

Transmission Modeling with Regression Adjustment for Analyzing Household-based Studies of Infectious Disease

Application to Tuberculosis

Forrest W. Crawford,^{a,b,c,d} Florian M. Marx,^e Jon Zelner,^f and Ted Cohen^e

Background: Household contacts of people infected with a transmissible disease may be at risk due to this proximate exposure, or from other unobserved sources. Understanding variation in infection risk is essential for targeting interventions.

Methods: We develop an analytical approach to estimate household and exogenous forces of infection, while accounting for individual-level characteristics that affect susceptibility to disease and transmissibility. We apply this approach to a cohort study conducted in Lima, Peru, of 18,544 subjects in 4,500 households with at least one active tuberculosis (TB) case and compare the results to those obtained by Poisson and logistic regression.

Results: HIV-coinfected (susceptibility hazard ratio [SHR] = 3.80, 1.56–9.29), child (SHR = 1.72, 1.32–2.23), and teenage (SHR = 2.00, 1.49–2.68) household contacts of TB cases experience a higher hazard of TB than do adult contacts. Isoniazid preventive therapy (SHR = 0.30, 0.21–0.42) and Bacillus Calmette–Guérin (BCG) vaccination (SHR = 0.66, 0.51–0.86) reduce the risk of disease among household contacts. TB cases without microbiological confirmation exert a smaller hazard of TB among their close contacts compared with smear- or culture-positive cases (excess hazard ratio = 0.88, 0.82–0.93 for HIV– cases and 0.82, 0.57–0.94 for HIV+ cases). The extra household force of infection results in 0.01 (95% confidence interval [CI] = 0.004, 0.028) TB cases per susceptible household contact per year and the rate of transmission between a microbiologically confirmed TB case and susceptible household contact at 0.08 (95% CI = 0.045, 0.129) TB cases per pair per year.

Conclusions: Accounting for exposure to infected household contacts permits estimation of risk factors for disease susceptibility and transmissibility and comparison of within-household and exogenous forces of infection.

Keywords: Cluster; Contagion, Force of infection; Household transmission

(*Epidemiology* 2020;31: 238–247)

Much of what is known about tuberculosis (TB) transmissibility comes from prospective household-based cohort studies focused on risk of infection or disease among household contacts of a newly diagnosed “index” patient.^{1–4} This study design is appealing because in addition to characterizing index cases and their coprevalent household contacts, it may also reveal the time dynamics of secondary infections and yield information helpful for designing better TB interventions. From prospective household-based cohort studies of TB infection, three levels of epidemiologic inference are needed to inform targeted TB testing, prevention, and treatment resources.

First, knowledge of risk factors affecting susceptibility to TB infection and progression to disease after infection would help identify contacts who are at most risk of TB disease.^{5–7} Young people may be especially susceptible to infection, and severe forms of disease^{8–13} and HIV+ individuals are more likely to progress to disease after infection. BCG vaccination and isoniazid preventive therapy (IPT) are associated with a decreased risk of TB disease.^{14–17}

Second, knowledge of factors affecting the risk of transmission by an infectious individual would help identify cases who are most likely to infect their susceptible contacts.^{18–20} Although HIV+ individuals are at higher risk for TB disease after infection, HIV+ people may transmit TB less often given the association between immunosuppression and extrapulmonary TB and, in some settings, shorter durations of infectiousness.^{21–27} Pulmonary disease with positive sputum smear status at the time of diagnosis is associated with higher risk of TB transmission to household contacts.^{12,28–30}

Submitted February 22, 2019; accepted November 11, 2019.

From the ^aDepartment of Biostatistics, Yale School of Public Health; ^bDepartment of Statistics & Data Science, Yale University; ^cDepartment of Ecology & Evolutionary Biology, Yale University; ^dYale School of Management; ^eDepartment of Epidemiology of Microbial Diseases, Yale School of Public Health, New Haven, CT; and ^fDepartment of Epidemiology, University of Michigan School of Public Health Ann Arbor, MI.

This work was supported by US National Institutes of Health grants NIAID U19 A1076217, NICHD 1DP2HD091799, and NIAID R01 AI112438.

The authors report no conflicts of interest.

Availability of code: Code is available by request to corresponding author.

SDC Supplemental digital content is available through direct URL citations in the HTML and PDF versions of this article (www.epidem.com).

Correspondence: Ted Cohen, Department of Epidemiology of Microbial Diseases, Yale School of Public Health, 60 College Street, New Haven, CT 06510. E-mail: theodore.cohen@yale.edu.

Copyright © 2019 Wolters Kluwer Health, Inc. All rights reserved.

ISSN: 1044-3983/20/3102-0238

DOI: 10.1097/EDE.0000000000001143

Third, knowledge of the relative importance of community and household forces of infection could guide case finding and intervention strategies at the community and household levels.^{7,31–33}

Analytic approaches for estimating transmission risk from household-based data generally fall into two categories. First, researchers have developed transmission models for infectious disease dynamics in small groups.^{32,34–40} But fitting these models to individual-level data in small populations, i.e., households, can be challenging because likelihoods may not be tractable. Furthermore, most transmission models do not accommodate multiple covariates, making it difficult to perform regression-style adjustment for individual-level characteristics.

Second, researchers routinely employ traditional models for regression adjustment to clustered infectious disease data. When disease outcomes are binary, generalized linear models are often used: Poisson, logistic, and log-binomial regression approaches are popular,^{16,31,41,42} often with household-level random intercept terms to account for unobserved heterogeneity. These methods permit estimation of risk or odds ratios but may not deliver estimates of important epidemiologic features of disease transmission: the extra household force of infection, and factors associated with susceptibility to disease, versus the infectiousness of an infected individual. Some researchers have attempted to adjust for measures of infection risk by assuming that an initial “index” case is the source of all infection risk experienced by household members.^{12,27} However, epidemiologists have warned that traditional approaches to estimation and regression adjustment for infectious disease outcomes may result in biased estimates because they cannot account for disease transmission.^{35,43–47}

More recently, hybrid transmission models that combine features of mechanistic and regression approaches have been introduced, permitting regression-style adjustment for individualistic correlates of susceptibility and infectiousness.^{39,48–56}

In this study, we adapt a hybrid regression/transmission model to estimate household and exogenous forces of TB infection, while adjusting for individual-level covariates associated with individual infectiousness and susceptibility to disease. We apply this approach to analyze outcomes among household contacts of TB cases in an prospective cohort study conducted in Lima, Peru, from 2009 to 2012.^{12,16,27,57}

METHODS

Data and Design of the Household Tuberculosis Cohort Study

We analyze data from a prospective cohort study of household contacts of adults (age ≥ 15 years) treated for active TB in Lima, Peru. Households were identified when a member (called the index case) was diagnosed with pulmonary TB at one of 106 participating health centers between September 2009 and August 2012. All household contacts of index cases

were eligible for this study and were recruited during home visits within 2 weeks following the index case diagnosis. In total, 18,544 subjects in 4,500 households with at least one active tuberculosis (TB) case were included. The study protocol was approved by the Harvard University Institutional Review Board (Ref. No 19332), by the Research Ethics Committee of the National Institute of Health of Peru.

Baseline, 6-month, and 12-month follow-up home visits were conducted by trained study nurses. During these visits, symptoms consistent with active TB were assessed among contacts via questionnaire; individuals with symptoms were then referred to a health center for clinical and bacteriologic evaluation. A detailed description of the methods for this cohort study has previously been published.^{12,16,27,57}

We analyze TB outcomes for household contacts who were TB disease free at the time of enrollment. Contacts with active disease during the baseline visit (i.e., coprevalent cases) were not included as incident cases in our analyses. We group index and coprevalent cases together as “primary cases” and restrict attention to the 3,446 households with at least one primary TB case and at least one disease-free household contact at enrollment. This leaves a total of 17,490 subjects, of whom 3,673 (21%) were primary TB cases and 13,817 were disease free at enrollment. A total follow-up time of 1 year was considered. Our outcome of interest was incident bacteriologically confirmed TB among household contacts. We refer to cases of incident active TB occurring among these subjects as secondary household TB.

For the 13,817 contacts free of active TB at enrollment, age, sex, HIV status, IPT status, and BCG vaccination status were recorded. From September 2009 to August 2011, HIV status was assessed using an enzyme immunoassay (EIA), with positive or ambiguous samples confirmed by immunofluorescence assay (IFA). From August 2011 to August 2012, HIV status was first assessed by rapid screening test, with positive or ambiguous samples confirmed by EIA and IFA.⁵⁸ Thirty-six of 3,673 = 1% of primary cases and 163/13,817 = 1% of at-risk subjects were missing HIV test results; we assumed these subjects had negative HIV status. Prior BCG vaccination in at-risk subjects was assessed by self-report (and by inspection for characteristic scarring⁵⁹); 4/13,817=0.03% were missing BCG information, and these subjects were assumed to have not received BCG. No at-risk subjects were missing IPT status. Every primary case had laboratory sputum smear or culture results immediately following diagnosis. A composite variable for microbiological confirmation (MC+), denoting either smear-positive or culture-positive TB disease was used.

Although the exact date of onset of active TB for these secondary cases is not observed, the last home visit before diagnosis and the date of diagnosis provide a time interval for each secondary TB case during which onset of active TB occurred. Section 3 of the eAppendix; <http://links.lww.com/EDE/B615>, describes the temporal pattern of active disease onset intervals for secondary cases.

Transmission Model

Hybrid transmission modeling and regression adjustment approaches can permit valid epidemiologic inference by accounting for both the structure of infectious disease transmission within households and covariates that may influence susceptibility and infectiousness.^{35,36,39,51–53,55,56,60,61} We develop a general model for infectious disease transmission that describes the instantaneous risk of infection for each susceptible household contact. The model permits simultaneous estimation of the exogenous and household forces of infection, while adjusting for covariates associated with progression to active disease, and infectiousness. For a collection of N households, where household i has n_i members, let x_{ij} be a $p \times 1$ vector of susceptibility-related baseline covariates for subject j in household i . Let z_{ij} be a $q \times 1$ vector of possibly different covariates that are related to infectiousness, given that j has active (transmissible) infection. Define $y_{ij}(t)$ to be the indicator that subject j in household i is infected just before time t , and define t_{ij} to be the time at which subject j in household i becomes infectious, or the end of the household follow-up window t_p , whichever comes first. Both times t_i and t_j are measured relative to enrollment of household i . Define $t_{ij}^{dx} > t_{ij}$ to be the date on which the incident case j was diagnosed with infection.

We construct a hazard model of infection by considering different sources of infectiousness to which a susceptible person j in household i is subjected, at time t following enrollment of household j . Assume person j is subjected to a constant force of infection (hazard) α from outside the household, measured as infections per unit time. Suppose further that the covariates x_{ij} are related to the risk of infection from outside the household per unit time by the hazard $\alpha \times \exp[x'_{ij}\beta]$. Assume j is also subject to a distinct and independent force of infection at time t from any other infectious members of household i . Suppose that person $k \neq j$ in household j is infectious. Then j is subject to a hazard of $\exp[x'_{ij}\beta] \times \exp[z'_{ik}\gamma]$ per unit time, where $\exp[z'_{ik}\gamma]$ is the force of infection due to person k . To incorporate loss of infectiousness following

diagnosis and onset of treatment of infection, we parameterize the baseline hazard of infection from k to a susceptible individual. Define $\omega \in [0,1]$ to be the proportion loss of infectiousness following onset of treatment, and let $w_{ik}(t) = 0$ if $t < t_{ik}$, 1 if $t \in (t_{ik}, t_{ik}^{dx})$, and ω if $t_{ik} > t_{ik}^{dx}$. The total hazard to subject j in household i at time t is therefore

$$\lambda_{ij}(t) = \exp[x'_{ij}\beta] \left(\alpha + \sum_{k=1}^{n_i} w_{ik}(t) \exp[z'_{ik}\gamma] \right)$$

We treat ω as fixed in the statistical analysis and conduct a sensitivity study of its effect on β and γ in the eAppendix; <http://links.lww.com/EDE/B615>. Figure 1 shows a schematic illustration of the infection hazard experienced by a susceptible subject over time, as other household members become infectious. A formal mathematical construction of the hazard model is given in Section 1 of the eAppendix; <http://links.lww.com/EDE/B615>.

Epidemiologic Meaning of Parameters

The scalar parameter α represents the exogenous force of infection experienced by a single susceptible subject. Formally, α is the adjusted baseline risk for an individual subject to no household infectiousness. In this analysis, α is assumed to be constant in time and across households, which may be appropriate in areas where TB is endemic. When the model is correctly specified, α can be interpreted as the expected number of new infections per unit time that a susceptible individual would experience in the absence of a household exposure. If the model is incorrectly specified, for example, by failing to accurately capture household exposure, α acts to absorb the extra risk not accounted for through the parameter describing within-household transmission. The household-level force of infection at time t is the sum of each infectious member's infectiousness at that time. The regression coefficient γ controls the infectiousness of person k according to z_{ik} . Likewise, the regression coefficient β controls the susceptibility of person j to exposure to disease according to their characteristics x_{ij} . The susceptibility hazard ratio comparing

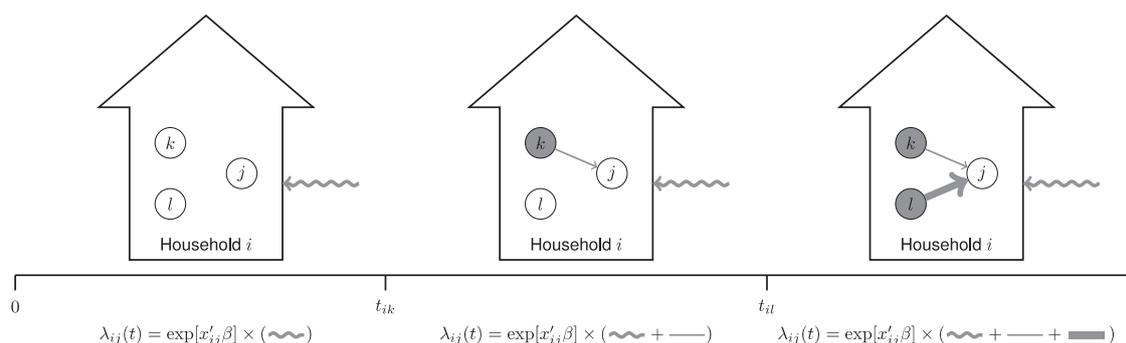


FIGURE 1. Schematic illustration of the transmission hazard model in a household of size three. The hazard to subject j in household i is proportional to the sum of the hazards they experience from infected household contacts and from the community. As more household members become infected, the hazard experienced by j increases.

the overall hazard at covariate level x^* with that at level x is $\text{SHR} = \exp[(x^* - x)\beta]$. Likewise, the infectiousness hazard ratio comparing the hazard of infection at level z^* with that at level x is $\text{IHR} = \exp[(z^* - z)\gamma]$. When the first element of z is equal to 1, the corresponding first element of the parameter vector γ , denoted γ_0 , is interpretable as the log baseline risk of transmission, absent in other infectiousness covariates.

These parameters also have familiar epidemiologic meaning in combination. The excess hazard due to a single exposure to disease at level z is the ratio of the infective hazard exerted by a single person with infectiousness covariates z to the overall hazard, $\text{EHE} = \exp[z'\gamma]/(\alpha + \exp[z'\gamma])$. The excess hazard due to exposure can change following initiation of treatment of a person with active disease. When $\omega < 1$, the excess hazard at level z following diagnosis of the infectious individual is $\text{EHE} = \omega \exp[z'\gamma]/(\alpha + \omega \exp[z'\gamma])$. The excess hazard due to exposure is a more specific measure of infective hazard than the ratio of infection odds in household contacts to controls (called the “community infection ratio”) that has previously been used to quantify TB risk due to household exposure.³³ The probability of exogenous infection for a particular incident secondary case j in household i is the ratio of the exogenous hazard to the total hazard experienced by person j at the time of infection, $C_{-ij} = \alpha / \left(\alpha + \sum_{k=1}^{n_i} w_{ik}(t_{ij}) \exp[z'_{ik}\gamma] \right)$.

Section 1.4 of the eAppendix; <http://links.lww.com/EDE/B615>, describes these composite estimands and their definitions in terms of the hazards $\lambda_{ij}(t)$ in greater detail.

Relationship to Other Transmission Models

The hazard model employed in this study is unlike traditional statistical models (e.g., logistic, Poisson) for regression analysis of binary infection outcomes because it incorporates the cumulative exposure to infective hazard experienced by each susceptible subject during the study. In particular, the hazard model does not assume that secondary cases are solely attributable to exposure from the index case.^{12,27,57} However, the model is closely related to several common constructs in epidemiology. First, the susceptible-infective model of infectious disease transmission in a closed population^{37,38,62} is a special case of the household-based hazard model obtained by setting $\alpha = 0$ and ignoring individual-level covariates. Second, the Wells-Riley equation^{63–67} describing the occurrence of new cases of an air-borne disease in an enclosed space can be derived from the expected number of infections in the household hazard model. Third, several researchers have proposed similar risk or hazard regression models for infectious disease outcomes.^{39,52–56} In particular, Zelner et al¹² propose a generalized additive mixed logistic model to estimate age-specific risk factors for latent TB infection; predictions from this model were used to estimate the community force of infection. Kenah⁵¹ proposes a semiparametric version of the hazard model presented above with $\alpha = 0$; time-varying transmission hazards can be estimated from this model.

Similar transmission modeling approaches have been used to successfully adjust for individual-level covariates in household transmission models of influenza.^{48–50} Section 1.3 of the eAppendix; <http://links.lww.com/EDE/B615>, provides formal derivations of these correspondences.

Statistical Analysis

The likelihood of the observed data under the hazard model takes a convenient form that is amenable to iterative maximization with respect to the unknown parameters $\theta = (\alpha, \beta, \gamma)$. We estimate θ by maximum likelihood, with a multiple imputation procedure⁶⁸ for dealing with unobserved infectiousness onset times, derived in Section 3 of the eAppendix; <http://links.lww.com/EDE/B615>. We compute susceptibility hazard ratio (SHR) for age, sex, HIV status, concurrent IPT, and BCG vaccination; IHR for HIV+ and microbiological TB confirmation; and the community force of infection α , the baseline infective hazard $\exp[\gamma_0]$, excess hazard due to exposure under different infectiousness parameters, and the probability of community transmission for each secondary case. We also compare the results to risk ratios obtained by modified Poisson regression.⁴²

RESULTS

Descriptive Results

Table 1 shows descriptive characteristics of subjects, stratified by their TB status: primary cases (index and secondary coprevalent) diagnosed at baseline, secondary cases diagnosed during the study, and subjects who were not diagnosed during the study. Primary cases at enrollment were 58% male, but the overall sex balance including all household contacts was 47% male. The average overall age was 28.3 ± 19.5 years; most primary cases were between 18 and 65, matching the enrollment criteria for index patients. HIV prevalence among primary cases at baseline was $123/3,673 = 3\%$ and $52/13,817 = 0.4\%$ among at-risk individuals at baseline. A microbiological confirmation of TB disease was made for $3,200/3,673 = 87\%$ of primary cases and $3,384/4,068 = 83\%$ of cases overall. Table 2 shows the number of primary and secondary TB cases across household sizes. A simulation analysis, presented in the eAppendix; <http://links.lww.com/EDE/B615>, shows that simulated TB outcomes from the transmission model match the frequency distribution of observed TB cases across household sizes in Table 2.

Risk Factors for Susceptibility and Transmission

Table 3 shows regression results for the hazard model of transmission with the proportion reduction in infectiousness following diagnosis $\omega = 0.2$ and composite estimands comparing the exogenous and household forces of infection. We set ω to reflect empiric evidence that treatment for TB is effective in reducing transmissibility very rapidly.^{69,70} The parameter ω is the subject of a sensitivity analysis presented in the eAppendix; <http://links.lww.com/EDE/B615>. The exogenous force of infection α estimated to be 0.01 (95% confidence interval

TABLE 1. Descriptive Characteristics of Subjects, Stratified by TB Status

| | Primary TB Count (%) | Secondary Count (%) | No TB Count (%) | All Count (%) |
|-------------|-------------------------|------------------------|--------------------|------------------|
| Age | | | | |
| 0–2 years | 4 (0) | 14 (4) | 688 (5) | 706 (4) |
| 2–12 years | 38 (1) | 94 (24) | 2,974 (22) | 3,106 (18) |
| 12–18 years | 297 (8) | 61 (15) | 1,507 (11) | 1,865 (11) |
| 18–65 years | 3,091 (84) | 210 (53) | 7,591 (57) | 10,892 (62) |
| >65 years | 243 (7) | 16 (4) | 662 (5) | 921 (5) |
| Male sex | 2,113 (58) | 200 (51) | 5,960 (44) | 8,273 (47) |
| HIV | 123 (3) | 5 (1) | 47 (0) | 175 (1) |
| IPT | 17 (0) | 42 (11) | 3,026 (23) | 3,085 (18) |
| BCG | 204 (6) | 318 (81) | 11,572 (86) | 12,094 (69) |
| TB dx | 3,673 (100) | 395 (100) | 0 (0) | 4,068 (23) |

TB dx indicates TB diagnosis.

TABLE 2. Number of Primary and Secondary Tuberculosis Cases in Households of Different Size

| | Household Size | | | | | | | | | | | | | | |
|------------------------|----------------|-----|-----|-----|-----|-----|-----|----|----|----|----|----|----|----|-----|
| | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | >15 |
| Number primary cases | | | | | | | | | | | | | | | |
| 1 | 489 | 651 | 624 | 473 | 335 | 227 | 134 | 92 | 64 | 48 | 30 | 21 | 12 | 7 | 37 |
| 2 | 22 | 24 | 19 | 35 | 16 | 18 | 8 | 7 | 5 | 6 | 4 | 3 | 0 | 1 | 8 |
| 3 | 0 | 2 | 5 | 3 | 3 | 1 | 3 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 2 |
| 4 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 5 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| 6 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| Number secondary cases | | | | | | | | | | | | | | | |
| 0 | 490 | 635 | 608 | 469 | 312 | 205 | 128 | 81 | 58 | 44 | 30 | 20 | 11 | 7 | 30 |
| 1 | 21 | 41 | 34 | 41 | 37 | 36 | 14 | 14 | 10 | 8 | 5 | 2 | 1 | 1 | 7 |
| 2 | 0 | 1 | 6 | 2 | 5 | 3 | 3 | 2 | 1 | 2 | 0 | 2 | 0 | 0 | 6 |
| 3 | 0 | 0 | 0 | 0 | 1 | 2 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 3 |
| 4 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2 |
| 5 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| 7 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

[CI] = 0.004, 0.03). Note that α is a raw hazard, not a hazard ratio, that incorporates features of infection and disease progression. Hazard ratios corresponding to susceptibility and infectiousness covariates (defined heuristically above and formally in Section 1.4 of the eAppendix; <http://links.lww.com/EDE/B615>) are given in the rows below.

Adjusted estimates of TB hazard experienced by household contacts in different age groups (adults age 19–65 are the reference group) exhibit patterns that are not evident in the unadjusted prevalence rates. Relative to adults (18–65 years), the susceptibility hazards by age are for infants (0–2 years) SHR = 1.04 (0.60–1.81); for children (3–12 years) SHR = 1.72 (1.32–2.23); for teens (13–18 years) SHR = 2.00 (1.49–2.68); and

for the elderly (>65 years) SHR = 0.75 (0.44–1.28). Figure 2 shows raw TB prevalence on the left-hand vertical axis by age, with a Kernel regression estimate and bootstrap pointwise CI. Estimated age susceptibility hazard ratios and CIs are overlaid, with the hazard ratio scale on the right-hand vertical axis. Male sex was not associated with increased hazard, SHR = 1.17 (0.96–1.43). Positive HIV status (measured at baseline) is associated with increased risk, SHR = 3.80 (1.56–9.29). Isoniazid preventive therapy (IPT) is associated with decreased hazard, SHR = 0.30 (0.21–0.42). BCG was associated with decreased hazard, SHR = 0.66 (0.51–0.86). The eAppendix; <http://links.lww.com/EDE/B615>, shows a sensitivity analysis for values of ω varying from 0.2 to 1.0.

TABLE 3. Hazards and Hazard Ratios for Susceptibility and Infectivity, with Comparisons of the Community and Within-Household Force of Infection for $\omega = 0.2$

| | Estimate | 2.5% | 97.5% |
|---------------------------------|----------|------|-------|
| Exogenous force of infection | 0.01 | 0.00 | 0.03 |
| Susceptibility SHR | | | |
| Age | | | |
| 0–2 years | 1.04 | 0.60 | 1.81 |
| 2–12 years | 1.72 | 1.32 | 2.23 |
| 12–18 years | 2.00 | 1.49 | 2.68 |
| >65 years | 0.75 | 0.44 | 1.28 |
| Male sex | 1.17 | 0.96 | 1.43 |
| HIV+ | 3.80 | 1.56 | 9.29 |
| IPT | 0.30 | 0.21 | 0.42 |
| BCG | 0.66 | 0.51 | 0.86 |
| Infectivity IHR | | | |
| Intercept | 0.08 | 0.05 | 0.13 |
| HIV | 0.60 | 0.20 | 1.83 |
| MC+ | 1.50 | 0.88 | 2.58 |
| Excess HR attributable exposure | | | |
| HIV–/MC– | 0.88 | 0.82 | 0.93 |
| HIV+/MC– | 0.82 | 0.57 | 0.94 |
| HIV–/MC+ | 0.92 | 0.84 | 0.96 |
| HIV+/MC+ | 0.87 | 0.64 | 0.96 |

The estimated “baseline” infectiousness hazard ($\exp[\gamma_0]$ in the hazard model), in the absence of smear- or culture-confirmed TB, and negative HIV status, is 0.08 (0.05–0.13). The baseline infectiousness is a raw hazard, with units of transmissions per infective-susceptible pair per year. TB cases with MC+ status upon diagnosis are not associated with increased infectiousness hazard compared with TB cases that were diagnosed on the basis of symptoms in the absence of microbiological confirmation [IHR = 1.50 (0.88–2.58)]. HIV+ status was not associated with infectiousness hazard, IHR = 0.60 (0.20–1.83) in this analysis.

Comparison of Exogenous and Household Forces of Infection

Comparing the baseline infectiousness hazard for a microbiologically unconfirmed case ($\exp[\gamma_0]$) to α directly, the hazard experienced by a susceptible subject from a single microbiologically unconfirmed HIV– household case (before that individual is diagnosed and treated) is almost eight times greater than the annual exogenous force of infection $\exp[\gamma_0]/\alpha = 7.7$ (2.4–24.2). The excess hazard due to exposure to a single household primary case (defined heuristically above and formally in Section 1.4 of the eAppendix; <http://links.lww.com/EDE/B615>) is useful for comparison of the hazard of infection from this household source with the hazard of infection from exogenous sources. Larger values of the excess hazard indicate that more infectiousness is due to a single infective

individual with the given characteristics than due to outside sources. TB cases without microbiological confirmation exert a smaller hazard of TB among their close contacts compared with smear- or culture-positive cases (excess hazard ratio = 0.88, 0.82–0.93 for HIV– cases and 0.82, 0.57–0.94 for HIV+ cases).

Comparison to Estimated Risk Ratios From Poisson Regression

Modified Poisson regression is a standard method for estimating risk ratios for clustered binary outcomes in prospective cohort studies. Table 4 shows estimated adjusted risk ratios for all susceptibility-related parameters: age, sex, HIV status, IPT, and BCG vaccination. We employ the standard “modified” Poisson regression⁴² fitting framework using generalized estimating equations and heteroscedasticity-robust standard errors under an exchangeable household working correlation structure.⁷¹ Section 4 of the eAppendix; <http://links.lww.com/EDE/B615>, gives similar results obtained by using logistic regression with the same outcome and susceptibility-related covariates. Estimates of susceptibility hazard ratios are similar to risk ratios for the same variables obtained by modified Poisson regression, with heteroscedasticity-robust standard errors, shown in Table 4. However, the proposed transmission modeling approach reveals much more about the transmission of TB than is available through traditional analysis of aggregate risk. Specifically, we report IHR estimates for transmission-related variables and the community force of infection, information that is not obtained from traditional regression analyses or existing transmission modeling approaches.

DISCUSSION

Transmission Modeling and Regression Analysis

Longitudinal household-based studies of infectious disease outcomes reveal features of susceptible subjects’ exposure to infectiousness, but this information is often ignored in statistical analyses. Hybrid transmission modeling and regression approaches have been proposed, but are not yet widely used. The approach we employed uses assessments of exposure over time to disentangle individual risk factors from exposure to infection, while simultaneously estimating the overall community force of infection. Estimates of susceptibility parameters under the transmission hazard model have some similarities with risk ratios from Poisson regression and previous analyses of the same study,^{12,16,27,57} but traditional regression analyses do not provide information about transmissibility or infectiousness of infected individuals.

The approach outlined in this article is subject to several limitations. We imposed parametric modeling assumptions to ensure interpretability and identifiability of parameters: constant exogenous force of infection, conditional independence of waiting times to transmission, and a log-linear model for infection hazards. When baseline enrollment of individuals happens simultaneously, it may be possible to estimate a

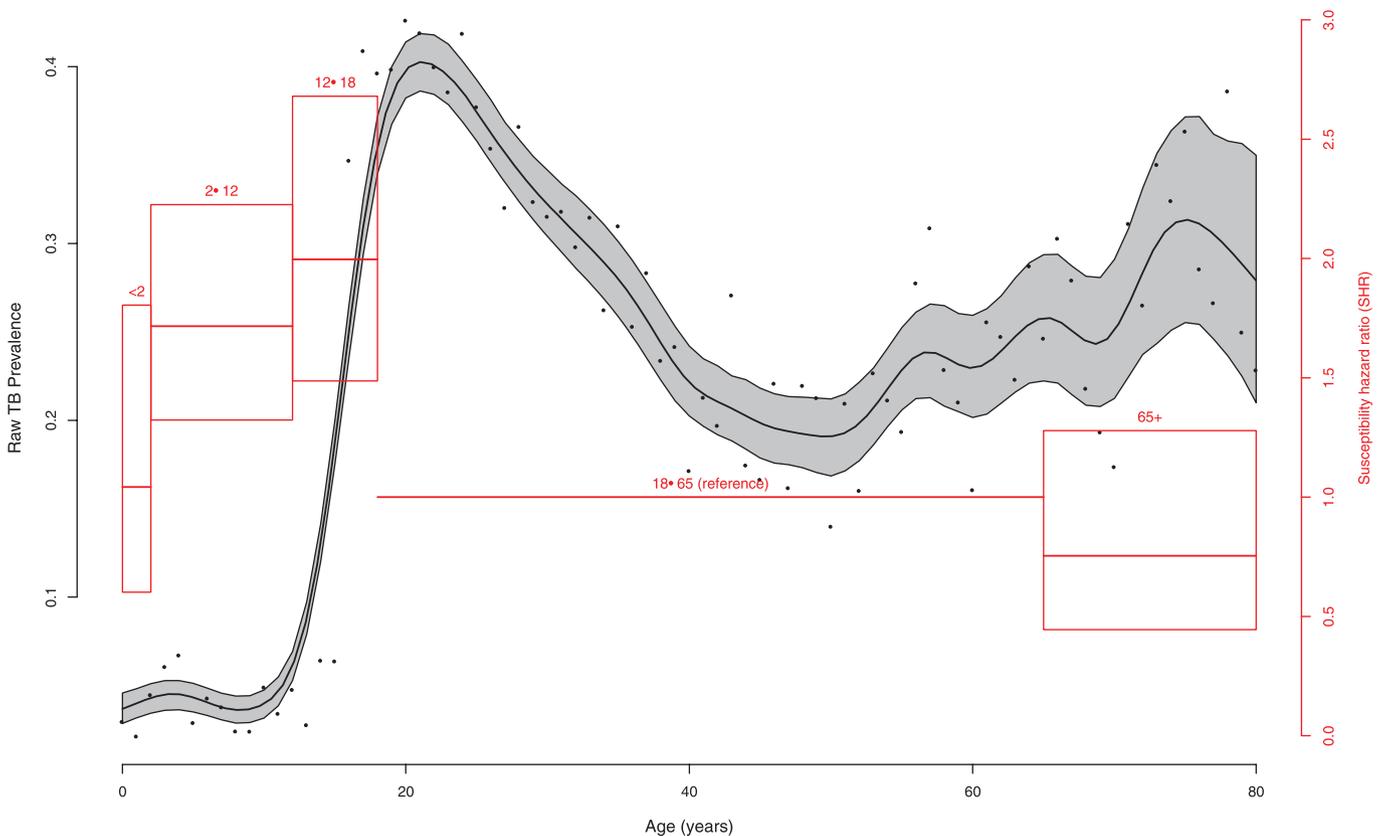


FIGURE 2. Raw TB prevalence and adjusted susceptibility hazard ratios by age. Raw prevalence (left-hand vertical axis) is estimated by Kernel regression and bootstrap pointwise confidence interval. Estimated age susceptibility hazard ratios and confidence intervals overlaid, with the hazard ratio scale on the right-hand vertical axis. Figure is available in color online.

time-varying exogenous force of infection $\alpha(t)$. Because all households in this study had at least one member with active TB at baseline, α may not be interpretable as the overall community force of infection if the within-household transmission model is misspecified. Instead, α could be interpreted as a marginal TB risk not accounted for by other covariates, conditional on at least one household member with active TB. We also employed a parametric model for the reduction in risk of transmission following onset of treatment at the time of diagnosis. Misspecification of these hazards or unmodeled heterogeneity in individual- or household-level risk could bias inferences of regression coefficients. Future extensions of this model could incorporate covariates and prior information to account for sources of variation in exogenous risk, as in Bayesian hierarchical regression models.

Implications of Findings for Tuberculosis

SHR estimates largely agree with previously published risk and odds ratios computed using longitudinal or cross-sectional (at baseline enrollment) TB outcomes. Younger people appear to be at increased risk of active TB disease when exposed,^{8–10,12,72} highlighting the difference between the raw TB prevalence and adjusted estimates shown in Figure 2. HIV+ individuals have greater susceptibility to

TABLE 4. Results from Modified Poisson Regression.

| | Risk Ratio | 2.5% | 97.5% |
|-------------|------------|------|-------|
| Intercept | 0.04 | 0.03 | 0.05 |
| Age | | | |
| 0–2 years | 1.00 | 0.58 | 1.73 |
| 2–12 years | 1.61 | 1.21 | 2.15 |
| 12–18 years | 1.93 | 1.44 | 2.57 |
| >65 years | 0.83 | 0.50 | 1.37 |
| Male sex | 1.21 | 1.00 | 1.46 |
| HIV+ | 3.60 | 1.40 | 9.24 |
| IPT | 0.30 | 0.20 | 0.44 |
| BCG | 0.67 | 0.52 | 0.87 |

TB.^{22–26,57,73} Although it is known that HIV+ people are less likely to transmit TB when infected,^{21,22,26,57} our estimated IHR for positive HIV status in this analysis is not different from one; previous analysis of these data suggested that the degree of HIV-associated immunosuppression (CD4) is associated with decreased transmission.⁵⁷ Furthermore, MC+ individuals may exert a slightly larger hazard of infection on their susceptible household members^{16,28–30} than individuals who are diagnosed without microbiological confirmation, which

is consistent with previous data that show a positive relationship between bacterial load in the sputum and risk of infection among contacts.

As compared with a traditional regression-based approach, our mechanistic model allows us to make sense of seemingly contradictory inferences about the relative importance of the exogenous force of TB infection experienced by these household contacts of known TB patients.⁷⁴ We estimate the exogenous force of infection to be approximately $\alpha = 0.01$ infection per susceptible person-year, which is substantially higher than the average exposure risk of community members in Lima at large where the annual TB incidence was between 150 and 300 cases/100,000 (0.0015–0.003) during the period of the study. This suggests that household contacts were at a higher than average risk of exogenous exposure (possibly as result of living in a relative transmission hotspot) or that our model did not accurately capture all of the risk associated with within-home exposure. One possibility is that household contacts of TB cases may share a higher risk of progression given infection due to factors not accounted for here, such as poor nutrition. A single infectious primary case (in the absence of other infectiousness effects) is estimated to exert a hazard of $\exp[\gamma_0] = 0.08$ on each susceptible in their household. When the infective individual is microbiologically confirmed and does not have HIV, this hazard is increased.

Our application of this novel method is subject to some additional caveats. Outcome misclassification could occur if some subjects were TB infected but not diagnosed at enrollment. Some subjects classified as not TB infected at baseline may have had undetected latent TB infection. In this case, the parameter α retains an interpretation as the adjusted overall risk of progressing to active TB disease in the absence of household-level infectiousness but may not accurately reflect the exogenous force of infection. The onset of active TB, and hence onset of infectiousness, was unobserved for secondary cases and is known only to lie within the time interval between the last TB-negative home visit and the date of diagnosis. Because we could not rule out possible dependence of clinic visit and diagnosis dates on the date of active disease onset, we chose not to impose the parametric conditional distribution for disease onset time, conditional on diagnosis date. Instead, we implemented a multiple imputation estimator that uniformly samples onset times within these intervals, which may result in a loss of statistical efficiency.

This modeling framework has applicability to diseases other than TB and to questions that extend beyond household-based transmission. For example, in the context of a public health response, e.g., the Ebola outbreak in West Africa, data are often collected from clustered environments (households, villages, schools, hospitals) in the course of contact investigation. In these contexts, when the exogenous infection rate may be the primary outcome of interest, it is still necessary to account for the role of transmission within clusters to obtain

accurate estimates of community exposure that make maximal use of available data.

ACKNOWLEDGMENTS

We are grateful to Chuan Chin Huang, Olga Morozova, Megan Murray, Laura F. White, and Zibiao Zhang for helpful comments on the manuscript. We also thank Mercedes Becerra, Leo Lecca, and all the members of *Socios En Salud* in Lima, Peru.

REFERENCES

- Teixeira L, Perkins MD, Johnson JL, et al. Infection and disease among household contacts of patients with multidrug-resistant tuberculosis. *Int J Tuberc Lung Dis*. 2001;5:321–328.
- Guwatudde D, Nakakeeto M, Jones-Lopez EC, et al. Tuberculosis in household contacts of infectious cases in Kampala, Uganda. *Am J Epidemiol*. 2003;158:887–898.
- Morrison J, Pai M, Hopewell PC. Tuberculosis and latent tuberculosis infection in close contacts of people with pulmonary tuberculosis in low-income and middle-income countries: a systematic review and meta-analysis. *Lancet Infect Dis*. 2008;8:359–368.
- van Schalkwyk C, Variava E, Shapiro AE, et al. Incidence of TB and HIV in prospectively followed household contacts of TB index patients in South Africa. *PLoS One*. 2014;9:e95372.
- Lienhardt C. From exposure to disease: the role of environmental factors in susceptibility to and development of tuberculosis. *Epidemiol Rev*. 2001;23:288–301.
- Tornee S, Kaewkungwal J, Fungladda W, Silachamroon U, Akarasewi P, Sunakorn P. Risk factors for tuberculosis infection among household contacts in Bangkok, Thailand. *Southeast Asian J Trop Med Public Health*. 2004;35:375–383.
- Middelkoop K, Bekker LG, Morrow C, Lee N, Wood R. Decreasing household contribution to TB transmission with age: a retrospective geographic analysis of young people in a South African township. *BMC Infect Dis*. 2014;14:221.
- Lienhardt C, Sillah J, Fielding K, et al. Risk factors for tuberculosis infection in children in contact with infectious tuberculosis cases in the Gambia, West Africa. *Pediatrics*. 2003;111(5 Pt 1):e608–e614.
- Singh M, Mynak ML, Kumar L, Mathew JL, Jindal SK. Prevalence and risk factors for transmission of infection among children in household contact with adults having pulmonary tuberculosis. *Arch Dis Child*. 2005;90:624–628.
- Middelkoop K, Bekker LG, Morrow C, Zwane E, Wood R. Childhood tuberculosis infection and disease: a spatial and temporal transmission analysis in a South African township. *S Afr Med J*. 2009;99:738–743.
- Wood R, Johnstone-Robertson S, Uys P, et al. Tuberculosis transmission to young children in a South African community: modeling household and community infection risks. *Clin Infect Dis*. 2010;51:401–408.
- Zelner JL, Murray MB, Becerra MC, et al. Age-specific risks of tuberculosis infection from household and community exposures and opportunities for interventions in a high-burden setting. *Am J Epidemiol*. 2014;180:853–861.
- Dodd PJ, Looker C, Plumb ID, et al. Age- and sex-specific social contact patterns and incidence of *Mycobacterium tuberculosis* infection. *Am J Epidemiol*. 2016;183:156–166.
- Colditz GA, Brewer TF, Berkey CS, et al. Efficacy of BCG vaccine in the prevention of tuberculosis. Meta-analysis of the published literature. *JAMA*. 1994;271:698–702.
- Mangtani P, Abubakar I, Ariti C, et al. Protection by BCG vaccine against tuberculosis: a systematic review of randomized controlled trials. *Clin Infect Dis*. 2014;58:470–480.
- Zelner JL, Murray MB, Becerra MC, et al. *Bacillus Calmette-Guérin* and isoniazid preventive therapy protect contacts of patients with tuberculosis. *Am J Respir Crit Care Med*. 2014;189:853–859.
- Nemes E, Geldenhuys H, Rozot V, et al.; C-040-404 Study Team. Prevention of *M. tuberculosis* infection with H4:IC31 vaccine or BCG revaccination. *N Engl J Med*. 2018;379:138–149.

18. Rodrigo T, Caylà JA, García de Olalla P, et al. Characteristics of tuberculosis patients who generate secondary cases. *Int J Tuberc Lung Dis*. 1997;1:352–357.
19. 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.
20. Crampin AC, Floyd S, Ngwira BM, et al. Assessment and evaluation of contact as a risk factor for tuberculosis in rural Africa. *Int J Tuberc Lung Dis*. 2008;12:612–618.
21. 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.
22. 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.
23. Mohammad Z, Naing NN, Salleh R, Ahmad N, Hamzah WM, Mahmud R. A preliminary study of the influence of HIV infection in the transmission of tuberculosis. *Southeast Asian J Trop Med Public Health*. 2002;33:92–98.
24. Pai M, McCulloch M, Colford JM Jr. Meta-analysis of the impact of HIV on the infectiousness of tuberculosis: methodological concerns. *Clin Infect Dis*. 2002;34:1285–1287.
25. Kenyon TA, Creek T, Laserson K, et al. Risk factors for transmission of *Mycobacterium tuberculosis* from HIV-infected tuberculosis patients, Botswana. *Int J Tuberc Lung Dis*. 2002;6:843–850.
26. Crampin AC, Glynn JR, Traore H, et al. Tuberculosis transmission attributable to close contacts and HIV status, Malawi. *Emerg Infect Dis*. 2006;12:729–735.
27. Huang CC, Tchetgen ET, Becerra MC, et al. Cigarette smoking among tuberculosis patients increases risk of transmission to child contacts. *Int J Tuberc Lung Dis*. 2014;18:1285–1291.
28. Shaw JB, Wynn-Williams N. Infectivity of pulmonary tuberculosis in relation to sputum status. *Am Rev Tuberc*. 1954;69:724–732.
29. Lutong L, Bei Z. Association of prevalence of tuberculin reactions with closeness of contact among household contacts of new smear-positive pulmonary tuberculosis patients. *Int J Tuberc Lung Dis*. 2000;4:275–277.
30. Ferrarini MA, Spina FG, Weckx LY, Lederman HM, De Moraes-Pinto MI. Rate of tuberculosis infection in children and adolescents with household contact with adults with active pulmonary tuberculosis as assessed by tuberculin skin test and interferon-gamma release assays. *Epidemiol Infect*. 2016;144:712–723.
31. Verver S, Warren RM, Munch Z, et al. Proportion of tuberculosis transmission that takes place in households in a high-incidence area. *Lancet*. 2004;363:212–214.
32. Brooks-Pollock E, Becerra MC, Goldstein E, Cohen T, Murray MB. Epidemiologic inference from the distribution of tuberculosis cases in households in Lima, Peru. *J Infect Dis*. 2011;203:1582–1589.
33. Madico G, Gilman RH, Checkley W, et al. Community infection ratio as an indicator for tuberculosis control. *Lancet*. 1995;345:416–419.
34. Akhtar S, Carpenter TE, Rathi SK. A chain-binomial model for intra-household spread of *Mycobacterium tuberculosis* in a low socio-economic setting in Pakistan. *Epidemiol Infect*. 2007;135:27–33.
35. Longini IM Jr, Koopman JS. Household and community transmission parameters from final distributions of infections in households. *Biometrics*. 1982;38:115–126.
36. Rampey AH Jr, Longini IM Jr, Haber M, Monto AS. A discrete-time model for the statistical analysis of infectious disease incidence data. *Biometrics*. 1992;48:117–128.
37. Anderson RM, May RM. *Infectious Diseases of Humans: Dynamics and Control*. New York: Oxford University Press; 1991.
38. Andersson H, Britton T. *Stochastic Epidemic Models and Their Statistical Analysis*. New York: Springer Science & Business Media; 2012.
39. Cauchemez S, Carrat F, Viboud C, Valleron AJ, Boëlle PY. A bayesian MCMC approach to study transmission of influenza: application to household longitudinal data. *Stat Med*. 2004;23:3469–3487.
40. Kasaie P, Andrews JR, Kelton WD, Dowdy DW. Timing of tuberculosis transmission and the impact of household contact tracing. An agent-based simulation model. *Am J Respir Crit Care Med*. 2014;189:845–852.
41. McNutt LA, Wu C, Xue X, Hafner JP. Estimating the relative risk in cohort studies and clinical trials of common outcomes. *Am J Epidemiol*. 2003;157:940–943.
42. Zou G. A modified poisson regression approach to prospective studies with binary data. *Am J Epidemiol*. 2004;159:702–706.
43. Morozova O, Cohen T, Crawford FW. Risk ratios for contagious outcomes. *J R Soc Interface*. 2018;15:20170696.
44. Halloran ME, Struchiner CJ. Causal inference in infectious diseases. *Epidemiology*. 1995;6:142–151.
45. Longini IM Jr, Koopman JS, Haber M, Cotsonis GA. Statistical inference for infectious diseases. Risk-specific household and community transmission parameters. *Am J Epidemiol*. 1988;128:845–859.
46. Koopman JS, Longini IM Jr, Jacquez JA, et al. Assessing risk factors for transmission of infection. *Am J Epidemiol*. 1991;133:1199–1209.
47. Eisenberg JN, Lewis BL, Porco TC, Hubbard AH, Colford JM Jr. Bias due to secondary transmission in estimation of attributable risk from intervention trials. *Epidemiology*. 2003;14:442–450.
48. Tsang TK, Cauchemez S, Perera RA, et al. Association between antibody titers and protection against influenza virus infection within households. *J Infect Dis*. 2014;210:684–692.
49. Tsang TK, Fang VJ, Chan KH, et al. Individual correlates of infectivity of influenza A virus infections in households. *PLoS One*. 2016;11:e0154418.
50. Cheung DH, Tsang TK, Fang VJ, et al. Association of oseltamivir treatment with virus shedding, illness, and household transmission of influenza viruses. *J Infect Dis*. 2015;212:391–396.
51. Kenah E. Semiparametric relative-risk regression for infectious disease transmission data. *J Am Stat Assoc*. 2015;110:313–325.
52. Auranen K, et al. Transmission of pneumococcal carriage in families: a latent Markov process model for binary longitudinal data. *Journal of the American Statistical Association*. 2000;95:1044–1053.
53. Cauchemez S, et al. Investigating heterogeneity in pneumococcal transmission: a Bayesian MCMC approach applied to a follow-up of schools. *Journal of the American Statistical Association* 2006;101:946–958.
54. Petrie JG, Eisenberg MC, Ng S, et al. Application of an individual-based transmission hazard model for estimation of influenza vaccine effectiveness in a household cohort. *Am J Epidemiol*. 2017;186:1380–1388.
55. Rhodes PH, Halloran ME, Longini IM Jr. Counting process models for infectious disease data: distinguishing exposure to infection from susceptibility. *Journal of the Royal Statistical Society. Series B (Methodological)*. 1996;58:751–762.
56. Tsang TK, Chen TM, Longini IM Jr, Halloran ME, Wu Y, Yang Y. Transmission of norovirus in Urban versus rural households in a large community outbreak in China. *Epidemiology*. 2018;29:675–683.
57. Huang CC, Tchetgen ET, Becerra MC, et al. The effect of HIV-related immunosuppression on the risk of tuberculosis transmission to household contacts. *Clin Infect Dis*. 2014;58:765–774.
58. Galea JT, Contreras C, Lecca L, et al. Rapid home-based hiv testing to reduce costs in a large tuberculosis Cohort Study. *Public Health Action*. 2013;3:172–174.
59. Floyd S, Ponnighaus JM, Bliss L, et al. BCG scars in northern Malawi: sensitivity and repeatability of scar reading, and factors affecting scar size. *Int J Tuberc Lung Dis*. 2000;4:1133–1142.
60. Haber M, Longini IM Jr, Cotsonis GA. Models for the statistical analysis of infectious disease data. *Biometrics*. 1988;44:163–173.
61. Kenah E. Nonparametric survival analysis of infectious disease data. *J R Stat Soc Series B Stat Methodol*. 2013;75:277–303.
62. Becker NG. *Analysis of Infectious Disease Data*. New York: Chapman and Hall; 1989.
63. Wells WF. Airborne contagion and air hygiene. An Ecological Study of droplet infections. Cambridge: Harvard University Press; 1955.
64. Riley RL. Airborne infection. *Am J Med*. 1974;57:466–475.
65. Riley EC, Murphy G, Riley RL. Airborne spread of measles in a suburban elementary school. *Am J Epidemiol*. 1978;107:421–432.
66. Beggs CB, Noakes CJ, Sleight PA, Fletcher LA, Siddiqi K. The transmission of tuberculosis in confined spaces: an analytical review of alternative epidemiological models. *Int J Tuberc Lung Dis*. 2003;7:1015–1026.
67. Noakes CJ, Sleight PA. Applying the wells-riley equation to the risk of airborne infection in hospital environments: the importance of stochastic and proximity effects. In *Indoor Air 2008: The 11th International Conference on Indoor Air Quality and Cl*. White Rose, Leeds; 2008.
68. Little RJ, Rubin DB. *Statistical Analysis With Missing Data*. Hoboken: John Wiley & Sons; 2014.
69. Crofton J. The contribution of treatment to the prevention of tuberculosis. *Bull Int Union Tuberc*. 1962;32:643–653.

70. Rouillon A, Perdrizet S, Parrot R. Transmission of tubercle *bacilli*: the effects of chemotherapy. *Tubercle*. 1976;57:275–299.
71. Liang KY, Zeger SL. Longitudinal data analysis using generalized linear models. *Biometrika* 1986;13–22.
72. 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.
73. Ayles HM, Sismanidis C, Beyers N, Hayes RJ, Godfrey-Faussett P. ZAMSTAR, The Zambia South Africa TB and HIV Reduction Study: design of a 2 x 2 factorial community randomized trial. *Trials*. 2008;9:63.
74. McCreesh N, White RG. An explanation for the low proportion of tuberculosis that results from transmission between household and known social contacts. *Sci Rep*. 2018;8:5382.