Journal of Infectious Diseases Advance Access published March 8, 2014

Integrating social contact and environmental data in evaluating tuberculosis transmission in a South African township

Jason R. Andrews[a], Carl Morrow[b], Rochelle P. Walensky[a], Robin Wood[b]
[a]Division of Infectious Diseases, Massachusetts General Hospital, Boston, MA, 02120
[b]Desmond Tutu HIV Centre, Institute of Infectious Disease and Molecular Medicine, University of Cape Town, Cape Town, Republic of South Africa
Correspondence: Dr. Jason R. Andrews, Division of Infectious Diseases, 50 Staniford Street, 9th Floor, Massachusetts General Hospital, Boston, MA 02114, Jandrews6@partners.org

Summary

Background: Population models of tuberculosis transmission have not accounted for social contact structure and the role of the environment in which tuberculosis is transmitted.

Methods: We utilized extensions to the Wells-Riley model of tuberculosis transmission, using exhaled carbon dioxide as a tracer gas, to describe transmission patterns in an endemic community. Drawing upon social interaction data and carbon dioxide measurements from a South African township, we created an age-structured model of tuberculosis transmission in households, public transit, schools and workplaces. We fit the model to local data on latent tuberculosis prevalence by age.

Results: Most tuberculosis infections (84%) were estimated to occur outside of one's own household. 50% of infections among young adults (ages 15-19) occurred in schools, due to high contact rates and poor ventilation. Despite lower numbers of contacts in workplaces, assortative mixing among adults with high rates of smear-positive tuberculosis contributed to transmission in this environment. Households and public transit were important sites of transmission between age groups.

Conclusions: Consistent with molecular epidemiologic estimates, a minority of tuberculosis transmission was estimated to occur within households, which may limit the impact of contact investigations. Further work is needed to investigate the role of schools in tuberculosis transmission.

Introduction

Tuberculosis transmission has been classically modeled as a mass-action process in a homogenous population, wherein individuals randomly contact one another and in which types and duration of contacts are not considered [1–3]. There is substantial evidence to suggest that the duration and environment in which contacts occur are key determinants of the risk of tuberculosis transmission [4]. Additionally, as with many other diseases, age-specific mixing patterns likely play an important role in tuberculosis transmission dynamics.

Tuberculosis is spread by droplet nuclei, which may remain suspended in the air for thirty minutes in the absence of air exchange. The impact of the indoor environment in tuberculosis transmission was first rigorously studied in the late 1950's by Wells and Riley, who exposed guinea pigs to air from a tuberculosis ward and measured infection rates under controlled conditions [5,6]. Models derived from the Wells-Riley findings have focused on single environments to characterize point-source outbreaks or nosocomial transmission [7,8]. Environmental settings have not been factored into broader models evaluating endemic transmission. This may be in part due to logistical difficulties in measuring ventilation, as well as characterizing contact patterns in various environments. The former obstacle has been surmounted in part due to the use of exhaled carbon dioxide as a natural tracer gas to evaluate air exchange [9], which can be done with increasingly low-cost and portable devices. Additionally, recent efforts to quantify social contact patterns have yielded rich data on age-assortativeness and variability of contacts between settings [10–12].

By integrating data on social interactions and environmental context, we developed a new approach to project the impact of age-specific contact patterns and estimate where tuberculosis transmission occurs. We utilized local data from a study of social interactions together with measurements of carbon dioxide in common indoor environments, to model tuberculosis transmission in a South African township.

Methods

Wells and Riley derived an equation describing the risk of tuberculosis infection (P) in an indoor environment as a function of the number of infectious individuals in a space (I), the breathing rate (p), the rate of generating infectious quanta (q; quanta/hour), the duration of exposure (t) and the room ventilation rate (Q) [6]:

$$P = 1 - \exp\left(-\frac{Ipqt}{Q}\right)$$

Rudnick and Milton extended this work to non-steady state situations and described the use of carbon dioxide as a natural tracer gas to overcome the need for resource-intensive measurements of conventional room ventilation analysis [13]. This approach involves estimating the proportion of air in a room that was expired by its occupants (the 'rebreathed fraction', f) by evaluating the excess CO_2 in the room over outdoor air CO_2. Using this rebreathed fraction, the probability of infection, as a function of time and number of infectious individuals (I) among all individuals (n) in a room, can be estimated through the following relationship (see **Supplementary Material**):

$$P = 1 - \exp\left(-\frac{\bar{f}Iqt}{n}\right)$$

We extended this work to describe endemic tuberculosis transmission in an age-structured model with multiple transmission environments.

We drew upon data from a 571 residents in a Cape Town township that contained data on the number of indoor contacts and time spent in various locations, stratified by ten age groups (0-4, 5-9, 10-14, 15-19, 20-24, 25-29, 30-34, 35-39, 40-44, ≥45 years) [14]. Over 97% percent of indoor contact-time occurred in five locations: own households, other households, schools, workplaces and public transit [14]. While the majority of time spent indoors was spent in households (Figure S1), most daily contacts occurred outside the home, with school accounting for the largest number of contacts among children and transit and workplaces serving as the environment for most contacts among adults (Figure S2). Compared with other settings, the number of contacts observed in public transit were highly dispersed with a long right tail; a number of individuals reported a very high contact number (>50) (Figure S1).

Additionally, using portable carbon dioxide (CO_2) detection devices (EasyView® 80 CO_2 analyser, Extech Instruments, Waltham, MA and custom-developed monitors using COZIR™ Ambient sensors, Gas Sensing Solutions Ltd, Glasgow), ambient air in four environments—public transit vehicles, schools/crèches, workplaces and households—was sampled by nine volunteers to assess mean and ranges for CO_2 concentration (Table 1). Volunteers collected 17,124 observations of CO_2 concentration in various locations throughout day and night. We used the mean and standard error of sets of concentration observations for each setting. We utilized tuberculosis notification data (29,478 cases reported in 2009) and estimated age-stratified incidence rates of smear-positive and smear-negative tuberculosis.

We assessed the risk of transmission in each environment as a function of: the number of individuals in the environment, the amount of time spent in the environment, the rebreathed fraction calculated from the carbon dioxide concentration, and the probability that there was an infectious individual present in the environment given age-specific tuberculosis prevalence, estimated from age-stratified notification data (Figure S3) and adjusted for population age structure estimated by census data (Figure S4). A contact matrix reflecting the age assortativeness of mixing was estimated by weighting reported number of contacts for each age group by the proportion of contact-time reported for each age group in each environment; in other words, the probability of a person with age i contacting a person with age j in environment k is proportionate to the fraction of total contact time in environment k that was reported by individuals of age j (See **Supplementary Materials)**. Therefore, exposure to other individuals was proportionate to the amount of time spent in each environment; the exception was interactions in schools, which were modeled such that all contact time among children was within their own 5-year age strata. This was done because mixing between age groups in schools was thought to be limited, as children spend most of their time in classrooms with other children of the same age. We weighted the infectiousness of smear-negative tuberculosis by 0.2, consistent with estimates from several contact investigation studies [15–17].

Using the risk derived from the modified Wells-Riley equation, we projected the annual risk of tuberculosis infection (ARI), by age, from the daily risk of infection in each environment. We then used these age-structured estimates for ARI to project the latent tuberculosis prevalence by age, according to the following equation [18]:

$$\text{Latent tuberculosis prevalence at age } a = 1 - \prod_{i=1}^{a} (1 - \text{ARI}(i))$$

We fit our model to age-structured data on latent tuberculosis prevalence from the same study community as the survey [19] by varying the quantum production rate and using a simplex descent algorithm to minimize the least-squares residuals between modeled and observed data.

To evaluate uncertainty in our projections resulting from our data, we used Latin Hypercube Sampling to draw from distributions of contacts and time spent in each environment for each age group and CO_2 measurements for each environment [20]. We drew from 1,000 sets of parameters to generate median and 95% credible intervals. We used broad one-way sensitivity analysis to evaluate two key parameters for which there is great uncertainty: the duration of tuberculosis prior to diagnosis and the proportion of contacts in each setting that are recurring. We generated estimates of the quantum production rate and the proportion of tuberculosis transmitted in each setting.

This model drew upon data that were collected as part of a study that was approved by the Human Research Ethics Committee of the University of Cape Town. Written informed consent was obtained from all participants. Parental/ guardian consent was obtained for participants under 18 years of age, and signed assent forms were obtained from adolescents aged 12–17 years.

Results

We found significant assortativeness of contact-time among children in schools and among adults in workplaces, in contrast to general mixing between children and adults that occurred in households and in public transit (Figure 1). Carbon dioxide levels

were highest in schools and public transit, followed by homes and workplaces (Table 1). There were higher levels of variation in CO_2 estimates from public transit compared with other environments.

We fit the modeled latent tuberculosis prevalence to empiric measurements from a recent community survey [19] (Figure 2). The projected age-weighted mean annual risk of tuberculosis infection overall was 6.6 % (95% credible interval: 3.1%-10.6%); the risk varied from 4.4% among children age 5-9 to 9.1% among young adults ages 15-19 (Figure S5). The majority of tuberculosis transmission was estimated to occur outside of one's own households, for all age groups (weighted mean percentage transmission within own households: 15.6%) (Figure 3). Schools/crèches accounted for about a quarter of infections among children aged 0-14; however, up to half of tuberculosis transmission in young adults age 15-19 was estimated to occur in schools. A substantial proportion of tuberculosis transmission among adults was estimated to take place in workplaces (weighted mean: 41.4%), where there was high level mixing among age groups with the highest prevalence of smear-positive tuberculosis (Figure S3). Owing to migration in recent years, working age adults represent the largest population in this community, which strongly influenced the weighted mean for transmission in workplaces. Public transit was estimated to account for 7.9-34.8% of transmission in all age groups (weighted mean 21.9%).

Among the most uncertain parameters in modeling transmission of tuberculosis through Wells-Riley approaches is the rate of infectious quanta production, *q*. We found that two factors, both of which are not well known, substantially impact the estimate of *q*. The first is the duration of infectiousness of undiagnosed tuberculosis, which has been

variably estimated as being from 4 months to over 18 months [21,22]. With shorter duration of infectiousness, a higher quantum production rate is required to achieve the same force of infection. We used a base case of 12 months as this reflected local data from this community [21]. Our estimate of *q* with this duration of infectiousness was 0.89 quanta/hour (table S1), but varied from 5.69 quanta/hour to 0.44 quanta/hour with duration of 4 to 18 months.

The second most influential factor in the estimate of *q* was the proportion of contacts that were recurrent. We evaluated three scenarios: 1) all contacts in all settings were recurrent; 2) no contacts were recurrent; 3) all contacts were recurrent in households, schools and workplaces, but not in public transit. Recurrent contacts implied that the same individuals were contacted each day; for example, if a person contacted three individuals per day in the household, it was the same three individuals every day (e.g., family members). We assumed the third scenario for our base case. Our estimate of *q* varied from 0.27 quanta/hour in scenario 2 (none recurrent) to 0.94 quanta/hour in scenario 1 (all recurrent). The scenario of no recurrent contacts had poor model fit compared with scenarios 1 and 3.

Despite the uncertainty wrought by these parameters, the main results in terms of where tuberculosis is transmitted were robust to reasonable uncertainty (Table S1). The percentage of tuberculosis cases transmitted in the household varied from 5.4% to 24.2% when varying duration of infectiousness from 4 to 18 months, respectively, and 59.5-15.5% when varying proportion of recurrent contacts from 0-100%.

Discussion

Mathematical models of tuberculosis transmission have relied upon abstract effective contact rates that do not account for the environment of contacts or the mixing patterns that characterize social interactions [1–3]. We have demonstrated how measurements of carbon dioxide in indoor environments can be combined with social mixing data to describe tuberculosis transmission in an endemic community. While this approach is subject to limitations in terms of data on transmission risks and simplifying assumptions concerning social interactions, it nevertheless offers a straightforward approach to understanding the importance of various environments and contact patterns in driving tuberculosis transmission. Moreover, this approach can be done using easily obtainable data and simple deterministic models, in contrast to detailed social network enumeration approaches [23,24], and could be applied to other infectious diseases spread by droplet nuclei, such as measles and varicella.

Epidemiologists have long cited crowding and population density in urban environments as key factors contributing to the spread of tuberculosis [25,26]; however, a detailed understanding of how these factors influence tuberculosis transmission has been lacking. Data used for this model, for example, demonstrate that close indoor contact rates in South African townships are substantially higher than Europe and rural Vietnam [10,11]. By integrating this data with duration of exposure and the ventilation characteristics of the environment, we estimate that a substantial proportion of tuberculosis transmission occurs outside of households. Indeed, molecular epidemiologic studies have suggested that only 19% of tuberculosis cases in South African townships arise from transmission within one's own household [27], while our model similarly

estimated that 16% of TB transmission occurs in this setting. Age-stratified molecular epidemiologic data on household transmission are not currently available, but as these data emerge in the future, they could be used to further validate models of this nature. However, because of delay between infection and onset disease in tuberculosis, it is difficult to use molecular epidemiologic approaches to evaluate the location of transmission, particularly between casual or non-recurring contacts. Our approach enables the prediction of where tuberculosis is transmitted, as well as which age groups are driving transmission. The use of detailed social network studies combined with recent advantages in whole genome sequencing of *Mycobacterium tuberculosis* may be useful in validating predictions from this approach [23,24,28,29].

By understanding where tuberculosis transmission occurs, we may be able to better project the impact of control interventions. For example, contact investigations targeting household members may fail to identify the majority of tuberculosis cases in this setting. In other settings, household contacts may account for a larger proportion of tuberculosis transmission, and simple data can provide insights to inform this. Schools are characterized by high social contact rates and high proportions of rebreathed air; we projected that this was a particularly high risk environment for transmission among young adults ages 15-19. Among younger children, who have a lower prevalence of smear-positive tuberculosis, infection risk in schools was estimated to be more limited. In contrast to schools, households and public transit are important areas for general mixing between age groups and likely drive transmission from adults to children. This modeling approach could be used to evaluate thresholds for CO_2 levels in congregate settings, such as schools, public transit, and workplaces and to project the impact of environmental

interventions, such as ventilation systems (mechanical or natural) or the use of upper room ultraviolet germicidal irradiation.

Among the most important and uncertain parameters in understanding the transmission of tuberculosis is the infectious quanta production rate. Wells and Riley estimated this in their original experiments to be 1.25 quanta/hour, derived from hospitalized patients at the beginning of their tuberculosis treatment [6]. Escombe and colleagues repeated these studies in a tuberculosis ward in Peru and reported considerably variability, from 0 quant/hour to over 200 quanta/hour [30]. Their mean estimate was 8.2 quanta/hour, which was affected by a few outliers. However, both sets of estimates were made by studying the infections induced among guinea pigs by hospitalized patients with a confirmed diagnosis of tuberculosis. The extent to which this can be generalized to individuals in a community setting with undiagnosed disease, many of whom may have few to no symptoms, is unclear. We estimated this rate by fitting our model to data on latent infections. Our estimates were heavily dependent upon assumptions about the recurrence of contacts and the duration of infectiousness, and ranged from 0.27 quanta/hour to 5.69 quanta/hour, with a base case estimate of 0.89 quanta/hour. Recurrence of contacts leads to local saturation of infections (by depletion of susceptible contacts of an infectious person) and requires a higher quantum generation rate to achieve a comparable population level of transmission. In general, however, these estimates were lower than those of Wells-Riley and Escombe and colleagues, which may indicate that individuals are less infectious earlier in their course of illness. More detailed enumeration of social contact patterns, including frequency of recurrent contacts, may lead to improved estimates of infectiousness.

Our approach and results are subject to limitations of our assumptions and available data. We fit our model to age-structured data on latent tuberculosis prevalence, which does not account for cohort effects of differences in tuberculosis transmission rates over time. Estimates of the annual risk of infection have remained fairly stable for at least 15 years in this area [31], though earlier differences in the annual risk of infection, if present, would lead to errors in estimating the annual risk of infection particularly in older adults. We projected latent tuberculosis prevalence by age according to annual risk of infection, as commonly done in tuberculosis epidemiology [18]; however, this doesn't account for mortality due to tuberculosis, which has a small impact on latent tuberculosis prevalence estimates. We did not consider 'self-quarantine' in this model, wherein individuals who feel ill due to undiagnosed disease stay at home, reducing their contact with others [32]. Given the disparity in duration of reported symptoms prior to tuberculosis diagnosis (around 4 weeks in most studies) and estimates of the duration of infectiousness (12 months), it is likely that self-quarantine does not account for a significant reduction in mixing. We did not have empirical data on age-assortativeness of contacts; therefore, we estimated age-assortativeness based on age-specific time reported in each setting. Additionally, there are currently no available data on the proportion of contacts that are recurring, or the frequency of recurrences; such data could be obtained through self-report or novel approaches to defining social networks that involve radiofrequency identification tags [33]. Nevertheless, our results remained fairly robust to reasonable assumptions concerning these uncertain parameters [27]. Children are likely much less infectious than adults with tuberculosis; in the absence of better data, we assumed that differences in infectiousness were explained by smear status. This led to

almost negligible transmission from the youngest children in this model. We did not incorporate heterogeneity in infectiousness, though multiple experimental data suggest tremendous inter-individual variability in infectiousness [6,30,34]. Furthermore, there is likely intra-individual variability in infectiousness over time, as individuals develop higher burdens of disease. Combined with skewed distributions of contact rates and contact time, these factors may help explain the observed phenomena of super-spreading [35]. As better data on variability in infectiousness become available, future work with agent-based models may build on this approach to yield more nuanced insights into transmission patterns. We did not attempt to adjust notification data, which was stratified by smear status, for dynamic changes in smear status that likely occur over time. There are conflicting data on the smear-adjusted relative infectiousness of individuals with HIV infection [36,37] and the relative duration of active tuberculosis [21,22]; we used data from a large survey done in this community that found the duration of undiagnosed tuberculosis to be equal among HIV-infected and -uninfected individuals. Our results demonstrate that the duration of infectiousness is a key influential parameter in understanding the social and environmental contributions to tuberculosis transmission; further epidemiological studies are greatly needed to provide additional estimates of this parameter in other settings and to elucidate the inter-individual variability in duration of infectiousness. We did not explicitly model dynamics of progression from tuberculosis infection to disease, but rather focused on modeling location and age structure of transmission assuming a stable epidemic. This approach could easily be extended to dynamic models. As with ‘mass action’ models of infectious disease transmission, this

model does not account for clustering of tuberculosis within local social networks over time, which could further drive heterogeneity of tuberculosis risk.

There have been a number of studies in recent years characterizing social contact patterns in various settings [10–12]. We have demonstrated how such information can be combined with simple, easily obtainable environmental data to describe tuberculosis transmission patterns and estimate where transmission is taking place. This approach may enable us to better understand the impact of crowding and population density in sustaining tuberculosis endemicity, as well as to target interventions to improve control of tuberculosis.

Supplementary Materials: Please visit:

http://web2.research.partners.org/cepac/JIDsupp.html

Conflicts of Interest: We declare that we have no conflicts of interest.

Acknowledgements: This work was supported by the National Institute of Allergy and Infectious Diseases (R01 AI093269, R01 AI058736, T32AI007433, U01 AI069924, U01 AI069519) and the Harvard Global Health Institute. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health, which played no role in the study design, methods, interpretation of results, the content of this manuscript, or the decision to submit it for publication.

Conferences: This work was presented in part at the XLIX Congress of the Brazilian Society of Tropical Medicine in Campo Grande, Brazil, August 2013.

Correspondence:
Dr. Jason R. Andrews
Division of Infectious Diseases
50 Staniford Street, 9th Floor
Massachusetts General Hospital
Boston, MA 02114
Jandrews6@partners.org

References

1. Blower SM, McLean AR, Porco TC, et al. The intrinsic transmission dynamics of tuberculosis epidemics. Nat. Med **1995**; 1:815–821.

2. Murray CJ, Salomon JA. Modeling the impact of global tuberculosis control strategies. Proc. Natl. Acad. Sci. U.S.A **1998**; 95:13881–13886.

3. Dye C, Garnett GP, Sleeman K, Williams BG. Prospects for worldwide tuberculosis control under the WHO DOTS strategy. Directly observed short-course therapy. Lancet **1998**; 352:1886–1891.

4. Sepkowitz KA. How contagious is tuberculosis? Clin. Infect. Dis **1996**; 23:954–962.

5. Sultan L, Nyka W, Mills C, O'Grady F, Wells W, Riley RL. Tuberculosis disseminators. A study of the variability of aerial infectivity of tuberculous patients. Am. Rev. Respir. Dis. **1960**; 82:358–369.

6. Riley RL, Mills CC, O'Grady F, Sulton LU, Wittstadt F, Shivpuri DN. Infectiousness of air from a tuberculosis ward. Ultraviolet irradiation of infected air: comparative infectiousness of different patients. Am. Rev. Respir. Dis. **1962**; 85:511–525.

7. Nardell EA, Keegan J, Cheney SA, Etkind SC. Airborne infection. Theoretical limits of protection achievable by building ventilation. Am. Rev. Respir. Dis. **1991**; 144:302–306.

8. Pankhurst LJ, Anaraki S, Lai KM. Combining environmental assessment and contact investigations to make tuberculosis screening decisions. Int. J. Tuberc. Lung Dis. **2012**; 16:1023–1029.

9. Persily A. Evaluating building IAQ and ventilation with indoor carbon dioxide. ASHRAE Transactions **1997**; 103:1–12.

10. Mossong J, Hens N, Jit M, et al. Social contacts and mixing patterns relevant to the spread of infectious diseases. PLoS Med **2008**; 5:e74.

11. Horby P, Pham QT, Hens N, et al. Social contact patterns in Vietnam and implications for the control of infectious diseases. PLoS ONE **2011**; 6:e16965.

12. Johnstone-Robertson SP, Mark D, Morrow C, 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.

13. Rudnick SN, Milton DK. Risk of indoor airborne infection transmission estimated from carbon dioxide concentration. Indoor Air **2003**; 13:237–245.

14. Wood R, Racow K, Bekker L-G, et al. Indoor social networks in a South african township: potential contribution of location to tuberculosis transmission. PLoS ONE **2012**; 7:e39246.

15. Behr MA, Warren SA, Salamon H, et al. Transmission of Mycobacterium tuberculosis from patients smear-negative for acid-fast bacilli. Lancet **1999**; 353:444–449.

16. 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**; 47:1135–1142.

17. 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**; 59:286–290.

18. Rieder H. Annual risk of infection with Mycobacterium tuberculosis. Eur Respir J **2005**; 25:181–185.

19. Wood R, Liang H, Wu H, et al. Changing prevalence of tuberculosis infection with increasing age in high-burden townships in South Africa. Int. J. Tuberc. Lung Dis. **2010**; 14:406–412.

20. Stein M. Large sample properties of simulations using latin hypercube sampling. Technometrics **1987**; 29:143–151.

21. Wood R, Middelkoop K, Myer L, et al. Undiagnosed tuberculosis in a community with high HIV prevalence: implications for tuberculosis control. Am. J. Respir. Crit. Care Med. **2007**; 175:87–93.

22. Corbett EL, Bandason T, Cheung Y-B, et al. Prevalent infectious tuberculosis in Harare, Zimbabwe: burden, risk factors and implications for control. Int. J. Tuberc. Lung Dis. **2009**; 13:1231–1237.

23. Gardy JL, Johnston JC, Sui SJH, et al. Whole-Genome Sequencing and Social-Network Analysis of a Tuberculosis Outbreak. New England Journal of Medicine **2011**; 364:730–739.

24. 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:e1001387.

25. Cantwell MF, McKenna MT, McCray E, Onorato IM. Tuberculosis and race/ethnicity in the United States: impact of socioeconomic status. Am. J. Respir. Crit. Care Med. **1998**; 157:1016–1020.

26. Munch Z, Van Lill SWP, Booysen CN, Zietsman HL, Enarson DA, Beyers N. Tuberculosis transmission patterns in a high-incidence area: a spatial analysis. Int. J. Tuberc. Lung Dis. **2003**; 7:271–277.

27. Verver S, Warren RM, Munch Z, et al. Proportion of tuberculosis transmission that takes place in households in a high-incidence area. The Lancet **2004**; 363:212–214.

28. Kato-Maeda M, Ho C, Passarelli B, et al. Use of whole genome sequencing to determine the microevolution of Mycobacterium tuberculosis during an outbreak. PLoS ONE **2013**; 8:e58235.

29. Walker TM, Ip CLC, Harrell RH, et al. Whole-genome sequencing to delineate Mycobacterium tuberculosis outbreaks: a retrospective observational study. Lancet Infect Dis **2013**; 13:137–146.

30. Escombe AR, Moore DAJ, Gilman RH, et al. The infectiousness of tuberculosis patients coinfected with HIV. PLoS Med. **2008**; 5:e188.

31. Kritzinger FE, den Boon S, Verver S, et al. No decrease in annual risk of tuberculosis infection in endemic area in Cape Town, South Africa. Trop. Med. Int. Health **2009**; 14:136–142.

32. Chen F, Jiang M, Rabidoux S, Robinson S. Public avoidance and epidemics: Insights from an economic model. Journal of Theoretical Biology **2011**; 278:107–119.

33. Stehlé J, Voirin N, Barrat A, et al. Simulation of an SEIR infectious disease model on the dynamic contact network of conference attendees. BMC Med **2011**; 9:87.

34. Fennelly KP, Jones-López EC, Ayakaka I, et al. Variability of Infectious Aerosols Produced during Coughing by Patients with Pulmonary Tuberculosis. Am. J. Respir. Crit. Care Med. **2012**; 186:450–457.

35. Lloyd-Smith JO, Schreiber SJ, Kopp PE, Getz WM. Superspreading and the effect of individual variation on disease emergence. Nature **2005**; 438:355–359.

36. Carvalho AC, DeRiemer K, Nunes ZB, et al. Transmission of Mycobacterium tuberculosis to contacts of HIV-infected tuberculosis patients. Am. J. Respir. Crit. Care Med. **2001**; 164:2166–2171.

37. Cruciani M, Malena M, Bosco O, Gatti G, Serpelloni G. The impact of human immunodeficiency virus type 1 on infectiousness of tuberculosis: a meta-analysis. Clin. Infect. Dis. **2001**; 33:1922–1930.

Table 1. Model parameters, data sources, and distributions.

Parameter	Symbol	Value	Range	Reference
Relative infectiousness, smear negative TB	NA	0.2	--	[15–17]
Duration of infectiousness	NA	12 months	(4-18)	[21,22]
Excess CO_2 concentration*, household	$[CO_2]_h$-$[CO_2]_o$	635 ppm	(453-817)	Observed
Excess CO_2 concentration, public transit	$[CO_2]_p$-$[CO_2]_o$	1464 ppm	(397-2531)	Observed
Excess CO_2 concentration, school/crèche	$[CO_2]_s$-$[CO_2]_o$	1404 ppm	(1227-1581)	Observed
Excess CO_2 concentration, workplace	$[CO_2]_w$-$[CO_2]_o$	538	(359-717)	Observed
Contacts per day in each location	C_k	variable by age	variable by age	[14]
Hours spent in each location	T_k	variable by age	variable by age	[14]
Infectious quanta production	q	fit	fit	--

*Excess CO_2 concentration refers to the difference in the indoor CO_2 and the outdoor CO_2, which varied by location and time of day. Range of values reflects 95% CI.
ppm: parts per million

Figure Legends

Figure 1. Estimated contact-hours between individuals of various age groups in four locations (a. own households; b. public transit; c. schools; d. workplaces), with Gaussian smoothing applied to the contour.

Figure 2. Modeled (blue line) and observed (red points) latent tuberculosis infection (LTBI) prevalence by age group, where duration of infectiousness is 12 months and all contacts were recurrent except those in public transit. The gray shaded area reflects 95% credible intervals for model estimates.

Figure 3. Estimated proportion of tuberculosis infections acquired in each setting, according to age group. HH: household.