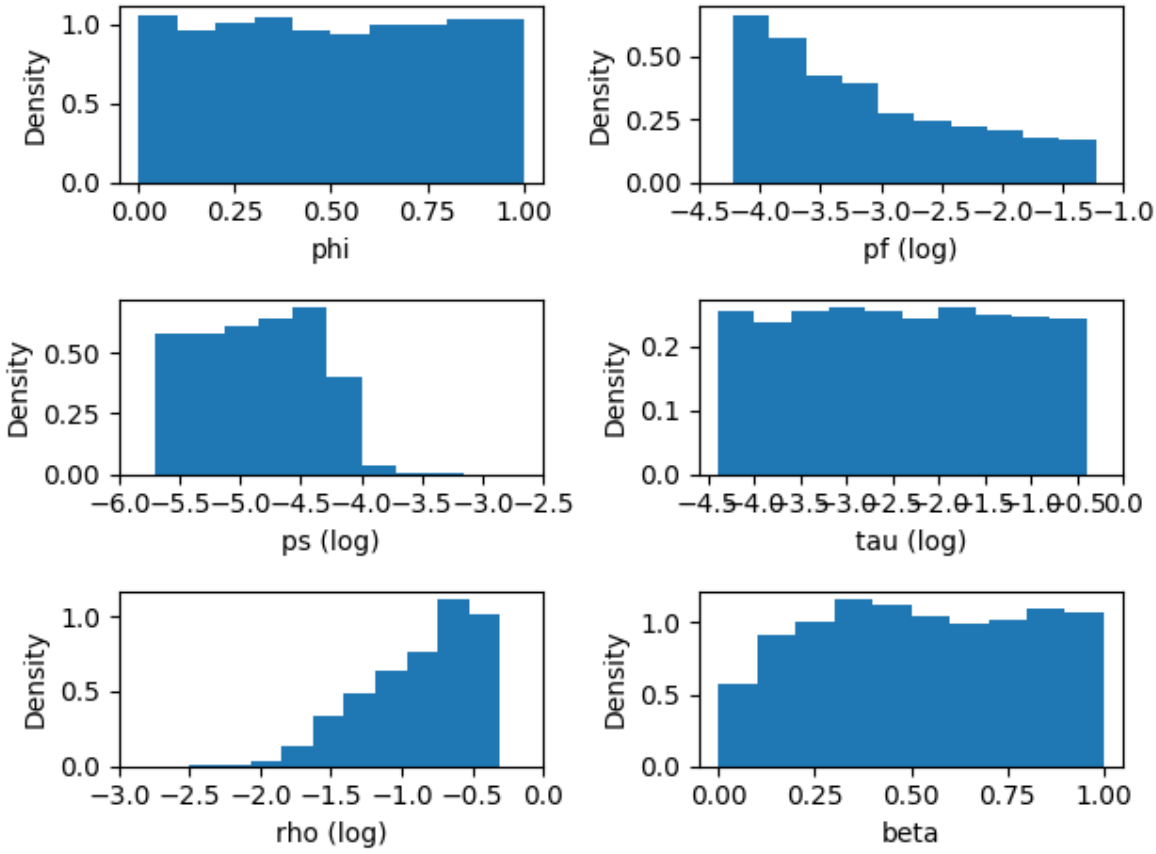


Figure 18: Posterior parameter distributions when fitted to prevalence

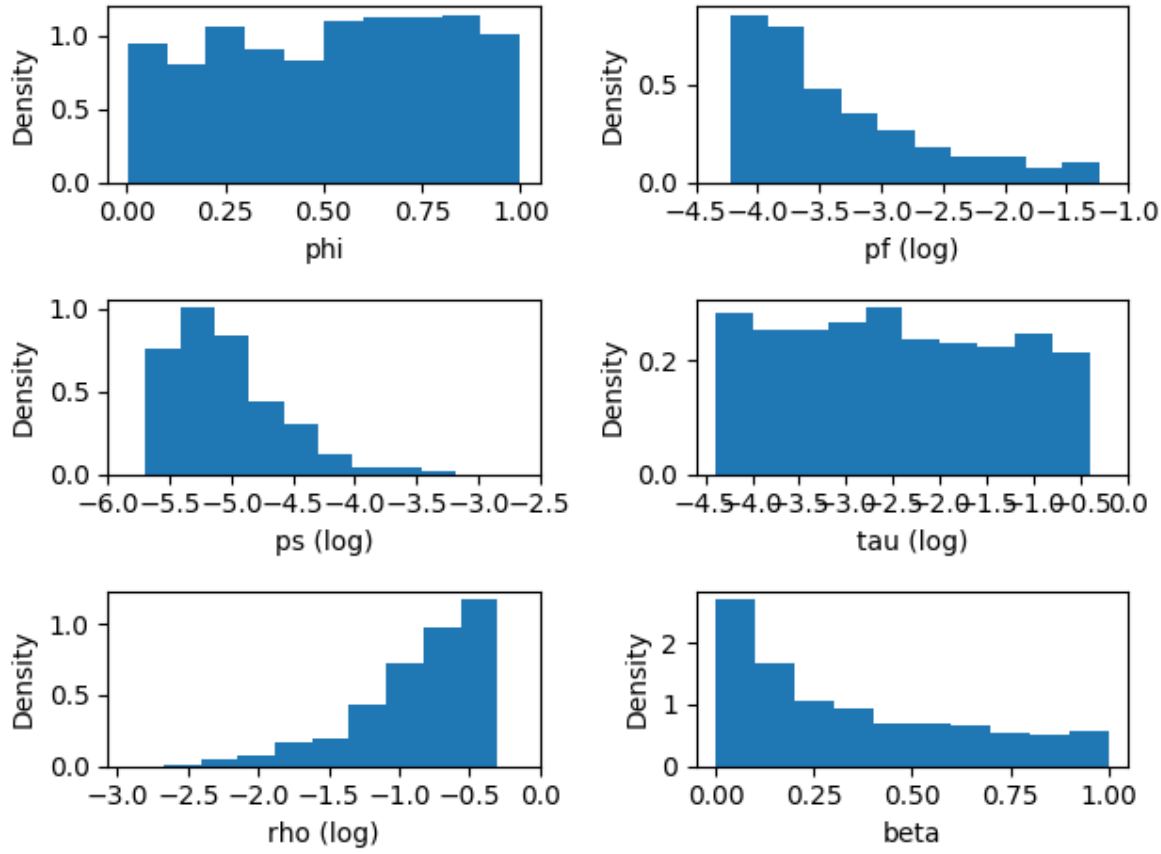


Figure 19: Posterior parameter distributions when fitted to incidence. This figure can be contrasted with figure 18, which showed the posterior distributions when the model was fitted to disease prevalence. In particular, there are large differences in β, p_f and p_s.

Table 4: Probability of no difference between the parameter distributions, according to the Kolmogorov-Smirnov test.

parameter	p-value that there is no difference in distributions
N	8.83×10^{-15}
p_f	7.06×10^{-108}
p_s	7.73×10^{-286}
τ	1.76×10^{-11}
ρ	4.95×10^{-10}
β	0
Ratio of transmission from cases with reactivation disease to transmission from cases with disease following recent infection	4.10×10^{-106}
Ratio of transmission from cases with reactivation disease to transmission from cases with disease following recent infection, per case	6.14×10^{-61}

Table 5: Median and 95% confidence intervals for the posterior distribution of each parameter, when fitted to incidence or prevalence.

variable	Fitted to prevalence			Fitted to incidence		
	median	Lower 95% CI	Upper 95% CI	median	Lower 95% CI	Upper 95% CI
ϕ	0.499	0.0248	0.979	0.540	0.0239	0.975
p_f	0.000489	6.54×10^{-5}	0.0421	0.00025	6.46×10^{-5}	0.0335
p_s	1.42×10^{-5}	2.22×10^{-6}	8.18×10^{-5}	7.54×10^{-6}	2.14×10^{-6}	0.000114
τ	0.00383	4.99×10^{-5}	0.301	0.00274	5.23×10^{-5}	0.287
ρ	0.166	0.0196	0.470	0.175	0.0107	0.478
β	0.518	0.0534	0.973	0.257	0.00658	0.953

Transmission ratios

When we fit to either prevalence or incidence (Figure 20 and Figure 21 respectively, Table 6), we see that the number of infections per case is greater for those with reactivation disease than it is for those with disease following recent (re)infection in almost all cases. This is true whether we include reinfection when estimating the ratio or not (i.e. the $T_{i,j,\text{reinfection}}$ terms), and whether we include transmission to casual contacts when estimating the ratio or not (i.e. the $T_{i,r}$ and $T_{i,r,\text{reinfection}}$ terms. The distribution of the ratio of the total number of infections varies whether we fit to incidence or prevalence, with the average value of the ratio higher when fitting to the incidence. This is likely to be, at least in part, because the prevalence of infection is lower when we fit to incidence, meaning that the proportion of contacts infected will be high irrespective of whether the index case had disease following recent infection. However, when we look at the per case ratio, the results are similar whether fitting to incidence or prevalence. Both the lower 95% confidence interval and the median of the ratio is 1.00 to 3 significant figures, when fitted to incidence or when fitted to prevalence. This implies that while those with reactivation disease do generate more infections per case, it is, for most of the posterior, an extremely small effect. The value of the per case ratio is most strongly correlated with τ whether we fit to incidence or prevalence, whether we include or exclude reinfection and whether we include or exclude casual contacts in the estimate of the ratio (Table 7).

However, if we exclude reinfection and fit to incidence, the value of the ratio is not particularly well correlated with any of the parameters.

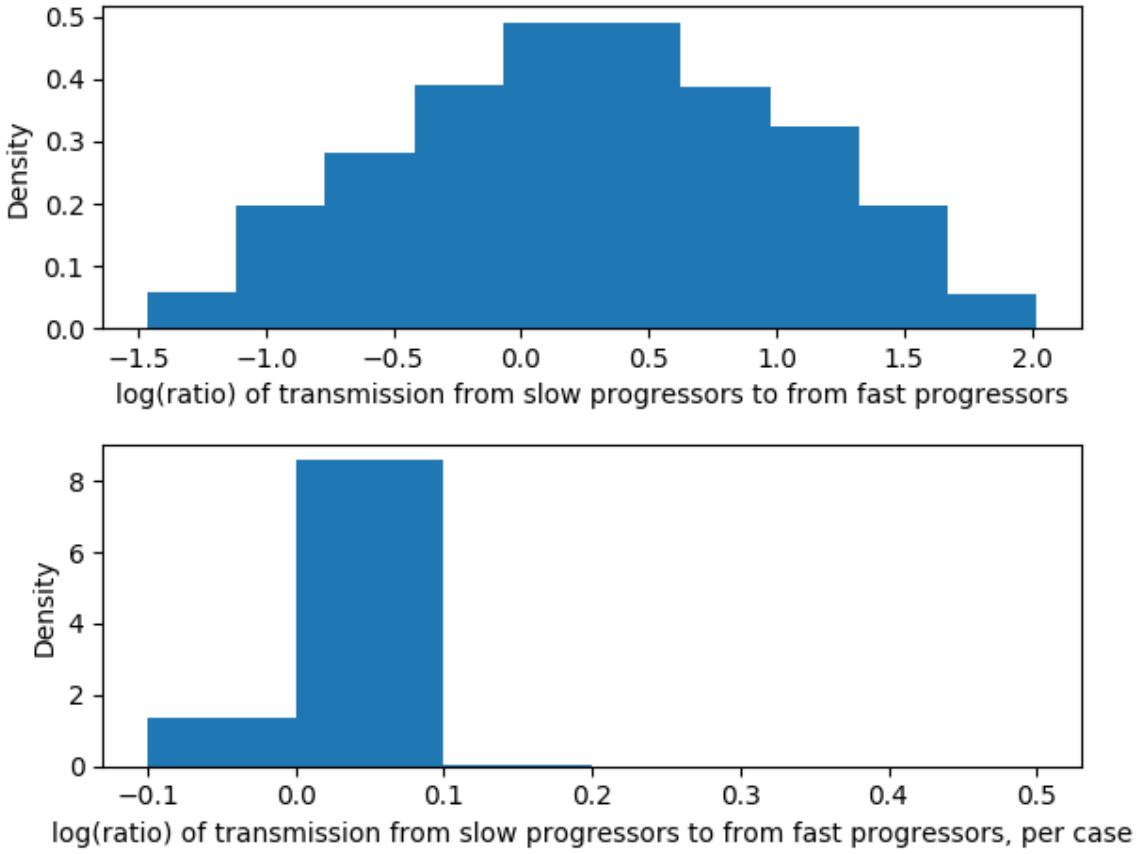


Figure 20: Posterior distribution when fitted to prevalence of the log ratio of the number of infections from those with reactivation disease to the number from those with disease following recent (re)infection (top panel), and the log ratio of the number of infections from those with reactivation disease per case to the number from those with disease following recent (re)infection per case(bottom panel). When these ratios are below zero, those with disease following recent (re)infection are generating more infections than those with reactivation disease

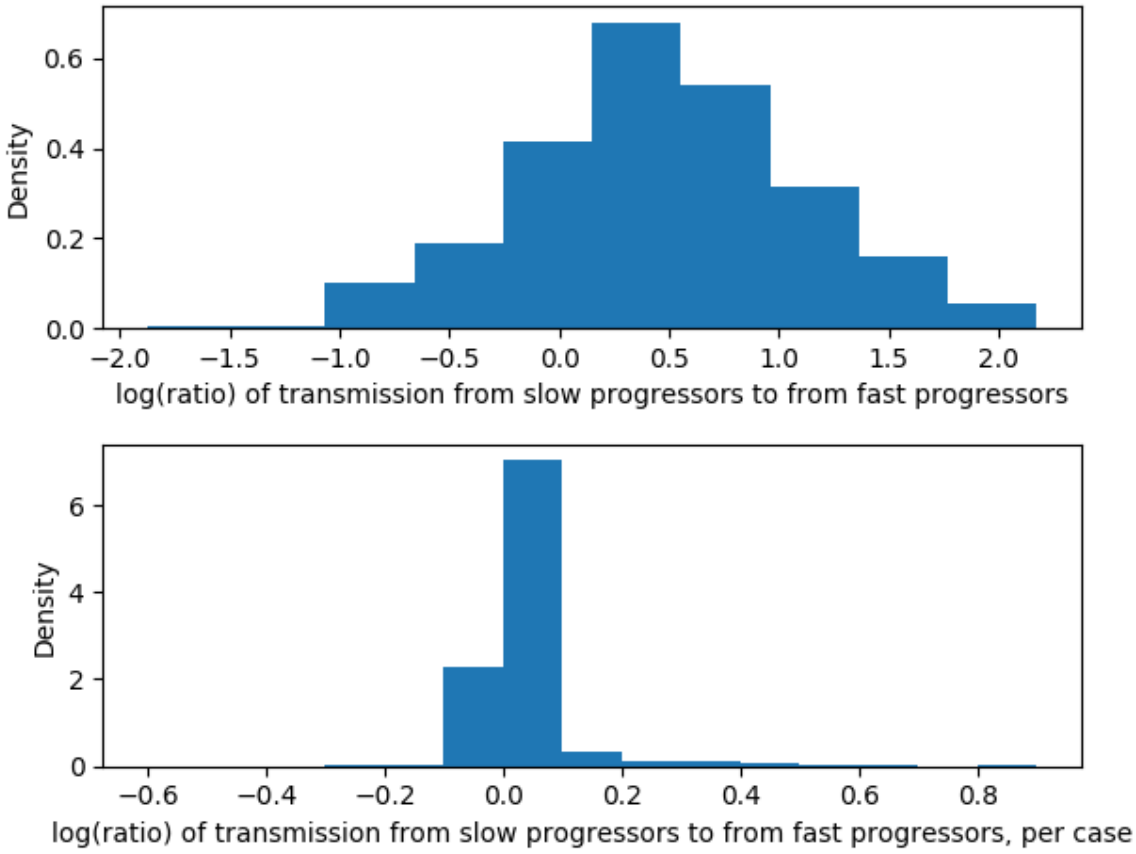


Figure 21: Posterior distribution when fitted to the incidence of the log ratio of the number of infections from those with reactivation disease to the number from those with disease following recent (re)infection (top panel), and the log ratio of the number of infections from those with reactivation disease per case to the number from those with disease following recent (re)infection per case(bottom panel). When these ratios are below zero, those with disease following recent (re)infection are generating more infections than those with reactivation disease

Table 6: Summary of results of the ratio of the total number of infections generated by those with reactivation disease to the number generated by those with disease following recent (re)infection, and of the ratio of the number of infections from those with reactivation disease per case to the number from those with disease following recent (re)infection per case. Posterior results are shown for when fitted to incidence and when fitted to prevalence. Note that while the figures used a log scale on the x-axis, to make the results more easily visible, this table uses the actual values

variable	Fitted to prevalence			Fitted to incidence		
	Median	Lower 95% CI	Upper 95% CI	Median	Lower 95% CI	Upper 95% CI
ratio	1.94	0.0841	42.0	3.04	0.135	56.4
ratio, per case	1.00	1.00	1.11	1.00	1.00	1.80
ratio, per case, excluding reinfection	1.00	1.00	1.23	1.00	1.00	2.85
ratio, per case, only	1.00	1.00	1.18	1.00	1.00	2.21

pairwise transmission ratio, per case, only pairwise transmission and no reinfection	1.00	1.00	1.49	1.00	1.00	3.93

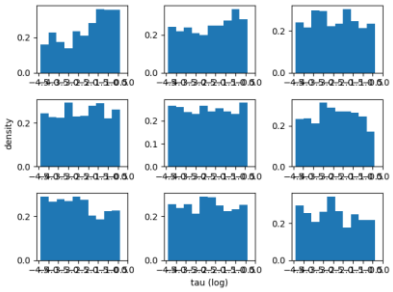
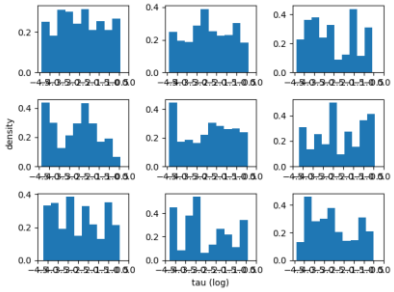
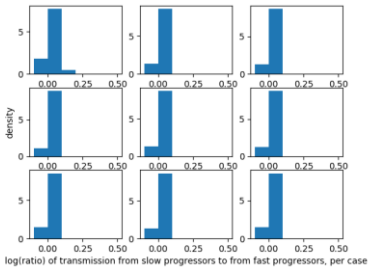
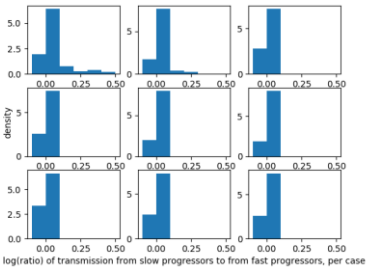
Table 7: Correlation of each parameter with the per case ratio, when fitted to either the prevalence or the incidence.

	ϕ	p_f	p_s	τ	ρ	β
prevalence	0.02	0	0	0.3	0	0.06
incidence	0.02	0	0	0.14	0	0.03
prevalence, no reinfection	0.01	0	0	0.16	0	0.03
incidence, no reinfection	0.02	0	0	0.05	0	0.03
prevalence, only pairwise transmission	0.04	0.01	0	0.44	0	0.05
incidence, only pairwise transmission	0.03	0	0	0.18	0	0.04
prevalence, no reinfection and only pairwise transmission	0.03	0	0	0.34	0	0.02
incidence, no reinfection and only pairwise transmission	0.02	0	0	0.05	0.01	0.03

Additional analyses

Stratified by β: Stratifying the results by values of β makes little difference to the estimates of τ (Table 8, 1st row), but makes a slight difference to the distribution of the estimates of the ratio (Table 8, 2nd row). However, the median value of the ratio stays roughly constant, even as the distribution changes as β increases.

Table 8: Results stratified by values ofβ. β was divided into nine equal width bands; in each figure, starting in the top left and reading horizontally first, in the nth image only values of beta between (n-1)/9 and n/9 were used.

Value	Fitted to prevalence	Fitted to incidence
τ		
Ratio of transmission from those with reactivation disease to those with disease following recent (re)infection.		

Including both prevalence and incidence of disease into the model fit: Including both the prevalence and the incidence in the model fit means that only two parameter sets are chosen, insufficient to generate a reliable spread of results. The two parameter sets that are chosen are shown in Table 9; one of the two parameter sets is much less likely than the other two. The two chosen parameter sets show consistency in the value of the recovery rate ρ, giving values around 0.0023 cases per day, and both choose relative small values of the casual contact rate β and above average values of the pairwise contact parameter τ, which has an expected value of 0.004. The most likely parameter set gives a value of the per case ratio of 1.05, suggesting a small saturation effect for this parameter set, and the other suggests no saturation effect.

Including infection prevalence into the model fit: Including the infection prevalence in the model fit alongside either prevalence or incidence means that only one or nine parameter sets (respectively) are chosen, insufficient to generate reliable results. In the former case (Table 10), the value of incidence of disease chosen is a factor of ten too great, and the parameter set chosen has a strong saturation effect, giving a per case ratio of 1.81. In the case of fitting to incidence of disease and prevalence of infection (Table 11), one parameter set dominates. This parameter set gives a per case ratio of 1.01, suggesting little or no saturation effect.

Table 9: Parameter sets chosen when fitting to both incidence and prevalence together, with associated values of incidence, prevalence and the per case ratio. The count column refers to the total number of times that parameter set was chosen, and is proportional to the likelihood.

Count	Prevalence	Incidence	ϕ	p_f	p_s	β	ρ	τ	Ratio per case
6302	0.000146	0.000132	0.193	7.38E-05	5.31E-06	0.00843	0.00229	0.00421	1.05
5402	0.000124	0.000111	0.0892	0.000274	3.18E-06	0.000502	0.00227	0.0107	1.00

Table 10: Parameter sets chosen when fitting to both prevalence of disease and of infection together, with associated values of incidence, prevalence and the per case ratio. The count column refers to the total number of times that parameter set was chosen, and is proportional to the likelihood.

Count	Prevalence	Incidence	ϕ	p_f	p_s	β	ρ	τ	Ratio per case
11704	0.000115	0.00128	0.906	0.000848	3.67E-05	0.0162	0.0303	0.00116	1.81

Table 11: Parameter sets chosen when fitting to both incidence of disease and prevalence of infection together, with associated values of incidence, prevalence and the per case ratio. The count column refers to the total number of times that parameter set was chosen, and is proportional to the likelihood.

Count	Prevalence	Incidence	ϕ	p_f	p_s	β	ρ	τ	Ratio per case
1	1.45E-06	0.000141	0.823056	0.000423	4.09E-06	0.003492	0.265196	0.970228	1.000002
11490	1.69E-06	0.000129	0.123072	0.000315	4.24E-06	0.120084	0.208552	0.505556	1.006928
57	2.66E-06	0.000137	0.048741	0.000167	5.95E-06	0.00013	0.140149	0.540094	0.999999
15	2.69E-06	0.000134	0.30826	0.000321	4.49E-06	0.000116	0.136163	0.534846	0.999999
4	1.88E-06	0.000112	0.467815	0.000403	2.61E-06	0.002132	0.162508	0.755808	1.000001
45	1.31E-06	0.000111	0.225446	6.70E-05	5.47E-06	0.188476	0.231942	0.641107	1.015996
1	4.22E-05	0.000141	0.724851	0.000427	4.02E-06	0.000867	0.008969	0.03054	0.999742
43	1.11E-05	0.000124	0.36487	0.000167	5.23E-06	0.002876	0.030495	0.117992	1.000145
37	1.61E-06	0.000111	0.164554	9.80E-05	5.22E-06	0.000323	0.188993	0.887149	1
11	1.16E-05	0.00011	0.990033	0.000138	3.92E-06	0.094977	0.02586	0.083185	1.243284

Including the proportion of all pairs containing at least one infectious case that are I-I into the model fit: Including this in the model fit alongside prevalence means that only 12 parameter sets are chosen, insufficient to generate reliable results. Including it alongside the incidence of disease means 267 parameter sets are chosen, no single one of which is chosen more than 683 times (out of 11704 parameter sets chosen overall). Whilst the parameter distributions are quite non-smooth (Figure 22), clearly higher values of p_f and ρ, values of τ centred on $10^{-\frac{3}{2}}$, around 0.03 transmissions per pair per day, are preferred, and low values of β and p_s. In this case, however, neither the prevalence of infection, nor the prevalence of disease, fit particularly well (Figure 23, Figure 24). The per case ratio is marginally above one in this case (Figure 25)

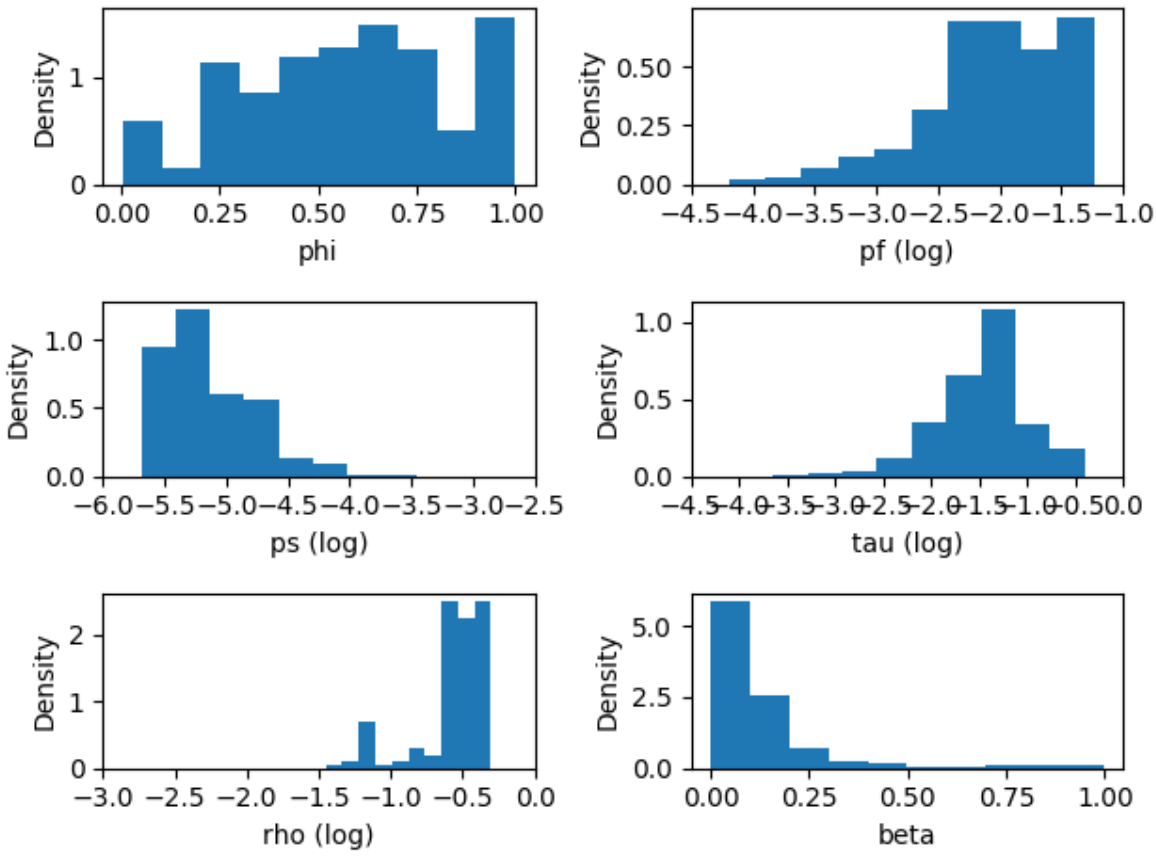


Figure 22: Posterior parameter distributions when fitted to incidence and the proportion of pairs containing at least one infectious case that are I-I.

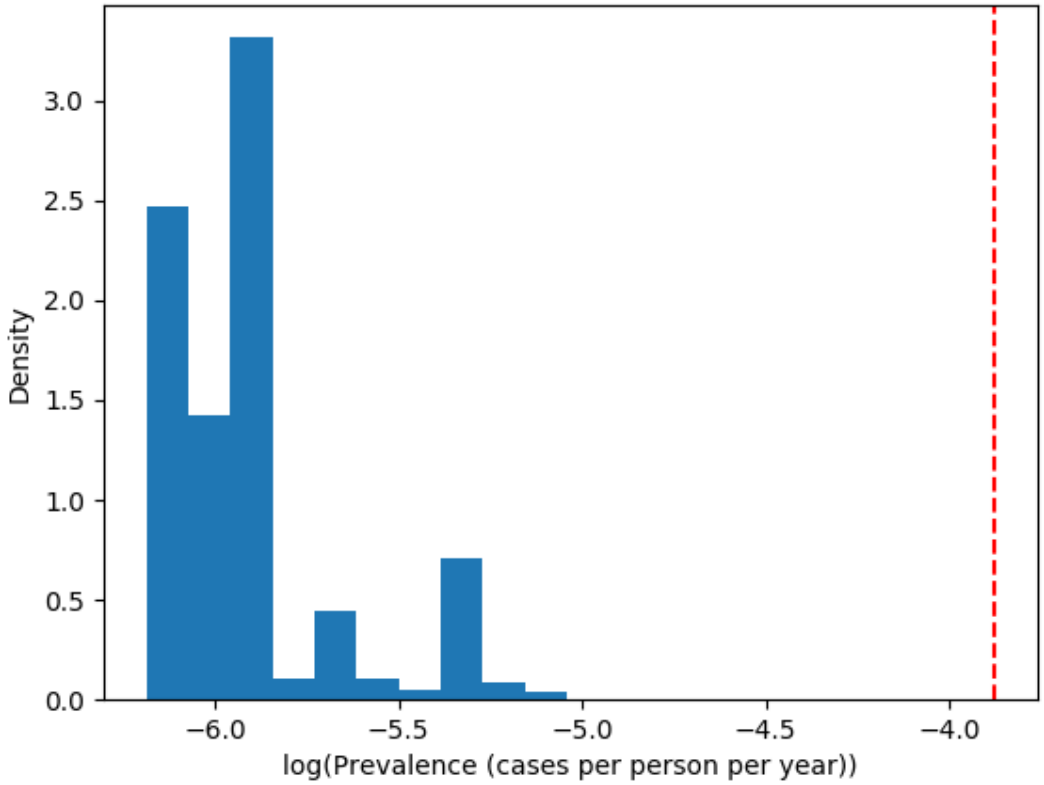


Figure 23: Distribution of prevalence when fitted to incidence and the proportion of pairs containing at least one infectious case that are I-I. The vertical dashed line is the target value. There are no values beyond the scale shown here.

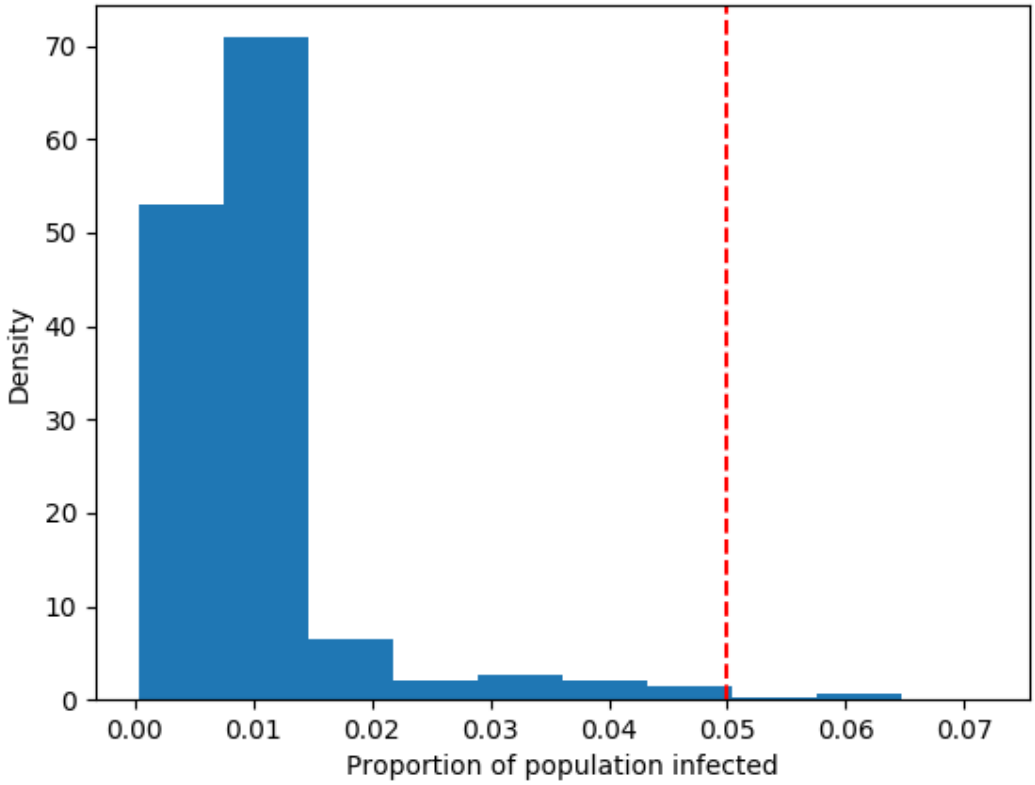


Figure 24: Prevalence of M. Tb infection when model is fitted to incidence of disease and the proportion of pairs containing at least one infectious case that are I-I. The red vertical line indicates 5%

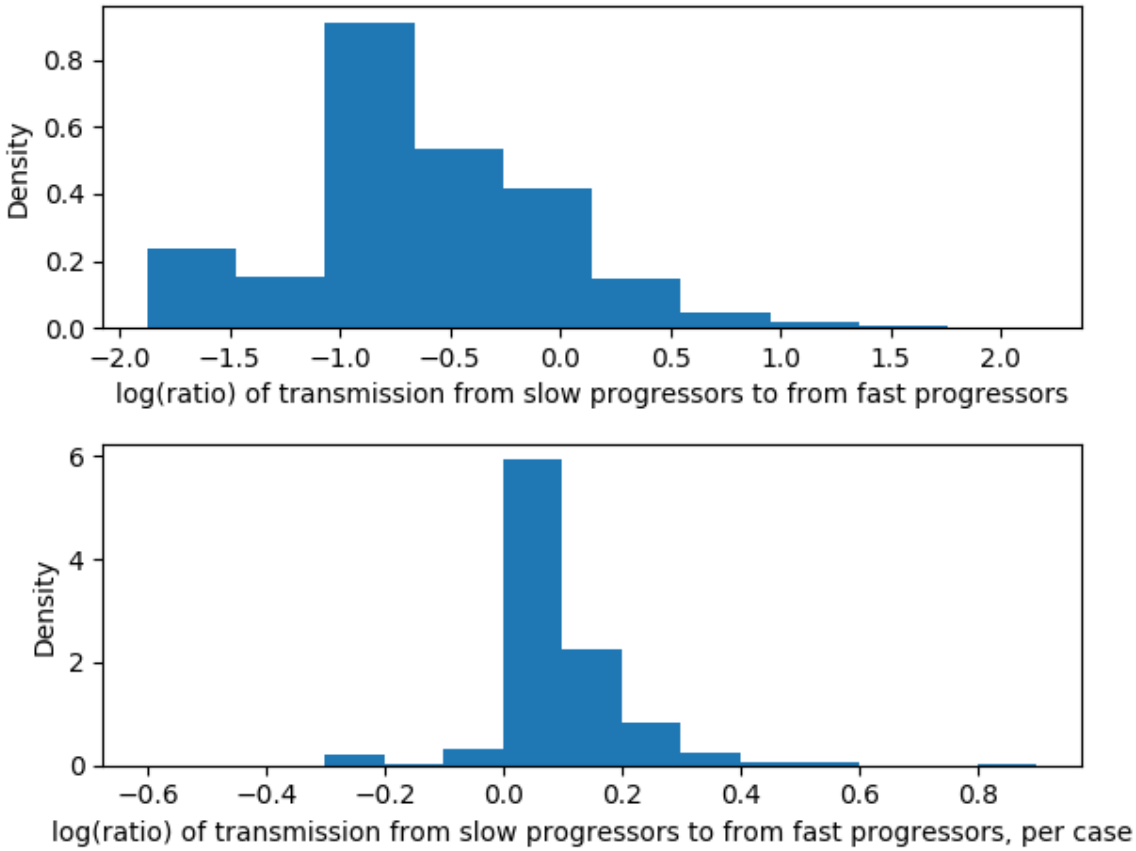
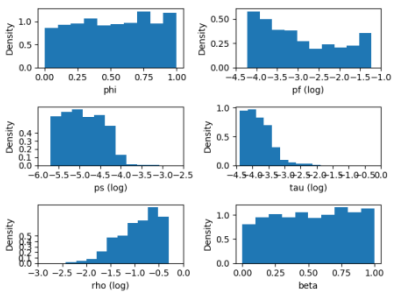
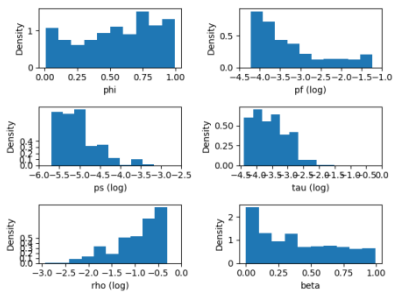
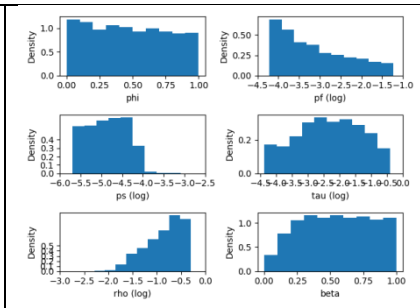
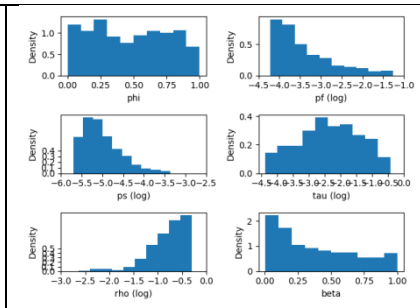
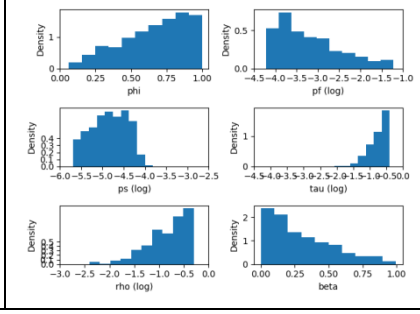
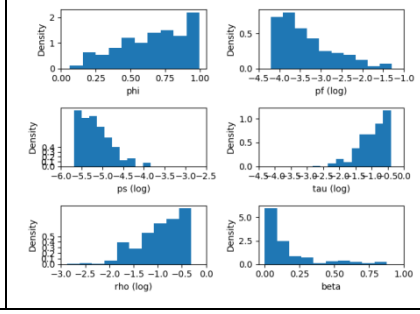


Figure 25: Posterior distribution of the log ratio of the number of infections from those with reactivation disease to the number from those with disease following recent (re)infection (top panel), and the log ratio of the number of infections from those with reactivation disease per case to the number from those with disease following recent (re)infection per case(bottom panel), when fitted to incidence of disease and the proportion of pairs containing at least one infectious case that are I-I. When these ratios are below zero, those with disease following recent (re)infection are generating more infections

Stratifying by values of the per case ratio: Whether fitting to incidence or prevalence, the per case ratio is higher, on average, for higher values of τ and lower values of β, and it is lower for lower values of τ and higher values of β (Table 12).

Table 12: How the parameter distributions vary with different values of the ratio of transmission per case from those with reactivation disease to transmission per case from those with disease following recent (re)infection.

	Fitted to Prevalence	Fitted to incidence
Ratio of transmission per case from those with reactivation disease to transmission per case from those with disease following recent (re)infection <1.0		

1.05> Ratio of transmission per case from those with reactivation disease to transmission per case from those with disease following recent (re)infection >1.0		
Ratio of transmission per case from those with reactivation disease to transmission per case from those with disease following recent (re)infection >1.05		

Fitting to the incidence and prevalence amongst the UK-born only does not really improve the combined fit: instead of two, eight parameter sets are chosen, of which one is chosen 10322 times (out of 11704), insufficient for a decent spread of results. The results when fitting to these separately do not improve the fit of infection prevalence when we fit to either incidence or prevalence, of incidence when we fit to prevalence or of prevalence when we fit to incidence. It also does not qualitatively affect the distribution of the ratio.

Fitting to the incidence in 2003 and the prevalence also does not really improve the combined fit; in this case only three parameter sets are chosen, of which two are chose more >1000 times (out of 11704).

MCMC fitting results. We started two separate MCMC chains from different regions of parameter space. The two chains seemed to converge to different regions of parameter space, selecting notably different parameter values, particularly for β and ρ (Figure 26). The model output from these parameter sets were also consequently different. The first chain estimated the prevalence of infection and the proportion of pairs with at least one infectious individual that had two infectious individuals (Figure 27). The second chain produced a good estimate disease prevalence, but por

estimates of the other three parameters (Figure 28). It should be noted that the log posterior density for the first chain (-6920) was much greater than for the second chain (-11084506), implying the first chain may be closer to the true parameter values. As we saw using the Bayesian melding algorithm, both chains were unable to produce values of incidence and prevalence that were close together. Using just the first chain, due to its higher posterior density, it appear that there is both much less transmission from those with disease following reactivation, and also slightly more transmission per case from that group, compare to those with disease following a recent (re-)infection (Figure 29).

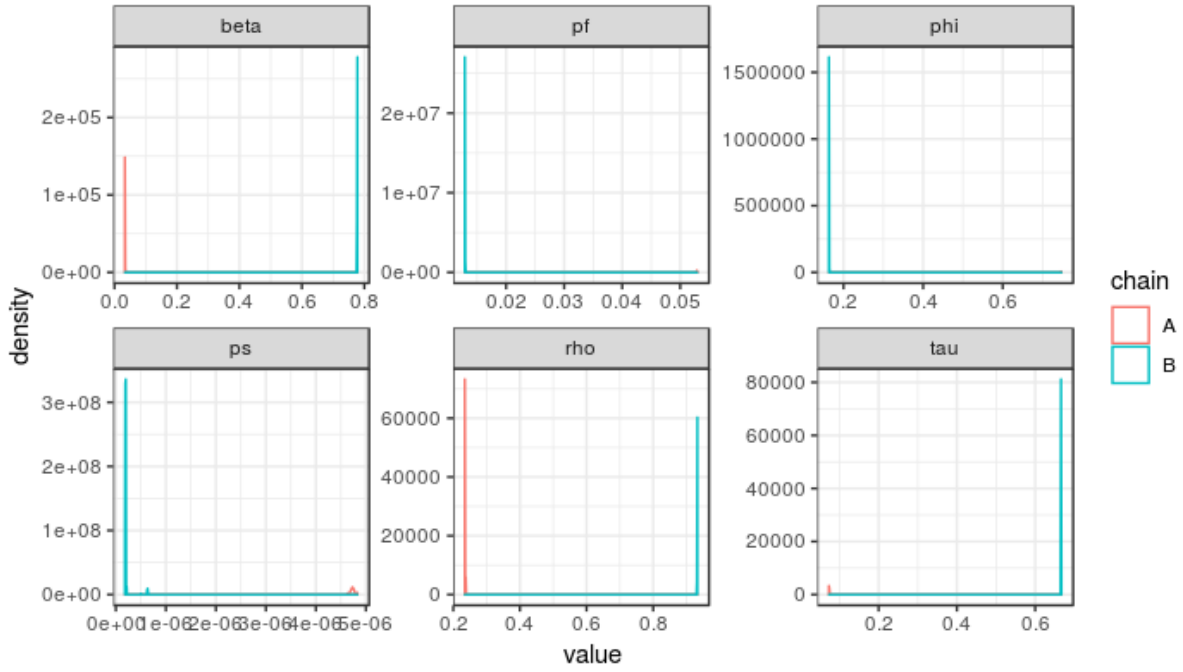


Figure 26: Density plots of parameter values for each parameter, using two chains started from different points in parameter space.

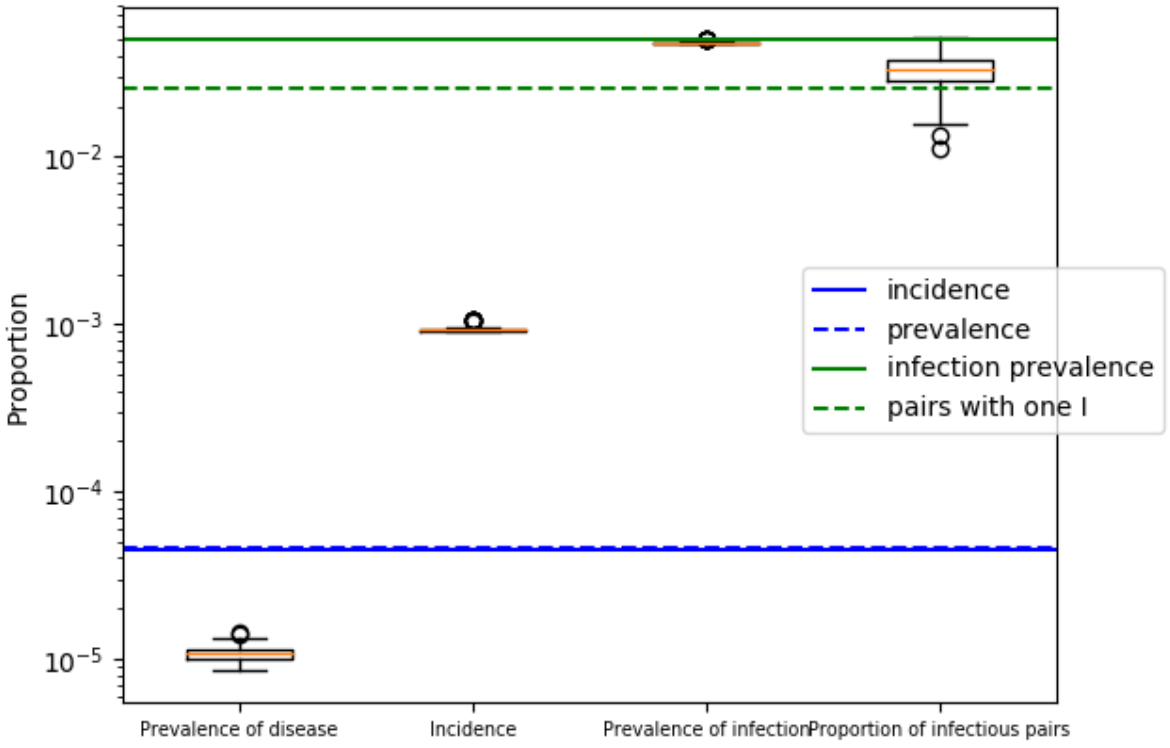


Figure 27: Distribution of model outputs using parameter distributions generated from the Markov chain Monte Carlo fitting procedure, using two chains started from different starting points. The targets for prevalence and incidence of disease were around 4/100000, the target for prevalence of infection was 5% and the target for the proportion of infectious-infectious pairs (of those with at least one infectious person) was 2.6%

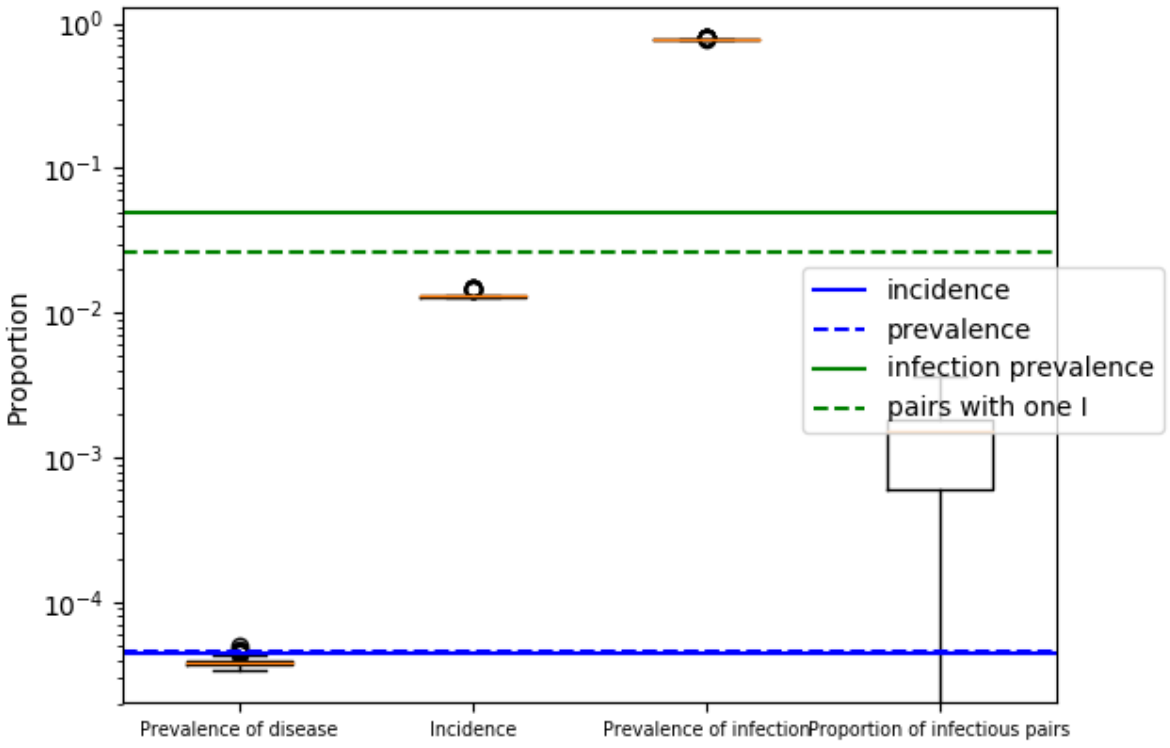


Figure 28: Distribution of model outputs using parameter distributions generated from the Markov chain Monte Carlo fitting procedure, using two chains started from different starting points. The targets for prevalence and incidence of disease were around 4/100000, the target for prevalence of infection was 5% and the target for the proportion of infectious-infectious pairs (of those with at least one infectious person) was 2.6%

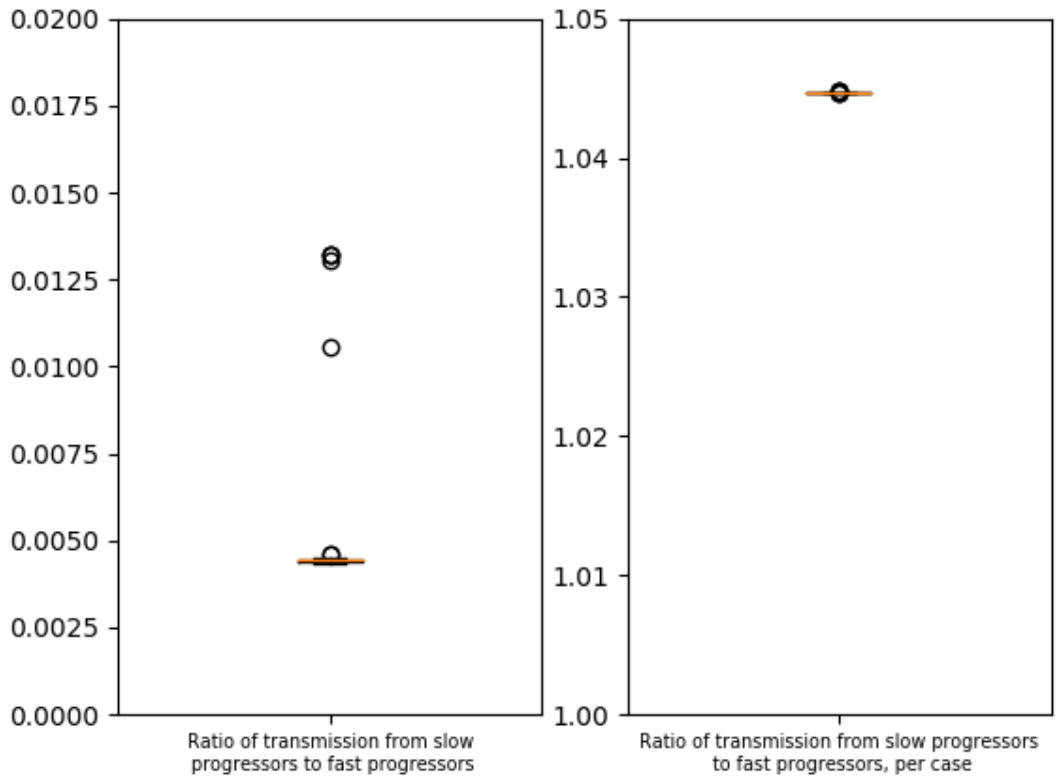


Figure 29: The ratio of transmission from those with disease following reactivation to from those with disease following recent (re-)infection, both overall (left) and per case (right)

Discussion

Primary findings: the primary finding of this work is that, while it appears to be possible to build a pairwise TB model, some, perhaps substantial, additions must be made in order to improve the realism of the model. The model is unable to fit to both prevalence and incidence of disease at the same time. Some of the issues with and potential improvements to the model are discussed below in further detail. A secondary finding is that less transmission occurs per case from those who developed disease following a recent (within one year) infection or reinfection than from those who developed reactivation disease. However, this ratio was, for most of the parameter distribution, very close to one. This implies that whilst there is a saturation effect upon transmission from cases who have been recently infected it is rather insignificant compared to other factors that may determine the probability of infection within a pair, such as whether one has smear positive or cavitary disease, or the pair live in the same room. However, caution should be employed in interpretation of this secondary finding, due to the poor fit of the model.

Strengths and weaknesses: This is the first study to employ a pairwise equation model to the study of TB. This enabled us to incorporate a rudimentary network structure within a model which is computationally efficient. The use of the SIR algorithm allows us to fit this model to surveillance data, similarly without too much computational cost.

A weakness of this study is the relatively poor model fit when trying to fit to more than one data point. It was difficult to fit to both the prevalence and incidence of disease simultaneously: the model typically underestimated the prevalence (compared to the level of prevalence observed in the data) when the model incidence matched the level observed in the data. This was not alleviated by using the prevalence and incidence amongst the UK born only, nor by using incidence from 2003, when the prevalence estimate was made. For similar reasons, as well as a lack of quality data, it was difficult to fit to the prevalence of latent infection – in this case, and when fitting to prevalence, the model predicted infection prevalence was much too high (Figure 16). This high prevalence of

infection when fitting to prevalence may reduce the size of the saturation effect, as contacts may be likely to be infected irrespective of whether or not the index case had disease following recent infection. On the other hand, when fitting to the incidence the prevalence of infection was too low, but the saturation effect was similar, and when fitting to the prevalence, infection prevalence and the size of the per case ratio are only weakly negatively correlated (R^2=0.027). Whilst we attempted to mitigate these by fitting separately to incidence and to prevalence, this then generated its own problem; in that fitting to one data point may not be sufficiently informative for each model parameter. For example, it was difficult to estimate both τ and β at the same time, as a low value of one can be to some extent compensated by a higher value of the other (Table 12). An exception was that we were able to fit to both the incidence of disease and the proportion of pairs containing at least one infectious case that were of the form I-I. In this case, we found smaller values of β and larger values of τ, centred on 0.03 transmissions per pair per day. However, even in this case, the fit to prevalence of disease and infection was very poor.

Using an MCMC algorithm also led to a poor, albeit slightly improved fit. However, this procedure was inconclusive as different starting points converged to different parameter sets. Both fitting procedures (MCMC and Bayesian melding) were unable to produce values of incidence and prevalence which were close together, as suggested by data. This result implies that model assumptions may be inconsistent with values of incidence and prevalence similar in value. However, for the first chain, the model was able to accurately reproduce the proportion of pairs with an infectious person in which both were infectious, and the prevalence of infection. In this scenario, it appeared that there was slightly more onward transmission from cases with disease following reactivation than from those with disease following a recent infection, commensurate with the idea of a small contact saturation effect. However, there was overall much less transmission from cases with disease following reactivation, suggesting that there are many more cases with disease following recent infection – an unrealistic prediction in the context of UK TB epidemiology.

A second limitation was that our model lacked a number of elements which would have made it more realistic, for instance an age structure. Part of the reason for this is the lack of data on age-structured TB contact patterns, although contact surveys that are non-specific to TB exist[88,89]. A second reason is that the number of model equations required grows with $O(n^2)$ for n the number of compartments in the model. Hence, including age-structure, or other model compartments, rapidly increases the complexity of the model. The consequence of this is that the model, as currently structured, is not really suitable for applied research questions, but rather is suited to a theoretical question as posed here. Another important facet of TB epidemiology in London that we excluded is immigration. Whilst attempting to fit to incidence and prevalence of disease in the UK born alone did not really improve the model fit, this approach ignores transmission from the non-UK born to the UK born. Incorporating immigration in a more realistic fashion, in terms of inflows into the model compartments may improve the ability of the model to match the data. It may also help to stratify the infectious population in terms of infectiousness (i.e. smear positive and smear negative), with the same caveat about the increase in complexity of the equation set.

Relation to other studies: As aforementioned, no previous studies have utilized a pairwise model of *M. tuberculosis* transmission. Previous pairwise models have examined measles[91] and sexually transmitted infections[93] and have used SI, SIR or SEIR model structures. These simpler model structures can greatly simplify the system of equations. Some of these studies have incorporated additional complexity in the form of an age-structured model[91] and heterogeneity in the number of contacts[92], both of which could feasibly be incorporated into this model. Many of the previous pairwise equation models have been used in theoretical contexts, and few attempt to fit their model to real data, although the Keeling et al model predicts quite well the number of fade-outs (3 or more consecutive weeks without infection) for measles epidemics in the UK, particularly for small population sizes. Eames et al. parametrize their STI model using detailed data on pairs, but don't compare their results to real data.

Previous *M. tuberculosis* transmission models have often been deterministic or individual based. Compared to deterministic models, the pairwise model has the advantage of enabling the examination of network features without sacrificing computational efficiency. However, a deterministic model is simpler to understand and to parameterize, and the complexity of the equation set is not as high. Individual based models have the advantage of directly incorporating stochastic elements, and may be more intuitive to understand, compared to pairwise models. However, they may be harder to set up and to interpret results, and would certainly be more computationally expensive to run. In theory, pairwise models should be easier to parametrize[129], although we encountered difficulties in this area.

As aforementioned, pairwise models are perhaps most suitable for STIs, due to the availability of data on network connections, and the fact that a pair forms the natural unit for transmission for these diseases[91]. However, a difficulty encountered when studying STIs with a pairwise model is the susceptible-infectious-susceptible dynamics these disease often exhibit[129]. Whilst moment closure approximations for pairwise models of unclustered susceptible-infectious-recovered dynamics have been shown to be exact, the situation for diseases which exhibit reinfection is complicated by local build-up of correlations between disease states that are poorly accounted for in a pairwise model[129]. Tuberculosis, then, suffers from a double difficulty of being an airborne disease, which hence allows chance transmission events between non-close contacts, and from exhibiting reinfection.

Interpretation of results: One of the difficulties we encountered when fitting the model were accurately reproducing prevalence of disease, incidence of disease and prevalence of infection simultaneously (Figure 13, Figure 14, Figure 16, Figure 17). In particular, the average values of the rate of progression following recent infection, p_f, the pairwise effective contact rate, τ, and the casual effective contact rate, β are higher when fitted to prevalence compared to incidence, whereas the reactivation rate, p_s, is lower (Table 5). This implies that when we fit to prevalence, more transmission occurs per capita and a greater proportion of cases are due to recent infection,

whereas when we fit to incidence a greater proportion of cases are due to reactivation. Values of the recovery rate, ρ, and the clustering proportion, ϕ, were similar whether fitting to incidence or prevalence, but the median recovery rate of around 0.17 cases per day in both cases, implying an average duration of infectiousness of 6 days, is much too high. In London, people are typically ill with TB for 3-4 months before they access care. The consequence of this is that, when we fit to prevalence, the predicted incidence is too high (Figure 13), whereas when we fit to incidence the prevalence is too low (Figure 14). In the former case, it may be that the greater amount of transmission, and the subsequent high risk following recent infection, means that too many incident cases are generated.

Whether fitting to incidence or prevalence, we saw more transmission from those with reactivation disease, compared to those with disease following recent (re)infection, when τ is higher and β is lower. This seems to make sense: we need a high $\beta$with a low τ (and vice versa) in order to obtain the required level of prevalence, but when τ is lower, the saturation effect is lessened as a greater proportion of transmissions are from random contacts. On the other hand, when τ is higher, a greater proportion of the transmissions are to close contacts, increasing the impact of the saturation effect and increasing the relative number of transmissions generated by those with reactivation disease. If we include reinfection when calculating the ratio, the number of parameter sets for which the ratio is below one increases (Figure 21), perhaps because of the very high prevalence of infection for many parameter sets (Figure 13, Figure 14) meaning that most pairs contain someone with latent infection, irrespective of whether recent transmission has taken place.

Throughout the experiments we keep the protection provided by recent infection fixed. Varying this may have an effect on the distributions of some of the other parameters (for instance, if more people are reinfected, then the progression rate following recent (re)infection may be smaller), and may also affect the prevalence of infection the model needs to fit to the incidence. This latter effect may serve to reduce the ratio of infections generated by those with reactivation disease to those

with disease following recent reinfection towards one, as contacts may be more likely to be infected irrespective of how recently the index case was infected.

Further research: Our initial feasibility study suggests that it is possible to use a pairwise model to study *M. Tb* transmission, but that significant work is needed to improve the realism of the model. One possible extension would be to incorporate immigration in the model. As transmission of *M.* Tb in London is driven by cases amongst immigrants, with 81% of cases amongst the non-UK born in 2016[8], not including immigration is probably a major reason why our model struggles to match the data. Whilst our rudimentary treatment of immigration, by fitting to the incidence of prevalence amongst UK born only, did not affect the results qualitatively, this approach does not account for transmission from non-UK born to the UK born. Good data exists in the UK on the prevalence of latent infection amongst new migrants, so it should not be too difficult to incorporate an additional inflow into the model to represent immigration. It may be more difficult to accurately allocate these immigrants to pairs, although a simplification could be to allocate them to pairs in the same proportion as the current distribution of pairs. Other possible extensions include heterogeneity in the number of contacts and age structure[91,93]. Improved data to help us calculate either τ or β might help us to disentangle the relative importance of each of these terms, although we could also separately calculate the amount of transmission generated by each term to get an insight into this. Data to inform the number of pairs of each type in London (rather than just those involving an infectious case) would likely improve parameterization of the model. Each of these additions should make the model more realistic, and allow it to be used to answer questions of more direct policy-relevance.

Supplementary material – additional details on MCMC

The MCMC algorithm for both chains seemed to get stuck in a small region of parameter space and did not effectively sample the whole space (Figures 29-30). The effective sample size of each parameter peaked after burning the first 100 iterations for the first chain and 150 for the second (Figure 31-32), so we burned this many iterations from the start of each chain.

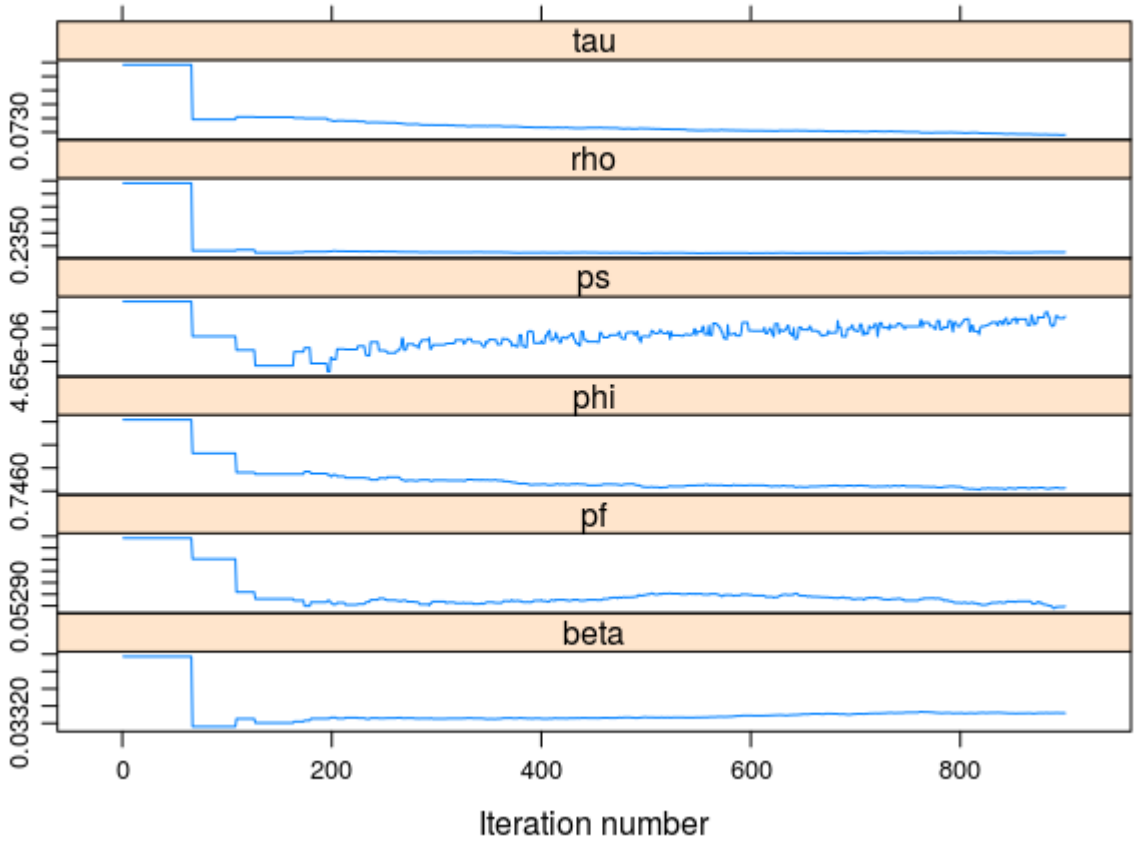


Figure 30: trace of first chain

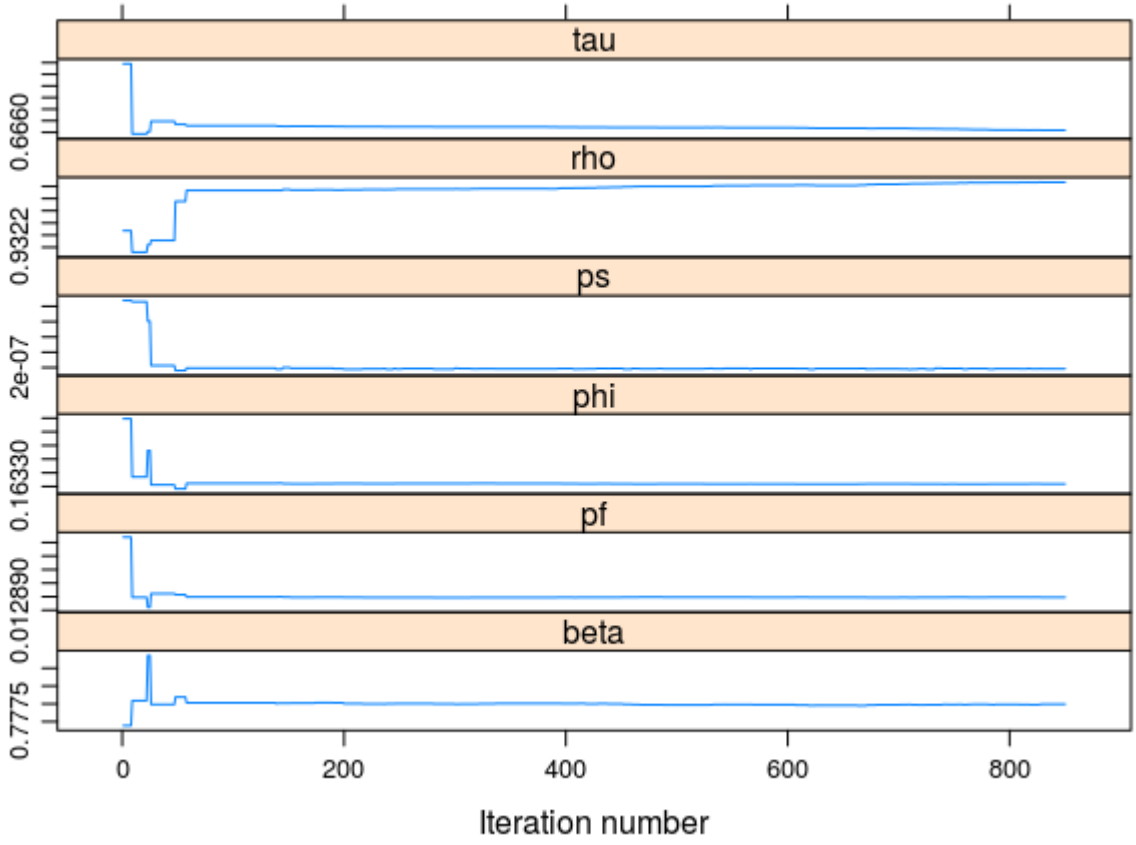


Figure 31: trace of second chain

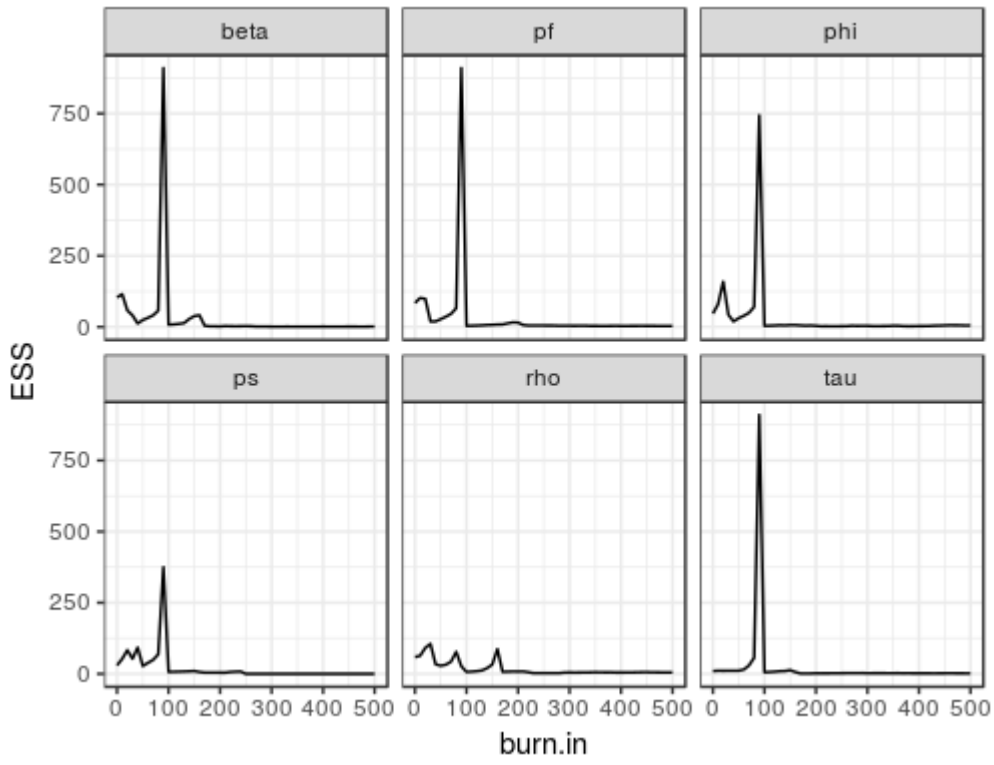


Figure 32: the effective sample size of the first chain after burning the number of elements shown on the x-axis at the start of the chain.

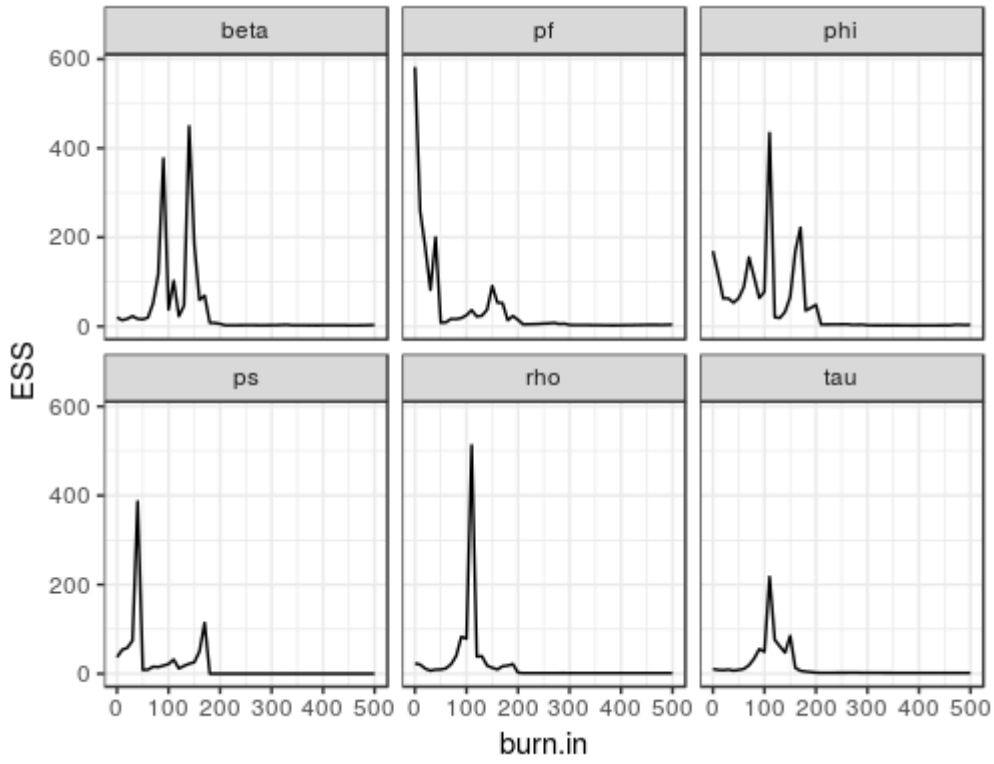


Figure 33: the effective sample size of the first chain after burning the number of elements shown on the x-axis at the start of the chain.

6. Discussion

6.1 Principal findings

The first paper of this thesis (chapter 2) determined that contact investigations in London were doing a good job of identifying contacts of pulmonary TB cases, and evaluating those contacts: 91% of pulmonary cases had at least one contact identified, and 86% of those identified were evaluated. These figures compare favourably with prior results in London, and results from other locations. That paper also showed that the proportion of evaluated contacts that are diagnosed with active TB or LTBI was approximately five times lower for contacts of non-pulmonary, non-laryngeal cases than it is for contacts of pulmonary cases. However, the proportion of contacts of non-pulmonary, non-laryngeal cases with active disease (0.70%) is still higher than the prevalence of active disease in the population (0.027%[114]). In the second paper (chapter 3), I found that 20% of contacts with TB found in contact investigations had a discordant isolate from their index case, suggesting they had reactivation disease or were infected by another source. This finding, combined with the high prevalence amongst contacts of ETB cases, suggests that there is a benefit to screening contacts, irrespective of whether or not the index case is infectious, and led to my questioning the decision in early 2016 to stop screening contacts of non-pulmonary, non-laryngeal cases. This in turn led to the analysis presented in the third paper (chapter 4), which showed that, whilst it is true that contact tracing of non-pulmonary, non-laryngeal cases reduces morbidity in contacts by about 2.6 years/year and prevents 5.5 cases/year, this was insufficient to make screening these contacts cost-effective at a £30000/QALY threshold. Hence, these results supported the changes to the NICE guidelines.

6.2 Strengths and limitations

This thesis addresses an area of great public health relevance, as contact tracing has been a core element of tuberculosis care and prevention for many years, is practised in most of the countries of the world[47], and is a key part of the national tuberculosis strategy in England[1]. A major strength is that it answers questions relating to contact tracing using a range of methodologies, including

statistical and economic analysis. A second strength is that it makes extensive use of the data available in London on contact tracing, giving the results practical relevance. Thirdly, it proposes a novel way to quantify the different impacts which contact tracing can have, by delineating the overall impact into reduced morbidity in contacts, prevented cases due to preventive therapy, reduced transmission from contacts, and reduced morbidity.

One limitation of this and many quantitative analyses is that it risks neglecting qualitative features that are not easily quantified. In this case, this could mean the impact that contact tracing activities might have on increasing awareness and reducing stigma of TB. A related issue for this study is our necessary neglect of elements for which little data was available. An example of this is the potential impact of visits to the home of index cases upon the engagement of contacts in the process, and the subsequent effect this may have on the number of contacts evaluated and on contact tracing yield. Whilst I included home visits as an explanatory variable in the first paper (chapter 2) this was done in a rudimentary fashion as no individual level data was available on home visits. Similarly, HIV status of cases was not available, and so was excluded from all analyses.

A difficulty encountered was the difficulty of obtaining a representative synthetic network on which to model tuberculosis contact tracing. Due to the endemic and chronic nature of the disease, tuberculosis dynamics evolve on a timescale which makes the changing household and community demographics important. This is heightened in the UK context for which immigration plays a key role in the dynamics. However, there is a lack of data on both the changing demographics of households in London and on non-household contact patterns. I avoided this issue in paper 3 (chapter 4) by not explicitly including a population-level transmission effect or a network structure. I examined pairwise equations as a way to introduce a network structure in chapter 5.

The second paper (chapter 3) relied on using MIRU-VNTR sequencing methods to understand transmission between index cases and contacts; it is possible that some cases were misattributed as transmission due to the resolution of this typing method. The recent rollout of whole genome

sequencing should improve the confidence in studies such as this one in the future. Both the second and third papers (chapters 3-4) made extensive use of the self-reported date of symptom onset of cases. Due to recall bias, it is likely that this underestimates the true length of time cases are symptomatic for. Finally, our analyses rely heavily on data from London, which, whilst having the aforementioned benefit of making results relevant to reality, also means that there are some settings to which our results are not easily transferred. This is particularly true of papers one and two (chapters 2-3), though the higher incidence of TB in London than other parts of the UK mean the qualitative results of paper three (chapter 4) probably hold in lower-incidence parts of the country. That is, it is probably not cost-effective to screen contacts of non-pulmonary, non-laryngeal cases anywhere in the country. However, national level data on contact tracing outcomes would certainly be useful for future research.

6.3 Relation to other studies and interpretation of results

The first paper (chapter 2) showed that contact tracing in London is being done well compared to that found in international studies[130,131] and a previous London-based study[51], when measured by the proportion of pulmonary cases with at least one contact identified and the proportion of identified contacts of pulmonary cases that are evaluated. The proportion of contacts of pulmonary cases with active TB in London (2.6%) was higher than Birmingham[2], another high incidence part of the UK, and higher than that in other high-income countries (1.4%) according to a systematic review[108] (in fact yield in London was closer to the yield of low- and middle-income countries in that review (3.1%)). The high yield in London could be due to the relatively high-prevalence of disease in London compared with other high-income countries, or because cases in London are infectious for longer on average than other high-income places (as they would then infect more of their contacts). Additionally, the second paper (chapter 3) showed that the proportion of contacts in London who have an isolate indistinguishable from their index cases' (80%) was higher than figures seen in the US[110,111] (70-71%), and higher than those in a study which calculated a similar figure for the whole of

the UK[27] (75%). So, in London, it seems that contact investigations find that a greater proportion of contacts have disease, and that for a greater proportion of those with TB, either transmission has occurred between case and contact, or case and contact share a common source. This may seem to imply that more transmission is occurring from index cases to their household contacts in London, as if the higher yield was solely due to the higher background prevalence, we would expect the proportion of pairs with matching isolates to be lower than other high-income countries, which does not appear to be the case. Cases may be transmitting to a greater number of their contacts if they are infectious for longer on average, or there is a greater proportion of smear positive cases, than in other high-income countries. I calculated the first of these figures, the mean symptomatic period in London in the third paper (chapter 4), and found that on average pulmonary cases in London are symptomatic for 110 days, as measured by self-reported date of symptom onset. Comparing this to a systematic review of diagnostic delay for tuberculosis[132] supports the hypothesis that index cases in London may be infectious for longer than in other high-incidence regions, as the London figure is higher than that found in the majority of studies in the review, and higher than all other high-income countries included.

In the first paper (chapter 2), I found that the prevalence of TB amongst contacts of non-pulmonary, non-laryngeal cases (0.70%) was high relative to the population prevalence in London[114] (0.027%), whilst paper two (chapter 3) showed that for 20% of contacts with TB no transmission had occurred. These observations led to the hypothesis that it may be worthwhile screening contacts of non-pulmonary, non-laryngeal cases, but this was refuted in the third paper (chapter 4), which showed that screening these contacts was probably not cost-effective. These cost-effectiveness estimates contained a large amount of uncertainty, but the result for non-pulmonary, non-laryngeal cases was clear in spite of this. The uncertainty did mean that I could not say anything very conclusive about the effectiveness of screening pulmonary or laryngeal cases. This uncertainty was largely driven by uncertainty in the estimates of the symptomatic period of non-pulmonary, non-laryngeal cases found through contact tracing.

All three papers (chapters 2-4) highlighted in different ways those TB cases with social risk factors (a history of imprisonment, homelessness, or drug or alcohol problems). The first paper (chapter 2) showed that former prisoners were more likely to have no contacts identified; the second paper (chapter 3) showed that 100% of case-contact pairs with a social risk factor had a likely transmission event, implying that more transmission is occurring between close contacts in this group than other groups; and the third paper (chapter 4) showed that when transmission from contacts is high, as may occur in a homeless shelter, or when the yield of non-pulmonary cases is high, as is the case for those with social risk factors in London, screening contacts of non-pulmonary cases may be worthwhile. Taken together, these things highlight the importance of additional screening beyond contact investigations in this group. In London, this is provided in part by the Find and Treat service, which effectively finds and treats those in the homeless population[56].

6.4 Further research and data requirements

In order to continue to monitor and to deepen our understanding of contact tracing in the United Kingdom, it is essential that the type of data available in London is collected and made available to researchers at a national level. It would also be extremely desirable to have some additional fields, whilst obviously bearing in mind that a larger number of fields may lead to poorer data completion. In particular, data on home visits and HIV status, improved linkage between cases and contacts (i.e. records for contacts found to have TB are accurately and consistently linked to their index case), and more data on those contacts identified but who don't have TB. In this latter case, only whether contacts are children or adults is recorded currently, unless they are found to have TB. Improving data on contacts, and/or improving linkage between contacts would enable an improved understanding of the transmission networks in London. This may also be complemented by whole-genome sequencing, which will be available nationwide in the coming years. If more data on home visits were available, it would be very useful to understand the potential for this intervention to increase the number of contacts identified and contact tracing yields.

The first paper (chapter 2) in this thesis quantified the indicators which are being used to monitor contact investigations in London. However, it did not look at whether these indicators correlated well with more fundamental goals, such as reduced transmission, and ultimately, perhaps, reduced incidence. Once a longer time-series of these, and other, indicators are available, it would be interesting to understand which ones best predict the desired population level outcomes.

Research to understand how better to engage with certain underserved groups would be of interest. In particular, whilst those with a history of homelessness are currently served by the Find and Treat service, results in this thesis also highlight that those with a prison history are not well-served by contact tracing. Research, which may be qualitative in nature, to understand why this group has poorer outcomes even after controlling for other social risk factors, and how better to engage with them, would have a positive public health impact. Qualitative research to understand some other aspects of contact tracing, such as addressing stigma, would also help.

This thesis has focussed almost exclusively on household contact investigations. However, there is a lack of understanding of the extent to which transmission in London, or the UK, is driven by household transmission or by community transmission. Whilst studies in high-income settings suggest that household contacts are significantly more likely to be infected than community contacts, these studies are from 1952 or earlier[133], and much has changed in the meantime, including the wide-spread availability of drugs and the influence of immigration on transmission. Research to understand where transmission takes place, could provide an upper limit on the effectiveness of contact tracing, and perhaps point the way for other screening interventions. We should also try to understand how the different types of screening currently in place in the UK (contact, pre-entry, find and treat) impact upon each other. For instance, pre-entry screening may mean that contact tracing becomes more focussed on the UK-born population in years to come.

Whilst I addressed one change to the NICE guidelines in paper three (chapter 4), that of whose contacts to screen, another change remains unaddressed: that of increasing the age-limit below

which people are offered LTBI therapy to 65 years. Research is needed to quantify how this will affect the number of cases prevented and rates of side-effects such as hepatotoxicity.

Finally, as discussed in the previous section, long diagnostic delays have a potential impact on the yield of contact investigations in the UK, as well as on the morbidity of cases. Uncertainty in diagnostic delays also causes much of the uncertainty in our estimates of cost-effectiveness in paper three (chapter 4). However, the estimation of diagnostic delay often relies upon the duration of the self-reported symptomatic period, which is likely to be underestimated due to recall bias. Development of a better methodology for estimating diagnostic delay would help greatly, perhaps by using the ratio of prevalence to incidence in the population. This in turn would require an improved estimate of prevalence in London, as the best current estimate dates from 2006 and uses a small sample. If a more robust analysis still shows that diagnostic delays in London or the UK are longer than average, then research to understand the reasons for the long delays, and how to improve them, would be very helpful.

6.5 Conclusions

Contact tracing is a key part of tuberculosis control in London and the UK, and effectively identifies and evaluates contacts of pulmonary cases in London, compared with past performance. It will be interesting, as the England TB strategy reaches the end of its implementation period in 2020, to compare progress to the indicator values presented here. However, while it appears to be cost-effective for contacts of pulmonary TB cases, my results suggest it is not so for non-pulmonary cases, supporting changes to NICE guidance.

References

1. Tuberculosis (TB): collaborative strategy for England - Publications - GOV.UK [Internet]. [cited 2017 Feb 15]. Available from: https://www.gov.uk/government/publications/collaborative-tuberculosis-strategy-for-england

2. Saunders MJ, Koh GCKW, Small AD, Dedicoat M. Predictors of contact tracing completion and outcomes in tuberculosis: a 21-year retrospective cohort study. Int J Tuberc Lung Dis Off J Int Union Tuberc Lung Dis. 2014 Jun;18(6):640–6.

3. Sepkowitz KA. How contagious is tuberculosis? Clin Infect Dis Off Publ Infect Dis Soc Am. 1996 Nov;23(5):954–62.

4. Rieder HL. The infectiousness of laryngeal tuberculosis: appropriate public health action based on false premises. Int J Tuberc Lung Dis Off J Int Union Tuberc Lung Dis. 2009 Jan;13(1):4–5.

5. Horowitz G, Kaslow R, Friedland G. Infectiousness of laryngeal tuberculosis. Am Rev Respir Dis. 1976 Jul;114(1):241–4.

6. Kruijshaar ME, Abubakar I. Increase in extrapulmonary tuberculosis in England and Wales 1999-2006. Thorax. 2009 Dec;64(12):1090–5.

7. Cowie RL, Sharpe JW. Extra-pulmonary tuberculosis: a high frequency in the absence of HIV infection. Int J Tuberc Lung Dis Off J Int Union Tuberc Lung Dis. 1997 Apr;1(2):159–62.

8. Tuberculosis (TB): regional and devolved administration reports - GOV.UK [Internet]. [cited 2017 Jun 22]. Available from: https://www.gov.uk/government/publications/tuberculosis-tb-regional-reports

9. Sotgiu G, Falzon D, Hollo V, Ködmön C, Lefebvre N, Dadu A, et al. Determinants of site of tuberculosis disease: An analysis of European surveillance data from 2003 to 2014. PLOS ONE. 2017 Nov 20;12(11):e0186499.

10. Musellim B, Erturan S, Sonmez Duman E, Ongen G. Comparison of extra-pulmonary and pulmonary tuberculosis cases: factors influencing the site of reactivation. Int J Tuberc Lung Dis Off J Int Union Tuberc Lung Dis. 2005 Nov;9(11):1220–3.

11. Esmail H, Barry CE, Young DB, Wilkinson RJ. The ongoing challenge of latent tuberculosis. Philos Trans R Soc Lond B Biol Sci. 2014;369(1645):20130437.

12. 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 Sep 29;190(9):1044–52.

13. Erkens CGM, Slump E, Verhagen M, Schimmel H, Cobelens F, Hof S van den. Risk of developing tuberculosis disease among persons diagnosed with latent tuberculosis infection in the Netherlands. Eur Respir J. 2016 Oct 6;ERJ-01157-2016.

14. Vynnycky E, Fine PE. The natural history of tuberculosis: the implications of age-dependent risks of disease and the role of reinfection. Epidemiol Infect. 1997 Oct;119(2):183–201.

15. Erkens CGM, Kamphorst M, Abubakar I, Bothamley GH, Chemtob D, Haas W, et al. Tuberculosis contact investigation in low prevalence countries: a European consensus. Eur Respir J. 2010 Oct 1;36(4):925–49.

16. Verver S, Warren RM, Beyers N, Richardson M, van der Spuy GD, Borgdorff MW, et al. Rate of reinfection tuberculosis after successful treatment is higher than rate of new tuberculosis. Am J Respir Crit Care Med. 2005 Jun 15;171(12):1430–5.

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

18. Rajagopalan S. Tuberculosis and aging: a global health problem. Clin Infect Dis Off Publ Infect Dis Soc Am. 2001 Oct 1;33(7):1034–9.

19. Horsburgh CRJ. Priorities for the Treatment of Latent Tuberculosis Infection in the United States. N Engl J Med. 2004 May 13;350(20):2060–7.

20. Steingart KR, Henry M, Ng V, Hopewell PC, Ramsay A, Cunningham J, et al. Fluorescence versus conventional sputum smear microscopy for tuberculosis: a systematic review. Lancet Infect Dis. 2006 Sep;6(9):570–81.

21. Behr MA, Warren SA, Salamon H, Hopewell PC, Leon AP de, Daley CL, et al. Transmission of Mycobacterium tuberculosis from patients smear-negative for acid-fast bacilli. The Lancet. 1999 Feb 6;353(9151):444–9.

22. Tuberculosis in England: annual report - GOV.UK [Internet]. [cited 2017 Jun 22]. Available from: https://www.gov.uk/government/publications/tuberculosis-in-england-annual-report

23. Wilson LG. The historical decline of tuberculosis in Europe and America: its causes and significance. J Hist Med Allied Sci. 1990 Jul;45(3):366–96.

24. Pareek M, Greenaway C, Noori T, Munoz J, Zenner D. The impact of migration on tuberculosis epidemiology and control in high-income countries: a review. BMC Med [Internet]. 2016 Mar 23 [cited 2017 Jun 15];14. Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4804514/

25. Hamblion EL, Menach AL, Anderson LF, Lalor MK, Brown T, Abubakar I, et al. Recent TB transmission, clustering and predictors of large clusters in London, 2010–2012: results from first 3 years of universal MIRU-VNTR strain typing. Thorax. 2016 Aug 1;71(8):749–56.

26. Yuen CM, Kammerer JS, Marks K, Navin TR, France AM. Recent Transmission of Tuberculosis - United States, 2011-2014. PloS One. 2016;11(4):e0153728.

27. Lalor MK, Anderson LF, Hamblion EL, Burkitt A, Davidson JA, Maguire H, et al. Recent household transmission of tuberculosis in England, 2010-2012: retrospective national cohort study combining epidemiological and molecular strain typing data. BMC Med. 2017 Jun 13;15(1):105.

28. Tuberculosis (TB): annual notification and mortality data (1913 onwards) - GOV.UK [Internet]. [cited 2017 Nov 17]. Available from: https://www.gov.uk/government/publications/tuberculosis-tb-annual-notifications-1913-onwards

29. Small PM, Hopewell PC, Singh SP, Paz A, Parsonnet J, Ruston DC, et al. The epidemiology of tuberculosis in San Francisco. A population-based study using conventional and molecular methods. N Engl J Med. 1994 Jun 16;330(24):1703–9.

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

31. Glynn JR, Vyonycky E, Fine PEM. Influence of Sampling on Estimates of Clustering and Recent Transmission of Mycobacterium tuberculosis Derived from DNA Fingerprinting Techniques. Am J Epidemiol. 1999 Feb 15;149(4):366–71.

32. Glynn JR, Bauer J, De Boer AS, Borgdorff MW, Fine PEM, Godfrey-Faussett P, et al. Interpreting DNA fingerprint clusters of Mycobacterium tuberculosis [Position Paper]. Int J Tuberc Lung Dis. 1999;3(12):1055–1060.

33. Vynnycky E, Fine PE. Interpreting the decline in tuberculosis: the role of secular trends in effective contact. Int J Epidemiol. 1999 Apr 1;28(2):327–34.

34. Tuberculosis: the green book, chapter 32 - GOV.UK [Internet]. [cited 2017 Jun 13]. Available from: https://www.gov.uk/government/publications/tuberculosis-the-green-book-chapter-32

35. Fine PEM. Variation in protection by BCG: implications of and for heterologous immunity. The Lancet. 1995 Nov 18;346(8986):1339–45.

36. Trunz BB, Fine P, Dye C. Effect of BCG vaccination on childhood tuberculous meningitis and miliary tuberculosis worldwide: a meta-analysis and assessment of cost-effectiveness. Lancet Lond Engl. 2006 Apr 8;367(9517):1173–80.

37. Zwerling A, Behr MA, Verma A, Brewer TF, Menzies D, Pai M. The BCG World Atlas: A Database of Global BCG Vaccination Policies and Practices. PLOS Med. 2011 Mar 22;8(3):e1001012.

38. Mangtani P, Abubakar I, Ariti C, Beynon R, Pimpin L, Fine PEM, et al. Protection by BCG vaccine against tuberculosis: a systematic review of randomized controlled trials. Clin Infect Dis Off Publ Infect Dis Soc Am. 2014 Feb;58(4):470–80.

39. Abubakar I, Pimpin L, Ariti C, Beynon R, Mangtani P, Sterne J a. C, et al. Systematic review and meta-analysis of the current evidence on the duration of protection by bacillus Calmette-Guérin vaccination against tuberculosis. Health Technol Assess Winch Engl. 2013 Sep;17(37):1–372, v–vi.

40. Mangtani P, Nguipdop-Djomo P, Keogh RH, Trinder L, Smith PG, Fine PE, et al. Observational study to estimate the changes in the effectiveness of bacillus Calmette-Guérin (BCG) vaccination with time since vaccination for preventing tuberculosis in the UK. Health Technol Assess Winch Engl. 2017 Jul;21(39):1–54.

41. Hart PD, Sutherland I. BCG and vole bacillus vaccines in the prevention of tuberculosis in adolescence and early adult life. Br Med J. 1977 Jul 30;2(6082):293–5.

42. Sutherland I, Springett VH. The effects of the scheme for BCG vaccination of schoolchildren in England and Wales and the consequences of discontinuing the scheme at various dates. J Epidemiol Community Health. 1989 Mar;43(1):15–24.

43. Tuberculosis | Guidance and guidelines | NICE [Internet]. [cited 2017 Jun 13]. Available from: https://www.nice.org.uk/guidance/ng33/chapter/Recommendations

44. Bothamley GH, Ditiu L, Migliori GB, Lange C, Contributors T. Active case finding of tuberculosis in Europe: a Tuberculosis Network European Trials Group (TBNET) survey. Eur Respir J. 2008 Oct 1;32(4):1023–30.

45. Veen J. Microepidemics of tuberculosis: the stone-in-the-pond principle. Tuber Lung Dis Off J Int Union Tuberc Lung Dis. 1992 Apr;73(2):73–6.

46. Society JTC of the BT, others. Control and prevention of tuberculosis in the United Kingdom: code of practice 2000. Thorax. 2000;55(11):887–901.

47. Rodriguez CA, Sasse S, Yuengling KA, Azzawi S, Becerra MC, Yuen CM. A systematic review of national policies for the management of persons exposed to tuberculosis. Int J Tuberc Lung Dis Off J Int Union Tuberc Lung Dis. 2017 Aug 1;21(8):935–40.

48. Mandal P, Craxton R, Chalmers JD, Gilhooley S, Laurenson IF, McSparron C, et al. Contact tracing in pulmonary and non-pulmonary tuberculosis. QJM Mon J Assoc Physicians. 2012 Aug;105(8):741–7.

49. Rubilar M, Brochwicz-Lewinski MJ, Anderson M, Leitch AG. The outcome of contact procedures for tuberculosis in Edinburgh, Scotland 1982-1991. Respir Med. 1995 Feb;89(2):113–20.

50. Munsiff SS, Ahuja SD, King L, Udeagu CC, Dorsinville M, Frieden TR, et al. Ensuring accountability: the contribution of the cohort review method to tuberculosis control in New York City. Int J Tuberc Lung Dis Off J Int Union Tuberc Lung Dis. 2006 Oct;10(10):1133–9.

51. Anderson C, White J, Abubakar I, Lipman M, Tamne S, Anderson SR, et al. Raising standards in UK TB control: introducing cohort review. Thorax. 2014 Feb 1;69(2):187–9.

52. Tuberculosis: pre-entry screening in the UK - GOV.UK [Internet]. [cited 2017 Jun 15]. Available from: https://www.gov.uk/government/publications/tuberculosis-pre-entry-screening-in-the-uk

53. Coker R, Bell A, Pitman R, Zellweger J-P, Heldal E, Hayward A, et al. Tuberculosis screening in migrants in selected European countries shows wide disparities. Eur Respir J. 2006 Apr 1;27(4):801–7.

54. Tuberculosis screening - GOV.UK [Internet]. [cited 2017 Jun 15]. Available from: https://www.gov.uk/guidance/tuberculosis-screening

55. Pareek M, Watson JP, Ormerod LP, Kon OM, Woltmann G, White PJ, et al. Screening of immigrants in the UK for imported latent tuberculosis: a multicentre cohort study and cost-effectiveness analysis. Lancet Infect Dis. 2011 Jun 1;11(6):435–44.

56. Jit M, Stagg HR, Aldridge RW, White PJ, Abubakar I. Dedicated outreach service for hard to reach patients with tuberculosis in London: observational study and economic evaluation. BMJ. 2011 Sep 13;343:d5376.

57. Ross R. The prevention of malaria. [Internet]. New York,; 1910. Available from: http://hdl.handle.net/2027/uc1.b4285338

58. Knight GM, Dharan NJ, Fox GJ, Stennis N, Zwerling A, Khurana R, et al. Bridging the gap between evidence and policy for infectious diseases: How models can aid public health decision-making. Int J Infect Dis. 2016 Jan 1;42:17–23.

59. Funk S, Camacho A, Kucharski AJ, Eggo RM, Edmunds WJ. Real-time forecasting of infectious disease dynamics with a stochastic semi-mechanistic model. Epidemics. 2016 Dec 16;

60. Bogoch II, Brady OJ, Kraemer MUG, German M, Creatore MI, Brent S, et al. Potential for Zika virus introduction and transmission in resource-limited countries in Africa and the Asia-Pacific region: a modelling study. Lancet Infect Dis. 2016 Nov;16(11):1237–45.

61. Vynnycky E, Yoshida LM, Huyen DTT, Trung ND, Toda K, Cuong NV, et al. Modeling the impact of rubella vaccination in Vietnam. Hum Vaccines Immunother. 2016;12(1):150–8.

62. 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 Oct 28;111(43):15520–5.

63. Bosch QA ten, Singh BK, Hassan MRA, Chadee DD, Michael E. The Role of Serotype Interactions and Seasonality in Dengue Model Selection and Control: Insights from a Pattern Matching Approach. PLoS Negl Trop Dis. 2016 May 9;10(5):e0004680.

64. Waaler H, Geser A, Andersen S. The Use of Mathematical Models in the Study of the Epidemiology of Tuberculosis. Am J Public Health Nations Health. 1962 Jun;52(6):1002–13.

65. Sutherland I, Svandová E, Radhakrishna S. The development of clinical tuberculosis following infection with tubercle bacilli. 1. A theoretical model for the development of clinical tuberculosis following infection, linking from data on the risk of tuberculous infection and the incidence of clinical tuberculosis in the Netherlands. Tubercle. 1982 Dec;63(4):255–68.

66. Harris RC, Sumner T, Knight GM, White RG. Systematic review of mathematical models exploring the epidemiological impact of future TB vaccines. Hum Vaccines Immunother. 2016 Nov;12(11):2813–32.

67. Waaler HT. Cost-benefit analysis of BCG-vaccination under various epidemiological situations. Bull Int Union Tuberc. 1968 Dec;41:42–52.

68. Bacaër N, Ouifki R, Pretorius C, Wood R, Williams B. Modeling the joint epidemics of TB and HIV in a South African township. J Math Biol. 2008 Oct;57(4):557–93.

69. Sumner T, Houben RMGJ, Rangaka MX, Maartens G, Boulle A, Wilkinson RJ, et al. Post-treatment effect of isoniazid preventive therapy on tuberculosis incidence in HIV-infected individuals on antiretroviral therapy. AIDS Lond Engl. 2016 May 15;30(8):1279–86.

70. Dodd PJ, Knight GM, Lawn SD, Corbett EL, White RG. Predicting the long-term impact of antiretroviral therapy scale-up on population incidence of tuberculosis. PloS One. 2013;8(9):e75466.

71. Fofana MO, Shrestha S, Knight GM, Cohen T, White RG, Cobelens F, et al. A Multistrain Mathematical Model To Investigate the Role of Pyrazinamide in the Emergence of Extensively Drug-Resistant Tuberculosis. Antimicrob Agents Chemother. 2017 Mar;61(3).

72. Knight GM, Colijn C, Shrestha S, Fofana M, Cobelens F, White RG, et al. The Distribution of Fitness Costs of Resistance-Conferring Mutations Is a Key Determinant for the Future Burden of Drug-Resistant Tuberculosis: A Model-Based Analysis. Clin Infect Dis Off Publ Infect Dis Soc Am. 2015 Oct 15;61Suppl 3:S147-154.

73. Castillo-Chavez C, Song B. Dynamical models of tuberculosis and their applications. Math Biosci Eng MBE. 2004 Sep;1(2):361–404.

74. Colijn C, Cohen T, Murray M. Mathematical models of tuberculosis: accomplishments and future challenges. In: BIOMAT 2006 [Internet]. WORLD SCIENTIFIC; 2007 [cited 2017 Jun 27]. p. 123–48. Available from: http://www.worldscientific.com/doi/abs/10.1142/9789812708779_0008

75. Lin H-H, Dowdy D, Dye C, Murray M, Cohen T. The impact of new tuberculosis diagnostics on transmission: why context matters. Bull World Health Organ. 2012 Oct 1;90(10):739–747A.

76. Dye C, Garnett GP, Sleeman K, Williams BG. Prospects for worldwide tuberculosis control under the WHO DOTS strategy. The Lancet. 1998 Dec 12;352(9144):1886–91.

77. Begun M, Newall AT, Marks GB, Wood JG. Contact Tracing of Tuberculosis: A Systematic Review of Transmission Modelling Studies. PLOS ONE. 2013 Sep 4;8(9):e72470.

78. Guzzetta G, Ajelli M, Yang Z, Merler S, Furlanello C, Kirschner D. Modeling socio-demography to capture tuberculosis transmission dynamics in a low burden setting. J Theor Biol. 2011 Nov 21;289:197–205.

79. Tuite AR, Gallant V, Randell E, Bourgeois A-C, Greer AL. Stochastic agent-based modeling of tuberculosis in Canadian Indigenous communities. BMC Public Health. 2017;17:73.

80. Guzzetta G, Ajelli M, Yang Z, Mukasa LN, Patil N, Bates JH, et al. Effectiveness of contact investigations for tuberculosis control in Arkansas. J Theor Biol. 2015 Sep 7;380:238–46.

81. Aparicio JP, Hernández JC. Preventive treatment of tuberculosis through contact tracing. In: Gumel AB, Castillo-Chavez C, Mickens RE, Clemence DP, editors. Contemporary Mathematics [Internet]. Providence, Rhode Island: American Mathematical Society; 2006 [cited 2017 Feb 15]. p. 17–29. Available from: http://www.ams.org/conm/410/

82. Tian Y, Osgood ND, Al-Azem A, Hoeppner VH. Evaluating the effectiveness of contact tracing on tuberculosis outcomes in Saskatchewan using individual-based modeling. Health Educ Behav. 2013;40(1 suppl):98S–110S.

83. Ziv E, Daley CL, Blower SM. Early Therapy for Latent Tuberculosis Infection. Am J Epidemiol. 2001 Feb 15;153(4):381–5.

84. 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 Feb 21;189(7):845–52.

85. Vynnycky E, White R. An Introduction to Infectious Disease Modelling. 1 edition. New York: Oxford University Press; 2010. 400 p.

86. Mellor GR, Currie CSM, Corbett EL. Incorporating Household Structure into a Discrete-event Simulation Model of Tuberculosis and HIV. ACM Trans Model Comput Simul. 2011 Sep;21(4):26:1–26:17.

87. Lum K, Swarup S, Eubank S, Hawdon J. The contagious nature of imprisonment: an agent-based model to explain racial disparities in incarceration rates. J R Soc Interface. 2014 Sep 6;11(98):20140409.

88. Mossong J, 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 Mar 25;5(3):e74.

89. Danon L, Read JM, House TA, Vernon MC, Keeling MJ. Social encounter networks: characterizing Great Britain. Proc R Soc B Biol Sci [Internet]. 2013 Aug 22 [cited 2017 May 18];280(1765). Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3712448/

90. Cohen T, Colijn C, Finklea B, Murray M. Exogenous re-infection and the dynamics of tuberculosis epidemics: local effects in a network model of transmission. J R Soc Interface. 2007 Jun 22;4(14):523–31.

91. Keeling MJ, Rand DA, Morris AJ. Correlation models for childhood epidemics. Proc Biol Sci. 1997 Aug 22;264(1385):1149–56.

92. Eames KTD. Modelling disease spread through random and regular contacts in clustered populations. Theor Popul Biol. 2008 Feb;73(1):104–11.

93. Eames KTD, Keeling MJ. Modeling dynamic and network heterogeneities in the spread of sexually transmitted diseases. Proc Natl Acad Sci U S A. 2002 Oct 1;99(20):13330–5.

94. Ross JV, House T, Keeling MJ. Calculation of disease dynamics in a population of households. PloS One. 2010 Mar 18;5(3):e9666.

95. House T, Keeling MJ. Deterministic epidemic models with explicit household structure. Math Biosci. 2008 May;213(1):29–39.

96. Black AJ, House T, Keeling MJ, Ross JV. Epidemiological consequences of household-based antiviral prophylaxis for pandemic influenza. J R Soc Interface [Internet]. 2013 Apr 6 [cited 2017 May 18];10(81). Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3627116/

97. NHS reference costs 2015 to 2016 - GOV.UK [Internet]. [cited 2017 Jun 28]. Available from: https://www.gov.uk/government/publications/nhs-reference-costs-2015-to-2016

98. Committee JF, editor. BNF 73 (British National Formulary) March 2017. 73rd Revised edition edition. London: Pharmaceutical Press; 2017. 1480 p.

99. Menzies NA, Gomez GB, Bozzani F, Chatterjee S, Foster N, Baena IG, et al. Cost-effectiveness and resource implications of aggressive action on tuberculosis in China, India, and South Africa: a combined analysis of nine models. Lancet Glob Health. 2016 Nov 1;4(11):e816–26.

100. Dasgupta K, Schwartzman K, Marchand R, Tennenbaum TN, Brassard P, Menzies D. Comparison of Cost-Effectiveness of Tuberculosis Screening of Close Contacts and Foreign-Born Populations. Am J Respir Crit Care Med. 2000 Dec 1;162(6):2079–86.

101. Tian Y, Alawami F, Al-Azem A, Osgood N, Hoeppner V, Dutchyn C. A System Dynamics Model of Tuberculosis Diffusion with Respect to Contact Tracing Investigation. In: Proceedings of the Winter Simulation Conference [Internet]. Phoenix, Arizona: Winter Simulation Conference; 2011 [cited 2018 Apr 4]. p. 1367–1378. (WSC '11). Available from: http://dl.acm.org/citation.cfm?id=2431518.2431682

102. Cavany SM, Sumner T, Vynnycky E, Flach C, White RG, Thomas HL, et al. An evaluation of tuberculosis contact investigations against national standards. Thorax. 2017 Apr 7;thoraxjnl-2016-209677.

103. Wingfield T, MacPherson P, Cleary P, Ormerod LP. High prevalence of TB disease in contacts of adults with extrapulmonary TB. Thorax. 2017 Nov 16;

104. Grzybowski S, Barnett GD, Styblo K. Contacts of cases of active pulmonary tuberculosis. Bull Int Union Tuberc. 1975;50(1):90–106.

105. Underwood BR, White VLC, Baker T, Law M, Moore-Gillon JC. Contact tracing and population screening for tuberculosis--who should be assessed? J Public Health Med. 2003 Mar;25(1):59–61.

106. Ansari S, Thomas S, Campbell IA, Furness L, Evans MR. Refined tuberculosis contact tracing in a low incidence area. Respir Med. 1998 Sep 1;92(9):1127–31.

107. 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 Jun;8(6):359–68.

108. Fox GJ, Barry SE, Britton WJ, Marks GB. Contact investigation for tuberculosis: a systematic review and meta-analysis. Eur Respir J. 2013 Jan;41(1):140–56.

109. WHO | Tuberculosis country profiles [Internet]. WHO. [cited 2018 Apr 4]. Available from: http://www.who.int/tb/country/data/profiles/en/

110. Bennett DE, Onorato IM, Ellis BA, Crawford JT, Schable B, Byers R, et al. DNA Fingerprinting of Mycobacterium tuberculosis Isolates from Epidemiologically Linked Case Pairs. Emerg Infect Dis. 2002 Nov;8(11):1124–9.

111. McNabb SJN, Kammerer JS, Hickey AC, Braden CR, Shang N, Rosenblum LS, et al. Added epidemiologic value to tuberculosis prevention and control of the investigation of clustered genotypes of Mycobacterium tuberculosis isolates. Am J Epidemiol. 2004 Sep 15;160(6):589–97.

112. Verver S, Warren RM, Munch Z, Richardson M, van der Spuy GD, Borgdorff MW, et al. Proportion of tuberculosis transmission that takes place in households in a high-incidence area. Lancet Lond Engl. 2004 Jan 17;363(9404):212–4.

113. Augustynowicz-Kopeć E, Jagielski T, Kozińska M, Kremer K, van Soolingen D, Bielecki J, et al. Transmission of tuberculosis within family-households. J Infect. 2012 Jun;64(6):596–608.

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

115. Hopewell PC, Fair E, Miller C, World Health Organization, Communicable Diseases Cluster, Stop TB Department. Recommendations for investigating contacts of persons with infectious tuberculosis in low- and middle-Income countries [Internet]. 2012 [cited 2017 Feb 15]. Available from: http://www.ncbi.nlm.nih.gov/books/NBK179062/

116. Guide to the methods of technology appraisal 2013 | Guidance and guidelines | NICE [Internet]. [cited 2017 May 31]. Available from:

https://www.nice.org.uk/process/pmg9/chapter/the-appraisal-of-the-evidence-and-structured-decision-making

117. Zammarchi L, Casadei G, Strohmeyer M, Bartalesi F, Liendo C, Matteelli A, et al. A scoping review of cost-effectiveness of screening and treatment for latent tuberculosis infection in migrants from high-incidence countries. BMC Health Serv Res. 2015 Sep 24;15:412.

118. Diel R, Nienhaus A, Schaberg T. Cost-effectiveness of isoniazid chemoprevention in close contacts. Eur Respir J. 2005 Sep;26(3):465–73.

119. Tan MC, Marra CA, Sadatsafavi M, Marra F, Morán-Mendoza O, Moadebi S, et al. Cost-effectiveness of LTBI treatment for TB contacts in British Columbia. Value Health J Int Soc Pharmacoeconomics Outcomes Res. 2008 Oct;11(5):842–52.

120. Mandalakas AM, Hesseling AC, Gie RP, Schaaf HS, Marais BJ, Sinanovic E. Modelling the cost-effectiveness of strategies to prevent tuberculosis in child contacts in a high-burden setting. Thorax. 2013 Mar;68(3):247–55.

121. Htet KKK, Liabsuetrakul T, Thein S. Cost-effectiveness of a new strategy to detect pulmonary tuberculosis in household contacts in Myanmar. Int J Tuberc Lung Dis. 2017 Feb 1;21(2):181–7.

122. Sekandi JN, Dobbin K, Oloya J, Okwera A, Whalen CC, Corso PS. Cost-effectiveness analysis of community active case finding and household contact investigation for tuberculosis case detection in urban Africa. PloS One. 2015;10(2):e0117009.

123. Eames KTD. Contact tracing strategies in heterogeneous populations. Epidemiol Infect. 2007 Apr;135(3):443–54.

124. Mears J, Vynnycky E, Lord J, Borgdorff MW, Cohen T, Crisp D, et al. The prospective evaluation of the TB strain typing service in England: a mixed methods study. Thorax. 2016 Aug 1;71(8):734–41.

125. 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 Mar 15;54(6):784–91.

126. Population Estimates for UK, England and Wales, Scotland and Northern Ireland - Office for National Statistics [Internet]. [cited 2017 Aug 2]. Available from: https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/datasets/populationestimatesforukenglandandwalesscotlandandnorthernireland

127. Mancuso JD, Diffenderfer JM, Ghassemieh BJ, Horne DJ, Kao T-C. The Prevalence of Latent Tuberculosis Infection in the United States. Am J Respir Crit Care Med. 2016 Aug 15;194(4):501–9.

128. Vynnycky E, Fine PE. Lifetime risks, incubation period, and serial interval of tuberculosis. Am J Epidemiol. 2000 Aug 1;152(3):247–63.

129. Keeling MJ, House T, Cooper AJ, Pellis L. Systematic Approximations to Susceptible-Infectious-Susceptible Dynamics on Networks. PLoS Comput Biol. 2016 Dec;12(12):e1005296.

130. 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 Jan 1;64(50–51):1369–74.

131. Borraccino A, Migliore E, Piccioni P, Baussano I, Carosso A, Bugiani M. Yield of tuberculosis contact investigation in a low-incidence country. J Infect. 2014 May;68(5):448–54.

132. Storla DG, Yimer S, Bjune GA. A systematic review of delay in the diagnosis and treatment of tuberculosis. BMC Public Health. 2008 Jan 14;8:15.

133. 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. Am J Epidemiol. 2017 Jun 15;185(12):1327–39.