We analyzed data from a large population-based prospective cohort study of household contacts of tuberculosis patients in Lima, Peru, to estimate the importance of within-household transmission relative to community-based transmission. We identified all adults (older than 15 years of age) who had incident pulmonary tuberculosis diagnosed at any of 106 public health centers in Lima from September 2009 to August 2012. A total of 14,041 household contacts of 3,446 index patients were assessed for tuberculosis infection and disease. We compared the prevalence of latent tuberculosis infection (LTBI) among persons who had received the Bacillus Calmette–Guérin vaccine in households with and without a microbiologically confirmed index case to estimate the age-specific risk of infection and excess risk of LTBI from household and community exposures. We found that the risk of infection from household and community sources increased from birth until 20 years of age. However, a large proportion of infections among child and young-adult household contacts could have been the result of household exposure. Excess infection risk associated with household exposure accounted for 58% (95% confidence interval: 47, 66) of LTBI prevalence among exposed children younger than 1 year of age, 48% (95% confidence interval: 39, 57) among 10-year-old children, and 44% (95% confidence interval: 34, 51) among 15-year-old adolescents. These findings suggest that expanded access to preventive therapy for older children and young-adult household contacts of known tuberculosis cases may be beneficial.

household transmission; preventive therapy; tuberculosis

Abbreviations: ARI, annual risk of infection; BCG, Bacillus Calmette–Guérin; CI, confidence interval; CPI, smear-negative, culture-positive index case; LTBI, latent tuberculosis infection; NI, smear-negative, culture-negative index case; ROI, risk of infection; SCPI, smear-positive, culture-negative index case; TST, tuberculin skin test.

In 2011, there were an estimated 8.7 million new cases of tuberculosis and 1.4 million deaths attributable to the disease worldwide (1), making it a leading infectious cause of morbidity and mortality. Despite recent gains, clear gaps between international targets and current tuberculosis trends have stimulated discussion about revising the central features of the global tuberculosis policy (1). For example, although the World Health Organization Stop TB strategy has emphasized passive case detection (i.e., waiting for persons with tuberculosis symptoms to self-present to a health provider), recent recommendations have focused on active approaches to community-based identification of tuberculosis cases (2). Although evidence in support of community-wide active case finding is limited (3), more intensive efforts to increase targeted screening among high-risk persons may help identify cases earlier in their course of disease, thereby improving treatment outcomes and reducing transmission (4).

One clear opportunity for targeted screening is in households of patients recently diagnosed with tuberculosis. Screening close contacts of tuberculosis patients is a global standard (5), and standardized approaches for these interventions have been outlined previously (2). Household-based contact tracing can identify undetected active tuberculosis cases (which are invariably more common within these households than in the rest of the community (6, 7)) and allow recently infected contacts at high risk of developing tuberculosis to receive...
preventive therapy. Although such interventions are deployed regularly in resource-rich, low-tuberculosis settings, they have yet to be rolled out broadly in resource-constrained, higher-incidence settings (1, 6). A better understanding of the relative contributions of household and community exposure may help guide routine contact investigations in high-burden settings.

A specific benefit of household-based contact tracing is in reducing the burden of pediatric tuberculosis (3, 8). Systematic reviews of household contact investigations have found a high risk of active disease among children in homes of known tuberculosis cases (4, 6, 9). Because of the relative difficulty of diagnosing pediatric tuberculosis, the rate of detection among children less than 15 years of age has consistently lagged behind that in adults (2, 10). Furthermore, because latently infected children in these households may have been infected very recently, preventive therapy may be especially useful in preventing pediatric tuberculosis among children with household exposure (11).

Understanding the potential benefit of household-based interventions to overall tuberculosis control efforts requires estimating the relative importance of household versus community transmission of tuberculosis. Infection with Mycobacterium tuberculosis, the causative agent of tuberculosis, may occur after a susceptible person inhales M. tuberculosis–laden droplet nuclei expired by an infectious person. Though the infectious dose is typically small (<10 mycobacteria) (6, 12), observational studies have indicated that even in high-risk environments, such as inpatient tuberculosis wards, prolonged exposure is required for transmission to occur (13). Household studies indicated that only approximately half of household contacts of persons diagnosed with the most transmissible forms of pulmonary tuberculosis show evidence of latent tuberculosis infection (LTBI), despite the fact that household exposure likely occurred for months before diagnosis (14–17). However, in tuberculosis-endemic settings, it is clear that household contacts of persons with infectious tuberculosis are also at substantial risk of infection in the community. Previous studies in which investigators used M. tuberculosis genotype data to identify transmission chains suggested that in some high-incidence settings, the majority of cases of secondary disease among household contacts may be due to exposures outside the home (18).

In the present study, we show a novel approach to estimating age-specific risks of infection (ROI) from household and community sources in Lima, Peru. Between September 2009 and August 2012, we identified all adults (>15 years of age) with incident pulmonary tuberculosis who were diagnosed at any of 106 participating public health centers located in our study catchment area, which contained approximately 3.3 million inhabitants. Within 1 month of diagnosis of tuberculosis in these “index patients,” a study nurse visited the patient’s home and invited exposed household contacts to participate in a baseline assessment of tuberculosis infection and disease. Of the 14,041 household contacts of 3,446 index cases, 2,851 contacts and 306 index cases were excluded from this analysis because of missing or incomplete records. Of the 2,851 contacts with missing data, 1,986 were excluded because they did not receive a tuberculosis skin test (TST) at baseline, 809 were excluded because we were missing data on whether they had received a Bacillus Calmette–Guérin (BCG) vaccination, and 56 were excluded because their household index case was excluded. Although BCG vaccination often leaves a characteristic scar, vaccinated persons may not scar (20). To ensure an accurate representation of the impact of vaccination, we use the person’s self-reported vaccination status rather than the presence of a BCG vaccine scar to estimate protective associations of BCG with disease. Of the 306 missing index cases, 79 were excluded because of missing sputum smear or culture results, whereas the remaining 227 were excluded either because they lived alone or because all of their household contacts were missing TST results. For results of a full analysis of the potential impact of these missing data on our results, see Web Appendix 1, Web Tables 1 and 2, and Web Figures 1–3, available at http://aje.oxfordjournals.org/.

In the present analysis, we took advantage of the fact that 22% of enrolled index cases who were diagnosed with tuberculosis based on clinical presentation and initiated treatment were eventually found to have negative sputum smear and culture results; that is, there was no microbiological evidence of M. tuberculosis after a thorough work-up. Although microbiologically unconfirmed tuberculosis cases are included in official notification statistics (21), these persons are substantially less infectious than smear-positive tuberculosis cases (15).

We isolated the age-specific associations of household exposure from cumulative community exposure by contrasting the infection status of household contacts of smear-positive/culture-positive index cases (SCPIs) and smear-negative/culture-positive index cases (CPIs) against those of persons of the same age who were exposed to a smear-negative/culture-negative index case (NIs). Consistent with previous observations (14–16), we hypothesized that SCPIs are the most infectious, followed by CPIs, and then NIs. Figure 1 illustrates the hypothesized levels of infectiousness from these types of contacts and Figure 2 compares the age-specific prevalence of LTBI among contacts of SCPIs and NIs. Because they were identified via the same passive surveillance process as were SCPIs, NIs are likely to have community exposure profiles more similar to SCPIs than community controls selected via another mechanism, such as random sampling and propensity matching. However, to ensure that our results were not confounded by systematic differences between SCPIs, CPIs, and NIs, all analyses were adjusted for the following factors: 1) household overcrowding, defined as having 3 or more persons per bedroom; 2) index case age, which is likely to provide

METHODS

Data

Our analysis is based on data that we collected during a large population-based prospective cohort study of households in Lima, Peru. Between September 2009 and August 2012, we identified all adults (>15 years of age) with incident pulmonary tuberculosis who were diagnosed at any of 106 participating public health centers located in our study catchment area, which contained approximately 3.3 million inhabitants. Within 1 month of diagnosis of tuberculosis in these “index patients,” a study nurse visited the patient’s home and invited exposed household contacts to participate in a baseline assessment of tuberculosis infection and disease. Of the 14,041 household contacts of 3,446 index cases, 2,851 contacts and 306 index cases were excluded from this analysis because of missing or incomplete records. Of the 2,851 contacts with missing data, 1,986 were excluded because they did not receive a tuberculosis skin test (TST) at baseline, 809 were excluded because we were missing data on whether they had received a Bacillus Calmette–Guérin (BCG) vaccination, and 56 were excluded because their household index case was excluded. Although BCG vaccination often leaves a characteristic scar, vaccinated persons may not scar (20). To ensure an accurate representation of the impact of vaccination, we use the person’s self-reported vaccination status rather than the presence of a BCG vaccine scar to estimate protective associations of BCG with disease. Of the 306 missing index cases, 79 were excluded because of missing sputum smear or culture results, whereas the remaining 227 were excluded either because they lived alone or because all of their household contacts were missing TST results. For results of a full analysis of the potential impact of these missing data on our results, see Web Appendix 1, Web Tables 1 and 2, and Web Figures 1–3, available at http://aje.oxfordjournals.org/.

In the present analysis, we took advantage of the fact that 22% of enrolled index cases who were diagnosed with tuberculosis based on clinical presentation and initiated treatment were eventually found to have negative sputum smear and culture results; that is, there was no microbiological evidence of M. tuberculosis after a thorough work-up. Although microbiologically unconfirmed tuberculosis cases are included in official notification statistics (21), these persons are substantially less infectious than smear-positive tuberculosis cases (15).

We isolated the age-specific associations of household exposure from cumulative community exposure by contrasting the infection status of household contacts of smear-positive/culture-positive index cases (SCPIs) and smear-negative/culture-positive index cases (CPIs) against those of persons of the same age who were exposed to a smear-negative/culture-negative index case (NIs). Consistent with previous observations (14–16), we hypothesized that SCPIs are the most infectious, followed by CPIs, and then NIs. Figure 1 illustrates the hypothesized levels of infectiousness from these types of contacts and Figure 2 compares the age-specific prevalence of LTBI among contacts of SCPIs and NIs. Because they were identified via the same passive surveillance process as were SCPIs, NIs are likely to have community exposure profiles more similar to SCPIs than community controls selected via another mechanism, such as random sampling and propensity matching. However, to ensure that our results were not confounded by systematic differences between SCPIs, CPIs, and NIs, all analyses were adjusted for the following factors: 1) household overcrowding, defined as having 3 or more persons per bedroom; 2) index case age, which is likely to provide

information on the cumulative exposure of the household; and 3) whether the household has an improved (i.e., metal or concrete) or unimproved (mud or thatch) roof. This reflects both the socioeconomic status and indoor environmental quality of the household. In addition, because transmission risk may scale with household size (22), we adjusted for whether the person lived in a large household. This was defined as 6 or more persons, which corresponds to the 75th percentile of the household size distribution in our sample. For a full analysis of the differences between exposure groups, see Web Appendix 2, Web Table 3, and Web Figures 4–6.

Models

Our analysis proceeded in 3 stages. First, we used a finite-mixture model to differentiate between latently infected and uninfected household contacts using the size (in millimeters) of the TST reaction. Second, we fitted a generalized additive mixed logistic regression model to these data to estimate age-specific risk factors for infection. Finally, we used predictions from this model to estimate household and community risks of infection, that is, the probability that an uninfected person would progress to LTBI after a single episode of household exposure or a year of community exposure.

Identifying latently infected persons

LTBI is usually diagnosed after the TST by measuring the diameter of the raised induration at the injection site 48–72 hours after subcutaneous injection of tuberculin. TST results are typically reported as binary outcomes in which indurations with a diameter of 10 mm or greater are considered positive, although a diameter of up to 15 mm may be used in some high-burden contexts (23). This cutoff is meant to minimize error in prediction of LTBI, but its diagnostic utility is not consistent across populations. We used a finite-mixture model fitted to the continuous TST responses of persons exposed to an NI to assign contacts to the latently infected and uninfected categories.
uninfected classes. To assess the sensitivity of our results to the choice of mixture model, we replicated our analysis using the 10-mm cutoff. A full description of the mixture model and results for the cutoff model are presented in Web Appendix 3, Web Appendix 4, and Web Figures 7–9.

Models to identify risk factors for latent infection

We use a generalized additive mixed model to estimate age-specific risk factors for LTBI. Generalized additive models are semiparametric regression models, similar to generalized linear models, except that they model an outcome as a smooth function of predictors rather than assuming a linear relationship (24). All models were fitted using the gamm4 package in R, version 3.02 (R Foundation for Statistical Computing, Vienna, Austria). The outcome for each person was the value of the LTBI status, \( y_i \), assigned by the mixture model from the previous section. We referred to the prevalence of TST-positivity for persons of age \( a \) without BCG vaccination who were exposed to an NI as the age-specific intercept for age \( a \). We also estimated age-specific odds ratios for CPI and SCPI exposure and BCG vaccination. Age-specific odds ratios were considered statistically significant if the 95% confidence interval for the association did not overlap the 95% confidence interval of the odds ratio of the corresponding age-specific intercept.

We included random intercepts because our sample was nested at 2 levels. First, persons were nested within 3,120 households, and index cases typically had more than 1 household contact. Each household fell within 1 of 106 different health center catchment areas in Lima, which corresponded to geographic areas of the city with varying average socioeconomic statuses and other tuberculosis risk factors. We accounted for correlated responses that arose from repeated measures of the association of exposure to a specific household index case, as well as unobserved household- and health center-level risks by allowing each household and health center to have a random intercept. We derived confidence intervals for model coefficients and quantities derived from these coefficients using the method described by Marra and Wood (25) and assessed goodness of fit for the ROI model using a posterior predictive check (Web Appendix 5, Web Figures 10 and 11).

Household and community risk of infection

We first calculated the ROIs posed by SCPIs and CPIs to their household contacts. We then estimated the age-specific community-associated ARIs among persons exposed to an NI and contrasted this with household ROIs for SCPIs to estimate the relative risk of infection associated with these 2 types of exposure.

We defined the age-specific household ROI as the probability that an uninfected person of age \( a \) would develop LTBI as a result of contact with a single household case. We denoted the ROI for exposure to an SCPI as \( \lambda^{SC} \) and the ROI for exposure to a CPI as \( \lambda^{C} \). Similar to Middelkoop et al. (26), we defined the community ROI (\( \psi \)) as the probability that an uninfected person exposed to an NI would become infected over the course of a year, reflecting the assumption that the rate of transmission from NIs within the household is negligible. We characterized household and community ROI using the marginal distributions of the fitted model parameters, setting random intercepts to 0 so that our estimates reflected expected ROIs across all spatial units. For simplicity, we present models in which all persons have received the BCG vaccination, live in an uncrowded household with an improved roof and fewer than 6 persons, and were exposed to an index case who was 30 years old, which was the mean index age in our sample. Because the prevalence of LTBI decreases beyond age 50 years (see Figure 2), we estimated age-specific ROI only for those persons younger than 50 years of age.

Excess risk

Excess risk was interpreted as the change in the proportion of the population with LTBI between the condition in which everyone was exposed to an SCPI/CPI and the condition in which no one had such an exposure, expressed as a fraction of the prevalence in the fully exposed condition (27). This provided a measure of the contribution of household transmission to the overall burden of infection in the community for each age group (27). For additional details on the calculation of ROI and excess risk, see Web Appendix 6.

RESULTS

TST mixture model

For uninfected persons with a positive TST result, the mean induration size \( \mu_E \) was 5.56 mm and the standard deviation was 1.99 mm. For the Gaussian mixture component for exposed noninfected persons (mean induration size = 2.55; standard deviation, 0.26; right). The dashed line illustrates the sum of the fitted densities from both mixture components.
deviation $\sigma_E$ was 1.99 mm. For persons with LTBI, the mean and standard deviation of the underlying Gaussian distribution for the lognormal component were 2.5 and 0.26, respectively, with a mean of 12.6 mm and standard deviation of 3.3 mm for the resulting lognormal distribution. Figure 3 illustrates the fit of the mixture model to the TST responses of the 1,356 persons with an induration greater than 0 mm who were exposed to an NI.
Risk factor of infection model

Because we modelled risk factors using smooth functions, we present parameter estimates graphically in Figure 4. Coefficients are expressed as log odds ratios relative to the age-specific intercept (Figure 4A). Solid lines in each panel illustrate point estimates by age, and the dashed lines are 95% confidence intervals at each point. If the confidence intervals for an age fall on both sides of the dotted line at zero, the association is considered nonsignificant at this age. An alternate model including the age-specific association of sex in addition to the covariates discussed here was estimated. This association was nonsignificant at all ages and therefore is omitted from the final model. We note that the odds ratio associated with BCG vaccination (Figure 4B), which shows a diminished

Figure 5. Household and community risks of infection, Lima, Peru, 2009–2012. A–B) Age-specific risk of infection for a household contact exposed to a culture-positive index case (A) or smear-positive/culture-positive index case (B). C) Change in annual risk of infection from community contact. D) Risk ratio comparing age-specific risk of infection from a single episode of household exposure to a smear-positive/culture-positive index case to community annual risk of infection for an uninfected person of the same age.
risk of LTBI associated with BCG vaccination among small children, likely reflects the high-risk population from which our sample was drawn rather than population-level associations of BCG with TST reactivity. For the potential impact of social and environmental factors, see Web Appendix 2.

Risk of infection

Household ROI. Figure 5A illustrates the age-specific household ROI from a CPI, which is 0.04 (95% confidence interval (CI): 0.01, 0.09) for children younger than 1 year of age and increases to 0.11 before becoming nonsignificant at age 37 years. Figure 5B illustrates age-specific household ROI from a SCPI. For children younger than 1 year of age, the ROI for contact with a SCPI is 0.11 (95% CI: 0.07, 0.17). This association grows to 0.35 at age 35 years (95% CI: 0.27, 0.41) before declining and becoming nonsignificant.

Community ARI. Figure 5C illustrates the age-specific community ARI, which increases from 0.009 (95% CI: 0.006, 0.013) at 1 year of age to 0.023 (95% CI: 0.016, 0.032) at 27 years of age, after which it declines and then becomes nonsignificant at 48 years of age. Figure 5D illustrates SCPI household ROI relative to community ARI at each age. For 1-year-old children, the risk posed by an episode of SCPI exposure was roughly 12 times greater than the risk from a year of exposure in the community (95% CI: 7.9, 15.6); this association increased with age, before becoming nonsignificant at 47 years of age.

Excess risk of infection due to household exposure. Figure 6 illustrates the excess risk associated with exposure to a CPI (Figure 6A) or SCPI (Figure 6B). For children younger than 1 year of age, the excess risk associated with CPI exposure was 34% (95% CI: 13, 51) of the total LTBI prevalence among exposed children; it was 27% (95% CI: 10, 40) among 10-year-old children and 22% (95% CI: 8, 35) among 15-year-old adolescents. It became nonsignificant at 37 years of age. For SCPI exposure, the excess risk was 58% (95% CI: 47, 66) of LTBI prevalence among exposed children younger than 1 year of age, 48% (95% CI: 39, 57) among 10-year-old children, and 44% (95% CI: 34, 51) among 15-year-old adolescents, and it remained significant among older age groups. These results show that although both community and household ROIs increase with age, a larger fraction of the total number of infections among younger household contacts is explained by household exposure, because younger persons are less likely to have been previously infected at home or in the community.

DISCUSSION

Our results suggest that the age-specific ROIs from contact with an SCPI and from community exposure increase from birth until approximately 30 years of age. Although prior studies have compared tuberculosis prevalence among persons with and without household exposure (19) and quantified the age-specific annual risk of tuberculosis infection in the community (26), none has addressed the age-specific household ROI. Our results show that comparing household and community ROIs provides a better understanding of age-specific risks in both contexts. For example, the similar
Among persons older than 50 years of age should not be interpreted to mean that these persons are protected against reinfection. Instead, this likely reflects the inability of TST to document reinfection events, which are expected to be more common among older persons, as well as decreasing TST reactivity with age (30, 31). However, these associations are likely to be less pronounced among the younger persons, who are the focus of our policy and practice recommendations. Nonetheless, future studies combining TST data with mechanistic models that account for reinfection, such as the one by Cohen et al. (32), may be useful for dealing with this limitation.

Because the household and community risk of infection have a more mechanistic interpretation than do other measures of relative household and community infection risk (19), our approach provides a blueprint for dynamic epidemiologic models to build on these findings using molecular data that can definitively identify transmission links (33). Such models will be useful for extending our findings to understand the role of household transmission in overall community risk and determining the optimal use of household screening in high-incidence, resource-constrained settings.

ACKNOWLEDGMENTS

Author affiliations: Department of Ecology and Evolutionary Biology, Princeton University, Princeton, New Jersey (Jonathan L. Zelner, Bryan T. Grenfell); Fogarty International Center, National Institutes of Health, Bethesda, Maryland (Jonathan L. Zelner, Bryan T. Grenfell); Department of Global Health and Social Medicine, Harvard Medical School, Boston, Massachusetts (Megan B. Murray, Mercedes C. Becerra); Department of Epidemiology, Harvard School of Public Health, Boston, Massachusetts (Megan B. Murray, Ted Cohen); Partners in Health, Boston, Massachusetts (Jerome Galea, Leonid Lecca, Roger Calderon, Rosa Yataco, Carmen Contreras); Socios En Salud, Lima, Peru (Jerome Galea, Leonid Lecca, Roger Calderon, Rosa Yataco, Carmen Contreras); and Division of Global Health Equity, Brigham and Women’s Hospital, Boston, Massachusetts (Zibiao Zhang, Ted Cohen).

J.L.Z. and B.T.G. were supported by the Research and Policy for Infectious Disease Dynamics program of the Science and Technology Directorate, Department of Homeland Security and the Fogarty International Center, National Institutes of Health. B.T.G. was funded by the Bill and Melinda Gates Foundation. M.B.M., M.C.B., J.G., L.L., R.C., R.Y., C.C., Z.Z., and T.C. were funded by the National Institutes of Health (grant U19 A1076217).

Conflict of interest: none declared.

REFERENCES

