A Case Control Study to Identify Community Venues Associated with Genetically Clustered Multidrug-Resistant Tuberculosis Disease in Lima, Peru

David P. Bui[1*], Eyal Oren[2], Denise J. Roe[1], Heidi E. Brown[1], Robin B. Harris[1], Gwenan M. Knight[3], Robert H. Gilman[4,5], Louis Grandjean[3,4]

1 The University of Arizona, Mel and Enid Zuckerman College of Public Health, Department of Epidemiology and Biostatistics, Tucson, Arizona, USA

2 San Diego State University, Graduate School of Public Health, San Diego, California, USA

3 London School of Hygiene and Tropical Medicine, London, United Kingdom

4 Universidad Peruana Cayetano Heredia, Lima, Peru

5 Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland, USA

*Corresponding Author: davidbui@email.arizona.edu

Summary

Multidrug-resistant tuberculosis (MDR-TB) transmission occurs in community venues. Identifying high-risk venues may direct community prevention efforts. We used whole-genome sequencing, venue-tracing interviews, and both conventional and social network analysis to identify several community venue types associated with MDR-TB disease.

Abstract

Background: The majority of tuberculosis transmission occurs in community settings. The primary aim of this study was to assess the association between exposure to community venues and multidrug-resistant tuberculosis (MDR-TB) disease. The secondary aim was to describe the social networks of MDR-TB cases and controls.

Methods: This case-control study was conducted in Lima, Peru. We recruited lab-confirmed MDR-TB cases and community controls matched on age and sex. Whole-genome sequencing was used to identify genetically-clustered cases. Venue-tracing interviews (nonblinded) were conducted to enumerate community venues frequented by participants. Logistic regression was used to assess the association between MDR-TB disease and person-time spent in community venues. A location-based social network was constructed with respondents connected if they reported frequenting the same venue and an exponential random graph model (ERGM) was fitted to model the network.

Results: We enrolled 59 cases and 65 controls. Participants reported 729 unique venues. Mean number of venues reported was similar in both groups (P=0.92). Cases reported spending more person-time (hours) in healthcare and transportation venues than controls (P<0.05). Person-time in healthcare venues (Adjusted Odds Ratio (OR)=1.67, P=0.01), schools (OR=1.53, P<0.01), and transportation (OR=1.25, P=0.03) was associated with MDR-TB disease. Healthcare venues, markets, cinemas, and transportation venues were commonly shared among clustered cases. The ERGM indicated significant community segregation between cases and controls. Case networks were more densely connected.

Conclusions: Exposure to healthcare venues, schools and transportation was associated with MDR-TB disease. Intervention across the segregated network of case venues may be necessary to effectively stem transmission.

Keywords

multidrug-resistant; tuberculosis; social network; genotyping; community transmission

Introduction

Multidrug-resistant tuberculosis (MDR-TB) requires longer treatment regimens, is associated with worse outcomes and threatens to derail global tuberculosis (TB) control efforts [1, 2]. Recent studies have found high levels of TB transmission between unrelated households in Peru and other high burden settings [3-6]. These findings suggest MDR-TB transmission may occur during exposure to infectious cases in community settings [6, 7]. Identification of community settings associated with MDR-TB disease may inform public health surveillance, enhance case finding and identify settings where environmental intervention, such as UV irradiation or improved air ventilation, may reduce exposure and transmission risk [8].

Prior investigations have used venue-tracing interviews, diaries and surveys of diagnosed TB patients to determine settings of possible transmission; however, these studies are predominantly descriptive and do not estimate measures of association between TB disease and exposure to community settings [9-13]. These prior studies also focus mainly on drug-sensitive TB and not MDR-TB and rarely leverage social network analysis to evaluate data derived from venue-tracing interviews [9, 10]. Gardy et al. used both whole-genome sequencing and social network questionnaires to investigate a TB outbreak in British Columbia; however, their social network investigation was limited to eight respondents and questionnaires were limited to known contacts [11]. Data from venue-tracing interviews can be used to construct social networks between respondents and unknown contacts in shared community venues, offering a more comprehensive view of a community's social network [14]. Social networks and contact patterns may influence the spread and epidemiology of infectious diseases [15-19] and may be studied to reveal important social network properties influencing the spread of MDR-TB in community settings.

The primary objective of this study was to investigate the association between exposure to community settings and the odds of MDR-TB disease. We hypothesized that MDR-TB cases would report more person-time exposure to community venues than controls. The secondary objective was

to visualize and statistically model and characterize the social networks of MDR-TB cases to identify social contact processes associated with MDR-TB disease.

Methods

A case-control study design was used to test the primary hypothesis. Cases were patients with confirmed MDR-TB residing in Callao and Lima Sur (Figure S1). These two regions cover extensive areas of Lima and consistently report a high proportion of incident MDR-TB cases [20].

Cases were recruited from a completed parent study which enrolled a cohort of MDR-TB patients between 2010-2013 in Callao and Lima Sur [21]. The parent study was designed to study the transmission of MDR-TB relative to drug-sensitive TB. All MDR-TB cases were whole-genome sequenced to identify genetically clustered cases [21]. Cases were considered clustered if their MDR-TB strain matched the strain of at least one other case within five or fewer single nucleotide polymorphisms (SNPs) [22].The majority of cases were genetically clustered (68%) and therefore likely to have arisen from recent transmission [23].

Community members living near cases with no history of TB were selected as controls. We recruited community health workers that worked in the health clinics in which cases were diagnosed. Additional community controls were recruited by referral from community health workers and sourced from churches, schools, restaurants and communal kitchens where the community health workers also worked. Controls were frequency matched to cases on study region, age (± 5 years) and sex.

The study protocol, consent forms and data collection instruments were reviewed and approved by the Institutional Committee of Ethics for Humans at La Universidad Peruana Cayetano Heredia.

We collected data through face-to-face, nonblinded interviews and used a venue-tracing interview guide to enumerate venues frequented by cases and controls. Cases were asked to name

venues visited during the month prior to MDR-TB diagnosis, defined as the date of sputum isolate collection prior to treatment, and healthy controls were asked about venues visited during the month prior to study enrollment. When cases were unavailable, deceased or too ill to respond, proxy interviews were conducted with available family members (e.g., spouses, parents). We asked participants to report the name, address, and type of venues visited, the average frequency of visits (daily, weekly, biweekly, etc.), and the average duration of visits in hours. Person-time per venue reported was estimated by multiplying the weekly frequency of visits by mean duration for each visit. For example, for someone frequenting a market three times a week for one hour per visit, the weekly person-time would be three hours.

Logistic regression was used to estimate the association between time spent in community venues and MDR-TB disease and report adjusted odds ratios (OR) as the measure of association. Education and income level were included in the final models as bivariate analysis showed they were associated with cases and person-time in community venues. The frequency matched variables (age, sex and region) were included in the final models. Likelihood ratio tests were used to assess interaction by study region and person-time and results were stratified by study region when a significant region by person-time interaction (P<0.10) was observed. Descriptive statistics and regression models were done in Stata 14 [24].

We used data from the venue-tracing interviews to construct two-mode social networks, showing respondents connected to venues they reported frequenting. The two-mode networks were visualized and reviewed to identify venues most visited by cases and controls, respectively. We reviewed the two-mode networks of each genetic cluster with >1 case to identify shared venues that may be considered epidemiologic links or sites of transmission.

Next, the two-mode networks were converted to one-mode "location-based social networks" (LSN), showing respondents connected to respondents when they reported frequenting the same venue. We calculated the network density (i.e., proportion of actual connections out of all

potential connections) of the one-mode networks of cases and controls to compare the difference in inter-connectedness among the two groups for each venue type [13]. In these LSNs, greater network density would indicate greater inter-connectedness among respondents through shared venues.

We used exponential random graph models (ERGMs) to model the one-mode LSNs to evaluate the extent to which cases are connected to other cases (i.e., homophily); in other words, the extent to which cases tend to frequent the same venues as other cases. Exponential random graph models (ERGMs) are a class of models for analyzing and modeling observed social networks [25]. ERGMs were fitted to the one-mode LSN assembled from the venue-tracing interviews with model terms for homophily by case status, age, sex, education, income and region. A term for geometrically-weighted edgewise shared partners (GWESP) was included to capture social network transitivity and improve model fit [26]. We used the 'igraph' and 'statnet' R packages for all network analyses [27-29].

Results

A total of 59 cases (30 in Callao and 29 in Lima Sur) and 65 controls (33 in Callao and 32 in Lima Sur) were enrolled. Forty (68%) cases were genetically clustered with at least one other case. Fifteen (27%) cases reported at least one prior episode of drug sensitive TB, of which 8 (53%) were genetically clustered; one case had one prior episode of MDR-TB.

Frequency matching between cases and controls was achieved by region, age and sex (Table 1). Eighteen (31%) of the case interviews were done by proxy. There were no significant demographic differences between primary cases (those directly interviewed) and proxy cases (those interviewed by proxy) (Table S1).

A total of 729 unique venues were reported; cases reported 352 unique venues, controls reported 427 unique venues and 50 venues were reported by both. Three quarters of venues reported by both cases and controls were healthcare venues (n=14, 28%) and markets (n=23, 46%);

ten (20%) of case-control shared venues included restaurants, cinemas, schools and transportation (Table S2). The mean number of venues reported was similar between cases (mean=8.3, sd=3.0) and controls (mean=8.4, sd=5.7) (*P*=0.92). There was no difference in the mean number of venues reported between proxy and primary cases (*P*=0.73).

Among cases, the most frequently reported venue types were healthcare venues, markets and stores (Figure 1A) and cases reported the greatest amount of weekly person-hours in healthcare venues, transportation venues, schools and markets (Figure 1B). Controls reported visiting markets, stores and restaurants most frequently and the majority of weekly person-hours was in markets, offices and stores (Figure 1B). Cases reported significantly more time in healthcare (*P*<0.001) and transportation venues (*P*=0.003) (Table 2).

The adjusted logistic regression models indicated that the odds of MDR-TB disease are increased by 15% for every additional ten weekly person-hours spent in community settings (OR=1.15, CI=1.02 - 1.30) (Table 2). Time in healthcare venues, schools and transportation were all significantly associated with MDR-TB disease (*P*<0.05). There was a significant interaction by region for time in store, transportation and bar/club venues (Table S3).

Visual representation of the two-mode networks (respondents linked to venues) and respective one-mode LSNs are shown in Figures 2 and 3. In Callao, six of the top ten venues most visited by controls were large markets whereas seven of the top ten venues visited by cases were hospitals or clinics (Table S4). In Lima Sur, clinics and markets were the most commonly shared venues among cases. One cinema in Lima Sur that was reported by eight (14%) cases (Table S4). Excluding healthcare venues, the most commonly reported venue type among cases were markets. Four markets were shared by two or more cases and controls (Table S5). Two-mode networks, stratified by cases and region, are provided in Figure S2.

There were eight unique genetic clusters ranging in size from two to seven cases and epidemiologic links (i.e., shared venues) were found in all but one cluster. The clusters are described

and visualized in Table S6. Healthcare venues were found in seven clusters. Markets, transportation and cinemas were also links among clustered cases. Two cases in one cluster reported spending time in the same prison. In one of two clusters with cross region cases, large markets and hospitals were the primary epidemiologic links (Table S6).

Among cases, network density was greatest through shared healthcare venues, markets and cinemas (Table 3). Among controls, network density was greatest through shared markets, followed by stores then healthcare venues. No cases were connected through shared stores and no controls were connected through shared bars/clubs. Network density based on cinemas was highest among clustered cases (Table 3). Network densities stratified by region are reported in Table S7.

The final ERGM for the LSN of both regions showed significant community segregation, as indicated by a statistically significant homophily effect among cases (Coef=2.03, P<0.0001) and controls (Coef=0.27, P=0.002) (Table S8). Model estimates indicate the probability that two cases in Callao with similar education and income levels frequenting the same venue would be 27.4%. In comparison, the probability two controls in Callao with similar education and income levels frequenting the same venue would be 6.1%. In other words, MDR-TB cases were 4.5 times more likely than controls to report frequenting the same community venue.

Discussion

Our analysis demonstrated that MDR-TB patients tended to spend significantly more time than matched controls in community settings, particularly in healthcare, schools and transportation venues (i.e., bus stops/lines, train stations and airports). Healthcare venues, markets and cinemas were the most common epidemiologic link between cases and may be potential sites of transmission. Markets were frequently shared by both cases and controls and potential sites for infectious-to-susceptible exposure. These findings highlight opportunities for health ministries to implement environmental improvements, such as enhanced ventilation and UV irradiation, to reduce the risk of community exposure and transmission at these venues. Reducing transmissibility

within MDR-TB hotspots by targeting environmental interventions in venues where exposure risk is high may have considerable community-wide benefits [30]. Our study demonstrates the utility of combining venue-tracing questionnaires with whole-genome sequencing to identify new areas of potential transmission.

Network density based on all venues was greater in the LSN of cases compared to that of controls, indicating that cases are more inter-connected through community venues than controls. The stratified analysis showed that healthcare venues were the primary venue connecting cases and are undoubtedly important venues for infection control. In addition to healthcare venues, LSNs based on markets, cinemas, bars, and transportation also had high case network density and may be candidate venues for MDR-TB infection control or case finding. Greater network density has been correlated with increased infection risk in simulation studies and may explain how outbreaks of TB tend to circulate within defined sub-populations [15].

Network analysis showed that case and control communities are highly segregated [31], with strong separation of cases and controls into distinct network communities. This finding suggests that environmental interventions should be applied across the distinct network of venues frequented by cases rather than in one site alone. Furthermore, resources should not be focused on venues frequented only by controls where the risk of exposure to MDR-TB may be very low. Social segregation may be a key driver of MDR-TB in this community, reflected in the observed difference in income and education levels between cases and controls. Segregation may have direct effects on the transmission of TB by concentrating impoverished populations into small geographic areas with poor housing and limited access to healthcare [32]. Spatial analyses of MDR-TB cases have found hotspots of localized transmission in areas of Lima, suggesting that transmission of MDR-TB occurs in small geographic areas that may have formed by socially segregated enclaves [33, 34].

Our results are supported by previous studies that have found healthcare venues to be high risk environments for MDR-TB transmission in Lima [33-35]. Venue-tracing studies in Uganda have

found that nearly all identified clusters of unrelated TB patients were only epidemiologically linked through shared healthcare venues [9, 13]. We are unable to confirm if cases were exposed to MDR-TB and infected during healthcare visits or if cases were infected elsewhere and visited clinics more frequently to address reactivated disease. However, the healthcare venues reported by cases were visited prior to diagnosis so the association between cases and healthcare utilization is unlikely to be confounded by TB treatment uptake [35].

Schools and transportation venues have been reported as sites of significant community contact where prolonged and close exposure to infectious cases may be likely [10, 36]. While our results are similar to previous studies of drug-sensitive TB cases [9, 13], we cannot definitively conclude the venues identified are generalizable to drug-sensitive TB populations given the use of a healthy TB-free control group; however, the advantage of using TB-free controls is the identification of venues where susceptible-to-infectious cross exposure is possible. MDR-TB is associated with greater catastrophic costs due to treatment than drug-sensitive TB and MDR-TB patients are more likely to have adverse outcomes, likely influencing mobility and community venue types frequented; however, none of the cases were hospitalized and sufficiently ambulant to visit clinics and other community venues. [2].

There were several limitations to this study. Proxy interviews were required for 18 (31%) of the cases because they were either too ill or had died by the time of interview; however, there were no demographic differences between proxy and primary cases indicating any potential bias may be minimal. Field data on participation refusals were unfortunately not digitized, precluding a comparison analysis of case participants and non-participants. The timeframes of interviews were also different for cases and controls; cases were interviewed about venues frequented during the month prior to diagnosis while controls were interviewed about venues frequented in the month prior to interview. Focusing on the month prior to diagnosis may have biased the frequency of healthcare utilization upwards as cases seek care for pre-diagnosis illness. Future studies might focus

on case recall periods before symptoms occur to ascertain venue exposure prior to any illness mediated changes in venue utilization habits. Nevertheless, since cases are highly infectious prior to treatment, focusing on venues visits prior to diagnosis is important for identifying places of concentrated infectious case activity where MDR-TB exposure may occur. Fixing recall periods for controls to the month prior to interview may also have introduced differential recall bias by fixing temporality and seasonality as compared to cases; differentials in the frequency of national holidays or vacation seasons could have biased reported venue utilization.

Statistical adjustment was used to account for the significant difference in income and education among cases and controls; however, given the use of broad binary indicators to control for socioeconomic status, residual confounding may still be a concern. Future studies should focus on detailed collection of socioeconomic data or control for socioeconomic status in the study design.

Several cases had prior episodes of TB, suggesting resistance could have been acquired rather than transmitted, limiting our ability to conclude cases were infected in reported venues. However, over half the cases with previous TB were genetically clustered, providing support for recent transmission and past studies have documented the high frequency of community-based MDR-TB transmission in Lima [3]. The accuracy of SNP thresholds to determine genetic clustering is dependent upon a number of factors, including strain diversity, mutation rates, complete case detection and identification of epidemiologic links between cases [37]. Strain diversity in Lima has been well documented, suggesting small SNP differences are likely instances of transmission [3, 38, 39]. The use of a 5 SNP threshold in our study was informed by Walker et al. who found epidemiologic links between cases with up to 5 SNP differences, as was the case in the present study [22]. Increasing the threshold to ≤10 SNPs did not have appreciable effect on the number of clustered cases or epidemiologic links found (Table S9). Finally, we did not complete an environmental assessment of each venue and are unable to comment on transmission risk in specific venues.

In summary, MDR-TB patients spent significantly more time in distinct and segregated healthcare, schools and transportation venues which were associated with increased odds of MDR-TB disease. The LSN analysis revealed that genetically clustered MDR-TB patients were distinct and significantly segregated from matched controls. This finding suggests the need to concentrate interventions at venues utilized and shared by high risk communities, which are not utilized by low risk communities, in order to have maximum impact on controlling disease spread.

Conflict of Interest

All authors have submitted the ICMJE Form for Disclosure of Potential Conflicts of Interest, and have no potential conflicts of interest to disclose.

References

1. Fauci AS, the NIAID Tuberculosis Working Group. Multidrug-Resistant and Extensively Drug-Resistant Tuberculosis: The National Institute of Allergy and Infectious Diseases Research Agenda and Recommendations for Priority Research. J Infect Dis **2008**; 197(11): 1493-8.

2. Wingfield T, Boccia D, Tovar M, et al. Defining Catastrophic Costs and Comparing Their Importance for Adverse Tuberculosis Outcome with Multi-Drug Resistance: A Prospective Cohort Study, Peru. PLoS Med **2014**; 11(7): e1001675.

3. Cohen T, Murray M, Abubakar I, et al. Multiple introductions of multidrug-resistant tuberculosis into households, Lima, Peru. Emerg Infect Dis **2011**; 17(6): 969-75.

4. Madico G, Gilman RH, Cabrera L, et al. Community infection ratio as an indicator for tuberculosis control. The Lancet **1995**; 345(8947): 416-9.

5. Otero L, Krapp F, Tomatis C, Zamudio C, Matthys F, Gotuzzo E. High prevalence of primary multidrug resistant tuberculosis in persons with no known risk factors. PLoS One **2011**; 6.

6. Yates TA, Khan PY, Knight GM, et al. The transmission of Mycobacterium tuberculosis in high burden settings. The Lancet Infectious Diseases **2016**; 16(2): 227-38.

7. Classen CN, Warren R, Richardson M, et al. Impact of social interactions in the community on the transmission of tuberculosis in a high incidence area. Thorax **1999**; 54(2): 136-40.

8. Yates TA, Tanser F, Abubakar I. Plan Beta for tuberculosis: it's time to think seriously about poorly ventilated congregate settings. The International Journal of Tuberculosis and Lung Disease **2016**; 20(1): 5-10.

9. Chamie G, Wandera B, Marquez C, et al. Identifying locations of recent TB transmission in rural Uganda: a multidisciplinary approach. Trop Med Int Health **2015**; 20(4): 537-45.

10. Wood R, Racow K, Bekker LG, et al. Indoor social networks in a South African township: potential contribution of location to tuberculosis transmission. PLoS One **2012**; 7(6): e39246.

11. Gardy JL, Johnston JC, Sui SJH, et al. Whole-Genome Sequencing and Social-Network Analysis of a Tuberculosis Outbreak. N Engl J Med **2011**; 364(8): 730-9.

12. Klovdahl AS, Graviss EA, Yaganehdoost A, et al. Networks and tuberculosis: an undetected community outbreak involving public places. Soc Sci Med **2001**; 52(5): 681-94.

13. Chamie G, Kato-Maeda M, Emperador DM, et al. Spatial overlap links seemingly unconnected genotype-matched TB cases in rural Uganda. PLoS One **2018**; 13(2): e0192666.

14. Frost SDW. Using sexual affiliation networks to describe the sexual structure of a population. Sex Transm Infect **2007**; 83(suppl 1): i37-i42.

15. Christley RM, Pinchbeck GL, Bowers RG, et al. Infection in social networks: using network analysis to identify high-risk individuals. Am J Epidemiol **2005**; 162(10): 1024-31.

16. Christakis NA, Fowler JH. Social Network Sensors for Early Detection of Contagious Outbreaks. PLoS One **2010**; 5(9).

17. Read JM, Eames KTD, Edmunds WJ. Dynamic social networks and the implications for the spread of infectious disease. Journal of the Royal Society Interface **2008**; 5(26): 1001-7.

18. Kitsak M, Gallos LK, Havlin S, et al. Identification of influential spreaders in complex networks. Nature Physics **2010**; 6(11): 888-93.

19. Newman MEJ. Spread of epidemic disease on networks. Physical Review E - Statistical, Nonlinear, and Soft Matter Physics **2002**; 66(1): 016128/11.

20. Alarcon V, Alarcon E, Figueroa C, Mendoza-Ticona A. Tubersulosis en el Perú: situación epidemiológica, avances y desafíos para su control. Rev Peru Med Exp Salud Publica **2017**; 34: 299-310.

21. Grandjean L, Gilman RH, Martin L, et al. Transmission of Multidrug-Resistant and Drug-Susceptible Tuberculosis within Households: A Prospective Cohort Study. PLoS Med **2015**; 12(6): e1001843.

22. Walker TM, Ip CLC, Harrell RH, et al. Whole-genome sequencing to delineate Mycobacterium tuberculosis outbreaks: a retrospective observational study. The Lancet Infectious Diseases **2013**; 13(2): 137-46.

23. Mathema B, Andrews JR, Cohen T, et al. Drivers of Tuberculosis Transmission. The Journal of Infectious Diseases **2017**; 216(suppl_6): S644-S53.

24. StataCorp. Stata Statistical Software: Release 14. College Station, TX: StataCorp LP, **2015**.

25. Robins G, Pattison P, Kalish Y, Lusher D. An introduction to exponential random graph (p) models for social networks. Social Networks **2007**; 29(2): 173-91.

26. Hunter DR, Goodreau SM, Handcock MS. Goodness of Fit of Social Network Models. Journal of the American Statistical Association **2008**; 103(481): 248-58.

27. Csardi G, Nepusz T. The igraph software package for complex network research. InterJournal **2006**; Complex Systems: 1695.

28. Handcock MS, Hunter DR, Butts CT, Goodreau SM, Morris M. statnet: Software Tools for the Representation, Visualization, Analysis and Simulation of Network Data. Journal of statistical software **2008**; 24(1): 1548-7660.

29. R Core Team. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing, **2017**.

30. Dowdy DW, Golub JE, Chaisson RE, Saraceni V. Heterogeneity in tuberculosis transmission and the role of geographic hotspots in propagating epidemics. Proc Natl Acad Sci U S A **2012**; 109(24): 9557-62.

31. Bojanowski M, Corten R. Measuring segregation in social networks. Social Networks **2014**; 39(Supplement C): 14-32.

32. Acevedo-Garcia D. Residential segregation and the epidemiology of infectious diseases. Soc Sci Med **2000**; 51(8): 1143-61.

33. Zelner JL, Murray MB, Becerra MC, et al. Identifying Hotspots of Multidrug-Resistant Tuberculosis Transmission Using Spatial and Molecular Genetic Data. J Infect Dis **2015**.

34. Shah L, Choi HW, Berrang-Ford L, et al. Geographic predictors of primary multidrug-resistant tuberculosis cases in an endemic area of Lima, Peru. The International Journal of Tuberculosis and Lung Disease **2014**; 18(11): 1307-14.

35. Escombe AR, Moore DAJ, Gilman RH, et al. The Infectiousness of Tuberculosis Patients Coinfected with HIV. PLoS Med **2008**; 5(9): e188.

36. Zamudio C, Krapp F, Choi HW, et al. Public Transportation and Tuberculosis Transmission in a High Incidence Setting. PLoS One **2015**; 10(2): e0115230.

37. Hatherell H-A, Colijn C, Stagg HR, Jackson C, Winter JR, Abubakar I. Interpreting whole genome sequencing for investigating tuberculosis transmission: a systematic review. BMC Med **2016**; 14(1): 21.

38. Iwamoto T, Grandjean L, Arikawa K, et al. Genetic Diversity and Transmission Characteristics of Beijing Family Strains of Mycobacterium tuberculosis in Peru. PLoS One **2012**; 7(11): e49651.

39. Sheen P, Couvin D, Grandjean L, et al. Genetic Diversity of Mycobacterium tuberculosis in Peru and Exploration of Phylogenetic Associations with Drug Resistance. PLoS One **2013**; 8(6): e65873.

Figure Legends

Figure 1. Total venue reported and person-hours reported by case and control, stratified by venue type. a) Bars indicate the total number of venues reported by cases and controls, stratified by venue type; grey bars are cases and white bars are controls. b) bars indicate the total number of person-time hours reported by cases and controls, stratified by venue type; grey bars are cases and white bars are controls.

Figure 2. Two-mode and one-mode (location-based social networks) of cases and controls across both regions. The two-mode networks represent the participants connected to venues they reported frequenting. Cases are represented as red circles and controls are represented as blue circles. Venues are represented as white boxes and with a numeric code corresponding to the venue type. In the one-mode network, participants are connected if they report frequenting the same venue. For clarity, only venues reported by at least two participants are visualized. The venue codes are: 1=Healthcare, 3=Market, 4=Store, 5=Church, 7=Restaurant, 9=Cinema, 11=School, 12=Transportation.

Figure 3. Two-mode and one mode (location-based social networks) of cases and controls stratified by region. The two mode networks represent the participants connected to venues they reported frequenting. Cases are represented as red circles and controls are represented as blue circles. Venues are represented as white boxes and with a numeric code corresponding to the venue type. In the one-mode network, participants are connected if they report frequenting the same venue. For clarity, only venues reported by at least two participants are visualized. The venue codes are: 1=Healthcare, 3=Market, 4=Store, 5=Church, 7=Restaurant, 9=Cinema, 11=School, 12=Transportation.

Table 1. Demographic and behavioral characteristics of MDR-TB cases and controls

	Cases	Controls	P-value*
	N=59	N=65	
Region, n (%)			P=0.99
Callao	30 (50.9)	33 (50.8)	
Lima Sur	29 (49.1)	32 (49.2)	
Age (years), mean (sd)	35.1 (15.1)	35.7 (14.4)	P=0.82
Male, n (%)	39 (66.1)	39 (60.0)	P=0.48
Household Size, mean (sd)	5.8 (2.7)	6.3 (2.8)	P=0.32
Marital Status, n (%)			P=0.42
Single	26 (44.1)	24 (36.9)	
Married	27 (45.8)	37 (56.9)	
Education Attained, n (%)			P<0.001
Secondary or more	35 (59.3)	61 (93.9)	
Primary or less	24 (40.7)	4 (6.2)	
Monthly Income, n (%)			P<0.001
> 1000 PEN	15 (25.4)	52 (80.0)	
≤ 1000 PEN	44 (74.6)	13 (20.0)	
Employment, n (%)			
Unemployed	10 (17.0)	6 (9.2)	P=0.37
Employed	42 (71.2)	48 (73.9)	
Other**	7 (11.8)	11 (16.9)	
Treatment Outcome, n (%)			
Completed	42 (71.2)		
Ongoing	2 (3.4)		
Abandoned	15 (25.4)		
Venues Reported, mean (sd)	8.3 (3.0)	8.4 (5.7)	P=0.92

* P-values based on t-tests (continuous) and chi-square tests (categorical)
** includes students and retired

Table 2. Mean weekly person-time (hours) at each venue type and adjusted odds ratios (OR) for the association between person-time (per 10 hours) and MDR-TB disease by venue type.

Venue Type	Cases (N=59) Mean Hours (sd)	Controls (N=63) Mean Hours (sd)		Logistic Regression Results†		
			P *	OR	95%CI	*P*
All Venues	81.7 (49.1)	68.7 (38.5)	**0.004**	1.15	1.02 - 1.30	**0.02**
Healthcare	15.3 (18.2)	4.9 (13.6)	**<0.001**	1.67	1.17 - 2.38	**0.01**
Transportation	15.1 (29.6)	5.8 (18.3)	**0.003**	1.25	1.02 - 1.53	**0.03**
School	14.3 (26.2)	5.9 (13.2)	0.28	1.53	1.16 - 2.01	**0.003**
Market	13 (22.7)	13.7 (20.3)	0.05	0.97	0.79 - 1.21	0.81
Store	4.8 (13.2)	11.1 (27.4)	0.20	0.80	0.61 - 1.06	0.12
Office	4.6 (17.6)	12.1 (24.9)	0.38	0.93	0.74 - 1.17	0.52
Restaurant	3.5 (13.2)	2.4 (7.3)	0.15	0.95	0.64 - 1.43	0.82
Residence	2.9 (8.3)	3.2 (10.7)	0.06	0.97	0.56 - 1.67	0.91
Bar/Club	2.8 (10.6)	0.7 (3.4)	0.08	1.66	0.60 - 4.61	0.33
Other	2.4 (10.7)	4.2 (13.6)	0.37	0.93	0.65 - 1.33	0.68
Parks/Recreation	1.7 (5.6)	1.4 (4.8)	0.17	1.42	0.61 - 3.27	0.42
Church	0.8 (1.5)	2.6 (7.5)	0.15	0.18	0.01 - 2.64	0.21
Cinema	0.6 (1.8)	0.8 (2.2)	0.91	1.11	0.11 - 11.31	0.93

* Based on ordinal logistic regression adjusted for income and education; bold P-values are significant (P<0.05)

† age, sex, region, education and income adjusted

Table 3. Network densities of location-based social networks by venue type and case type. Density is measured as the proportion of actual ties among nodes in a network divided by the number of all total possible ties in a network. Network densities of location-based social networks represent the proportion of a network connected through shared venues.

Both Regions					
Venue Type*	**All**	**Controls**	**Cases**	**Clustered Cases**	**Non-clustered Cases**
All Venues	12.5%	8.3%	28.8%	27.9%	32.2%
Healthcare	9.1%	1.1%	27.6%	26.4%	30.4%
Market	3.0%	5.5%	2.3%	2.4%	5.8%
Store	0.4%	1.3%			
Church	0.1%	0.4%	0.1%		
Bar/Club	0.1%		0.2%	0.4%	
Restaurant	0.2%	0.3%	0.1%	0.1%	
Cinema	0.6%	0.3%	1.7%	2.1%	0.6%
School	0.1%	0.2%	0.1%	0.1%	
Transportation	0.3%	0.3%	0.2%	0.4%	

* Office, Residence, Other and Parks/Rec omitted because they did not connect participants (i.e., network densities were zero). Empty cells represent zero network density.