SUPPLEMENTARY MATERIALS AND METHODS

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

Jason R. Andrews, Carl Morrow, Rochelle P. Walensky, Robin Wood

Overview

We adapted the Wells-Riley approach for assessing the infection risk in indoor air environments to evaluate tuberculosis transmission in an endemic setting, by evaluating contact time in multiple environments. We utilized an age-structured model, drawing upon demographic, epidemiologic, and social contact data, together with carbon dioxide measurements in common indoor environments, from a peri-urban township near Cape Town, South Africa.

Model

The risk of tuberculosis transmission in an indoor space was first rigorously studied by Wells and Riley [1], who exposed guinea pigs to air from a tuberculosis ward and measured infection rates under controlled conditions. The Wells-Riley equation relates tuberculosis infection probability to the number of infectious individuals in a space (I), the breathing rate (p), the rate of generating infectious quanta (q), the duration of exposure (t) and the room ventilation rate (Q). The probability of at least one infection (P) occurring during time t follows the Poisson probability mass function:

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

Among the difficulties in applying this equation broadly in epidemiology and public health fields has been the effort required to evaluate the room ventilation rate, particularly under conditions in which rooms are occupied. The approach of using exhaled carbon dioxide (CO_2) as a natural tracer gas with which to evaluate building ventilation has been described in the Air Quality literature for many years [2]. Rudnick and Milton [3] demonstrated how carbon dioxide measurements could be applied to modify the Wells-Riley equation [1] for evaluating respiratory infectious disease transmission in an indoor environment. For this approach, they introduced the 'rebreathed fraction' (*f*), or proportion of breathed air that was exhaled from other room occupants, as a means to model transmission without measuring air flow explicitly. The rebreathed fraction—the volume of expired air (V_e) in the total room air volume (V)—can be estimated from the ambient CO_2 concentration ($[CO_2]_b$), the outdoor air CO_2 concentration ($[CO]_o$), and the concentration of CO_2 in expired air ($[CO]_a$) by the following relationship:

$$f = \frac{V_e}{V} = \frac{[CO]_b - [CO]_O}{[CO]_a}$$

The fraction of air rebreathed from infectious individuals is this rebreathed fraction (*f*) multiplied by the proportion of infectious individuals in the room (I/n). Rudnick and Milton demonstrated that the rate of exposure to infectious quanta is therefore given by multiplying the respiration rate (pt), the fraction of air that was expired from infectious individuals (fI/n) and the concentration of infectious quanta in exhaled breath (q/p). The Wells-Riley equation can then be reformulated as:

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

Here n is the number of individuals in the room, $\bar{f}$ is the averaged rebreathed fraction, and the other parameters are the same as formulated by Wells and Riley (above). We draw upon this formulation to model tuberculosis transmission across multiple environments using data on CO_2 concentration in these settings along with contact data structured by age and setting.

The probability of being infected by an individual of age group a while in environment k is a function of the time spent in that environment ($T_{a,k}$) and the number of contacts in that environment ($C_{a,k}$). The median time per encounter is multiplied by the number of non-contiguous encounters reported per age group a in environment k ($E_{a,k}$). The probability that a person of age group a is infected in environment k is a function of the amount of time spent in that environment per encounter, the number of encounters, and the probability of there being an infectious TB case there. Following the modified Wells-Riley equation above, the probability of infection per encounter in which I infectious persons among n total individuals (C contacts + 1, the referent individual) are present is:

$$P_{a,k}(I) = 1 - \exp\left(-\frac{\bar{f_k} I q T_{a,k}}{C_{a,k} + 1}\right)$$

The prevalence of TB (Y) in compartment k is the product of the tuberculosis prevalence (B) and the population weight (W) for each age group, and the weighted contact time spent in each location:

$$Y_k = \sum_j \frac{C_{j,k} T_{j,k} E_{j,k} B_j W_j}{\sum_i C_{i,k} T_{i,k} E_{i,k}}$$

The population weight is defined by the proportion of population in age group *j*. For the own household and other household environments, because some contacts in one's own household represent contacts reported as "other household" from their counterpart, and vice versa, we further weighted the prevalence by the amount of contact time spent in that setting, by age group. Because social mixing between age groups in schools is thought to be limited, all contact time between children in schools was assortative according to their five year age strata. The probability ($\boldsymbol{\theta}$) of there being *m* infectious contacts of an individual in age group *a* of age *j* is assumed to be Poisson distributed:

$$\theta_{a,k}(m) = \frac{(C_{a,k}Y_k)^m \exp(-C_{a,k}Y_k)}{m!}$$

We modeled up to two simultaneous infectious contacts, as the probability of having more than two simultaneous infectious contacts in one environment was <0.001. For each contact an individual has, it may be either a one-time contact or recurring. For recurring contacts, the probability of being infected in *d* days in location *k* is a function of the probability of having m infectious contacts, $\boldsymbol{\theta}$(m), among the recurring contacts multiplied by the probability of infection given m infectious contacts, *P(m),* over d*E encounters:

$$R_{a,k}(d) = 1 - \prod_m (1 - \theta_{a,k}(m) * (1 - \left(1 - P_{a,k}(m)\right)^{dE_{a,k}}))$$

For non-recurring contacts, the probability of being infected in *d* days is governed by an equation for the probability of infection if there are m contacts, *P*(m) times the probability of having m contacts, $\boldsymbol{\theta}$(m) over the course of d*E encounters:

$$N_{a,k}(d) = 1 - \prod_{m} \left(1 - P_{a,k}(m) * \theta_{a,k}(m)\right)^{dE_{a,k}}$$

The overall probability for an individual in age group *a* of being infected depends proportion of contacts that are recurrent (r) in each setting, and is given by:

$$P_a = 1 - \prod\nolimits_k (1 - r_k R_{a,k})(1 - (1 - r_k) N_{a,k})$$

The latent tuberculosis prevalence at age (L_a) is therefore given by the following equation:

$$L_a = 1 - \prod_{i=1}^{a} (1 - P_i)$$

Data

Using portable carbon dioxide (CO_2) detection devices (EasyView® 80 CO2 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. The outdoor air CO_2 concentration (C_o) was also sampled, and was cross-referenced with city estimates. The concentration of CO_2 in breath, C_a, is determined by Dubois body surface area (A_D), respiratory quotient (RQ), breathing rate (r) in L/min and level of physical activity (mets, M) [2].

The number of daily contacts in each location and time spent in each location were elicited from a diary study involving 571 residents of a South African township [4]. These data are stratified by age group in years (0-4,5-9,10-14,15-19,20-24,25-29,30-34,35-39,40-44,>45). Individuals documented number of indoor contacts and amount of time spent in each setting, using blocks of time: 0-15 minutes, 15-60 minutes, 1-2 hours, 2-4 hours, 4-8 hours, 8-12 hours, and > 12 hours. If multiple non-contiguous encounters occurred in the same setting, these were reported as separate encounters. We utilized median contacts and time spent per encounter, and we used triangular distributions defined by the median and interquartile range in uncertainty analysis.

We utilized data on the population age distribution from a recent census in this community (Figure S4), and we estimated tuberculosis incidence by age from a study of tuberculosis notifications in Cape Town, South Africa (Figure S3) [5]. We estimated tuberculosis prevalence from incidence by multiplying notifications by the estimated duration of undiagnosed tuberculosis in this community. We excluded individuals with extrapulmonary tuberculosis from transmission calculations, which eliminated most tuberculosis in young children. Because the estimate of infectious quanta generation was derived from smear positive individuals, we weighted smear negative individuals by 0.2 (smear positive individuals were weighted by 1.0)., the estimated relative infectiousness of smear negative patients compared with smear positive patients [6,7], to compose an 'effective' tuberculosis prevalence. We assumed that the average duration of infectiousness was one year, based on a

published estimate from this community [8], and we examined a broad range of estimates in sensitivity analysis (Table S1).

Data on the proportion of contacts in each setting that were one-time versus recurrent were not available from the contact diary study. We therefore examined a range of recurrence rates of contact from 0% to 100%. For our base case, we assumed that all contacts in households, schools and workplaces were recurrent, and that no contacts in public transit were recurrent. We examined the alternative scenarios in sensitivity analysis.

Parameter Estimation

The rate at which individuals generate infectious particles (q) is highly uncertain. The two published estimates (1.25 quanta/hour and 8.5 quanta/hour, respectively) were among symptomatic, smear-positive, hospitalized patients, which may be higher than the general population of undiagnosed tuberculosis patients [1,9]. We therefore estimated q by fitting the model to observed data on latent tuberculosis prevalence in various age groups in this community [10]. We used a simplex descent algorithm to find q that minimized the least-squares differences between modeled and observed data. We also estimated q for varying estimates of the duration of undiagnosed active tuberculosis.

Uncertainty and Sensitivity Analyses

For the main model results, we performed uncertainty analysis by drawing from a normal distribution defined by the mean and standard error of CO_2 estimates for each setting and triangular distributions defined by the median and interquartile

range for contacts and hours spent in each setting. We used a Latin Hypercube sampling strategy [11], drawing from 1,000 sets of parameters to create median and 95% uncertainty intervals for model estimate.

References

1. 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 Apr;85:511–25.

2. Persily A. Evaluating building IAQ and ventilation with indoor carbon dioxide. ASHRAE Transactions. 1997;103(2):1–12.

3. Rudnick SN, Milton DK. Risk of indoor airborne infection transmission estimated from carbon dioxide concentration. Indoor Air. 2003 Sep;13(3):237–45.

4. Wood R, Racow K, Bekker L-G, Morrow C, Middelkoop K, Mark D, et al. Indoor social networks in a South african township: potential contribution of location to tuberculosis transmission. PLoS ONE. 2012;7(6):e39246.

5. Wood R, Lawn SD, Caldwell J, Kaplan R, Middelkoop K, Bekker L-G. Burden of new and recurrent tuberculosis in a major South African city stratified by age and HIV-status. PLoS ONE. 2011;6(10):e25098.

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

7. Tostmann A, Kik SV, Kalisvaart NA, Sebek MM, Verver S, Boeree MJ, et al. Tuberculosis transmission by patients with smear-negative pulmonary tuberculosis in a large cohort in the Netherlands. Clin. Infect. Dis. 2008 Nov 1;47(9):1135–42.

8. Wood R, Middelkoop K, Myer L, Grant AD, Whitelaw A, Lawn SD, et al. Undiagnosed tuberculosis in a community with high HIV prevalence: implications for tuberculosis control. Am. J. Respir. Crit. Care Med. 2007 Jan 1;175(1):87–93.

9. Escombe AR, Moore DAJ, Gilman RH, Pan W, Navincopa M, Ticona E, et al. The infectiousness of tuberculosis patients coinfected with HIV. PLoS Med. 2008 Sep 30;5(9):e188.

10. Wood R, Liang H, Wu H, Middelkoop K, Oni T, Rangaka MX, et al. Changing prevalence of tuberculosis infection with increasing age in high-burden townships in South Africa. Int. J. Tuberc. Lung Dis. 2010 Apr;14(4):406–12.

11. Stein M. Large sample properties of simulations using latin hypercube sampling. Technometrics. 1987 May;29:143–51.

Supplementary Tables

Table S1. Sensitivity analysis of contact recurrence and duration of infectiousness on estimates of the quantum production rate and percentage of tuberculosis transmission that occurred in own households.

Parameters	Quantum Production Rate (q)	%TB Transmitted in Households	Sum of Squares
All contacts recurrent	0.94 quanta/hour	15.5%	1.0e-2
No contacts recurrent	0.27 quanta/hour	59.5%	2.1e-2
All contacts recurrent, except in public transit	0.89 quanta/hour	15.6%	1.0e-2
Duration of Infectiousness, 4 months	5.69 quanta/hour	5.4%	1.3e-2
Duration of Infectiousness, 18 months	0.44 quanta/hour	24.2%	1.0e-2

For the base case scenario, we assumed that all contacts were recurrent except in public transportation, and that the duration of infectiousness was 12 months.

Supplementary Figures Legend

Figure S1. Median contacts and hours spent in each of five key locations (a. households; b. public transit; c. school; d. workplaces; e. other households). Each point represents one observation (individual). Marginal histograms for contacts and hours shown above and to the right, respectively, of each figure.

Figure S2. Median number of indoor contacts by age and location.

Figure S3. Tuberculosis notifications (cases per 100,000 population) in Cape Town by age and smear status.

Figure S4. Age distribution of residents living in study township (2010 Census).

Figure S5. Modeled annual risk of infection by age group, with 95% credible intervals.

Figure S1a.

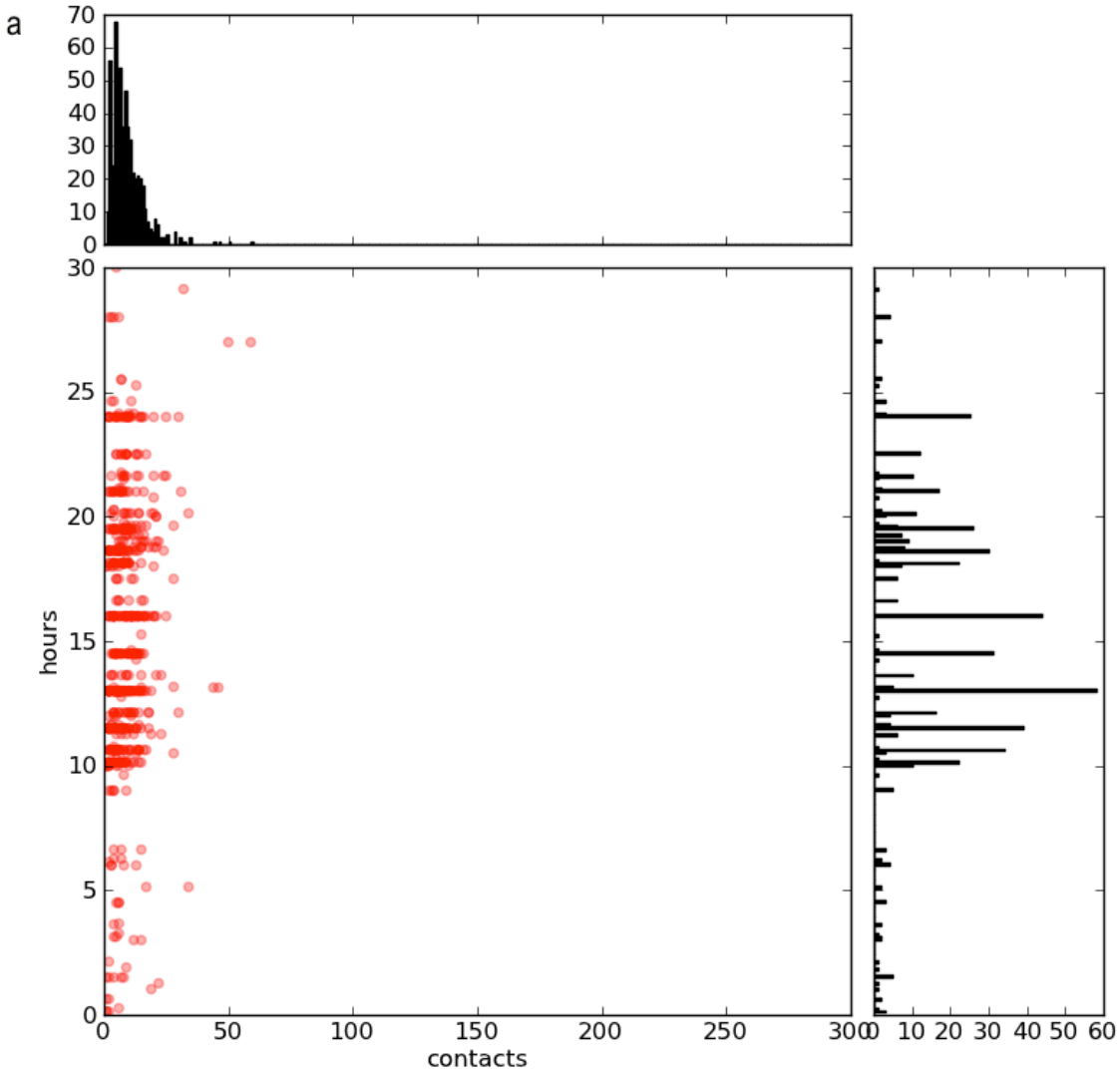

a
70
60
50
40
30
20
10
0
30
25
20
15
10
5
0
hours
0
50
100
150
200
250
300
contacts
0 10 20 30 40 50 60

Figure S1b.

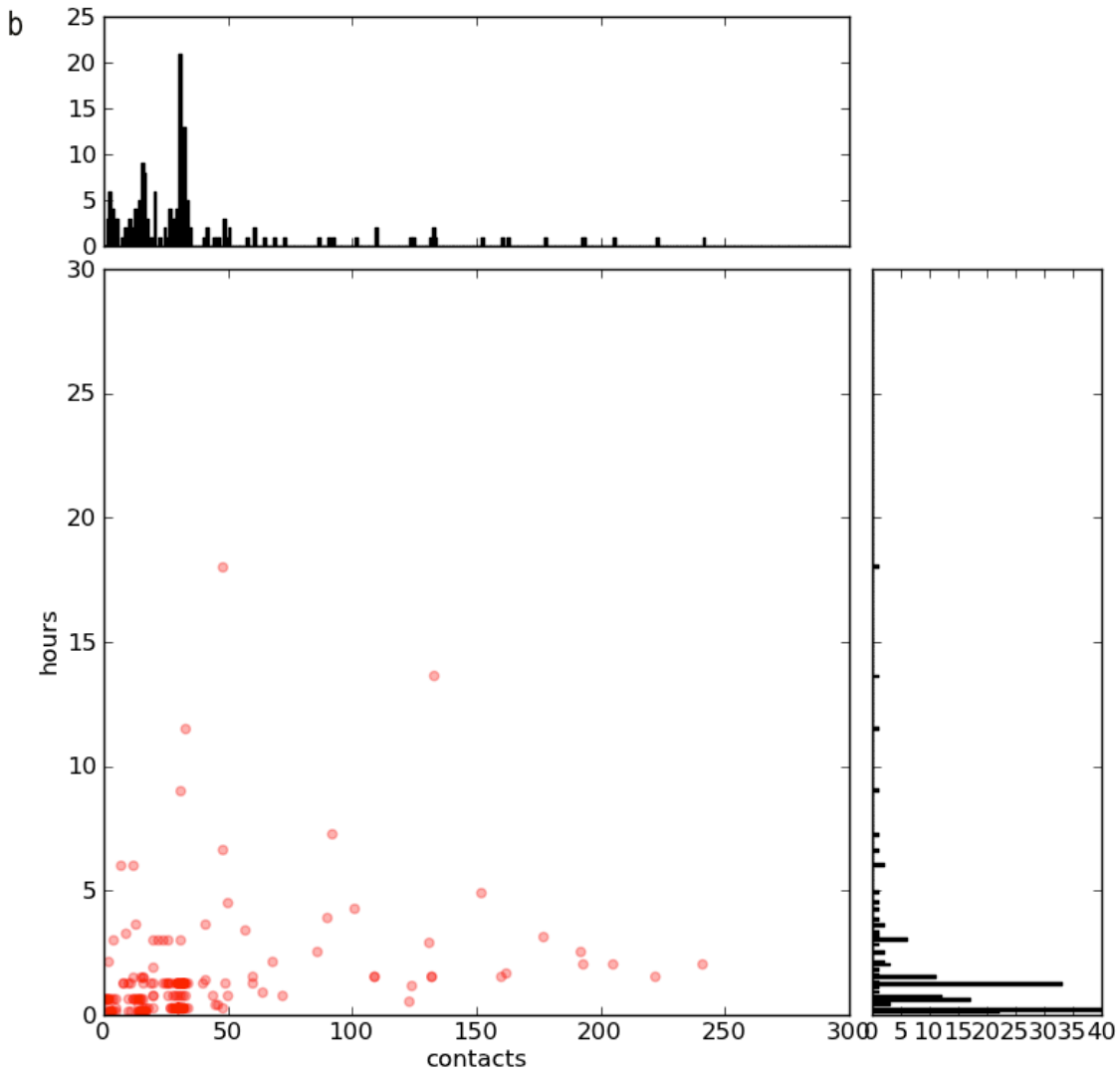


Figure S1c.

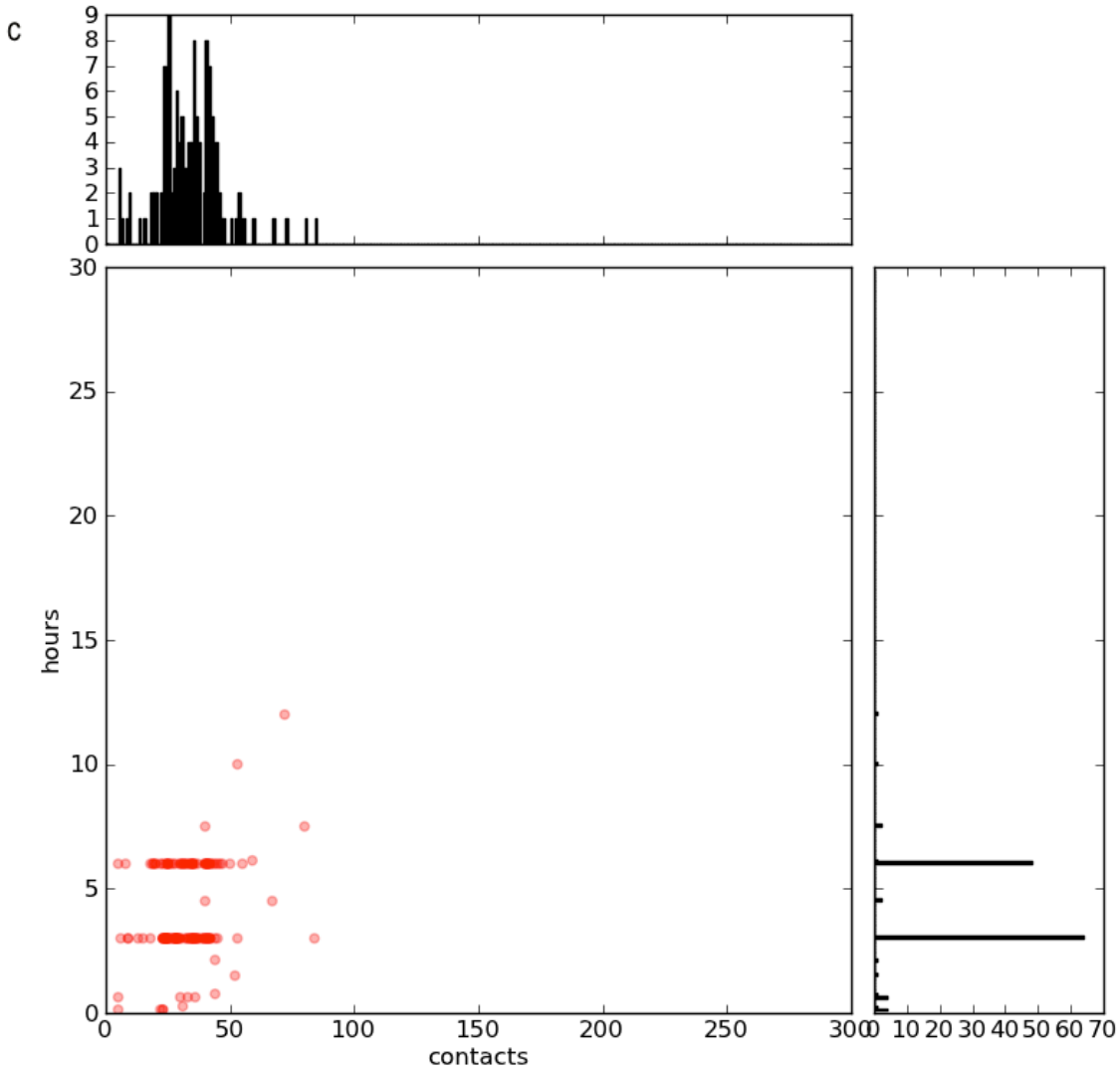

c
hours
contacts
0
1
2
3
4
5
6
7
8
9
0
5
10
15
20
25
30
0
50
100
150
200
250
300
0
10
20
30
40
50
60
70

Figure S1d.

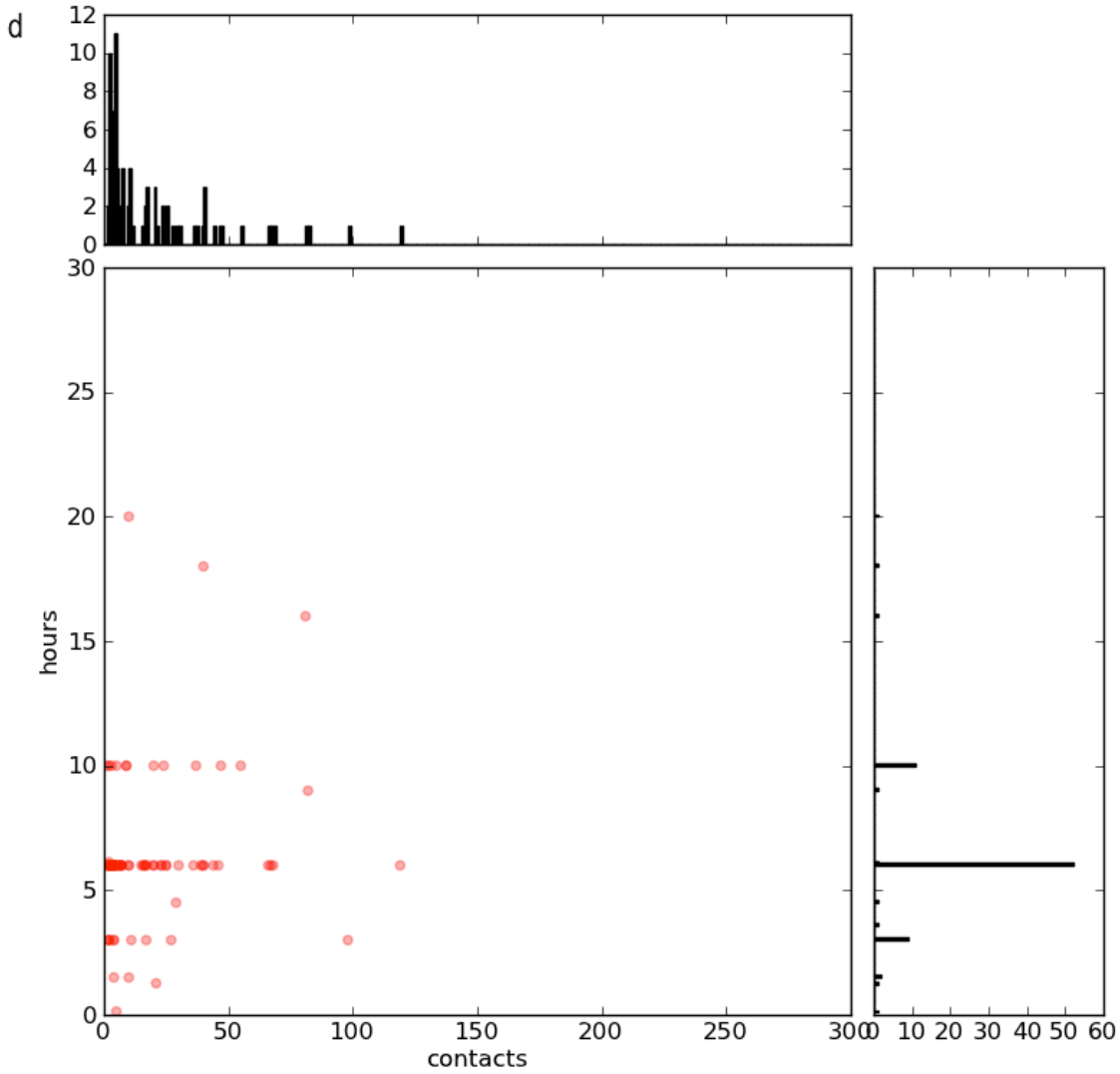

d
hours
contacts
0
2
4
6
8
10
12
0
5
10
15
20
25
30
0
50
100
150
200
250
300
0
10
20
30
40
50
60

Figure S1e.

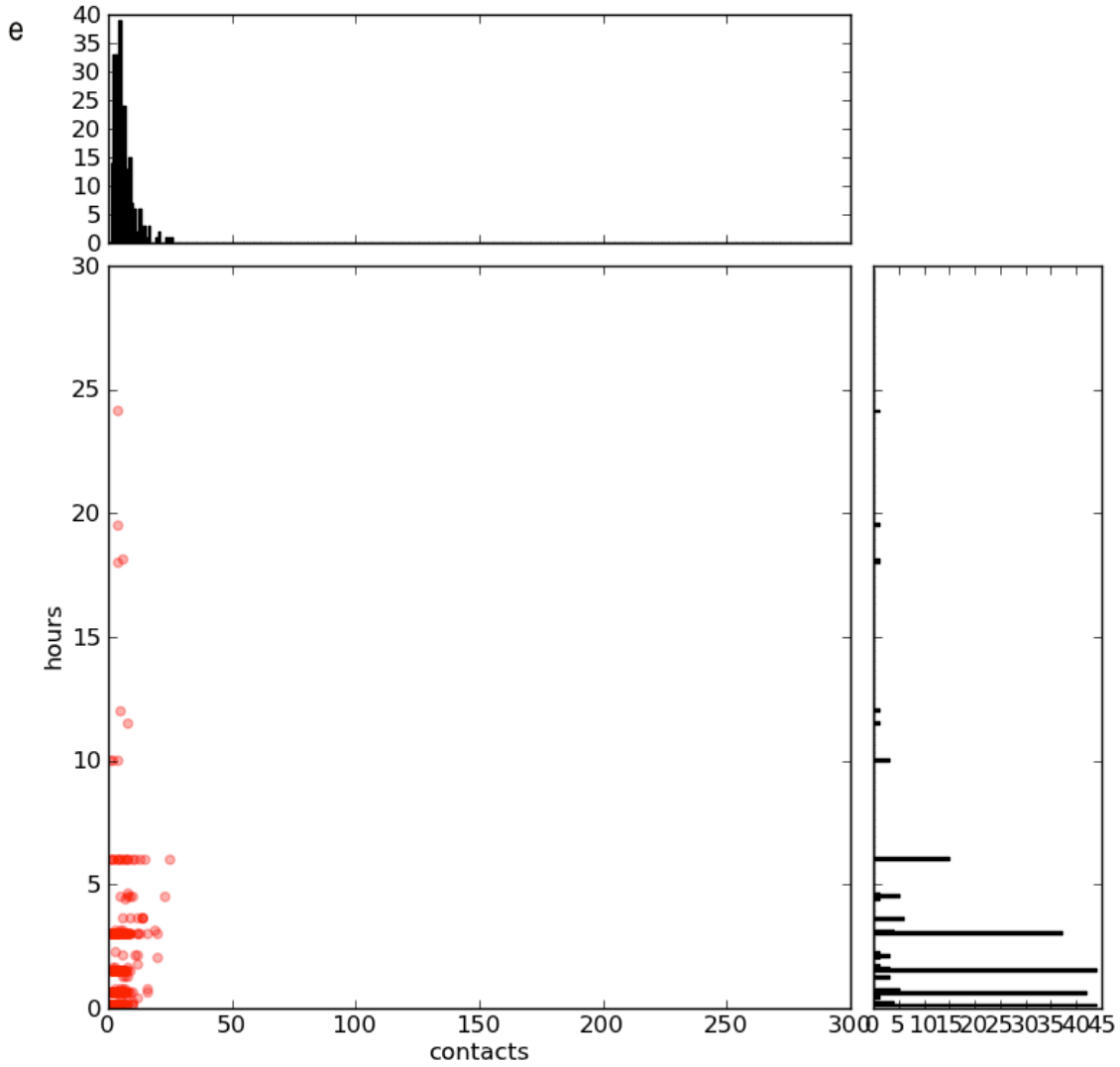


Figure S2.

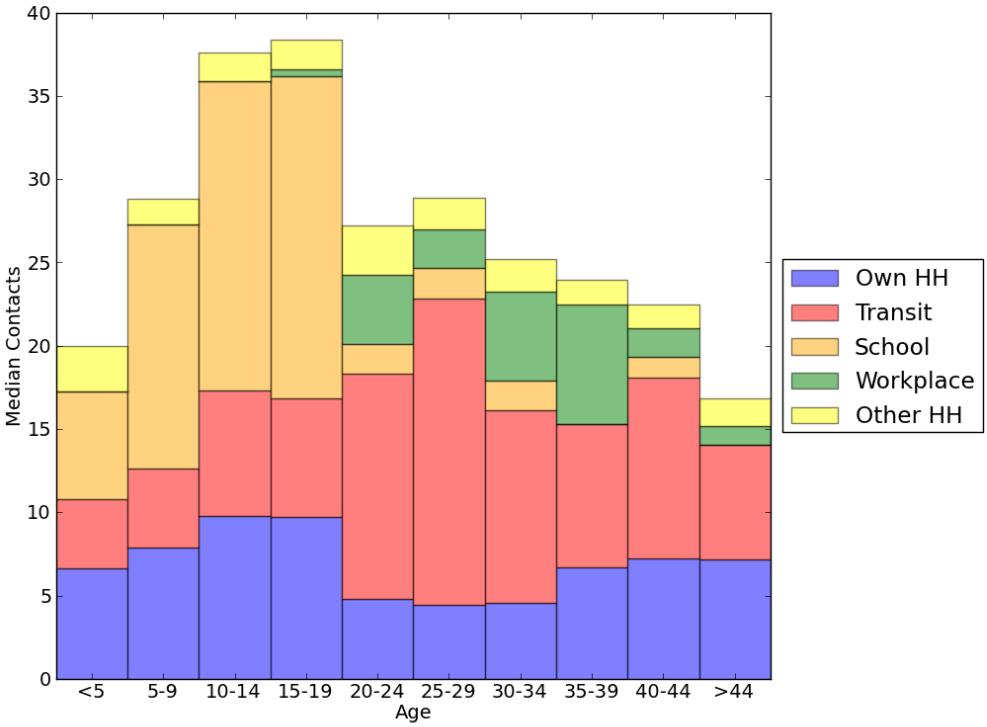

Median Contacts
Age
<5
5-9
10-14
15-19
20-24
25-29
30-34
35-39
40-44
>44
0
5
10
15
20
25
30
35
40
Own HH
Transit
School
Workplace
Other HH

Figure S3.

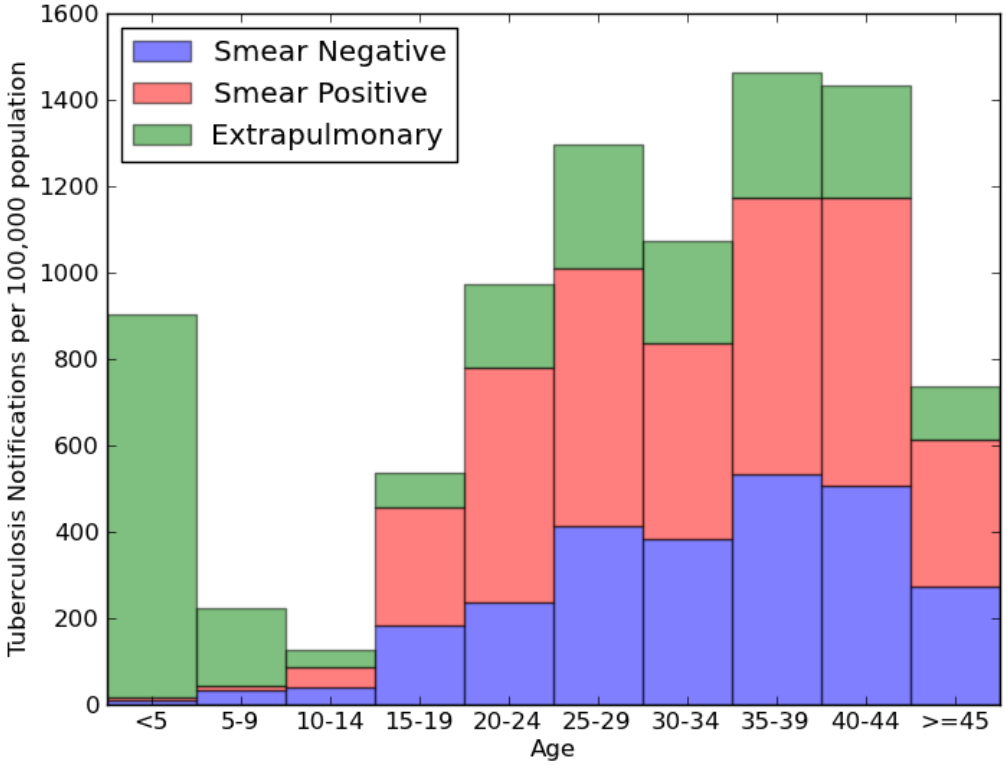

Smear Negative
Smear Positive
Extrapulmonary
Tuberculosis Notifications per 100,000 population
0
200
400
600
800
1000
1200
1400
1600
<5
5-9
10-14
15-19
20-24
25-29
30-34
35-39
40-44
>=45
Age

Figure S4.

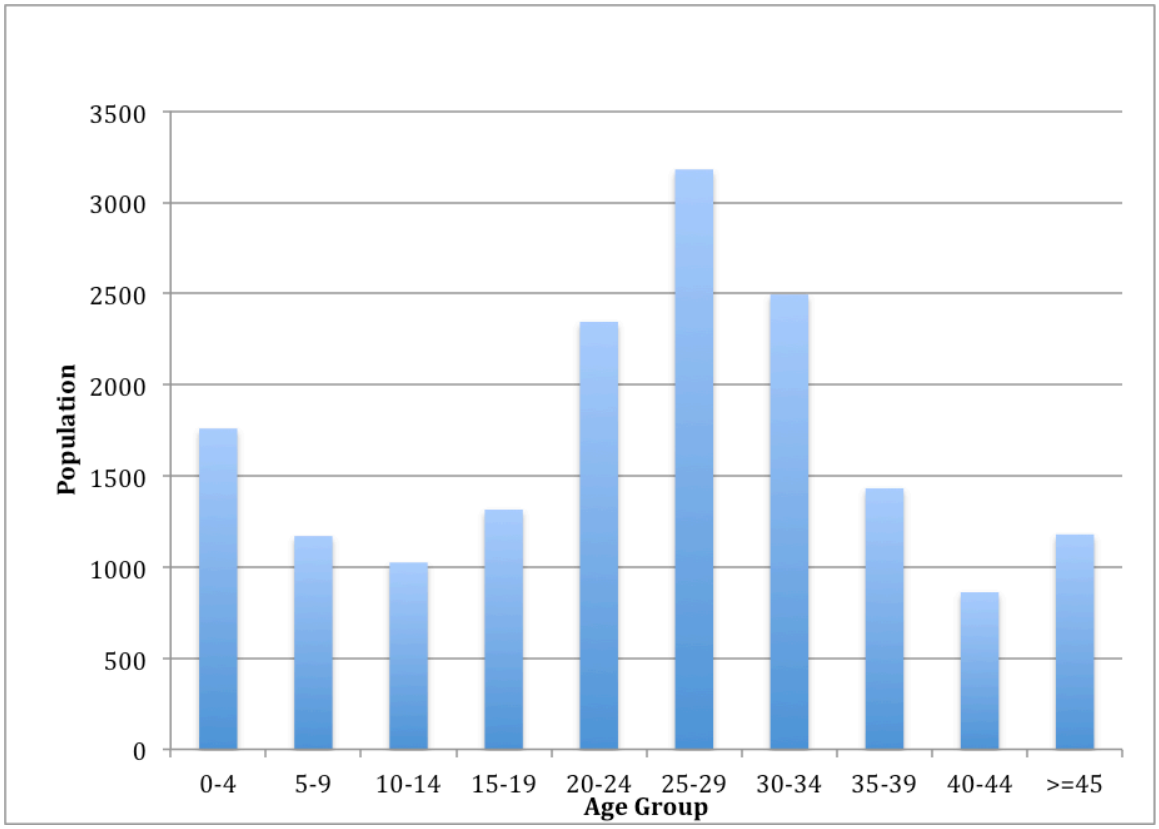
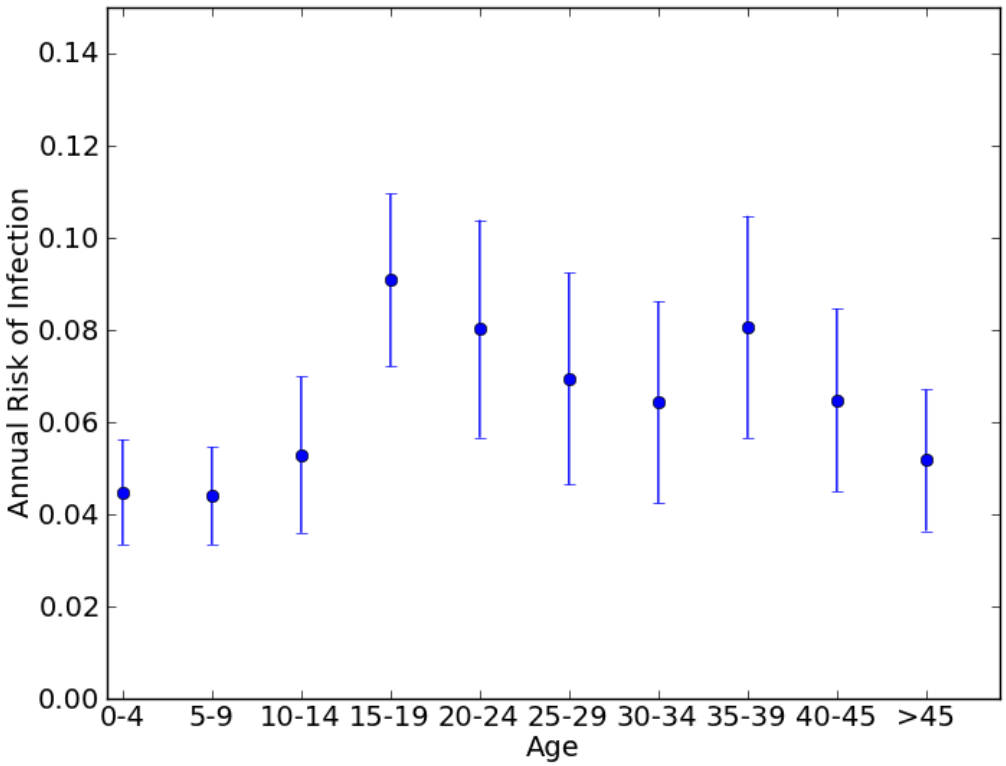

3500
3000
2500
2000
1500
1000
500
0
Population
0-4
5-9
10-14
15-19
20-24
25-29
30-34
35-39
40-44
>=45
Age Group

Figure S5.


Annual Risk of Infection
0.14
0.12
0.10
0.08
0.06
0.04
0.02
0.00
0-4
5-9
10-14
15-19
20-24
25-29
30-34
35-39
40-45
>45
Age