ORIGINAL ARTICLE

Who Transmits Tuberculosis to Whom
A Cross-Sectional Analysis of a Cohort Study in Lima, Peru

Letizia Trevisi[1], Meredith B. Brooks[2], Mercedes C. Becerra[1,3], Roger I. Calderón[4], Carmen C. Contreras[4,5], Jerome T. Galea[1,6], Judith Jimenez[4], Leonid Lecca[1,4], Rosa M. Yataco[4], Ximena Tovar[1], Zibiao Zhang[3], Megan B. Murray[1,3,7], and Chuan-Chin Huang[1,3,7]

[1]Department of Global Health and Social Medicine, Harvard Medical School, Boston, Massachusetts; [2]Department of Global Health, Boston University School of Public Health, Boston, Massachusetts; [3]Division of Global Health Equity, Brigham and Women's Hospital, Boston, Massachusetts; [4]Socios en Salud, Lima, Peru; [5]Harvard Global Health Institute, Cambridge, Massachusetts; [6]College of Behavioral and Community Sciences, School of Social Work, University of South Florida, Tampa, Florida; and [7]Center for Communicable Disease Dynamics, Department of Epidemiology, Harvard T. H. Chan School of Public Health, Boston, Massachusetts

ORCID ID: 0000-0002-8775-3755 (C-C.H.).

Abstract

Rationale: The persistent burden of tuberculosis (TB) disease emphasizes the need to identify individuals with TB for treatment and those at a high risk of incident TB for prevention. Targeting interventions toward those at high risk of developing and transmitting TB is a public health priority.

Objectives: We aimed to identify characteristics of individuals involved in TB transmission in a community setting, which may guide the prioritization of targeted interventions.

Methods: We collected clinical and sociodemographic data from a cohort of patients with TB in Lima, Peru. We used whole-genome sequencing data to assess the genetic distance between all possible pairs of patients; we considered pairs to be the result of a direct transmission event if they differed by three or fewer SNPs, and we assumed that the first diagnosed patient in a pair was the transmitter and the second was the recipient. We used logistic regression to examine the association between host factors and the likelihood of direct TB transmission.

Measurements and Main Results: Analyzing data from 2,518 index patients with TB, we identified 1,447 direct transmission pairs. Regardless of recipient attributes, individuals less than 34 years old, males, and those with a history of incarceration had a higher likelihood of being transmitters in direct transmission pairs. Direct transmission was more likely when both patients were drinkers or smokers.

Conclusions: This study identifies men, young adults, former prisoners, alcohol consumers, and smokers as priority groups for targeted interventions. Innovative strategies are needed to extend TB screening to social groups such as young adults and prisoners with limited access to routine preventive care.

Keywords: pulmonary tuberculosis; transmission; whole-genome sequencing; interventions

At a Glance Commentary

Scientific Knowledge on the Subject: The medical and research community is actively exploring the best strategies to identify and treat individuals with tuberculosis (TB) disease. Genomic studies have aided in identifying the primary source of TB infection by examining connections between patients in transmission clusters. However, existing research reveals a gap in current TB strategies regarding the effective detection and tracking of transmission sources, as well as understanding how assortative mixing patterns may influence transmission dynamics.

What This Study Adds to the Field: In this study, we propose a novel approach for evaluating factors that influence TB transmission dynamics within a community setting, while also considering assortative mixing patterns. By integrating data from a cohort of more than 2,500 patients with TB who were systematically enrolled and whole-genome sequencing data of *Mycobacterium* tuberculosis, we found that age, sex, tobacco use, alcohol use, and a history of incarceration may modify the TB transmission dynamics at the population level. Moreover, we also revealed that assortative mixing patterns play an important role in TB transmission dynamics. By identifying the characteristics of individuals involved in TB transmission, we can gain insights into who should be prioritized for targeted interventions.

The World Health Organization estimates that 10.6 million people developed tuberculosis (TB) disease in 2021 and that 1.6 million people died of the disease, making TB the single infectious disease that killed the most people in that year after COVID-19 (1). These rates halted a consistent downward trend in TB incidence and death between 2005 and 2020 and returned the TB death rate back to the level observed in 2017 (1). Much of this rise has been attributed to the disruptions in TB programs during the COVID-19 pandemic; in particular, it was challenging to identify and treat people with TB disease during this period and thereby interrupt ongoing transmission of *Mycobacterium tuberculosis* (Mtb), the causative agent of the disease (2). This struggle has amplified the need for active case-finding strategies that target those who will most benefit from these interventions.

Targeted approaches to TB control include providing preventive therapy for individuals with latent TB and active case finding and treatment of those with TB disease (3). Previous screening strategies have often focused on hotspots where TB transmission takes place, using spatially targeted approaches to implement interventions in geographical areas of high transmission (4). An alternate strategy is to focus interventions on those most likely to develop and transmit TB disease. Most often, subgroups targeted for interventions are those with specific risk factors for TB infection and disease, such as HIV infection (5) or diabetes mellitus (DM) (6), but this approach does not specifically focus on those most likely to "transmit" or "receive" the Mtb bacilli. To do that, it is necessary to identify the characteristics of pairs of people involved in transmission events (7).

In previous work, we have used whole-genome sequencing (WGS) of Mtb isolates to study spatial patterns of TB transmission in a well-defined catchment area of Lima, Peru. In that study, we found that area-specific TB prevalence did not closely correlate with estimates of local transmission derived from genomic epidemiology (8). This finding suggests that the prevalence of TB disease, which can result from recent transmission or reactivation of an earlier infection, in a specific risk group may not be the best proxy for the risk of transmission. Because of the challenges of directly measuring TB exposure and transmission in TB high-burden settings, we used proxy indicators to identify direct transmission events. This required combining information on Mtb genetic distance and time of diagnosis. Identification of these individual attributes may inform who should be prioritized for targeted interventions.

Methods

Study Setting and Population
Between September 2009 and August 2012, we recruited patients with TB and their household contacts as part of a longitudinal cohort study in Lima, Peru. The study design, methods, and setting have been described previously in detail (9). The study took place in a catchment area of 20 districts within urban Lima, where nearly 3.3 million people live. In brief, we identified index patients with pulmonary TB aged 16 or older diagnosed at any of the 106 participating health centers that diagnose and treat TB and provide routine health care. These centers follow national guidelines to diagnose patients with pulmonary TB, which requires at least one positive sputum smear or a chest radiograph consistent with TB. We invited patients with pulmonary TB who could give informed consent to participate in the study. We ascertained TB disease by performing routine diagnostic microbiology, including sputum smear microscopy and mycobacterial culture. Culture-positive sputum samples underwent drug susceptibility testing and genotyping by mycobacterial interspersed repetitive units. We collected the following clinical and sociodemographic information from index patients: age, sex assigned at birth, height, weight, education level, socioeconomic status (SES) (10), alcohol and tobacco use, housing information, TB symptoms, history of TB disease, bacillus Calmette-Guérin vaccination history, frequency of public transportation use, work outside the home, incarceration status, and comorbidities including HIV and DM. We collected blood samples for host genetic analysis, which we conducted using a customized array (LIMAArray) with 712,200 markers (11). We used these data to estimate the proportion of Peruvian ancestry for each participant, and we then categorized the Peruvian genetic ancestry variable into a dichotomous variable using the mean proportion as the cutoff (12). We defined patients with TB as having preexisting DM if they were previously diagnosed with DM or were being treated with hypoglycemic drugs (13–15).

WGS
We obtained whole-genome sequences from a subset of Mtb isolates collected at the time of diagnosis using the Illumina HiSeq system in paired-end mode with a read length of 100–150 base pairs and mean coverage of at least 30-fold. Following standard procedures, we mapped the paired-end raw sequencing data to the H37Rv reference genome using the BWA-MEM algorithm (16, 17) and used SAMtools and Pilon to identify the single nucleotide polymorphisms (SNPs) and

insertions and deletions using a coverage-based approach (17, 18). Isolates with <95% genome mapping coverage were excluded. We assigned a nucleotide call of an isolate as missing if the valid depth of coverage at a specific site was less than 10% of mean coverage of this isolate, if the mean read mapping quality at the site did not reach 10, or if the alternative alleles occupied less than 85% of the valid coverage of the site. We excluded the variants in the proline-glutamic acid/proline-proline-glutamic acid family because the error rate of short-read sequencing methodology is high in highly repetitive regions (19). We also excluded the isolates with evidence of mixed strain infection using the barcode method (20).

Characterization of Pairwise Patients

Mtb genetic distance. Given the potential differences in transmission dynamics within households compared with community settings, our analysis focused exclusively on index patients with TB. Household contacts who were diagnosed with TB at enrollment or developed incident TB during the follow-up were excluded from our analyses. We calculated the Mtb genetic distance between the Mtb isolates from all possible pairs of patients with TB in our dataset by estimating the number of SNPs that differentiated the two Mtb isolates cultured from each patient pair. In the parent cohort study, we identified 78 within household, and more than 93% of the within-household direct transmission patient pairs had a genetic distance less than or equal to three SNPs (9). Therefore, we considered paired patients to have been in a direct transmission link (one-step direct transmission link) if the genetic distance between their Mtb strains was less than or equal to three SNPs. We also assumed that the person diagnosed first in a pair—indicating an earlier diagnosis than the second—was the one transmitting the infection, and the second person was considered to be the one receiving it. Henceforth, we refer to these pairs as direct transmission or DT pairs. We note that transmitters and receivers are dynamic states, so a patient identified as a recipient in one pair may serve as a transmitter in another pair. The binary variable indicating whether a transmission pair had a genetic distance less than or equal to three SNPs ($Y = 1$) or not ($Y = 0$) will serve as the primary outcome for the analysis.

Host factors. We characterized each pair on the basis of baseline variables that may modify an individual's contact pattern, susceptibility to TB infection or disease progression, or transmissibility after the development of active TB. These variables include demographic variables such as age, sex, or native Peruvian ancestry; social variables such as employment outside the home, frequency of public transportation use, SES, history of incarceration, and alcohol and tobacco use; and biological variables such as HIV infection and DM. Within a pair, we defined the patient with the earliest date of diagnosis as the transmitter (denoted by [T]) and the patient with the later date of diagnosis as the receiver (denoted by [R]). For example, for pairs with one male patient and one female patient, we further categorized them into two categories—male transmitter/female receiver pairs (denoted by Male[T]→Female[R]) and female transmitter/male receiver pairs (denoted by Female[T]→Male[R]). We considered the possibility that the first diagnosed case may not always be the transmitter, so we excluded pairs with diagnosis dates less than 60 days apart.

A three-category individual-level variable has nine possible configurations (three types of assortative pairing and six types of disassortative pairing [$P(3,2) = 9$]), which may complicate the interpretation of findings. Therefore, we categorized all individual-level host factors into two-category variables. For each pair, we considered four possible configurations: two types of assortative pairing in which both members or neither member of a pair had a specific attribute and two types of disassortative pairing in which either the transmitter or the receiver had an attribute that was not shared by the other member of a pair. The variables affected by dichotomization in this study include age, frequency of public transportation, SES, smoking, alcohol use, and coughing. Detailed information on how we categorized the individual-level variables are shown in the METHODS section of the online supplement. We then tested the hypothesis that these attribute configurations predicted a direct transmission link.

Regression Analyses

Demographic, social, and biological variables. We assessed the contribution of the characteristics of all possible pairs in our dataset to the outcome of direct transmission using univariate and multivariate logistic regression. Covariate configurations with a global likelihood ratio test *P* value ≤0.1 in the univariate analyses entered a backward stepwise algorithm to construct the multivariate models. We retained covariates with a *P* value <0.1 in the multivariate model. The detailed information on how to interpret the odds ratios of the regression models is shown in Table E1 in the online supplement. To evaluate whether our findings were sensitive to the choice of cutoff in determining the direct transmission pairs, we performed two sensitivity analyses. In the first sensitivity analysis, we repeated the analyses using a two-SNP difference as the cutoff in determining a direct transmission pair. In the second sensitivity analysis, instead of using a cutoff, we used fractional logistic regression, assigning pairs with a zero- to six-SNP difference a probability of being DT pairs ($Y = 1$ for zero or one SNP, $Y = 0.5$ for two or three SNPs, and $Y = 0.1$ for four to six SNPs), based on a distribution of 78 index/household-contact patient pairs with a within-household transmission (Figure E1). We also considered that some of the defined direct transmission pairs may not represent true direct transmission pairs, because patients with TB may be considered as recipients in multiple pairs, which could result in misclassification. To address this, we performed an additional sensitivity analysis in which we retained the pairs with the shortest genetic distance when these pairs had the same recipients.

Clinical variables. We evaluated whether our analytic framework can also detect the well-known clinical risk factors that modify the TB transmission dynamics. Although TB-related clinical variables measured at the time a transmitter was diagnosed with TB may indicate his transmissibility, the clinical variables of his recipient collected at the time when the recipient was diagnosed with TB should have no impact on the transmission event between a transmitter-recipient pair, because the transmission event occurred when the recipient was TB-free. Therefore, we considered only the direction of transmission and the characteristics of transmitters (but not recipients) in characterizing transmission pairs. For example, for a binary cavitary TB variable, we categorized pairs into two groups: "Yes[T]" (transmitter) to any recipients and "No[T]" to any recipients.

Pairs with missing data were excluded from the analysis. We chose this complete data analysis approach because less than 13% of pairs were excluded in the multivariable model due to missing data. All analyses were conducted using RStudio and SAS version 9.4 (SAS Institute).

Research Ethics

Before study participation, all study participants provided voluntary, written informed consent. The Harvard School of Public Health and Peru's Research Ethics Committee of the National Institutes of Health provided institutional review board approval.

Results

Study Samples

Throughout the study period, 5,266 patients with TB ≥16 years old were newly diagnosed with TB. Among the 4,500 index patients with TB enrolled in the study, 3,427 were culture-positive, and 2,573 underwent WGS. After excluding 55 isolates with evidence of mixed infection or poor raw read quality, we included 2,518 index patients with TB, among whom there were 3,168,903 possible pairs (Figures 1 and 2). After excluding 538,793 pairs in which patients within the same pair were diagnosed with TB less than 60 days

apart, we retained 2,630,110 pairs for the data analyses. Of these, 1,447 (0.06%) pairs had a genetic distance of three or fewer SNPs and were therefore considered to be DT pairs (Table 1 and Figures 1 and E2). Among these DT pairs, we observed a median time of TB transmission of 6.9 months, with an interquartile range of 6.9 months. Out of the 2,518 index patients with TB, we identified 589 (23.4%) individuals as TB transmitters.

Regression Analyses

Demographic, biological, and clinical variables. In the univariate analyses, we found that age, sex, frequency of using public transportation, SES, incarceration history, alcohol use, tobacco use, and DM status of members of pairs were associated with the likelihood of DT (Figure 3 and Table E3). In the multivariate model, we found that all covariates except use of public transportation, SES, and DM status remained associated with the outcome. Regardless of the status of

receivers, we found that compared with non-DT pairs, the transmitters within DT pairs were more likely to be younger than 35 years old (>34 yr old$^T\rightarrow\leqslant34$ yr old^R vs. $\leqslant34$ yr old$^T\rightarrow\leqslant34$ yr old^R: odds ratio [95% confidence interval], 0.64 [0.55–0.75]; >34 yr old$^T\rightarrow>34$ yr old^R vs. $\leqslant34$ yr old$^T\rightarrow\leqslant34$ yr old^R: 0.73 [0.59–0.91]), to be male (Male$^T\rightarrow$Female^R vs. Female$^T\rightarrow$Female^R: 1.34 [1.08–1.67]; Male$^T\rightarrow$Male^R vs. Female$^T\rightarrow$Female^R: 1.32 [1.07–1.62]), and to be a smoker (Smoker$^T\rightarrow$Smoker^R vs. Nonsmoker$^T\rightarrow$Nonsmoker^R: 2.81 [1.24–6.34]; Smoker$^T\rightarrow$Nonsmoker^R vs. Nonsmoker$^T\rightarrow$Nonsmoker^R: 1.33 [1.04–1.71]). We found that compared with pairs both of whose transmitters and recipients did not have a history of incarceration, the two disassortative pairings had an increased risk of being in a DT pair (Yes$^T\rightarrow$No^R vs. No$^T\rightarrow$No^R: 2.39 [1.98–2.87]; No$^T\rightarrow$Yes^R vs. No$^T\rightarrow$No^R: 2.17 [1.74–2.71]). We noted that, conditional on one of the two
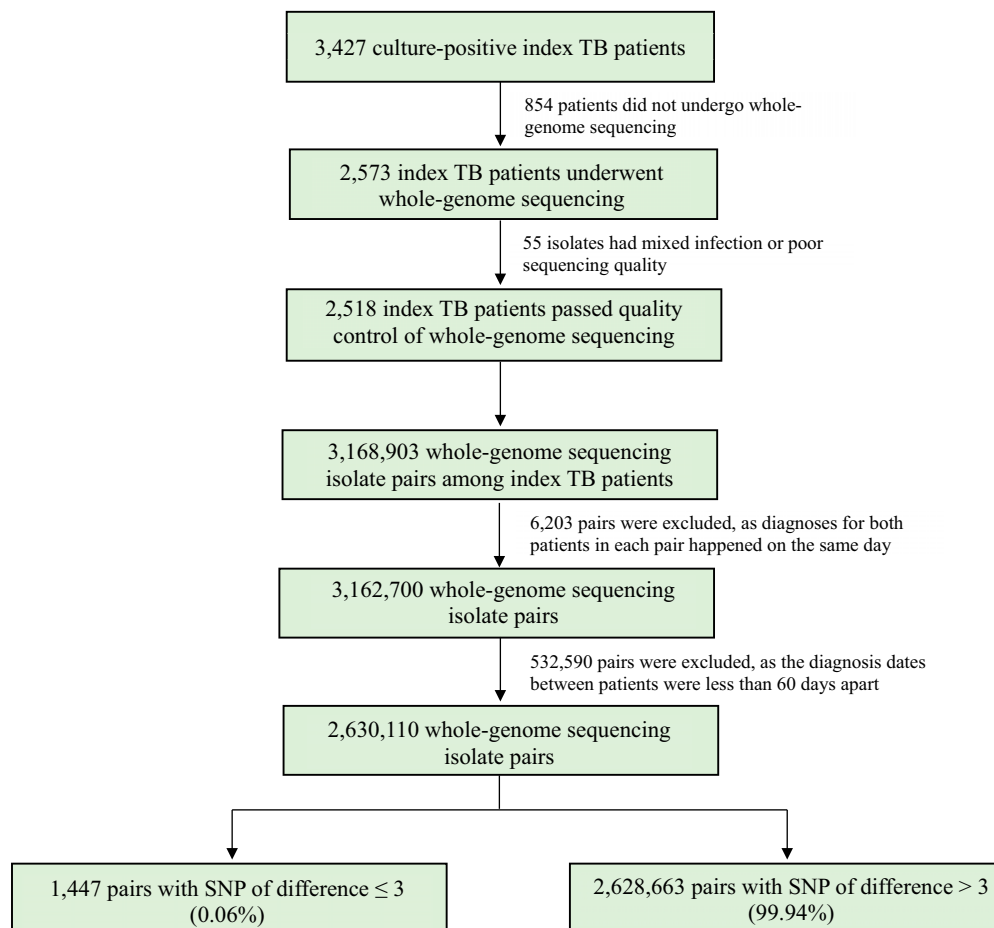

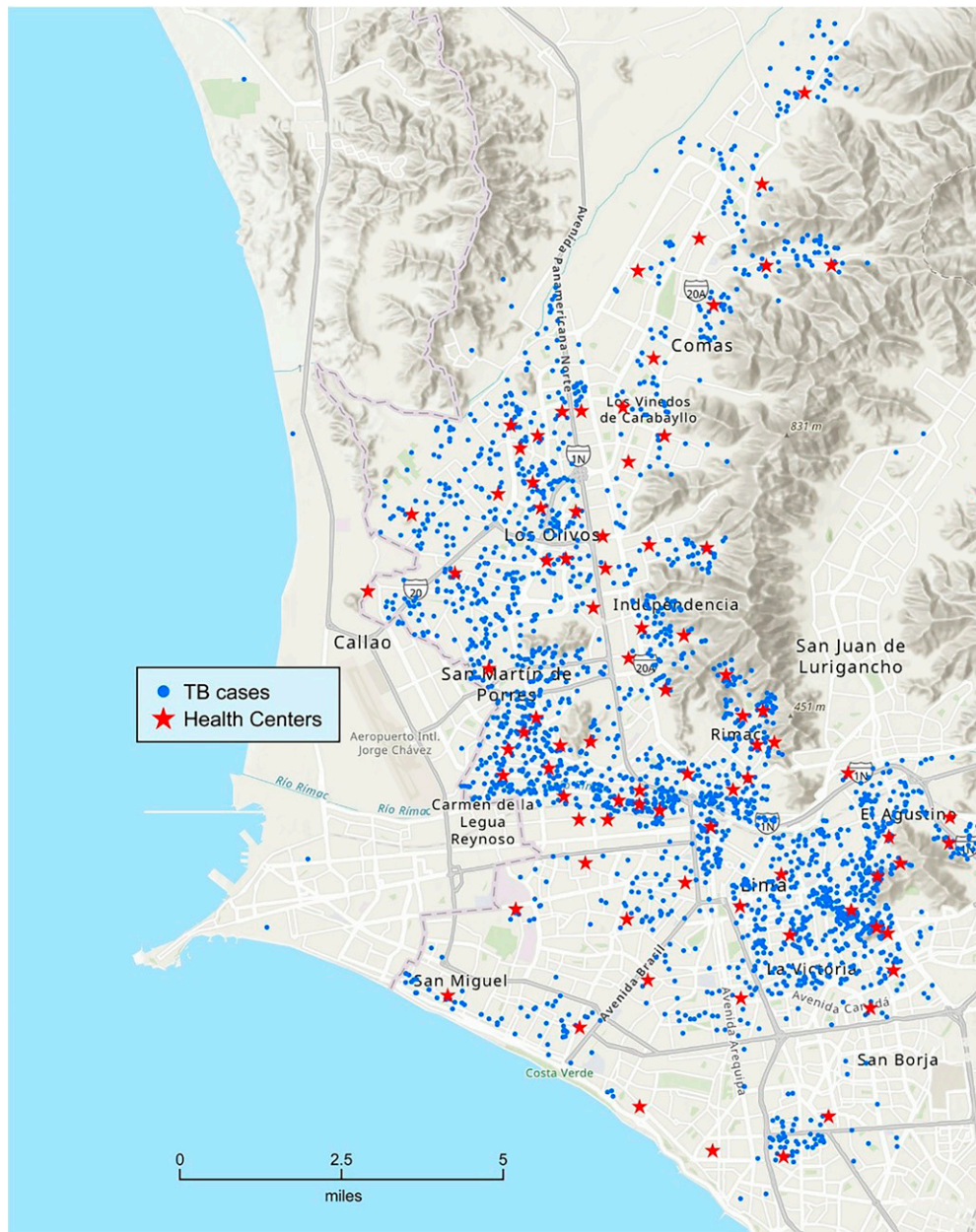**Figure 1.** Flowchart for patient pairs. TB = tuberculosis.

Figure 2. Geographic distribution of 2,518 patients. TB = tuberculosis.

patients in a pair having a history of incarceration, the likelihood of being a DT pair further increased by 4.5-fold when the other patient also had a history of incarceration ($Yes^T \rightarrow Yes^R$ vs. $No^T \rightarrow No^R$: 10.73 [7.46–15.41]) (Figure 4 and Table E4). The likelihood of being a DT pair increased when both the transmitter and the receiver of a pair were drinkers ($Drinker^T \rightarrow Drinker^R$ vs. $Nondrinker^T \rightarrow Nondrinker^R$: 1.21 [1.01–1.44]). We also found that direct transmission occurred less often between a nonsmoker transmitter and a smoker recipient than between a nonsmoker transmitter and a nonsmoker recipient ($Nonsmoker^T \rightarrow Smoker^R$ vs. $Nonsmoker^T \rightarrow Nonsmoker^R$: 0.68 [0.46–1.00]). In two separate sensitivity analyses in which we used two SNPs as the cutoff in determining the DT and kept the pairs with the shortest genetic distance when these pairs had the same recipients, we found that the effective size of most point estimates increased. Among several point estimates that had a reduced effect size, they were reduced by less than 20% (Tables E5 and E6). In a sensitivity analysis in which we used a fractional logistic regression model, we found a stronger association for all comparisons except an insignificant finding in the $Smoker^T \rightarrow Nonsmoker^R$ versus $Nonsmoker^T \rightarrow Nonsmoker^R$ comparison ($P = 0.105$) (Table E7).

Clinical variables. In the univariate analyses for TB-related clinical variables, we found that pairs had an increased likelihood of being DT when the transmitters had cavitary disease ($Yes^T \rightarrow Everybody\ else^R$ vs. $No^T \rightarrow Everybody\ else^R$: 1.15 [1.03–1.29]), a

Table 1. Characteristics of Patient Pairs ($N = 2{,}630{,}110$ Pairs)

Characteristics of Pairs (Transmitter-Receiver)	Pairs with SNP Difference ≤3 n (%) (n = 1,447)	Pairs with SNP Difference >3 n (%) (n = 2,628,663)	Total Pairs n (%) (N = 2,630,110)
Age			
≤34[T]→≤34[R]	755 (52.18)	1,201,020 (45.69)	1,201,775 (45.69)
≤34[T]→>34[R]	319 (22.05)	508,854 (19.36)	509,173 (19.36)
>34[T]→≤34[R]	250 (17.28)	643,521 (24.48)	643,771 (24.48)
>34[T]→>34[R]	123 (8.5)	275,268 (10.47)	275,391 (10.47)
Sex			
Female[T]→Female[R]	123 (8.5)	369,609 (14.06)	369,732 (14.06)
Female[T]→Male[R]	263 (18.18)	635,647 (24.18)	635,910 (24.18)
Male[T]→Female[R]	339 (23.43)	596,904 (22.71)	597,243 (22.71)
Male[T]→Male[R]	722 (49.9)	1,026,503 (39.05)	1,027,225 (39.06)
Peruvian ancestry			
No[T]→No[R]	182 (12.58)	273,138 (10.39)	273,320 (10.39)
No[T]→Yes[R]	159 (10.99)	211,524 (8.05)	211,683 (8.05)
Yes[T]→No[R]	163 (11.26)	254,290 (9.67)	254,453 (9.67)
Yes[T]→Yes[R]	130 (8.98)	197,038 (7.5)	197,168 (7.5)
Missing	813 (56.19)	1,692,673 (64.39)	1,693,486 (64.39)
Work outside the house			
No[T]→No[R]	562 (38.84)	1,005,698 (38.26)	1,006,260 (38.26)
No[T]→Yes[R]	348 (24.05)	583,716 (22.21)	584,064 (22.21)
Yes[T]→No[R]	323 (22.32)	638,049 (24.27)	638,372 (24.27)
Yes[T]→Yes[R]	203 (14.03)	370,741 (14.1)	370,944 (14.1)
Missing	11 (0.76)	30,459 (1.16)	30,470 (1.16)
Public transportation			
Light[T]→Light[R]	449 (31.03)	750,676 (28.56)	751,125 (28.56)
Light[T]→Heavy[R]	300 (20.73)	624,697 (23.76)	624,997 (23.76)
Heavy[T]→Light[R]	360 (24.88)	619,327 (23.56)	619,687 (23.56)
Heavy[T]→Heavy[R]	288 (19.9)	514,618 (19.58)	514,906 (19.58)
Missing	50 (3.46)	119,345 (4.54)	119,395 (4.54)
Socioeconomic status			
Low[T]→Low[R]	177 (12.23)	300,329 (11.43)	300,506 (11.43)
Low[T]→Not Low[R]	286 (19.77)	599,267 (22.8)	599,553 (22.8)
Not low[T]→Low[R]	279 (19.28)	528,934 (20.12)	529,213 (20.12)
Not low[T]→Not low[R]	634 (43.81)	1,057,515 (40.23)	1,058,149 (40.23)
Missing	71 (4.91)	142,618 (5.43)	142,689 (5.43)
Incarceration status			
No[T]→No[R]	1,081 (74.71)	2,318,736 (88.21)	2,319,817 (88.2)
No[T]→Yes[R]	120 (8.29)	119,423 (4.54)	119,543 (4.55)
Yes[T]→No[R]	186 (12.85)	138,794 (5.28)	138,980 (5.28)
Yes[T]→Yes[R]	47 (3.25)	7,122 (0.27)	7,169 (0.27)
Missing	13 (0.9)	44,588 (1.7)	44,601 (1.7)
Smoking status			
Nonsmoker[T]→Nonsmoker[R]	1,227 (84.8)	2,370,469 (90.18)	2,371,696 (90.17)
Nonsmoker[T]→Smoker[R]	27 (1.87)	71,766 (2.73)	71,793 (2.73)
Smoker[T]→Nonsmoker[R]	75 (5.18)	78,019 (2.97)	78,094 (2.97)
Smoker[T]→Smoker[R]	6 (0.41)	2,341 (0.09)	2,347 (0.09)
Missing	112 (7.74)	106,068 (4.04)	106,180 (4.04)
Drinking status			
Nondrinker[T]→Nondrinker[R]	305 (21.08)	745,730 (28.37)	746,035 (28.37)
Nondrinker[T]→Drinker[R]	267 (18.45)	518,183 (19.71)	518,450 (19.71)
Drinker[T]→Nondrinker[R]	387 (26.74)	667,186 (25.38)	667,573 (25.38)
Drinker[T]→Drinker[R]	310 (21.42)	467,360 (17.78)	467,670 (17.78)
Missing	178 (12.3)	230,204 (8.76)	230,382 (8.76)
HIV infection			
No[T]→No[R]	1,322 (91.36)	2,388,694 (90.87)	2,390,016 (90.87)
No[T]→Yes[R]	46 (3.18)	88,004 (3.35)	88,050 (3.35)
Yes[T]→No[R]	41 (2.83)	92,487 (3.52)	92,528 (3.52)
Yes[T]→Yes[R]	2 (0.14)	3,311 (0.13)	3,313 (0.13)
Missing	36 (2.49)	56,167 (2.14)	56,203 (2.14)
Diabetes mellitus			
No[T]→No[R]	1,315 (90.88)	2,353,578 (89.54)	2,354,893 (89.54)
No[T]→Yes[R]	76 (5.25)	105,401 (4.01)	105,477 (4.01)
Yes[T]→Everybody else[R]	40 (2.76)	124,015 (4.72)	124,055 (4.72)

(Continued)

Table 1. (*Continued*)

Characteristics of Pairs (Transmitter-Receiver)	Pairs with SNP Difference ≤3 *n* (%) (*n* = 1,447)	Pairs with SNP Difference >3 *n* (%) (*n* = 2,628,663)	Total Pairs *n* (%) (*N* = 2,630,110)
Missing	16 (1.11)	45,669 (1.74)	45,685 (1.74)
Cavity			
No[T]→Everybody else[R]	965 (66.69)	1,810,313 (68.87)	1,811,278 (68.87)
Yes[T]→Everybody else[R]	460 (31.79)	749,739 (28.52)	750,199 (28.52)
Missing	22 (1.52)	68,611 (2.61)	68,633 (2.61)
Cough			
≤1 mo[T]→Everybody else[R]	862 (59.57)	1,687,557 (64.2)	1,688,419 (64.2)
>1 mo[T]→Everybody else[R]	493 (34.07)	841,719 (32.02)	842,212 (32.02)
Missing	92 (6.36)	99,387 (3.78)	99,479 (3.78)
TB history			
No[T]→Everybody else[R]	966 (66.76)	2,086,200 (79.36)	2,087,166 (79.36)
Yes[T]→Everybody else[R]	476 (32.9)	532,157 (20.24)	532,633 (20.25)
Missing	5 (0.35)	10,306 (0.39)	10,311 (0.39)
Smear			
No[T]→Everybody else[R]	297 (20.53)	679,699 (25.86)	679,996 (25.85)
Yes[T]→Everybody else[R]	1,115 (77.06)	1,925,569 (73.25)	1,926,684 (73.25)
Missing	35 (2.42)	23,395 (0.89)	23,430 (0.89)

Definition of abbreviations: R = receiver; T = transmitter; TB = tuberculosis.
Data are represented as *n* (%) of pairs.

cough lasting more than 1 month before the TB diagnosis (>1 mo[T]→Everybody else[R] vs. ≤1 mo[T]→Everybody else[R]: 1.15 [1.03–1.28]), had previously been diagnosed with TB (Yes[T]→Everybody else[R] vs. No[T]→Everybody else[R]: 1.93 [1.73–2.16]), or had a sputum smear positive when they were diagnosed with TB (Yes[T]→Everybody else[R] vs. No[T]→Everybody else[R]: 1.33 [1.17–1.51]) (Table E8).

Discussion

Using the whole-genome sequences from Mtb isolates of a population-based cohort of patients with TB, we were able to evaluate the impact of demographic, biological, and social factors on TB transmission dynamics through a multivariate adjusted model. We found that younger patients with TB were more likely to be transmitters and to transmit TB to older patients. Male patients with TB generated more secondary TB cases than female patients with TB did. TB transmissions were also more likely to occur among people who drank alcohol, smoked, and had a history of incarceration. We also found that our method was able to identify the well-known TB-related clinical risk factors that may modify TB transmission dynamics.

Our finding that younger adults were more likely to be the source cases is consistent with the results of two similar studies from the Netherlands. ten Asbroek and colleagues used *IS 6110*–based restriction fragment length polymorphism typing to identify 69 source-secondary case pairs and found that 45% of source cases were aged 15–34 years (21). Using the same genotyping method and the diagnosis dates of more than 6,000 patients with TB in the Netherlands, Borgdorff and colleagues reported that younger patients generated most of the secondary cases (7). This association may be the result of heterogeneous social contact patterns in which younger patients have more social contacts than older ones; multiple studies that have evaluated age-specific social contact patterns in low- and middle-income countries, including one in Peru, have reported that young adults have an increased number of social contacts to all age groups compared with older individuals (22–26). Despite a well-documented association of TB transmission with younger patients, likely because of increased social contacts and the fact that younger adults are less likely to seek care until the disease progresses to a very serious level, few TB programs are designed to specifically target this at-risk population.

Many previous studies have also shown that TB is diagnosed more often in men than in women (27, 28). However, few studies have evaluated sex differences in the transmission of TB. The study cited above conducted by Borgdorff and colleagues found that TB due to recent transmission was mainly attributed to male source cases (7). Two studies from Uganda examined the social contact patterns of patients with TB and people recently infected with TB; both found that these participants had proportionally more adult male contacts (29, 30). Using a modeling approach based on the results of a survey of social contact patterns and local TB burden data, Dodd and colleagues estimated that the overall percentage of TB infections due to contact with adult men was 31% and 15% higher than due to contact with adult women in Zambia and South Africa, respectively (23). Although these studies did not directly measure the transmissibility of male patients with TB, the results suggested that a substantial proportion of TB infections may be generated by male patients with TB, a finding that is highly consistent with our results.

Tobacco and heavy alcohol use are also well-known risk factors for TB infection and disease. Several studies have also demonstrated that patients with TB who are smokers are more likely to transmit TB than those who are not (31–33). Here, we found that TB transmissions were more likely to occur only when both transmitters and receivers smoked or drank alcohol. In addition to an increased risk of transmission due to the biological effects of smoking and drinking, our finding is also likely to be driven by assortative social mixing patterns. In particular, our observation that transmissions occurred less often between a nonsmoker transmitter and a smoker recipient was likely due to a reduced nonassortative mixing
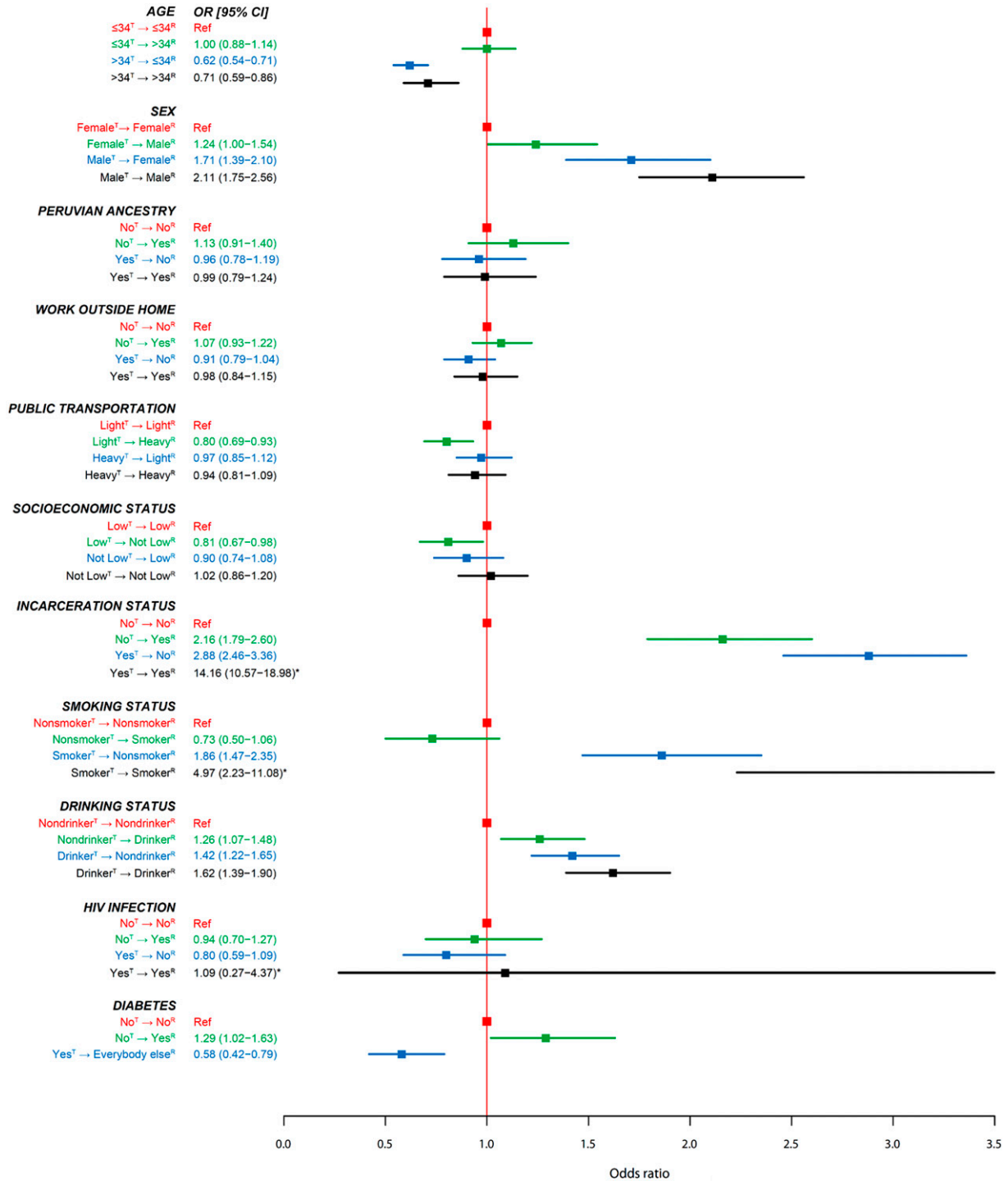
Figure 3. Univariate associations between risk factors and direct transmission (SNP difference, ≤3). 95% CI = 95% confidence interval; OR = odds ratio; R = receiver; Ref = reference; T = transmitter. *The value is out of scale and has not been included or entirely represented.

pattern between smokers and nonsmokers. In Finland, Saari and colleagues studied an age cohort born in 1979. In 2008, it was found that participants whose close friends were smokers were more than fivefold more likely to be smokers than those whose close friends were nonsmokers (34). Multiple high-transmission TB outbreaks have been traced to individuals frequenting bars, where people can drink and smoke in shared indoor spaces for prolonged periods of time (35). In South Africa, Classen and colleagues evaluated the Mtb restriction fragment length polymorphism genotyping of 84 patients with TB (from 26 households) and their social contact networks and found that patients with TB from 12 households formed community clusters that were likely attributable to
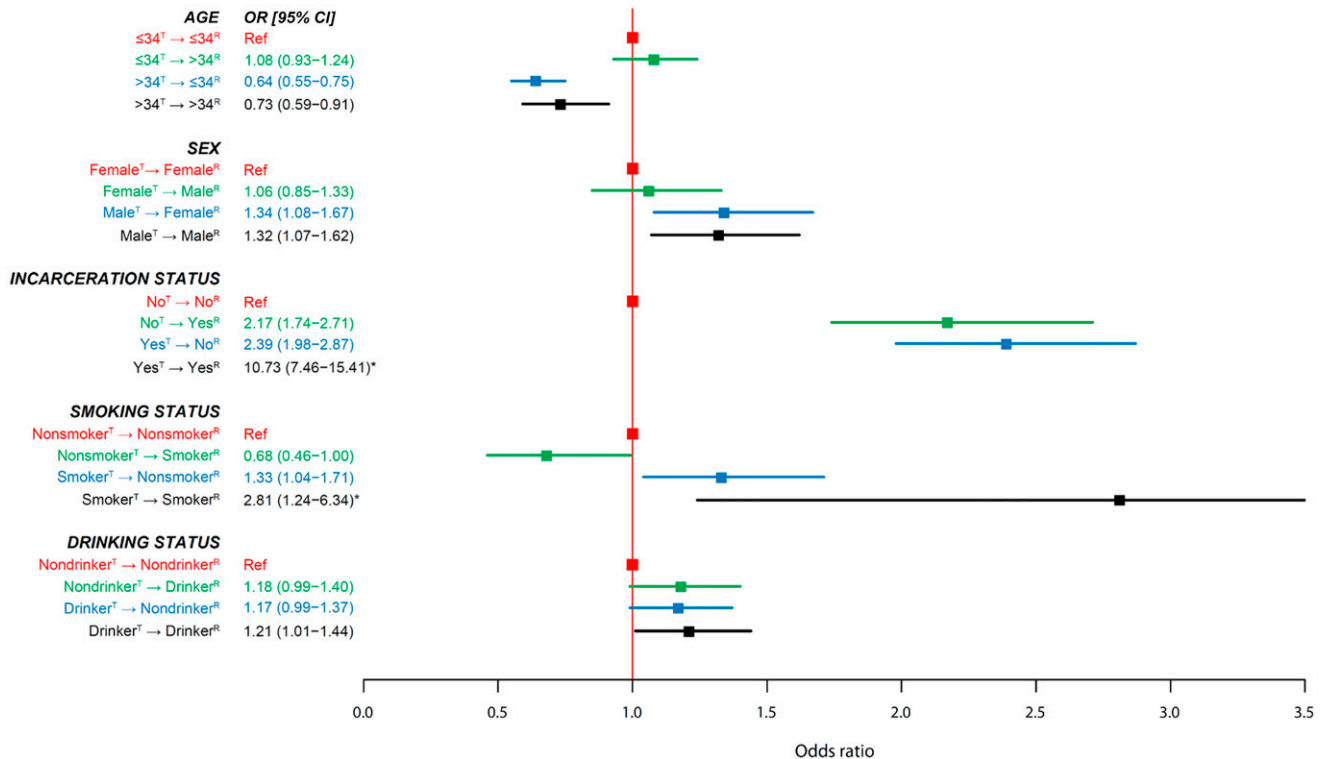
Figure 4. Multivariate associations between risk factors and direct-transmission (SNP difference ≤3). 95% CI = 95% confidence interval; OR = odds ratio; R = receiver; Ref = reference; T = transmitter. *The value is out of scale and has not been included or entirely represented.

drinking in social groups (36). In another study from South Africa, Booi and Breetzke reported that the spatial density pattern of TB was consistent with the location of the alcohol outlets in the town of Manelodi (37). In a study conducted in Texas, Agarwal and colleagues used molecular epidemiology to show that patients who were excessive alcohol users were more likely to form genetic clusters than those who were not (38). These results suggest that assortative social mixing patterns among smokers and drinkers may independently contribute to TB transmission events.

Multiple previous studies have demonstrated that prisons are hotspots of TB transmission (39, 40). Within low- and middle-income countries, the prevalence of TB among incarcerated individuals has been reported to be more than 10-fold greater than in the general population (41). In Peru, Salazar-De La Cuba and colleagues found that the self-reported TB prevalence among incarcerated individuals from 66 Peruvian correctional facilities was 2,510 in 100,000, at least 25-fold higher than the prevalence in the general population (42, 43). In another study from Peru, Warren and colleagues used the Mtb genetic data collected from patients with

TB in a prison and in surrounding neighborhoods to demonstrate a clear prison spillover effect (44). In Thailand, Miyahara and colleagues used Mtb WGS data to show that patients in large transmission clusters (defined as a cluster with >10 patients with TB) were more than 10-fold more likely to have a history of incarceration than those who were unclustered or whose Mtb isolates belonged to a smaller cluster (40). Our finding that direct transmission pairs often included at least one patient with a history of incarceration is consistent with these results. This suggests that prisons are a reservoir for TB and that TB programs within prisons are essential to reduce the TB burden, both in prisons and in the surrounding community. In addition to the transmissions that occurred in the prisons, an assortative mixing pattern for patients with a history of incarceration in the community may also contribute to the exalted risk of direct transmission between pairs that included both patients with a history of incarceration, because people with a history of incarceration may be more likely to congregate in the community than those without such a history.

Although individuals living with HIV face an increased risk of TB reactivation or TB

disease progression after recent infection, a study has indicated that in a low HIV prevalence setting, the elevated risk of TB disease associated with HIV is attributed primarily to reactivation rather than to recent primary infection (45). In our previous study, we also demonstrated that patients with TB/HIV with advanced immunosuppression exhibited reduced infectiousness (46). Therefore, from the perspective of the direction of associations, although lacking statistical significance, our observation that patients with TB coinfected with HIV were less likely to act as either receivers (0.94 [0.70–1.27]) or transmitters (0.80 [0.59–1.09]) aligns with these previous findings. The lack of statistical significance can be attributed to the small sample size of patients with TB coinfected with HIV, because the prevalence of HIV among patients with diagnosed TB in our study cohort was only 3%.

Many previous studies have reported a strong association between TB transmission and social activity and SES (47, 48). Although our findings indicated that the frequency of using public transportation and SES were associated with the likelihood of being DT pairs in the univariate analyses, these significant associations no longer held after

multivariate adjustment. Although we cannot rule out the possibility that our categorization of these variables as dichotomous may have oversimplified their complexity, our findings suggest that a substantial portion of the associations between social activity/SES and TB transmissions was confounded by other demographic and social factors.

In the univariate analysis, we found that patients with TB with DM were more likely to be recipients but less likely to be transmitters. Although these associations disappeared after multivariate adjustment, these findings may indicate that the way DM modifies TB transmission dynamics is more complex than previously believed. The first finding, that patients with DM were more likely to be recipients, is consistent with many previous studies that have demonstrated an increased risk of TB disease among patients with DM (49). The second finding suggests that patients with DM/TB may be less infectious than other patients with TB without DM. It remains unclear how DM modifies the infectiousness of patients with TB and thereby alters TB transmission dynamics (49, 50). A systematic review and meta-analysis that assessed the Mtb molecular clustering of patients with TB with DM found that patients with TB with DM were less likely to be clustered than patients without DM (global odd ratio, 0.84) (51). However, another study that used a phylogenetic modeling approach to identify TB transmission chains in Spain found that DM was enriched in transmitters compared with nonclustered patients with TB (52).

Limitations
Our study has several limitations. First, household-based cohort studies or TB outbreak studies have shown that although the Mtb genetic distances between two patients with a direct transmission relationship were usually less or equal to 3, it could range between 0 and 12 base SNPs. Therefore, it is possible that the classification of DT pairs was nondifferentially misclassified. Second, within a pair, we do not always know which patient was the source case, because many patients with TB can remain infectious for long periods before they

come to diagnosis, and thus a source patient may have been diagnosed after the patient to whom they transmitted infection. Although we have excluded pairs whose dates of TB diagnosis were apart less than 60 days for our analyses, we cannot rule out the possibility that some pairs that remained had been nondifferentially misclassified. Third, we considered all possible patient combinations because a patient may serve as a transmitter for multiple recipients. Consequently, we recognized that the same patient with TB could be classified as a recipient in different pairs with different transmitter patients. As a result, some of the identified direct transmission pairs may not accurately represent true direct transmission events, whereas a patient could serve as a transmitter in multiple DT pairs, and a misclassification must occur when a patient was considered as a recipient in multiple pairs. All of these types of misclassifications could have driven our results toward the null and could have reduced the true effect of the pair's attributes. This statement is also supported by the fact that the effect size of most point estimates increased in the sensitivity analyses when we used more stringent settings to reduce the likelihood of misclassifying a non-DT pair as a DT pair. Fourth, although we used data from patients with pulmonary TB enrolled systematically through protocol-driven processes, the study protocol specified the exclusion of children diagnosed with TB disease because they were often unable to produce sputum. Furthermore, among the adult patients enrolled in the cohort study who had a positive culture test result, WGS was completed for 75% of the isolates because of financial constraints. Patients with TB that was not bacteriologically confirmed were also excluded from this study because of a lack of WGS data for Mtb. Therefore, it is possible that we may have missed transmission events. Consequently, it is not clear whether our findings can be generalized to children or whether the inferences described in our study would have been altered if we had we been able to incorporate all patients with TB into our study. Fifth, the smoking, drinking status, or DM status of patients with TB may be misclassified if patients were reluctant to

report potentially stigmatizing behaviors or had preexisting undiagnosed DM. These nondifferential misclassifications may also drive our results toward the null. Sixth, in our study, we did not collect information on the duration of incarceration among patients with TB with a history of incarceration. Therefore, it is uncertain whether the observed increased risk of direct transmission among patients with a history of incarceration is due to an elevated risk of TB infection during their time of incarceration or their congregation in the community.

Conclusions
In conclusion, we implemented a newly developed method that used a population-based cohort of patients with TB and WGS data of Mtb to concurrently assess the influence of demographic, biological, and social factors on TB transmission dynamics using a multivariate adjusted model, accounting for potential assortative mixing patterns. We demonstrated that, in addition to clinical and biological factors, demographic and social attributes also play an important role in transmission dynamics in a community setting. These demographic and social groups include men, younger adults, former prisoners, drinkers, and smokers who could benefit the most from TB intervention programs such as active case finding. Our data also reinforce the need to further strengthen TB control measures in prisons and among released prisoners. They also indicate the need to devise creative ways to extend TB screening to social groups, such as younger adults, who may be less likely to access routine preventive care. ∎

References

1. World Health Organization. Global tuberculosis report 2022 [accessed 2024 Feb 7]. Available from: https://www.who.int/teams/global-tuberculosis-programme/tb-reports/global-tuberculosis-report-2022.
2. Pai M, Kasaeva T, Swaminathan S. Covid-19's devastating effect on tuberculosis care—a path to recovery. *N Engl J Med* 2022;386: 1490–1493.
3. Brooks MB, Jenkins HE, Puma D, Tzelios C, Millones AK, Jimenez J, *et al.* A role for community-level socioeconomic indicators in targeting tuberculosis screening interventions. *Sci Rep* 2022;12:781.
4. Khundi M, Carpenter JR, Nliwasa M, Cohen T, Corbett EL, MacPherson P. Effectiveness of spatially targeted interventions for control of HIV, tuberculosis, leprosy and malaria: a systematic review. *BMJ Open* 2021; 11:e044715.
5. Date A, Modi S. TB screening among people living with HIV/AIDS in resource-limited settings. *J Acquir Immune Defic Syndr* 2015;68: S270–S273.
6. Nyirenda JLZ, Bockey A, Wagner D, Lange B. Effect of tuberculosis (TB) and diabetes mellitus (DM) integrated healthcare on bidirectional screening and treatment outcomes among TB patients and people living with DM in developing countries: a systematic review. *Pathog Glob Health* 2023;117:36–51.
7. Borgdorff MW, Nagelkerke NJD, de Haas PEW, van Soolingen D. Transmission of *Mycobacterium tuberculosis* depending on the age and sex of source cases. *Am J Epidemiol* 2001;154:934–943.
8. Huang CC, Trevisi L, Becerra MC, Calderón RI, Contreras CC, Jimenez J, *et al.* Spatial scale of tuberculosis transmission in Lima, Peru. *Proc Natl Acad Sci USA* 2022;119:e2207022119.
9. Becerra MC, Huang C-C, Lecca L, Bayona J, Contreras C, Calderon R, *et al.* Transmissibility and potential for disease progression of drug resistant *Mycobacterium tuberculosis*: prospective cohort study. *BMJ* 2019;367:l5894.
10. Odone A, Calderon R, Becerra MC, Zhang Z, Contreras CC, Yataco R, *et al.* Acquired and transmitted multidrug resistant tuberculosis: the role of social determinants. *PLoS One* 2016;11:e0146642.
11. Luo Y, Suliman S, Asgari S, Amariuta T, Baglaenko Y, Martínez-Bonet M, *et al.* Early progression to active tuberculosis is a highly heritable trait driven by 3q23 in Peruvians. *Nat Commun* 2019;10:3765.
12. Asgari S, Luo Y, Huang CC, Zhang Z, Calderon R, Jimenez J, *et al.* Higher native Peruvian genetic ancestry proportion is associated with tuberculosis progression risk. *Cell Genomics* 2022;2:100151.
13. Magee MJ, Khakharia A, Gandhi NR, Day CL, Kornfeld H, Rhee MK, *et al.* Increased risk of incident diabetes among individuals with latent tuberculosis infection. *Diabetes Care* 2022;45:880–887.
14. Eckold C, Kumar V, Weiner J, Alisjahbana B, Riza AL, Ronacher K, *et al.*; TANDEM consortium. Impact of intermediate hyperglycemia and diabetes on immune dysfunction in tuberculosis. *Clin Infect Dis* 2021; 72:69–78.
15. Faurholt-Jepsen D, Range N, Praygod G, Kidola J, Faurholt-Jepsen M, Aabye MG, *et al.* The role of diabetes co-morbidity for tuberculosis treatment outcomes: a prospective cohort study from Mwanza, Tanzania. *BMC Infect Dis* 2012;12:165.
16. Li H, Durbin R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* 2010;26:589–595.
17. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, *et al.*; 1000 Genome Project Data Processing Subgroup. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 2009;25: 2078–2079.
18. Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, *et al.* Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One* 2014;9: e112963.
19. Comas I, Chakravartti J, Small PM, Galagan J, Niemann S, Kremer K, *et al.* Human T cell epitopes of *Mycobacterium tuberculosis* are evolutionarily hyperconserved. *Nat Genet* 2010;42:498–503.
20. Coll F, McNerney R, Guerra-Assunção JA, Glynn JR, Perdigão J, Viveiros M, *et al.* A robust SNP barcode for typing *Mycobacterium tuberculosis* complex strains. *Nat Commun* 2014;5:4812.
21. ten Asbroek AH, Borgdorff MW, Nagelkerke NJ, Sebek MM, Devillé W, van Embden JD, *et al.* Estimation of serial interval and incubation period of tuberculosis using DNA fingerprinting. *Int J Tuberc Lung Dis* 1999;3:414–420.
22. Trauer JM, Dodd PJ, Gomes MGM, Gomez GB, Houben RMGJ, McBryde ES, *et al.* The importance of heterogeneity to the epidemiology of tuberculosis. *Clin Infect Dis* 2019;69:159–166.
23. Dodd PJ, Looker C, Plumb ID, Bond V, Schaap A, Shanaube K, *et al.* Age- and sex-specific social contact patterns and incidence of *Mycobacterium tuberculosis* infection. *Am J Epidemiol* 2016;183:156–166.
24. Mousa A, Winskill P, Watson OJ, Ratmann O, Monod M, Ajelli M, *et al.* Social contact patterns and implications for infectious disease transmission—a systematic review and meta-analysis of contact surveys. *eLife* 2021;10:e70294.
25. 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;5:e74.
26. Johnstone-Robertson SP, Mark D, Morrow C, Middelkoop K, Chiswell M, Aquino LDH, *et al.* Social mixing patterns within a South African township community: implications for respiratory disease transmission and control. *Am J Epidemiol* 2011;174:1246–1255.
27. Horton KC, Hoey AL, Béraud G, Corbett EL, White RG. Systematic review and meta-analysis of sex differences in social contact patterns and implications for tuberculosis transmission and control. *Emerg Infect Dis* 2020;26:910–919.
28. Horton KC, Sumner T, Houben RMGJ, Corbett EL, White RG. A Bayesian approach to understanding sex differences in tuberculosis disease burden. *Am J Epidemiol* 2018;187:2431–2438.
29. Miller PB, Zalwango S, Galiwango R, Kakaire R, Sekandi J, Steinbaum L, *et al.* Association between tuberculosis in men and social network structure in Kampala, Uganda. *BMC Infect Dis* 2021;21:1023.
30. Marquez C, Chen Y, Atukunda M, Chamie G, Balzer LB, Kironde J, *et al.* The association between social network characteristics and tuberculosis infection among adults in 9 rural Ugandan communities. *Clin Infect Dis* 2023;76:e902–e909.
31. Chu AL, Lecca LW, Calderón RI, Contreras CC, Yataco RM, Zhang Z, *et al.* Smoking cessation in tuberculosis patients and the risk of tuberculosis infection in child household contacts. *Clin Infect Dis* 2021; 73:1500–1506.
32. Huang CC, Tchetgen ET, Becerra MC, Cohen T, Galea J, Calderon R, *et al.* Cigarette smoking among tuberculosis patients increases risk of transmission to child contacts. *Int J Tuberc Lung Dis* 2014;18: 1285–1291.
33. Godoy P, Caylà JA, Carmona G, Camps N, Álvarez J, Alsedà M, *et al.*; Grupo de Trabajo de Estudios de Contactos de Tuberculosis de Cataluña. Smoking in tuberculosis patients increases the risk of infection in their contacts. *Int J Tuberc Lung Dis* 2013;17:771–776.
34. Saari AJ, Kentala J, Mattila KJ. The smoking habit of a close friend or family member—how deep is the impact? A cross-sectional study. *BMJ Open* 2014;4:e003218.
35. Kline SE, Hedemark LL, Davies SF. Outbreak of tuberculosis among regular patrons of a neighborhood bar. *N Engl J Med* 1995;333: 222–227.
36. Classen CN, Warren R, Richardson M, Hauman JH, Gie RP, Ellis JHP, *et al.* Impact of social interactions in the community on the transmission of tuberculosis in a high incidence area. *Thorax* 1999;54:136–140.
37. Booi L, Breetzke GD. The spatial relationship between tuberculosis and alcohol outlets in the township of Mamelodi, South Africa. *Afr Health Sci* 2022;22:162–168.
38. Agarwal S, Nguyen DT, Teeter LD, Graviss EA. Spatial-temporal distribution of genotyped tuberculosis cases in a county with active transmission. *BMC Infect Dis* 2017;17:378.
39. Dara M, Acosta CD, Melchers NVSV, Al-Darraji HAA, Chorgoliani D, Reyes H, *et al.* Tuberculosis control in prisons: current situation and research gaps. *Int J Infect Dis* 2015;32:111–117.
40. Miyahara R, Piboonsiri P, Chiyasirinroje B, Imsanguan W, Nedsuwan S, Yanai H, *et al.* Risk for prison-to-community tuberculosis transmission, Thailand, 2017-2020. *Emerg Infect Dis* 2023;29:477–483.
41. Haeusler IL, Torres-Ortiz A, Grandjean L. A systematic review of tuberculosis detection and prevention studies in prisons. *Glob Public Health* 2022;17:194–209.
42. Salazar-De La Cuba AL, Ardiles-Paredes DF, Araujo-Castillo RV, Maguiña JL. High prevalence of self-reported tuberculosis and

associated factors in a nation-wide census among prison inmates in Peru. *Trop Med Int Health* 2019;24:328–338.

43. Partners in Health. Peru: study aims to reveal how TB spreads. 2015 [accessed 2024 Feb 7]. Available from: https://www.pih.org/article/peru-unprecedented-study-aims-to-reveal-how-tb-spreads.

44. Warren JL, Grandjean L, Moore DAJ, Lithgow A, Coronel J, Sheen P, *et al.* Investigating spillover of multidrug-resistant tuberculosis from a prison: a spatial and molecular epidemiological analysis. *BMC Med* 2018;16:122.

45. Winter JR, Smith CJ, Davidson JA, Lalor MK, Delpech V, Abubakar I, *et al.* The impact of HIV infection on tuberculosis transmission in a country with low tuberculosis incidence: a national retrospective study using molecular epidemiology. *BMC Med* 2020;18:385.

46. Huang C-C, Tchetgen ET, Becerra MC, Cohen T, Hughes KC, Zhang Z, *et al.* The effect of HIV-related immunosuppression on the risk of tuberculosis transmission to household contacts. *Clin Infect Dis* 2014; 58:765–774.

47. Tadokera R, Bekker LG, Kreiswirth BN, Mathema B, Middelkoop K. TB transmission is associated with prolonged stay in a low socio-economic, high burdened TB and HIV community in Cape Town, South Africa. *BMC Infect Dis* 2020;20:120.

48. Hargreaves JR, Boccia D, Evans CA, Adato M, Petticrew M, Porter JDH. The social determinants of tuberculosis: from evidence to action. *Am J Public Health* 2011;101:654–662.

49. Baker MA, Harries AD, Jeon CY, Hart JE, Kapur A, Lönnroth K, *et al.* The impact of diabetes on tuberculosis treatment outcomes: a systematic review. *BMC Med* 2011;9:81.

50. Goldhaber-Fiebert JD, Jeon CY, Cohen T, Murray MB. Diabetes mellitus and tuberculosis in countries with high tuberculosis burdens: individual risks and social determinants. *Int J Epidemiol* 2011;40: 417–428.

51. Blanco-Guillot F, Delgado-Sánchez G, Mongua-Rodríguez N, Cruz-Hervert P, Ferreyra-Reyes L, Ferreira-Guerrero E, *et al.* Molecular clustering of patients with diabetes and pulmonary tuberculosis: a systematic review and meta-analysis. *PLoS One* 2017; 12:e0184675.

52. Xu Y, Cancino-Muñoz I, Torres-Puente M, Villamayor LM, Borrás R, Borrás-Máñez M, *et al.* High-resolution mapping of tuberculosis transmission: whole genome sequencing and phylogenetic modelling of a cohort from Valencia Region, Spain. *PLoS Med* 2019;16: e1002961.