

# How much do smear-negative patients really contribute to tuberculosis transmissions? Re-examining an old question with new tools

Leyla Asadi,<sup>a</sup> Matthew Croxen,<sup>b</sup> Courtney Heffernan,<sup>a</sup> Mannat Dhillon,<sup>a</sup> Catherine Paulsen,<sup>a</sup> Mary Lou Egedahl,<sup>a</sup> Greg Tyrrell,<sup>b</sup> Alexander Doroshenko,<sup>a</sup> and Richard Long,<sup>a\*</sup>

<sup>a</sup>The Department of Medicine, Faculty of Medicine and Dentistry, University of Alberta, Room 8334A, 3rd Floor, Aberhart Centre, 11402 University Avenue NW, Edmonton, Edmonton, AB T6G 2J3, Canada

<sup>b</sup>The Department of Laboratory Medicine and Pathology, Faculty of Medicine and Dentistry, University of Alberta, Edmonton, AB, Canada

## Summary

**Background** Sputum smear microscopy is a common surrogate for tuberculosis infectiousness. Previous estimates that smear-negative patients contribute 13–20% of transmissions and are, on average, 20 to 25% as infectious as smear-positive cases are understood to be high. Herein, we use an ideal real-world setting, a comprehensive dataset, and new high-resolution techniques to more accurately estimate the true transmission risk of smear-negative cases.

**Methods** We treated all adult culture-positive pulmonary TB patients diagnosed in the province of Alberta, Canada from 2003 to 2016 as potential transmitters. The primary data sources were the Alberta TB Registry and the Provincial Laboratory for Public Health. We measured, as primary outcomes, the proportion of transmissions attributable to smear-negative sources and the relative transmission rate. First, we replicated previous studies by using molecular (DNA) fingerprint clustering. Then, using a prospectively collected registry of TB contacts, we defined transmission events as active TB amongst identified contacts who either had a 100% DNA fingerprint match to the source case or a clinical diagnosis. We supplemented our analysis with genome sequencing on temporally and geographically linked DNA fingerprint clusters of cases not identified as contacts.

**Findings** There were 1176 cases, 563 smear-negative and 613 smear-positive, and 23,131 contacts. Replicating previous studies, the proportion of transmissions attributable to smear-negative source cases was 16% (95% CI, 12–19%) and the relative transmission rate was 0.19 (95% CI, 0.14–0.26). With our combined approach, the proportion of transmission was 8% (95% CI, 3–14%) and the relative transmission rate became 0.10 (95% CI, 0.05–0.19).

**Interpretation** When we examined the same outcomes as in previous studies but refined transmission ascertainment with the addition of conventional epidemiology and genomics, we found that smear-negative cases were ~50% less infectious than previously thought.

**Funding** Alberta Innovates Health Solutions

**Copyright** © 2021 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

**Keywords:** Transmission of TB; Smear-negative TB

## Introduction

Prior to the COVID-19 pandemic, globally, tuberculosis (TB) was the number one infectious disease cause of death. With 10 million new cases and 1.45 million deaths annually, TB has appropriately, albeit belatedly, become the focus of high-level commitments and concerted elimination efforts.<sup>1</sup> However, while TB

elimination requires an end to transmission, there are still many unanswered questions about the transmission of TB.<sup>2</sup>

The number of infectious droplet nuclei per volume of air<sup>3</sup> determines TB infectiousness. Varying patient, pathogen, and environmental factors can create the requisite conditions for a transmission. In terms of patient-level factors, infectiousness of a person with TB depends on disease type (laryngeal vs pulmonary vs extra-pulmonary), radiographic findings, symptomology

\*Corresponding author.

E-mail address: [richard.long@ualberta.ca](mailto:richard.long@ualberta.ca) (R. Long).

**EClinicalMedicine**  
2022;43: 101250  
Published online xxx  
<https://doi.org/10.1016/j.eclinm.2021.101250>

### Research in context

#### *Evidence before this study*

To gather evidence on the studies examining sputum smear microscopy status and transmission, we searched PubMed/Medline between 1996 and 2020, with the keywords “tuberculosis” and “transmission” or “smear microscopy” or “smear negative” or “smear positive”. Three studies (published between 1999 and 2007) used molecular epidemiology to estimate two outcomes describing transmissions: the proportion of TB transmissions attributable to smear-negatives and the relative transmission rate. These studies estimated that the proportion of transmissions arising from smear-negative patients was ~17% and the relative transmission rates was 0.22–0.24.

#### *Added value of this study*

When we exclusively used molecular epidemiology, our estimates of infectiousness corresponded with previous results. However, when we combined molecular epidemiology with conventional epidemiology (namely, rigorous, systematic contact tracing) and genome sequencing, we found that smear-negative cases were 50% less infectious than reported in these prior studies. Our study shows the value of conventional epidemiology and whole genome sequencing and also highlights the challenges of accurately discerning chains of transmissions.

#### *Implications of all the available evidence*

Our findings reflect the value of a comprehensive strategy of discerning chains of transmissions and the variable results that may ensue from different methods. Our study also suggests that smear-negative patients are less infectious than previously thought. These updated estimates may be used in TB transmission modelling, and smear-negative patients may be considered a lower priority for contact tracing and a lesser infection control concern.

(most notably cough), delays in diagnosis, treatment characteristics, and, importantly, sputum smear-status.<sup>3</sup>

Sputum smear microscopy remains the most commonly used surrogate marker for estimating infectiousness.<sup>4</sup> Historical studies observed that household contacts of smear-positive patients were between 2 and 12 times more likely than household contacts of smear-negative patients to be infected with *Mycobacterium tuberculosis* (*M. tb*).<sup>3</sup> More recent studies have examined the question of transmission from smear-negative patients by comparing clusters of DNA fingerprints arising from smear-negative vs smear-positive source patients.<sup>5–7</sup> These studies, themselves now between 10 and 20 years old, estimated that the relative transmission rate of smear-negative compared with smear-

positive patients was 0.22–0.24, or roughly 20 to 25% the likelihood of transmission.

There is no gold standard for identifying TB transmissions. In some studies in low-incidence settings, conventional epidemiology alone, without DNA fingerprinting, may underestimate transmissions.<sup>8</sup> By contrast, DNA fingerprint clustering alone appears to overestimate the number of recent transmissions.<sup>9,10</sup> When France et al. compared field-based (conventional and molecular investigation) evidence of recent transmission with genotype-based (molecular) estimates, they found that genotype-based estimates could overestimate recent transmission by as much as 75%.<sup>11</sup> However, it is also well-established that compared with traditional molecular methods, genome sequencing allows for greater resolution in identifying transmissions.<sup>12–14</sup> Unfortunately, for most jurisdictions, genome sequencing remains prohibitively expensive.

Given the reliance on sputum smear microscopy for triaging contact-tracing efforts and establishing infection control protocols, estimates of the infectiousness of smear-negative cases need to be updated. Therefore, we re-examined the contribution of smear-negative source patients to tuberculosis transmissions in a low TB incidence, low HIV prevalence, high-income setting. We compared previously described DNA fingerprint clustering techniques with an approach combining molecular and conventional epidemiology supplemented with genome sequencing.

## Methods

### Population

We included all culture-positive pulmonary TB patients diagnosed in the province of Alberta, Canada (population 4251,900 and crude TB incidence 5.5/100,000 population in 2016) from January 2004 to December 2016 as notified in the Alberta TB Registry.<sup>15</sup> For each patient, we extracted routinely and systematically collected (in real-time) demographic, clinical, laboratory, and contact-tracing information from the Integrated Public Health Information System—the location of the provincial tuberculosis database.

### Smear-Status definition

We used auramine-rhodamine stains for screening respiratory sputum samples and followed this with confirmatory Ziehl-Neelsen staining. A smear-positive individual had at least one positive respiratory sample prior to the initiation of anti-tuberculosis therapy. A smear-negative individual had to have at least three respiratory samples submitted for analysis and all samples submitted prior to initiation of therapy had to be negative. All mycobacteriology in the province is performed in a single laboratory: Provincial Laboratory for Public Health.

### Molecular genotyping

From July 2003 to June 2016, isolates of *Mycobacterium tuberculosis* from all culture-positive cases of TB diagnosed in Alberta were routinely DNA fingerprinted by use of standardised restriction fragment-length polymorphism (RFLP), supplemented in isolates with five or fewer copies of the insertion sequence *6110*, by spoligotyping.<sup>16,17</sup> From January 2014 onwards, isolates had 24 loci mycobacterial interspersed repetitive units (MIRU) typing.<sup>18</sup>

### Genome sequencing of *M. tb*

*M. tb* was grown on Lowenstein-Jensen media, and heat killed at 90 °C for 30 min prior to DNA extraction. Routine methods of extraction, sequence quality control and assessment were then performed.

Following bead beating with 0.5 mm glass beads (Sigma Z763748) for 5 min, lysozyme (Sigma L6876) was added and incubated at 37 °C for 1 h. Further lysis of the cells was done with the MagaZorb<sup>®</sup> DNA Mini-prep kit proteinase K, lysis buffer (Promega MB1004) and RNase A (Qiagen #19,101) digestion at 60 °C overnight. Extraction was completed using the MagaZorb<sup>®</sup> kit on the Kingfisher mL Purification System. Extracted DNA was checked on a 1% agarose gel and quantified using the Qubit BR kit.

Illumina-compatible libraries were generated using the Illumina Nextera DNA Flex Library Prep Kit (Illumina 20,018,705), and sequenced on an Illumina MiSeq with the 600-cycle MiSeq Reagent Kit v3 (Illumina MS-102-3003).

### Sequence quality control and assessment

Raw Illumina sequences were assessed for quality with FastQC v0.11.8 (bioinformatics.babraham.ac.uk/projects/fastqc/) and MultiQC v1.7.<sup>19</sup> Adapters were trimmed with trimmomatic v0.39, as well as quality filtering and keeping any sequences greater than 75 bp in length. Kraken 2.0.8-beta<sup>20</sup> with minikraken2\_v2\_8GB\_201,904\_UPDATE was used to classify the trimmed reads, allowing us to identify and keep reads classified as *M. tb* complex (MTBC) or the genus *Mycobacterium* – all other reads were discarded. Of the 49 *M. tb* genomes sequenced, two (Mtb34 and Mtb42) had significant non-*Mycobacterium* sequences (26% and 41%, respectively) that were removed by this taxonomic filtering. The remaining 47 genomes were > 99% classified as *Mycobacterium*. These quality-filtered classified sequences were used in downstream analysis. Estimated depth of coverage after quality filtering ranged from 40 to 454x. Biohansel 2.4.0<sup>21</sup> was used to determine the lineages of each data set, indicating a mix of lineage 1, 2 and 4.

### Phylogenetic tree and pairwise single nucleotide variant distances

H37Rv (NC\_000962.3) was used as a reference for mapping and single nucleotide variant (SNV) calling with Snippy 4.5.0 (github.com/tseemann/snippy) using default settings (minimum mapping quality 60, minimum base quality 13, minimum depth 10). Core SNVs were determined with Snippy, masking the variable regions with the packaged masking file resulting in 93–95% coverage of the H37Rv reference genome. A core SNV phylogeny was generated using IQ-TREE 1.6.12<sup>22</sup> with a generalized-time reversible model and with discrete Gamma rate heterogeneity (GTR+G4), one thousand bootstraps,<sup>23,24</sup> one thousand single branch test replicates,<sup>25</sup> and using constant sites as determined by snp-sites 2.5.1.<sup>26</sup> The resulting phylogenetic tree was visualized with FigTree v1.4.4 (github.com/rambaut/figtree), displayed using a midpoint-root. Pairwise SNV distances were calculated using snp-dists 0.6.3 (github.com/tseemann/snp-dists).

Since three different lineages were used to generate the pairwise SNV distances and the phylogenetic trees, to provide further resolution on the pairwise distances between known MIRU pairs, we regenerated the core SNVs for each individual pair using Snippy and re-calculated the pairwise distance using snp-dists.

### Outcome

The primary outcomes of interest were the proportion of TB transmission attributable to smear-negative cases and the relative transmission rate. The proportion of TB transmission attributable to smear negative sources was defined as: [number of transmissions arising from smear-negative sources]/[total number of transmissions]. The relative transmission rate is defined as: [Smear-negative transmission events/total number of smear-negative patients with culture-positive pulmonary TB]/[smear-positive transmission events/total number of smear-positive patients with culture-positive pulmonary TB].

### Statistical methods

The Wald method was used to estimate 95% confidence intervals (CI) for the proportions of transmission attributable to smear-negative cases; we also calculated 95% CI for relative transmission rate using the Clopper-Pearson (exact) method. Chi-square testing was used to examine whether there was a statistically significant difference in the proportion of transmissions attributable to smear negative cases between method #1 (molecular epidemiology using DNA fingerprint clusters) and method #2 (combined conventional epidemiology, molecular epidemiology and genome sequencing). Chi-square testing was also used to examine whether there was global difference between method #1, method #2,

and the two sensitivity analyses. SAS version 9.4 was used for statistical analysis.

#### **Method #1: molecular epidemiology (DNA fingerprint cluster)**

The goal of this analysis was to replicate the methodology of previous studies.<sup>5–7</sup> The analysis was carried out on all TB patients from July 2003 to June 2016. Given there was no RFLP typing after June 2016, clustering could no longer be determined beyond this date. We defined a cluster as two or more patients having 100% matched DNA fingerprints and ordered them chronologically. Within the cluster, if the first patient had only extra-pulmonary TB, they were excluded and the subsequent patient (ordered chronologically), would be considered the index patient. We employed the methodology described in the seminal paper by Behr et al.: “Secondary cases that were preceded only by cases of smear-negative TB were attributed to smear-negative transmission.”<sup>6</sup> All cases that occurred after any case of smear-positive TB were attributed to smear-positive transmission.” Tostmann et al. also used this method.<sup>7</sup> We also carried out an analysis where we restricted the time elapsed between a transmission event to 2-years.

#### **Method #2: combined conventional and molecular epidemiology supplemented with genome sequencing**

For this method, we identified all contacts of pulmonary TB patients diagnosed from January 2004 to December 2016. The province of Alberta has a rigorous, prospective, routine contact tracing system that relies upon the stone-in-pond principle (i.e., contact tracing begins with close or high-risk contacts before casual or low-risk contacts).<sup>3</sup> As per the Canadian TB standards, for smear-negative cases, household and other high-priority contacts are assessed in the initial round of contact investigations; however, medium-priority contacts (e.g. other close non-household contacts) are only assessed if there is evidence of transmission in the first round.<sup>3</sup> For smear-positive cases, both high and medium priority are assessed from the outset. Contacts in places of social aggregation (i.e. educational or work settings, places of worship, or homeless shelters) are routinely identified.<sup>27</sup>

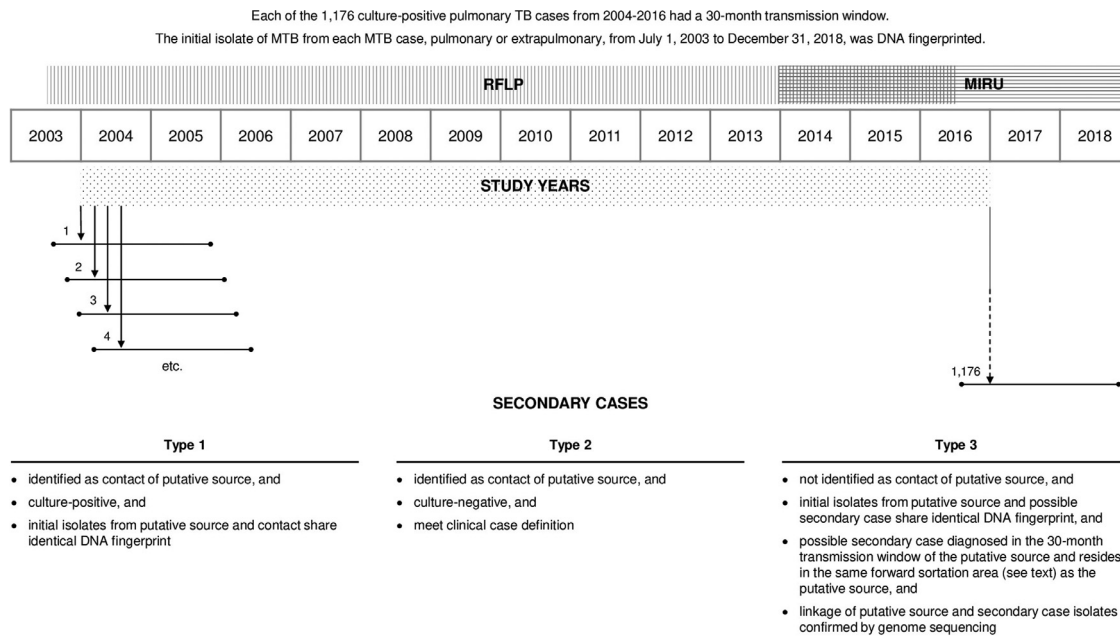
Once the contact list for each case was assembled, we matched contacts by name, date-of-birth, and their unique TB registry number to a list of all known cases of TB in the province that occurred 6 months before or 24 months after the diagnosis of the putative source case. The diagnosis date was defined as the start date of treatment. As previously reported<sup>28,29</sup> secondary cases were described as follows (see Fig. 1). “Type-1” secondary cases were contacts diagnosed in the 30 month window who had identical (100% matched) DNA fingerprints via RFLP or MIRU (if the putative source was diagnosed after June 2014). “Type-2” secondary

cases were contacts of the source case who were diagnosed with active TB within the 30-month transmission window but were culture-negative. This classification captures primarily paediatric cases from whom it may be difficult to obtain a specimen for culture. A paediatric chest radiologist independently verified that the paediatric Type-2 secondary cases met radiographic criteria for the diagnosis of primary pulmonary TB.<sup>30</sup> If a secondary case was diagnosed in the 6 months prior to the source case, it was only considered a transmission event if the secondary case had primary disease.

Otherwise, we ordered conventional epidemiologically linked cases with identical DNA fingerprints chronologically and assumed the first case was the source case. As in the DNA fingerprint clustering analysis, “secondary cases that were preceded only by cases of smear-negative TB were attributed to smear-negative transmission. All cases that occurred after any case of smear-positive TB were attributed to smear-positive transmission”.<sup>6</sup>

Despite rigorous contact tracing, some secondary cases may still not be identified. To identify secondary cases that had not been identified as contacts, we first identified all “non-contacts” in the 30-month transmission window that had an isolate with an identical DNA fingerprint. We then determined whether these “non-contacts” were spatially linked to the putative source – i.e. lived in the same forward sortation area, a geographic unit associated with a postal facility that is determined by the first three digits of their postal code. These were referred to as “Type-3” secondary cases. To further confirm that putative source and Type-3 secondary cases were linked, their isolates were sequenced and required to show no more than 10 SNVs differences.

We also undertook four sensitivity analyses. First, we recognized that the assumption that all cases occurring after a smear-positive case were attributable to the smear-positive source could, in some chains of transmission, lead to an underestimation of transmissions arising from smear-negative cases. For instance, underestimation could occur where the first case is smear-positive and the second and third cases are smear-negative and the third case is a contact of both the first and the second case. Therefore, we undertook a sensitivity analysis where cases that were linked to both a smear-positive and smear-negative case but occurred chronologically immediately after the smear-negative case were attributed to the smear-negative case. That is, in the sensitivity analysis, all cases that occurred after the smear-positive TB case were *not* attributed to the smear-positive individual. Next, we undertook a sensitivity analysis where we excluded all transmissions in contacts who had been diagnosed within 6 months of arrival to Canada. This was done to account for the possibility that the contact may have acquired their disease prior to coming to Canada. That is, their disease may have been acquired from a source who was not residing in Alberta. For the



**Figure 1.** A summary of the steps involved in measuring TB transmission by “Method #2”.

final sensitivity analysis, we only considered transmissions that arose in close contacts.

The STROBE checklist was reviewed and, where possible and relevant to our study design, we adhered to STROBE guideline recommendations.

### Study approval

The University of Alberta Health Research Ethics Board (HREB) Biomedical Panel for review of non-invasive studies involving humans provided approval for this study, protocol ID Pro00088408. Patient consent was not obtained as routinely collected data was used.

### Role of the funding source

The funding agency had no role in the study design, data collection, data analysis, data interpretation or writing of the report.

## Results

### Method #1: molecular epidemiology (DNA fingerprint cluster)

There were 1767 patients diagnosed with culture-positive TB (pulmonary or extra-pulmonary) between July 2003 and June 2016. From this cohort, 633 individuals were included in 151 unique clusters, ranging in size from 2 to 90. The cluster of 90 individuals was due to a

low-copy number and these isolates underwent spoligotyping. Upon spoligotyping, there were 29 unique transmission chains—all of which were initiated by a smear-positive case. 62% of clusters consisted of only 2 individuals. There were 372 secondary events. From these secondary events, 314 arose from smear-positive patients and 58 from smear-negative patients. That is, the proportion of cases arising from a smear-negative source was 16% (58/372), 95% CI, 12–19%. In this cohort, there were 1135 pulmonary TB cases; 554 were smear-negative and 581 were smear-positive. Therefore, the relative transmission rate was (58/554)/(314/581) = 0.19, 95% CI, 0.14–0.26 (see Table 1). These findings were unchanged when we restricted transmissions to a 2-year window.

### Method #2: combined conventional and molecular epidemiology supplemented with genome sequencing

We then looked for secondary cases amongst 23,131 contacts of 1176 adult pulmonary TB cases diagnosed from January 2004 to December 2016. The characteristics of smear-negative vs smear-positive patients are described in Table 2. Smear-negative patients were more likely to be foreign-born (84 vs 76%), less likely to present with cavitation (7% vs 44%), and had fewer total contacts ( $6 \pm 9$  vs  $32 \pm 64$ ) but the same number of paediatric close contacts under 5 years of age ( $0.5 \pm 1.5$  vs  $0.7 \pm 1.7$ ).

	Relative Transmission Rate	Proportion of Transmissions Attributable to Smear-Negative Source
Replication of DNA fingerprint clustering methods (Method #1)	0.19 (95% CI, 0.14–0.26)	16% (95% CI, 12–19%)
Molecular and conventional epidemiology supplemented with whole genome sequencing (Method #2)	0.10 (95% CI, 0.05–0.19)	8% (95% CI, 3–14%)
Sensitivity analysis #1 (attributed transmission to a smear-negative case even if the secondary case had also been exposed to a smear-positive individual)	0.13 (95% CI, 0.07–0.24)	11% (95% CI, 5–17%)
Sensitivity analysis #2 (excluded contacts who had only been in Canada less than 180 days)	0.08 (95% CI, 0.04–0.17)	7% (95% CI, 2–12%)
Behr et al. 1999	0.22	17%
Tostmann et al. 2008	0.24	13%
Hernandez-Garduno et al. 2004	n/a	17–41%

**Table 1: A comparison of the relative transmission rate and proportion of transmissions<sup>a</sup> attributable to smear-negative sources when replicating previous methodology versus our combined approach.**

<sup>a</sup> Test comparing Method #1 vs Method #2,  $p = 0.048$ . Global test comparing Method #1 vs Method #2 vs Sensitivity analysis #1 vs Sensitivity analysis #2,  $p = 0.038$ .

In addition to Type-1 and Type-2 secondary cases there were 23 putative sources with potentially 28 Type-3 secondary cases. Using core SNV analysis on these isolates, we were able to determine that 13 were consistent with transmission (see Fig. 2). From these 13 Type-3 secondary cases, one transmission was attributed to a smear-negative and 12 to a smear-positive source. Thus, once we included geographically and temporally linked DNA fingerprint clusters that were confirmed to be linked by core SNVs, the proportion of cases attributable to smear-negative cases remained at 8% (9/110), 95% CI, 3–14%,  $p = 0.05$ , and the relative transmission rate at 0.10 (9/563)/(101/613), 95% CI, 0.05–0.19.

In the first sensitivity analysis, out of the 110 transmissions which occurred, only three more cases would be re-classified to be attributable to smear-negative patients. The proportion of cases attributable to smear-negative patients would be 11% (12/110), 95% CI, 5–17%, and the relative transmission rate would be 0.13 (12/563)/(98/613), 95% CI, 7–24%. In the second sensitivity analysis, we excluded eight contacts who were diagnosed with TB within 6 months of arrival to Canada. All excluded cases would have constituted a Type-2 transmission event. Therefore, when considering transmission to known contacts, the proportion of cases attributable to smear-negative cases was further reduced

	Smear-Negative N = 563 (%)	Smear-Positive N = 613 (%)
Female	264 (47)	251 (41)
Age (years)	49 (SD 21)	48 (SD 20)
Cavitation (missing $n = 1$ from smear-negative; $n = 1$ from smear-positive)	41 (7)	270 (44)
HIV (missing $n = 4$ )	35 (6)	35 (6)
Relapse/re-treatment	47 (8)	59 (10)
Resistance to any first-line agents	71 (13)	69 (11)
Ethnicity <sup>a</sup>		
CB, non-Indigenous	32 (6)	48 (8)
Indigenous	62 (11)	105 (17)
Foreign-Born	469 (84)	460 (76)
Total number of contacts	3409	19,376
Total number of close contacts	2670	6153
Close contacts/case	4.7 (SD 6.7)	10 (SD 15)
	Median=4 (IQR 4)	Median=5 (IQR 8)
All contacts/case	6 (SD 9)	32 (64)
(Close contacts + casual contacts)/case	Median=4 (IQR 5)	Median=12 (IQR 25)
Number of close paediatric contacts (<5 years of age)	0.5 (1.5)	0.7 (1.7)

**Table 2: Characteristics of the 1176 patients 14 years of age or older with culture-positive pulmonary TB.**

<sup>a</sup> CB=Canadian-born; Indigenous refers to First Nations, Métis, or Inuit peoples according to the Constitution Act of 1982.

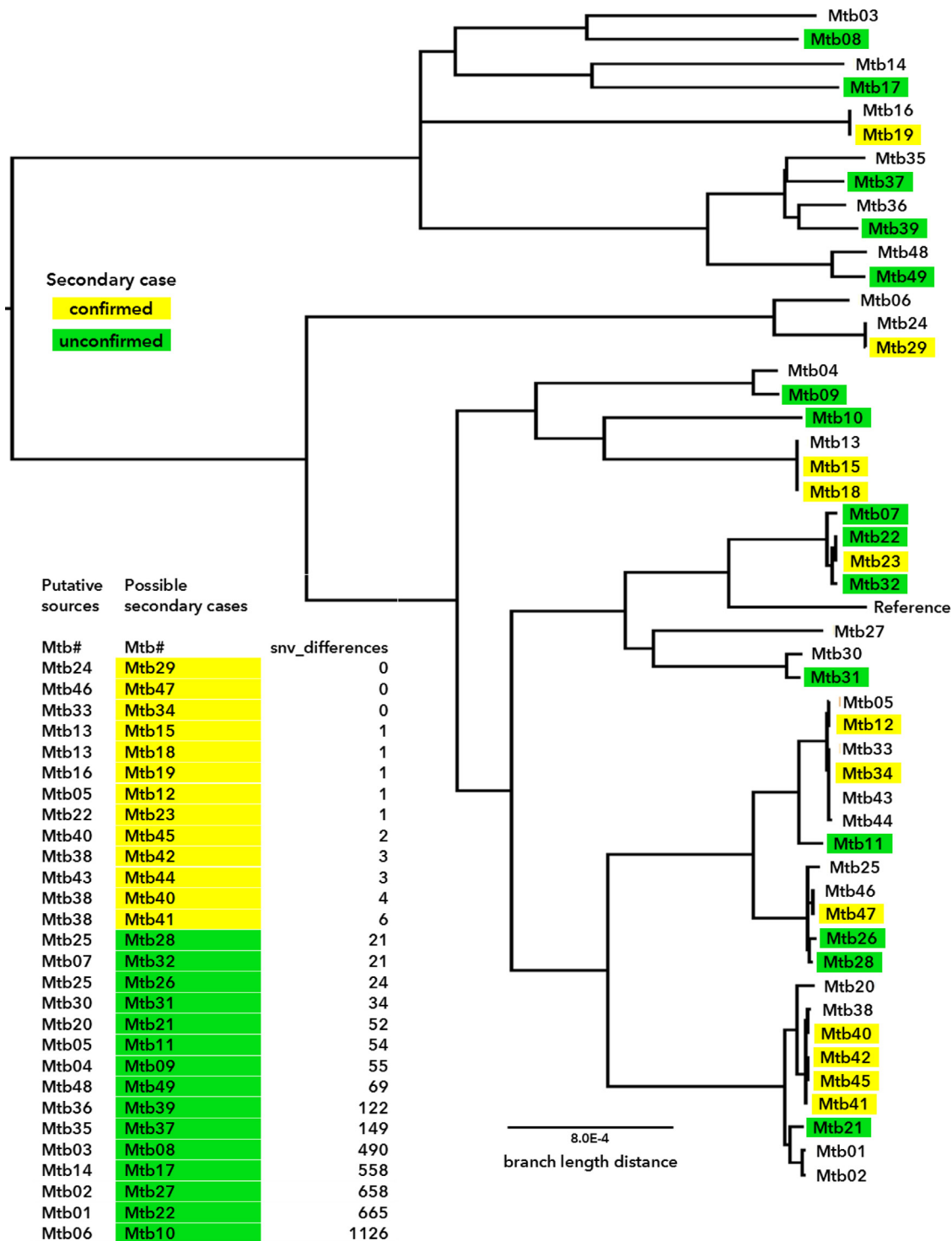


Figure 2. Phylogenetic tree from core SNVs of 49 *M. tuberculosis* isolates.

to 7% (7/102), 95% CI, 2–12%, and the relative transmission rate decreased to 0.08 (7/563)/(95/613), 95% CI, 0.04–0.17. There was a statistically significant difference in the global analysis when comparing

proportions of transmissions attributable to smear-negative sources as determined by the various methods and sensitivity analyses (Method #1, Method #2, and the two sensitivity analyses) (Table 1). When looking only at

	Smear-Negative	Smear-Positive
<b>Type-1 (N = 64)</b>	<b>N = 4 (%)</b>	<b>N = 60 (%)</b>
Days to diagnosis of contact (mean, SD)	243 (185)	119 (143)
Close contact	3 (75)	52 (87)
Household contact	2 (50)	35 (58)
Average age (SD)	33 (10)	33 (19)
Median age (IQR)	31 (12)	30 (11)
<5 years old (all contacts)	0	2 (3)
Between 5–14 years old	0	8 (13)
HIV (+)	1 (25)	3 (5)
Foreign-born	2 (50)	18 (30)
Positive airway secretion smear-status	1 (25)	16 (27)
Diagnosed within 6 months of arrival to Canada	0	0
<b>Type-2 (N = 33)</b>	<b>N = 4 (%)</b>	<b>N = 29 (%)</b>
Days to diagnosis of contact (mean, SD)	33 (30)	47 (49)
Close contact	4 (100)	36 (100)
Household contact	4 (100)	23 (79)
Average age (SD)	20 (26.4)	8 (9.5)
Median age (IQR)	7 (22)	4 (9)
<5 years (all contacts)	2 (50)	16 (55)
Between 5–14 years old	1 (20)	9 (31)
HIV (+)	0	0
Foreign-born contact	4 (100)	9 (31)
Diagnosed within 6 months of arrival to Canada	2 (50)	6 (21)
<b>Type-3 (N = 13)</b>	<b>N = 1 (%)</b>	<b>N = 12 (%)</b>
Days to diagnosis of contact (mean, SD)	279	293 (200)
Average age (SD)	55	40 (10)
Median age (IQR)		39 (11)
<5 years	0	0
Between 5–14 years old	0	0
HIV (+)	0	0
Foreign-born contact	0	3 (25)
Diagnosed within 6 months of arrival to Canada	0	0
Positive airway secretion smear-status	1 (100)	6 (50)

**Table 3: Characteristics of secondary cases, according to source case smear-status.**

Alberta's prospective, routine contact tracing consists of the gathering of information about the number, type (close or casual), tuberculin skin test (TST), and disease status of contacts for all pulmonary TB cases. We defined close and casual contacts as per the Canadian TB Standards. Assessment of contacts included a symptom enquiry and TST 8–12 weeks after the final contact with the source case (if the contact was not already TST positive), a chest radiograph if symptomatic or TST positive, and sputum for acid-fast bacilli smear and culture if symptomatic or if chest radiograph was abnormal.

transmissions to close contacts, the relative transmission rate was  $[8/563]/[75/613]=0.12$  (95% CI 0.06–0.25).

The characteristics of all secondary cases are described in [Table 3](#).

### Discussion

Using the DNA fingerprint clustering technique reported in three separate studies,<sup>5–7</sup> we found that the proportion of TB transmitted from smear-negative patients was 16% and the relative transmission rate was 0.19, a finding in line with these previous studies. However, when we combined both molecular and conventional epidemiology and supplemented it with core SNV

analysis, we found that smear-negative cases were ~50% less infectious than previously thought.

When Tostmann et al. looked at transmission based on epidemiologic data alone, only 6% (26/417) of source cases were smear-negative.<sup>7</sup> They attributed the discrepancy in their findings to epidemiologic links failing to reveal all contacts. While relying only on epidemiologic data may miss certain transmission events, it is also possible that DNA fingerprint clustering techniques in the absence of epidemiologic data techniques overestimate transmission events.<sup>9,10</sup>

In most settings, accounting for TB transmission is very challenging.<sup>2</sup> However, our resource-rich, low-incidence setting provides an ideal real-world condition for studying transmission biology. First, there is a high



proportion of foreign-born patients. These individuals are often re-activating imported strains of *M. tb* rather than acquiring new strains. This means there is a higher diversity of strains, fewer large clusters, and an enhanced ability to discern transmissions. In our cohort, only 35% of cases had clustered DNA fingerprints and 85% of the clusters included five or fewer cases. Furthermore, we routinely implement rigorous contact tracing and incorporate social network analysis as necessary. The comprehensiveness of our routine contact tracing enhances our confidence in the assessment of transmissions.

Our study also includes Type-2 secondary cases (culture-negative active TB diagnoses). These patients were diagnosed based on a clinical case definition that included radiographic findings or rarely, sputum smear-positive but culture-negative instances. Only 5/33 of the Type-2 secondary cases were adult-type cases (over 14 years old). To our knowledge, no other study comparing transmissions from smear-negative to smear-positive patients has included such secondary cases. Because young children are highly vulnerable and therefore may serve as sensitive sentinels of transmission even when the bacillary burden of the source case is low, we believe their inclusion is crucial. In fact, the tendency to rely on only molecular or even genomic techniques to determine transmission may lead to a neglect of paediatric TB in transmission analyses.

Despite the strengths of our study, there are also several limitations. First, smear-negative source cases have fewer close and casual contacts. Based on the Canadian TB standards,<sup>3</sup> our protocols dictate that contact tracing of a smear-negative source can begin with only household contacts whereas for smear-positive sources, all close contacts (including non-household) are included in the first round of contact tracing. This likely accounts for the large difference in close-contacts for smear-negatives ( $4.7 \pm 6.7$ ) versus smear-positives ( $10 \pm 15$ ). While this can introduce bias, it is mitigated by the fact that if we identify tuberculin skin test (TST) conversions or secondary cases in the household, contact tracing is expanded to non-household close contacts and beyond. Furthermore, if it is assumed that contacts under the age of 5-years-old are close household contacts (which unfortunately are not captured separately from other close contacts), then close household contacts are similar for smear-negatives ( $0.5 \pm 1.5$ ) vs ( $0.7 \pm 1.7$ ). Also, given that 91% of the transmissions within the whole cohort occurred amongst close contacts and 63% amongst household contacts, the first “concentric circle” of household contacts should be identifying most cases and triggering further contact tracing, as necessary. Finally, the inclusion of the temporally and geographically linked cases that were confirmed by core SNV analysis should help to identify transmissions amongst non-identified contacts. When we conducted a sensitivity analysis which only considered transmission arising in

close contacts, the relative transmission rate was 0.12 (95% CI 0.06–0.25). However, we think that considering only close contacts is a less accurate representation of transmission risk posed by the two groups because an important feature of smear-positivity is precisely the fact that there is the potential for “superspreading” to casual (or distant) contacts. Given the outcome measure we chose to examine, which was selected out of a desire to easily compare our findings to previous studies, we could not adjust for number of contacts or clustering. Future studies should look at the odds of transmission arising from a patient with smear-negative disease as their main outcome, and consider accounting for household clustering, symptoms, and other clinical parameters. Importantly, though, even in such a study, the outcome (number of transmissions) would still rely on the contact-tracing methodology.

Another limitation is that for the combined analysis “Method 2”, we defined a transmission event as having occurred within a 30-month transmission window. While there may be cases of very delayed disease onset,<sup>31</sup> our method has been previously published and described<sup>28,32,33</sup> with sensitivity analyses corroborating its appropriateness. Findings from other studies<sup>31,34,35</sup> also suggest that this length of time is likely to capture the vast majority of transmission events. It is also possible that some contacts may have left the province and developed TB outside of Alberta or Canada; unfortunately, there are no national registries to allow us to consider secondary cases reported outside of the provincial jurisdiction. However, there is no reason to think that there should be a differential bias in migration out of Alberta between contacts of patients with smear-negative vs smear-positive disease. In addition, because 38 cases occurred in contacts who were foreign-born, it is possible that their TB exposure and thus the true transmission event occurred outside of Canada. However, when we restricted analyses to only include those who developed TB after 6 months since arrival to Canada, this only served to further reduce the relative contribution of cases from smear-negative sources.

Finally, while we are the first group to use core SNV analysis in ascertaining transmissions attributable to smear-negative patients, given the known higher resolution of genome sequencing over DNA fingerprinting, we would ideally, financial constraints aside, have sequenced all clusters. Nevertheless, there remains a high degree of certainty around the validity of the transmission when there is both a clear epidemiologic link and 100% matched DNA fingerprint by conventional methods.<sup>36</sup>

We recognize the growing body of literature exploring the role of subclinical (asymptomatic) disease in transmission of tuberculosis<sup>37,38</sup> and the evolving understanding of the dynamic nature of TB disease progression.<sup>38</sup> This could mean that we may be systematically underestimating the contribution of subclinical

cases,<sup>37</sup> the vast majority of which are smear-negative.<sup>39</sup> Mechanistically, if you consider the significant aerosol production arising from routine activities like speaking or singing, cough would not be required for transmission.<sup>40</sup> While we believe our findings are an accurate reflection of transmission ascertainment reliant on routine contact tracing practice and while we attempt to account for unidentified secondary cases, the evolving literature around the role of transmission from subclinical cases shows the challenge in attributing precise transmission rates based on smear-status. Developing the most accurate picture of transmission and attack rates requires the use of a variety of methodologies ranging from mathematical modelling to conventional epidemiology reliant on contact tracing.

The aim of this study was to replicate previous techniques while enhancing the previously used molecular clustering technique with conventional epidemiology and genomics. Our findings, therefore, are an update of these previous techniques. Based on this update, we believe that smear-negative cases result in a lower burden of transmissions than previously estimated. And our exploration of different methodologies of ascertaining transmission highlights the continued importance of conventional field epidemiology and the importance of revisiting questions about TB biology with the assistance of new tools. By showing that smear-negative cases contribute to fewer transmissions than previously thought, our findings can update estimates used in TB disease modelling and may assist in triaging of resources in increasingly resource constrained public health programs.

### Contributors

LA was involved in all parts of the study conceptualization and design, data collection and analysis, data interpretation, writing and editing, funding acquisition and the decision to submit for review. RL supervised the project and contributed to conceptualization, funding acquisition, methodology and writing and was involved in the decision to submit for review. MD, CP, and ME contributed to data curation and validation. MC was involved in data collection and analysis, methodology, and writing. CH contributed to data curation, project administration and writing. GT was involved in data collection and analysis, methodology, and writing. AD contributed to formal analysis, methodology and writing. RL, LA, and CH were responsible for the raw data associated with this study.

### Funding

This project was funded by Alberta Innovates Health Solutions.

### Data sharing statement

Anonymized, de-identified data will be available upon written request and provision of a detailed statistical analysis plan to the authors.

### Declaration of Competing Interest

We declare no competing interests.

### Acknowledgements

We are very grateful to the staff of the TB Prevention and Care Program of Alberta and the Provincial Laboratory for Public Health, Alberta Health Service, for their assistance with data collection. We are also grateful to the staff at the TB Program Evaluation & Research Unit for their assistance in preparing the manuscript.

### References

- 1 World Health Organization (WHO). Global tuberculosis report 2020. Geneva: WHO; 2020. Available from: <https://apps.who.int/iris/bitstream/handle/10665/336069/9789240013131-eng.pdf> (accessed: July 9, 2021).
- 2 Auld SC, Kasmar AG, Dowdy DW, et al. Research roadmap for tuberculosis transmission science: where do we go from here and how will we know when we're there? *J Infect Dis* 2017;216(suppl\_6): S6628.
- 3 Canadian Lung Association/Canadian Thoracic Society and Public Health Agency of Canada (PHAC). Canadian tuberculosis standards. 7th ed. Ottawa: PHAC; 2014. Available from: <https://www.canada.ca/en/public-health/services/infectious-diseases/canadian-tuberculosis-standards-7th-edition.html> (accessed: July 5, 2021).
- 4 Migliori GB, Nardell E, Yedilbayev A, et al. Reducing tuberculosis transmission: a consensus document from the world health organization regional office for Europe. *Eur Respir J* 2019;53(6):1900391.
- 5 Hernández-Garduño E, Cook V, Kunimoto D, Elwood RK, Black WA, FitzGerald JM. Transmission of tuberculosis from smear negative patients: a molecular epidemiology study. *Thorax* 2004;59(4):286–90.
- 6 Behr MA, Warren SA, Salamon H, et al. Transmission of Mycobacterium tuberculosis from patients smear-negative for acid-fast bacilli. *Lancet* 1999;353(9151):444–9.
- 7 Tostmann A, Kik SV, Kalisvaart NA, et al. Tuberculosis transmission by patients with smear-negative pulmonary tuberculosis in a large cohort in the Netherlands. *Clin Infect Dis* 2008;47(9):1135–42.
- 8 Small PM, Hopewell PC, Singh SP, et al. The epidemiology of tuberculosis in San Francisco. A population-based study using conventional and molecular methods. *N Engl J Med* 1994;330(24):1703–9.
- 9 Lalor MK, Anderson LF, Hamblion EL, 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;15(1):105.
- 10 Guthrie JL, Kong C, Roth D, et al. Molecular epidemiology of tuberculosis in British Columbia, Canada: a 10-year retrospective study. *Clin Infect Dis* 2018;66(6):849–56.
- 11 France AM, Grant J, Kammerer JS, Navin TR. A field-validated approach using surveillance and genotyping data to estimate tuberculosis attributable to recent transmission in the United States. *Am J Epidemiol* 2015;182(9):799–807.
- 12 Walker TM, Ip CL, Harrell RH, et al. Whole-genome sequencing to delineate mycobacterium tuberculosis outbreaks: a retrospective observational study. *Lancet Infect Dis* 2013;13(2):137–46.
- 13 Roetzer A, Diel R, Kohl TA, et al. Whole genome sequencing versus traditional genotyping for investigation of a mycobacterium tuberculosis outbreak: a longitudinal molecular epidemiological study. *PLoS Med* 2013;10(2):e1001387.
- 14 Gurjav U, Outhred AC, Jelfs P, et al. Whole genome sequencing demonstrates limited transmission within identified mycobacterium tuberculosis clusters in New South Wales, Australia. *PLoS ONE* 2016;11(10):e0163612.
- 15 Vachon J, Gallant V, Siu W. Tuberculosis in Canada, 2016. *Can Commun Dis Rep* 2018;44(3–4):75–81.
- 16 Dale JW, Brittain D, Cataldi AA, et al. Spacer oligonucleotide typing of bacteria of the mycobacterium tuberculosis complex: recommendations for standardised nomenclature. *Int J Tuberc Lung Dis* 2001;5(3):216–9.
- 17 van Embden JD, Cave MD, Crawford JT, et al. Strain identification of mycobacterium tuberculosis by DNA fingerprinting:

- recommendations for a standardized methodology. *J Clin Microbiol* 1993;31(2):406–9.
- 18 Supply P, Allix C, Lesjean S, et al. Proposal for standardization of optimized mycobacterial interspersed repetitive unit-variable-number tandem repeat typing of mycobacterium tuberculosis. *J Clin Microbiol* 2006;44(12):4498–510.
  - 19 Ewels P, Magnusson M, Lundin S, Källner M. MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics* 2016;32(19):3047–8.
  - 20 Wood DE, Lu J, Langmead B. Improved metagenomic analysis with Kraken 2. *Genome Biol* 2019;20(1):257.
  - 21 Labbé G, Kruczkiewicz P, Robertson J, Mabon P, Schonfeld J, Kein D, Rankin MA, Gopez M, Hole D, Son D, Knox N, Laing CR, Bessonov K, Taboada EN, Yoshida C, Ziebell K, Nichani A, Johnson RP, Van Domselaar G, Nash JHE. Rapid and accurate SNP genotyping of clonal bacterial pathogens with BioHansel. *Microb Genom*. 2021 Sep;7(9). <https://doi.org/10.1099/mgen.0.000651>.
  - 22 Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol* 2015;32(1):268–74.
  - 23 Minh BQ, Nguyen MA, von Haeseler A. Ultrafast approximation for phylogenetic bootstrap. *Mol Biol Evol* 2013;30(5):1188–95.
  - 24 Hoang DT, Chernomor O, von Haeseler A, Minh BQ, Vinh LS. UFBoot2: improving the ultrafast bootstrap approximation. *Mol Biol Evol* 2018;35(2):518–22.
  - 25 Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol* 2010;59(3):307–21.
  - 26 Page AJ, Taylor B, Delaney AJ, et al. SNP-sites: rapid efficient extraction of SNPs from multi-FASTA alignments. *Microb Genom* 2016;2(4):e000056.
  - 27 Cook VJ, Sun SJ, Tapia J, et al. Transmission network analysis in tuberculosis contact investigations. *J Infect Dis* 2007;196(10):1517–27.
  - 28 Asadi L, Heffernan C, Menzies D, Long R. Effectiveness of Canada's tuberculosis surveillance strategy in identifying immigrants at risk of developing and transmitting tuberculosis: a population-based retrospective cohort study. *Lancet Public Health* 2017;2(10):e450–7.
  - 29 Long R, Asadi L, Heffernan C, et al. Is there a fundamental flaw in Canada's post-arrival immigrant surveillance system for tuberculosis? *PLoS ONE* 2019;14(3):e0212706.
  - 30 Dhawan V, Bown J, Lau A, et al. Towards the elimination of paediatric tuberculosis in high-income, immigrant-receiving countries: a 25-year conventional and molecular epidemiological case study. *ERJ Open Res* 2018;4(2):00131–2017.
  - 31 Reichler MR, Khan A, Sterling TR, et al. Risk and timing of tuberculosis among close contacts of persons with infectious tuberculosis. *J Infect Dis* 2018;218(6):1000–8.
  - 32 Parhar A, Gao Z, Heffernan C, Ahmed R, Egedahl ML, Long R. Is early tuberculosis death associated with increased tuberculosis transmission? *PLoS ONE* 2015;10(1):e0117036.
  - 33 Lau A, Barrie J, Winter C, Elamy AH, Tyrrell G, Long R. Chest radiographic patterns and the transmission of tuberculosis: implications for automated systems. *PLoS ONE* 2016;11(4):e0154032.
  - 34 Ferebee SH. Controlled chemoprophylaxis trials in tuberculosis. A general review. *Bibl Tuberc* 1970;26:28–106.
  - 35 Sloot R, Schim van der Loeff MF, Kouw PM, Borgdorff MW. Risk of tuberculosis after recent exposure. A 10-year follow-up study of contacts in Amsterdam. *Am J Respir Crit Care Med* 2014;190(9):1044–52.
  - 36 Guthrie JL, Delli Pizzi A, Roth D, et al. Genotyping and whole-genome sequencing to identify tuberculosis transmission to pediatric patients in British Columbia, Canada, 2005–2014. *J Infect Dis* 2018;218(7):1155–63.
  - 37 Kendall EA, Shrestha S, Dowdy DW. The epidemiological importance of subclinical tuberculosis. A critical reappraisal. *Am J Respir Crit Care Med* 2021;203(2):168–74. <https://doi.org/10.1164/rccm.202006-2394PP>. Jan.
  - 38 Drain PK, Bajema KL, Dowdy D, Dheda K, Naidoo K, Schumacher SG, Ma S, Meermeier E, Lewinsohn DM, Sherman DR. Incipient and subclinical tuberculosis: a clinical review of early stages and progression of infection. *Clin Microbiol Rev* 2018;31(4). <https://doi.org/10.1128/CMR.00021-18>. Jul 18e00021-18.
  - 39 Min J, Chung C, Jung SS, Park HK, Lee SS, Lee KM. Clinical profiles of subclinical disease among pulmonary tuberculosis patients: a prospective cohort study in South Korea. *BMC Pulm Med* 2020;20(1):316. <https://doi.org/10.1186/s12890-020-01351-z>. Dec 2.
  - 40 Wilson NM, Marks GB, Eckhardt A, Clarke AM, Young FP, Garden FL, Stewart W, Cook TM, Tovey ER. The effect of respiratory activity, non-invasive respiratory support and facemasks on aerosol generation and its relevance to COVID-19. *Anaesthesia* 2021;76(11):1465–74. <https://doi.org/10.1111/anae.15475>. Mar 30;10.1111/anae.15475.