

Epidemiologic Inference From the Distribution of Tuberculosis Cases in Households in Lima, Peru

Ellen Brooks-Pollock,^{1,a} Mercedes C. Becerra,^{2,3} Edward Goldstein,¹ Ted Cohen,^{1,3} and Megan B. Murray^{1,3,4}

¹Department of Epidemiology, Harvard School of Public Health, ²Department of Global Health and Social Medicine, Harvard Medical School, ³Division of Global Health Equity, Brigham and Women's Hospital, and ⁴Division of Infectious Diseases, Massachusetts General Hospital, Boston

Background. Tuberculosis (TB) often occurs among household contacts of people with active TB. It is unclear whether clustering of cases represents household transmission or shared household risk factors for TB.

Methods. We used cross-sectional data from 764 households in Lima, Peru, to estimate the relative contributions of household and community transmission, the average time between cases, and the immunity afforded by a previous TB infection.

Results. The distribution of cases per household suggests that almost 7 of 10 nonindex household cases were infected in the community rather than in the household. The average interval between household cases was 3.5 years. We observed a saturation effect in the number of cases per household and estimated that protective immunity conferred up to 35% reduction in the risk of disease.

Conclusions. Cross-sectional household data can elucidate the natural history and transmission dynamics of TB. In this high-incidence setting, we found that the majority of cases were attributable to community transmission and that household contacts of case patients derive some immunity from household exposures. Screening of household contacts may be an effective method of detecting new TB cases if carried out over several years.

Tuberculosis (TB) is a debilitating infectious disease that is responsible for ~2 million deaths worldwide annually [1, 2]. Infection occurs through the transmission of respiratory droplets from individuals with TB; therefore persons in close and sustained proximity to TB cases are at high risk of infection. Accordingly, much of our basic understanding of TB epidemiology is based on household studies of disease transmission that typically measure the risk of infection [3–5] or disease [6–8] among those living with an index case. Such studies have estimated the burden of TB among close contacts of TB cases [9], the effects of putative risk factors for infection and disease conditional on

exposure [10–12], and the time course over which secondary cases arise [9, 13]. Epidemiologic inference from household studies of TB, however, is complicated by the need to differentiate between transmission that occurs within the household of an index case and transmission from community exposure [14, 15]. In addition, exposure to *Mycobacterium tuberculosis* that does not result in disease may provide protective immunity from future infection or disease [16–18], further complicating the measurement of risk in exposed contacts. Researchers have addressed such methodologic issues by using molecular fingerprinting to identify transmission chains [15] and by nesting household studies within prospective studies that ascertain infection status at baseline [9, 19]. Although these studies have helped to elucidate TB transmission dynamics, they can be costly and take years to complete. We propose that program data may inform estimates of the relative risks of infection from household and community exposures, the degree of partial immunity conferred by a past infection, and the time between cases within households. This type of easily available data could be routinely collected and would provide alternative, cost-efficient methods to improve

Received 5 June 2010; accepted 30 December 2010.

Potential conflicts of interest: none reported.

^aPresent address: Department of Veterinary Medicine, University of Cambridge, UK

Correspondence: Ellen Brooks-Pollock, PhD, Department of Veterinary Medicine, University of Cambridge, Madingley Road, Cambridge, CB3 0ES, UK (ellen.brooks.pollock@gmail.com).

The Journal of Infectious Diseases 2011;203:1582–9

© The Author 2011. Published by Oxford University Press on behalf of the Infectious Diseases Society of America. All rights reserved. For Permissions, please e-mail: journals.permissions@oup.com

0022-1899 (print)/1537-6613 (online)/2011/20311-0012\$14.00

DOI: 10.1093/infdis/jir162

our understanding of disease dynamics and our ability to select disease control strategies.

The cross-sectional distribution of infection within households has been used to study the dynamics of acute respiratory diseases, such as influenza [20–23] and other diseases [24, 25]. Despite the limitations of cross-sectional data, they are more readily obtained than those from prospective studies of exposed cohorts. In the absence of household transmission, the number of cases within a household should reflect a random sample of community prevalence. Deviations from this pattern may be associated with household structure resulting from, for example, clustering of risk factors or household transmission. In this study, we extend the methods of Longini et al [20, 21] to study the distribution of cases of chronic infectious TB occurring in households in Lima, Peru, an area of high disease burden and a growing epidemic of highly drug-resistant disease [1, 26]. We use data from household contacts of index cases with multidrug-resistant TB (disease resistant to at least isoniazid and rifampin, the 2 most powerful antibiotics used to treat TB) to estimate the risk and time course of nonindex cases given a household exposure, the risk of disease acquired through community rather than household transmission, and the protective effect of previous exposure to TB.

METHODS

Study Population

We assessed the pattern of TB cases (persons with active disease) within households that were included in a retrospective cohort study conducted within metropolitan Lima, Peru [27]. Study households were identified by a household member who had been diagnosed and initiated treatment for laboratory-confirmed multidrug-resistant or extensively drug-resistant TB between 1996 and 2002. On enrollment, households were visited and basic demographic and clinical data for household members were collected, including birthdates, relationships between household members, and detailed histories of prior TB disease episodes and treatments received. Coprevalent TB cases were identified by sputum microscopic screening of household contacts. For the purposes of this analysis, the index case in a household will be defined as the first TB diagnosis among the people residing in a household. All subsequent cases will be defined as nonindex cases. We assume that all members of the study cohort have been exposed to TB by living with someone with active disease. We note that the index case in the household is not necessarily the drug-resistant TB case that prompted the household to be enrolled into the study, which we term the enrollment case for clarity.

Peru had an estimated incidence of 150 cases per 100,000 persons between 2004 and 2006 [1, 2] with a strong national TB program with high detection rates (~93%) and high levels of BCG vaccination at birth. The prevalence of human immunodeficiency

virus (HIV) infection was low in Lima and in our study population, with HIV coinfection in <.9% of enrollment cases. Notably, adult household contacts of known TB cases do not receive treatment for latent infection.

Data Analysis: Age-Specific Risks and Intrinsic Time Scales

We estimated the proportion of participants diagnosed with TB by age, and the distribution of ages at first diagnosis for the pooled participant data (irrespective of household affiliation). In household of ≤ 7 adults with >1 case, we estimated the generation time of cases within households from the distribution of times between the first and second diagnoses within households. The generation time within households provides only a crude estimate of the true serial interval, given that nonindex cases in the household may have been infected from an external source. Nevertheless, this quantity provides information for minimum lengths of time for increased symptom monitoring of household contacts once an index case has been diagnosed. We used the secondary attack rate (SAR) within households, calculated as $100 \times (\text{cumulative incidence} - 1) / (\text{total adults in household} - 1)$, as a first approximation of transmission risk [20].

Model

In each household, we measured the total number of adults (k) and the cumulative incidence of adults with disease (j). We used the date that TB treatment was started as a proxy for the disease onset, justified by the high detection rate and strong national TB program in Peru [1, 2]. For each household member, we defined a risk of community-acquired disease ($1-B$) and a risk of household-acquired disease from a single infectious household contact ($1-Q$), as described in [21]. Implicit in the model is the assumption that the transmission process began with the first diagnosed case.

We computed the unconditional probabilities p_{jk} of having j cases in a household of k adults, using the recursive formulae from Longini et al [21], as follows:

$$p_{jk} = \binom{k}{j} B^{k-j} Q^{j(k-j)} p_{jj} \text{ for } 0 \leq j < k \quad (1a)$$

$$p_{jj} = 1 - \sum_{r=0}^{j-1} p_{rj} \text{ otherwise.} \quad (1b)$$

Then, we condition on the probability that there was ≥ 1 case in the household, because this was a condition for entry into the study. The probabilities for the model are as follows:

$$\hat{p}_{jk} = P(j \text{ cases} | 1 \leq \text{cases}) = \frac{p_{jk}}{1 - p_{0k}} \text{ for } 1 \leq j \leq k \quad (2)$$

The SAR within a household (the probability that each household contact develops disease) increases with household size

if the probability for within-household transmission $(1-Q)$ is non-zero [21, 28] (for comparison, see [29]).

We estimated the community and household transmission parameters (B and Q , respectively) using Markov Chain Monte Carlo (MCMC) with Metropolis-Hastings sampling [23]. For each set of proposed parameters, the probability of observing the data given the model was computed for a log-likelihood

function, such that $L(B, Q) = \sum_{i=1}^{764} H_i(B, Q)$, where $H_i(B, Q)$ is the

logarithm of the likelihood of observing household number i , given parameters B and Q . The proposed parameters were accepted with a tolerance that ensured good mixing. Two million MCMC samples were taken to approximate the posterior density, with a burn-in of 100,000 samples. We used a χ^2 -test to measure goodness of fit.

Parameter Interpretation

The parameter B corresponds to the average probability of community-acquired disease during a person's adult life in this study setting and time period. The median age of all participants was 30.9 years, and the age distributions did not differ significantly between households of different sizes or with different numbers of cases. In addition to estimating B and Q for all diagnosed cases, we also estimated B and Q for cases diagnosed since 1996 to investigate the effect of including different time periods in the analysis.

Protective Immunity

The baseline model above includes the assumption that the probability of disease after a single exposure is independent of previous exposures to disease, and so does not account for any associated protective immunity. We looked for evidence of protective immunity by allowing the probability of disease to vary with the number of previous exposures. The baseline model was fitted to subcollections of households with at least a specified number of cases, thereby measuring the change in risk of disease in households as the cumulative number of cases increased.

For each number of cases $m \geq 1$, we restricted our attention to the collection C_m of households within the study cohort that had at least m cases. For example, C_4 is the collection of all households with ≥ 4 cumulative TB cases and C_1 is the entire data set (because households were entered in the study only after the enrollment case was identified). The probability that a household with k adults has j cases in the restricted collection C_m is

$$\tilde{p}_{jkm} = P(j \text{ cases} | \text{at least } m \text{ cases}) = \frac{p_{jk}}{1 - \sum_{r=0}^{m-1} p_{rk}} \quad (3)$$

From equation (3), we inferred the transmission parameters (B_m, Q_m) in the restricted group C_m , using MCMC as above. We

used the probabilities \hat{p}_{jk} and \tilde{p}_{jkm} to predict the expected number of cases in households with at least m cases for the 2 models. Given that the restricted model is conditioned on there being at least m prior household cases, we estimated the efficacy of any protective effect of prior exposure by taking the ratio of expected number of cases under each model:

$$\text{immunity}(m) = 1 - \frac{\sum_{C_m} \sum_{i=m+1}^k \tilde{p}_{ikm} (i-m)}{\sum_{C_m} \sum_{i=m+1}^k \hat{p}_{ik} (i-m)}$$

where k is the number of adults in a household in C_m .

RESULTS

In the 764 participant households, the mean number of individuals per household was 7, with a range between 1 and 27 household members.

Age-Specific Risks

Of individuals in the cohort with previous diagnoses of TB, 78% were between 14 and 35 years of age at first diagnosis (Figure 1a). The cumulative incidence of TB increased by 1% per year of life between the ages of 0 and 20 years, to an average cumulative incidence of $\sim 35\%$ ($R^2 = .80$, Figure 1b). There was no clear pattern of disease associated with age for older age groups in the cohort.

Time Between Diagnoses in Households

The mean waiting time between first and second diagnoses in households with ≤ 7 adults was 3.5 years, and the distribution was skewed toward shorter times, with a median interval between cases of 1.65 years (Figure 1c).

Modeling Chains of Infection Within Households

The data suggest that the SAR within households increased with the number of adults for households of ≤ 7 adults (Figure 1d). The median age of adults within households did not vary with household size, suggesting that the increase in SAR was associated with transmission. For larger households, rates of pair-wise contacts between individuals presumably decreased. Therefore, for parameter inference, we used only households containing ≤ 7 adults.

There was good agreement between distributions of cases per household produced using the baseline model and the data for households containing 2–7 adults (Figure 2). The median of the posterior density provided the best point estimate, yielding the lowest χ^2 fit, with $\chi^2 = \{0.83, 0.64, 2.69, 5.78, 0.58, 5.95\}$ and P values = $\{0.36, 0.73, 0.44, 0.22, 0.99, 0.43\}$ with 1–6 degrees of freedom. A visual comparison between the model and the data for households with 2–7 adults is shown in Figure 2. In general, the model overestimated the proportion of larger households with intermediate levels of infection.

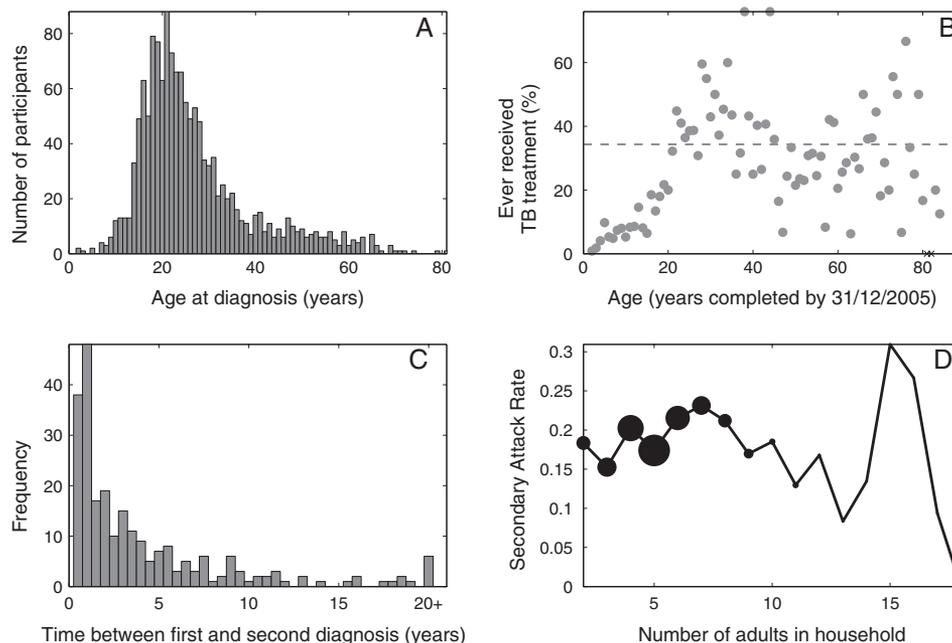


Figure 1. Characteristics of cohort study. *A*, Distribution of ages at first diagnosis. *B*, Proportion of participants by age (as of 31 December 2005) who had been treated for active tuberculosis (TB); x's toward right end of horizontal axis denote ages for which there were no data; dashed horizontal line, mean cumulative incidence in persons aged ≥ 20 years. *C*, Time lag between diagnoses in the first 2 individuals in the same household with active disease, among households with ≤ 7 adults; data are binned in 6-month intervals. *D*, Secondary attack rate, calculated as (average no. of cases in household $- 1$)/(household size $- 1$). Sizes of markers are proportional to numbers of households of each size.

Probability of Community Versus Household Infection

The expected value for the probability of community-acquired disease using all cases was $(1-B)=0.19$, and for the probability of acquiring disease from a single infectious household member it was $(1-Q)=0.1$, illustrating that most household contacts did not develop disease (Figure 3). On average, in households with multiple cases, we estimate that 7 of 10 nonindex cases were actually infected in the community rather than by their infectious household contact. The marginal posterior density for household transmission has a tighter distribution (variance, .004) than the marginal posterior density for community transmission (variance, .008) indicating that the probability of household transmission is better described by the data than the probability of community transmission (Figure 3a).

Restricting the analysis to cases diagnosed since 1996 reduces the expected risk of community- and household-acquired disease to $(1-B)=0.12$ and $(1-Q)=0.07$, respectively, and the percentage of nonindex cases that acquired disease in the community to 65%.

The joint posterior density demonstrates that both household and community transmission contributed to the observed distributions of cases within households located in this high-incidence area (Figure 3b). The joint posterior density is most consistent with a higher risk of community acquisition than household acquisition among household contacts (62% of the best-fit region), but the posterior density also suggests that

household transmission is important (8% of the best-fit region where there is an equal role of community and household transmission and 30% of the best-fit region where household transmission is actually more important than community transmission). Although we estimate a similar risk of transmission from the community and from within the household to the contacts living with only a single TB case patient, nearly half of the households in this study (49.4%) had a documented presence of ≥ 2 infectious cases. In these households, the increased exposure within the house increased the probability that subsequent cases were due to household transmission. In the most extreme example, a household where TB had been diagnosed in 11 of 16 adults, the model predicted that >4 of 5 subsequent cases in the home would have been due to household transmission.

Multiple Reinfections and Immunity

The goodness of fit improved when we allowed the probability of disease to vary with the number of cumulative cases within a household (ratio of χ^2 values between the restricted model equation (3) and the baseline model equation (2): $(m=2)0.99$, $(m=3)0.87$, $(m=4)0.61$). By comparing the expected number of future cases in each model, we measured a protective effect associated with exposure that increased with the number of previous cases within a household. We estimated a maximum efficacy of $<35\%$ (table 1).

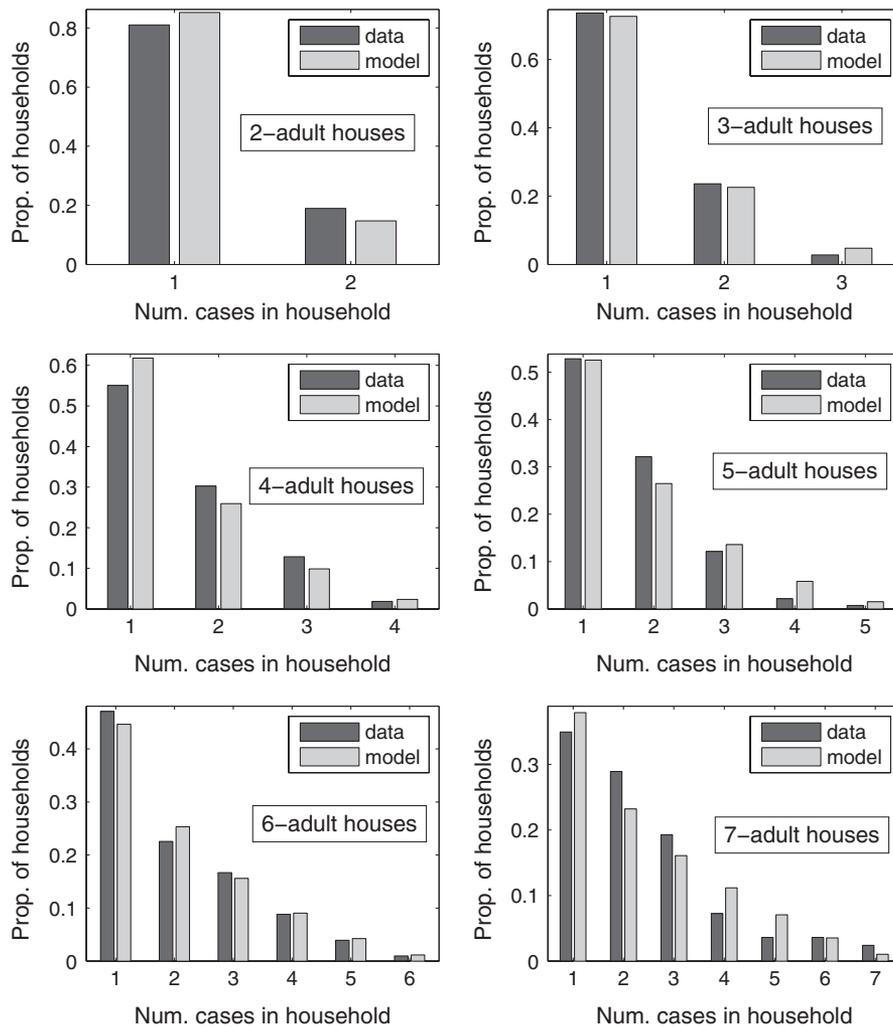


Figure 2. Comparison between data and models for households with 2–7 adults. Point estimates for B and Q were taken as median values from posterior densities. Num., number of; Prop., proportion.

DISCUSSION

In this study, we analyzed the cross-sectional distribution of TB cases within households in urban Lima, Peru, to estimate the relative contributions of household and community transmission, the average time between household cases, and the protective immunity conferred by a previous exposure. Our findings rely on data that are often routinely collected within well-functioning TB programs, and our parameter estimates are consistent with previous estimates derived from more labor-intensive approaches such as cohort studies and molecular epidemiology studies. The results of this study will contribute to the practical design of case-finding strategies and help elucidate TB transmission dynamics and parameterize more accurate epidemic models.

Consistent with findings from other high burden settings, we find that for the time covered by this analysis a high proportion of nonindex case patients in households with a previously identified

case were probably infected in the community rather than through household transmission. Our finding that >65% of the nonindex cases were due to community transmission is comparable with the results of a recent molecular epidemiologic study in Cape Town, South Africa, which compared DNA fingerprints of *M. tuberculosis* isolates from households with ≥ 2 cases of TB. This study found that 61% of paired isolates had mismatching fingerprints, indicating that *at least* 61% of secondary cases were due to community transmission [15]. This result would underestimate the true proportion of cases due to community transmission if some community transmission events involved the same strain that infected a household index case.

One of the challenges in TB epidemiology is the long time scales of transmission, infection, and disease during which other factors such as disease prevalence or household composition may have changed. The methods presented here, like other techniques for quantifying TB epidemiology (such as [15]), therefore result in parameter estimates that represent averages

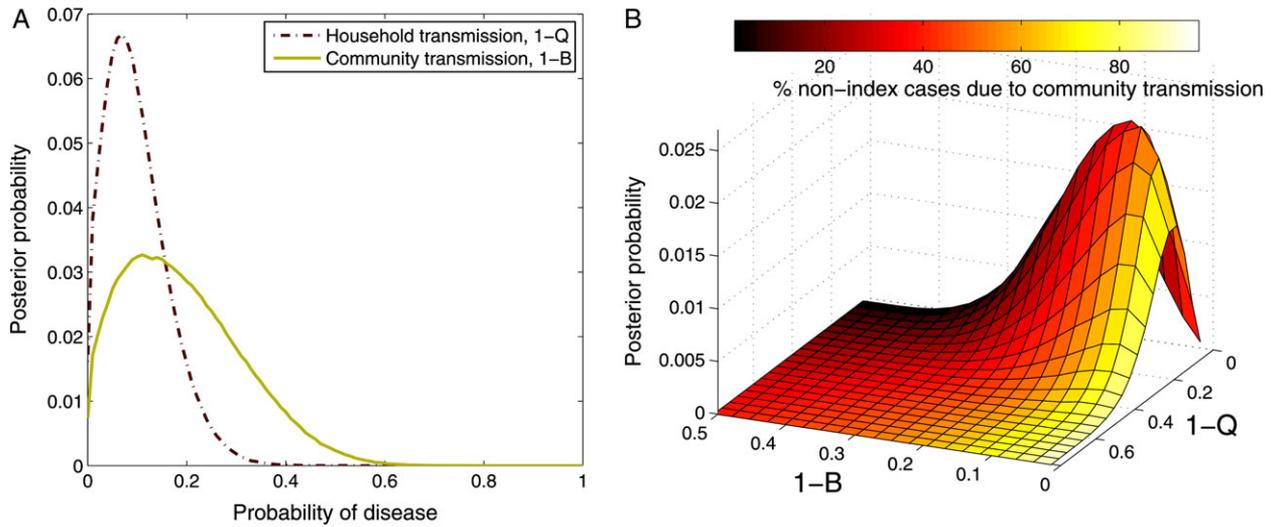


Figure 3. Posterior probability densities for household and community disease parameters. *A*, Marginal posterior densities for probabilities of community-acquired disease (*solid line*), $(1 - B)$, and household-acquired disease (*dashed line*), $(1 - Q)$. *B*, Joint posterior density for household and community transmission. Vertical axis indicates probability of observing a particular combination of household $(1 - Q)$ and community $(1 - B)$ transmission. Color of the surface indicates the percentage of nonindex cases in a household that are due to community transmission, rather than household transmission, for each parameter combination; darker colors indicate majority household transmission, and lighter colors, majority community transmission. Both figure parts were produced using 2 million samples for the posterior density.

over the time period of the analysis. In Peru, the estimated incidence of all forms of TB has fallen by ~60% since 1990 [2]; this decreasing incidence is expected to result in decreases in the risk of community-acquired disease. By restricting the time period of the analysis, we observed a decrease in both the risk of community- and household-acquired disease, as well as a small decrease in the percentage of cases resulting from community transmission, from 70% to 65%. We note that the percentage of community-acquired cases may also have been overestimated if household members with active TB died before the study.

The elevated risk of TB disease in household contacts of index cases did not scale proportionately with the number of cases in the household. This is consistent with the idea that previous TB exposure provides some protection against future disease with rechallenge [30–32]. By using residence with a single case as a unit of exposure, we were able to estimate the degree of

protection afforded by a previous exposure. In future studies, additionally collecting the infection status of household members would allow more detailed conclusions about the role of latent infection in the development of disease. Our estimate of 35% protection is lower than that of Sutherland and colleagues, who used trends in population level estimates of infection and disease in the Netherlands to estimate that previous exposure afforded 65% protection [30, 31], but our results are consistent with data from England and Wales in the early part of the 20th century [32]. Similarly, the incidence rate of disease in individuals who had positive skin test results at baseline in the control arm of a BCG vaccine trial suggested a 50% protective effect associated with infection [33, 34]. The reduced protection found in our study might reflect a biologic difference due to multiple reinfections in this high-burden area, or the difference between protective immunity conferred by exposure to drug-

Table 1. Number of Subsequent Cases After the First *m* Cases in the Data and in the Baseline and Restricted Models

Preexisting cases in household, no.	Subsequent cases, no.	Subsequent cases in baseline model, expected no.	Subsequent cases in restricted model, expected no.	Protective effect
1	704	705.04	705.04	0
2	333	339.88	330.12	.03
3	155	170.79	156.96	.08
4	75	74.46	64.89	.17
5	30	46.09	32.10	.32
6	12	19.44	13.23	.35

NOTE. The baseline model assumption was that the probability of disease was independent of number of previous exposures; in the restricted model, the probability of disease was allowed to vary with the number of previous exposures. The protective effect of *m* exposures was calculated by comparing the baseline and restricted models.

sensitive and drug-resistant disease. Observed a saturation effect in the number of cases per household, suggesting that even with high levels of exposure some individuals do not develop disease.

We propose that these types of household data may be used to inform the minimum duration that household contacts should be monitored and actively assessed for disease after the discovery of an infectious household case. We estimated a median time between diagnoses within households of 3.5 years, although the wide variance in time between successive cases indicates that secondary cases may occur many years after the primary case. This estimate lies on the upper end of population-wide estimates of 1–7 years calculated by Blower et al [35]. The extended tail of our distribution is quantitatively similar to previous estimates from population-level data in Europe [35–37] and South Africa [38] and reflects a small number of secondary cases occurring decades after an index case. In these situations, molecular fingerprinting can distinguish between reinfection and reactivation.

The model used in this study necessitated a number of simplifying assumptions. First, we modeled the transmission process as if it began with the first case of active disease in the household. In reality, household contacts may have had a prior infection from community exposure that provided some immunity to infection from their household contact. This assumption may alter our estimate of within household transmission by overestimating the impact of the first case, although the effect of this assumption is reduced if there is widespread and uniform exposure in the community. The effect of immunity was estimated from a ratio of 2 models with the same assumption, and therefore it should also be robust.

In this analysis, we used the cumulative number of cases among adults within a household to infer transmission parameters. We excluded persons <15 years of age because of the different natural history of disease in children. A potential extension of the model could include children to understand the relative risks faced by children in households with TB cases and their role in transmission [39]. Additionally, stratifying adults by age could provide more detailed estimates of community transmission over time.

We parameterized our model with data from a cohort study of households living with multidrug-resistant patients. This means that our estimate of community-acquired disease will not reflect community prevalence and will be greater than population-wide estimates. Repeating this type of analysis among household contacts of patients with drug-sensitive TB would lead to improved estimates of prevalence and an increased understanding of the mechanisms involved in the development and transmission of drug-resistant strains.

Funding

This work was supported by the National Institutes of Health (grant U19 AI0176217 [PI: MBM] to E. B. P., T. C. and M. B. M.) and the Charles H. Hood Foundation (M. B.).

Acknowledgments

We thank the Charles H. Hood Foundation for funding data collection. We are also grateful to Leon Danon for helpful discussions and to two anonymous referees for their constructive comments.

References

1. Dye C, Scheele S, Dolin P, Pathania V, Raviglione MC, for the WHO Global Surveillance and Monitoring Project. Global burden of tuberculosis: estimated incidence, prevalence, and Mortality by Country. *JAMA* **1999**; 282:677–86.
2. World Health Organization. Global Tuberculosis Control: Epidemiology, Strategy, Financing: WHO Report 2009. Geneva: WHO Press, **2009**.
3. Teixeira L, Perkins M, Johnson J, Keller R. Infection and disease among household contacts of patients with multidrug-resistant tuberculosis. *Int J Tuberc Lung Dis* **2001**; 5:321–8.
4. Guwatudde D, Nakakeeto M, Jones-Lopez EC, et al. Tuberculosis in household contacts of infectious cases in Kampala, Uganda. *Am J Epidemiol* **2003**; 158:887–98.
5. Morrison J, Pai M, Hopewell P. 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**; 8:359–68.
6. Frost W. Risk of persons in familial contact with pulmonary tuberculosis. *Am J Public Health* **1933**; 23:426–32.
7. Puffer R, Zeidberg LD, Dillon A, Gass R, Hutcheson R. Tuberculosis attack and death rates of household associates. *Am Rev Tuberc* **1952**; 65:111–27.
8. Claessens NJM, Gausi FF, Meijnen S, Weismuller MM, Salaniponi FM, Harries AD. High frequency of tuberculosis in households of index TB patients. *Int J Tuberc Lung Dis* **2002**; 6:266–9.
9. Bayona J, Chavez-Pachas A, Palacios E, Llaro K, Sapag R, Becerra M. Contact investigations as a means of detection and timely treatment of persons with infectious multidrug-resistant tuberculosis. *Int J Tuberculosis Lung Dis* **2003**; 7:S501–9.
10. Lienhardt C. From exposure to disease: the role of environmental factors in susceptibility to and development of tuberculosis. *Epidemiol Rev* **2001**; 23:288–301.
11. Lienhardt C, Fielding K, Sillah J, et al. Risk factors for tuberculosis infection in Sub-Saharan Africa: a contact study in the Gambia. *Am J Respir Crit Care Med* **2003**; 168:448–55.
12. Salinas C, Capelastegui A, Altube L, et al. Longitudinal incidence of tuberculosis in a cohort of contacts: factors associated with the disease. *Arc Bronconeumol* **2007**; 43:317–23.
13. Teale C, Cundall DB, Pearson SB. Time of development of tuberculosis in contacts. *Respir Med* **1991**; 85:475–7.
14. Madico G, Gilman R, Checkley W, Cabrera L. Community infection ratio as an indicator for tuberculosis control. *Lancet* **1995**; 345:416–9.
15. Verver S, Warren R, Munch Z, et al. Proportion of tuberculosis transmission that takes place in households in a high-incidence area. *Lancet* **2004**; 363:212–4.
16. van-Rie A, Warren R, Richardson M, et al. Exogenous reinfection as a cause of recurrent tuberculosis after curative treatment. *N Engl J Med* **1999**; 341:1174.
17. Flynn JL, Chan J. Tuberculosis: latency and reactivation. *Infect Immun* **2001**; 69:4195–201.
18. Lillebaek T, Dirksen A, Baess I, Strunge B, Thomsen V, Andersen AB. Molecular evidence of endogenous reactivation of *Mycobacterium tuberculosis* after 33 years of latent infection. *J Infect Dis* **2002**; 185:401–4.
19. Mitnick C, Bayona J, Palacios E, Shin S, Furin J. Community-based therapy for multidrug-resistant tuberculosis in Lima, Peru. *N Engl J Med* **2003**; 348:119–48.
20. Longini I, Koopman JS. Household and community transmission parameters from final distributions of infections in households. *Biometrics* **1982**; 38:115–26.

21. Longini I, Koopman J, Monto A, Fox J. Estimating household and community transmission parameters for influenza. *Am J Epidemiol* **1982**; 115:736–51.
22. Ball F, Mollison D, Scalia-Tomba G. Epidemics with two levels of mixing. *Ann Appl Probab* **1997**; 7:46–89.
23. Cauchemez S, Carrat F, Viboud C, Valleron AJ, Boëlle PY. A Bayesian MCMC approach to study transmission of influenza: application to household longitudinal data. *Statist Med* **2004**; 23:3469–87.
24. Blake IM, Burton MJ, Bailey RL, et al. of Ocular *Chlamydia trachomatis*. *PLoS Negl Trop Dis* **2009**; 3:e401.
25. Melegaro A, Gay NJ, Medley GF. Estimating the transmission parameters of pneumococcal carriage in households. *Epidemiol Infect* **2004**; 132:433–41.
26. World Health Organization/International Union Against Tuberculosis and Lung Disease. Anti-tuberculosis drug resistance in the world: fourth global report. Geneva: WHO Press, **2008**.
27. Becerra MC, Appleton SC, Franke MF, et al. Tuberculosis burden in households of patients with multidrug-resistant and extensively drug-resistant tuberculosis: a retrospective cohort study. *Lancet* **2011**; 377:147–52.
28. Dietz K. Epidemics: the fitting of the first dynamic models to data. *J Contemp Mathemat Anal* **2009**; 44:97–104.
29. 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.
30. Sutherland I, Bleiker MA, Meijer J, Stýblo K. The risk of tuberculosis infection in the Netherlands from 1967 to 1979. *Tubercle* **1983**; 64:241–53.
31. Canetti G, Sutherland I, Svandova E. Endogenous reactivation and exogenous reinfection: their relative importance with regard to the development of non-primary tuberculosis. *Bull Int Union Tuberc* **1972**; 47:116–34.
32. Vynnycky E, Fine PE. The natural history of tuberculosis: the implications of age-dependent risks of disease and the role of reinfection. *Epidemiol Infect* **1997**; 119:183–201.
33. Hart PD. Efficacy and applicability of mass B.C.G. vaccination in tuberculosis control. *Br Med J* **1967**; 1:587–92.
34. 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.
35. Blower SM, McLean AR, Porco TC, et al. The intrinsic transmission dynamics of tuberculosis epidemics. *Nat Med* **1995**; 1:815–21.
36. Asbroek AT, Borgdorff MW, Nagelkerke NJD, et al. Estimation of serial interval and incubation period of tuberculosis using DNA fingerprinting. *Int J Tuberculosis Lung Dis* **1999**; 3:414–20.
37. Vynnycky E, Fine PEM. Lifetime risks, incubation period, and serial interval of tuberculosis. *Am J Epidemiol* **2000**; 152:247–63.
38. Fine PEM. The interval between successive cases of an infectious disease. *Am J Epidemiol* **2003**; 158:1039–47.
39. van der Spuy G, Kremer K, Ndabambi S, et al. Changing *Mycobacterium tuberculosis* population highlights clade-specific pathogenic characteristics. *Tuberculosis (Edinb)* **2009**; 89:120–5.
40. Schaaf HS, Michaelis IA, Richardson M, et al. Adult-to-child transmission of tuberculosis: household or community contact? *Int J Tuberc Lung Dis* **2003**; 7:426–31.