Web Appendix 1. Sensitivity to Missing Data

Of the 14,041 individuals in our original sample, 2,851 were excluded from the present analysis because they did not receive a tuberculin skin test (TST) at the time of the intake observation, did not respond to the question asking whether they had received a BCG vaccination, or because their household index case was excluded for missing smear or culture status information. To ensure that our results are not impacted by the exclusion of these cases, we assessed the similarity between these excluded individuals and those with complete records and imputed skin test results for those with a missing baseline TST. We then used information on observable BCG scarring, for which only 4 individuals had missing responses, to impute values of missing BCG questionnaire responses. Using this reconstructed dataset, we then performed a sensitivity analysis to determine how our results would change if these excluded cases were systematically different than those with complete records.

Differences between cases with and without missing data

Web Table 1 shows the results of a logistic regression analysis of age, socioeconomic status (SES), and household exposure factors predicting whether an individual did not receive a TST at enrollment. Web Table 2 shows the results of a similar analysis predicting the probability that an individual would have a missing BCG questionnaire response. This model shows that children and young adults were less likely to have a missing baseline TST than adults. In addition, individuals living in households with an unimproved roof were more likely to be missing from the analysis. Individuals in large households (> 6 individuals) were less likely to have a missing TST than those in smaller

ones, as were those exposed to an SCPI, although this result was slightly outside of our cutoff for statistical significance ($p = 0.06$).

Assigning missing BCG responses

While 11,482 of the 12,716 (> 99%) individuals in our sample who self-reported BCG vaccination also had a BCG scar, 75% of individuals without a BCG scar (1,234/1,637) still reported having been vaccinated. To assign BCG vaccination statuses to those missing this information, we fit a logistic regression model predicting the probability of a positive response for BCG vaccination as a function of the presence of a BCG scar and individual age. This model reflects the strong association between a BCG scar and self-reported vacation (OR=33.1, 95%CI=26.6, 41.08), and a slowly declining prevalence of vaccination with age (OR=0.98/yr, 95%CI=0.97, 0.98). We tested an additional model with an interaction between BCG scar and age, but found that this was non-significant. To assign BCG vaccination statuses in the synthetic dataset to the 810 individuals with missing values, we drew a Bernoulli random variate with success probability equal to that predicted by the logistic regression model given their BCG scar status and age.

Sensitivity Analysis

Because our results for household force of infection and excess risk rely on the difference in prevalence between SCPI-exposed and NI-exposed individuals under the age of 20, we constructed a sensitivity analysis that would allow us to assess the impact of the worst-case scenario in which NI-exposed, excluded individuals were disproportionately more likely to be TST-positive than NI-exposed individuals included in the analysis. This would reduce the gap in prevalence between SCPI and NI-exposed individuals of the

same age, which could impact our estimates of the excess risk and force of infection associated with SCPI exposure.

To do this, we used the fitted values of the generalized additive mixed logistic regression model presented above to generate a synthetic dataset for the 2,851 individuals with missing or incomplete records. Individuals with SCPI or CPI exposure were assigned a TST status by drawing a Bernouilli random variate with success probability equal to the probability of TST positivity predicted by the GAMM model. This ensures that these individuals have prevalence similar to other individuals of the same age and exposure. To ensure that the excluded NI-exposed individuals in our synthetic dataset had significantly elevated risk as compared to otherwise similar individuals, we multiplied the predicted probabilities for these individuals by two and drew a Bernouilli random variate based on these adjusted probabilities. We then re-fit the GAMM model to a dataset comprised of both the set of individuals with complete records and those in our synthetic dataset. Results from this analysis are shown in Web Figures 1 & 2. We then repeated this analysis for the extreme case in which the excluded NI-exposed cases were assumed to be four times more likely to have been TST positive than the original model predictions. Excess risk results from this model are presented in Web Figure 3.

Web Figure 1 shows the fitted smooth terms of the GAMM model, which still show an increased age-specific odds ratio of infection for younger SCPI-exposed individuals. Web Figure 2 shows the predicted excess risk attributable to household exposure from this model, which is slightly diminished as compared to the results for complete cases only (Figure 5, main text), but qualitatively the same. Web Figure 3 shows that, while the effect of household exposure is significantly diminished by the extreme assumption that

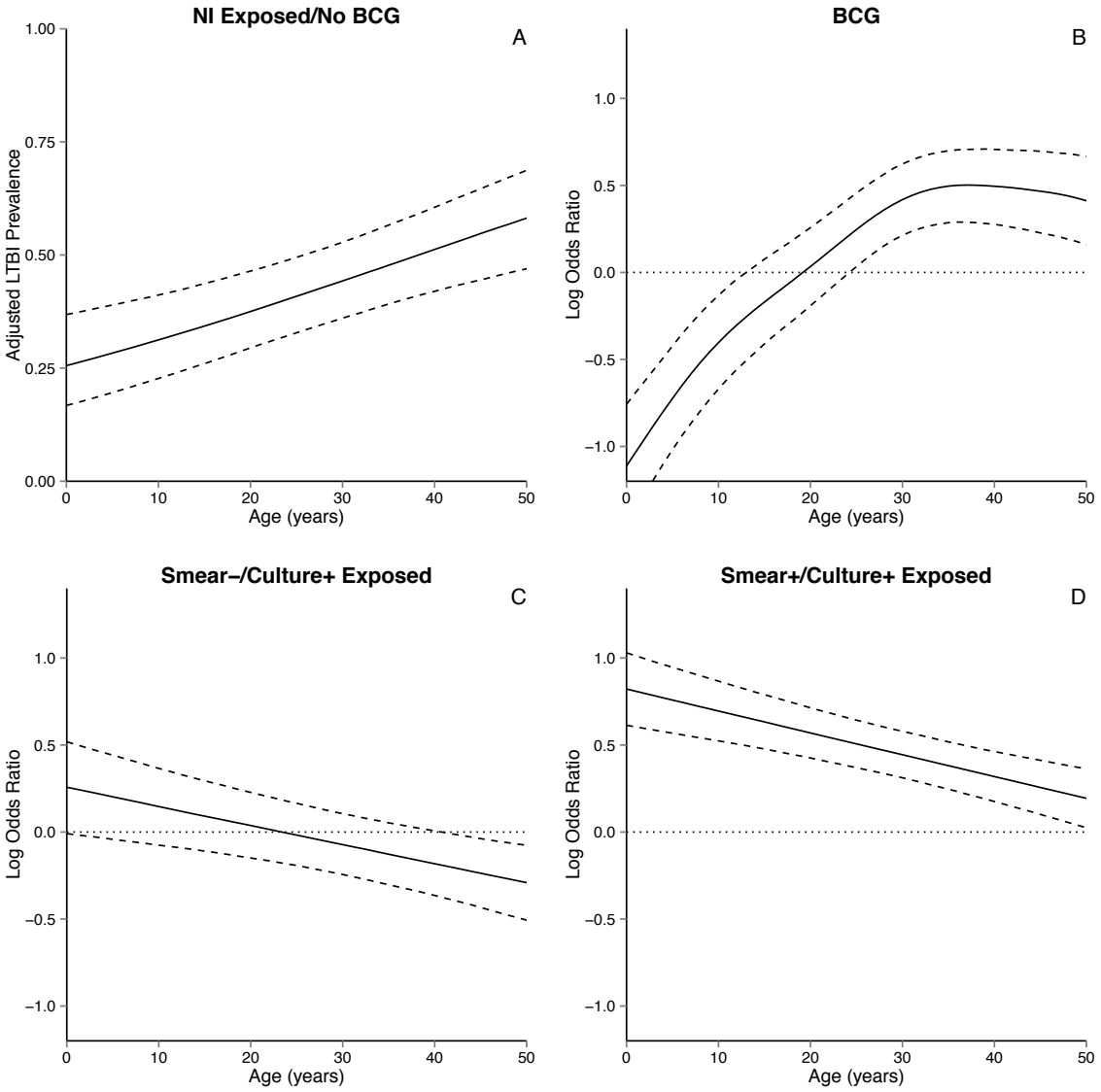
the excluded NI-exposed individuals are four times more likely to have been TST positive than otherwise-equivalent included individuals, SCPI-exposed individuals still experience elevated and detectable excess risk of infection from household SCPI-exposure exposure. Taken together, these results suggest that the exclusion of these cases was unlikely to impact the overall conclusions of our analysis.

	Variable	OR	95% CI
	Intercept	0.10	(0.07, 0.15)
Vaccination	BCG	0.97	(0.85, 1.12)
Age Groups	0 to 4 Years	**2.51**	**(2.11, 2.98)**
	5 to 9 Years	**1.70**	**(1.41, 2.06)**
	10 to 19 Years	**1.37**	**(1.16, 1.62)**
	20 to 29 Years	REF	-
	30 to 39 Years	0.94	(0.77, 1.15)
	40 to 49 Years	0.85	(0.68, 1.06)
	50+ Years	1.02	(0.84, 1.23)
SES	Overcrowded	1.10	(0.99, 1.23)
	Poor Roof	**1.42**	**(1.07, 1.88)**
	Large Household	**0.77**	**(0.69, 0.86)**
Exposure	CPI	1.08	(0.93, 1.25)
	SCPI	0.90	(0.80, 1.02)

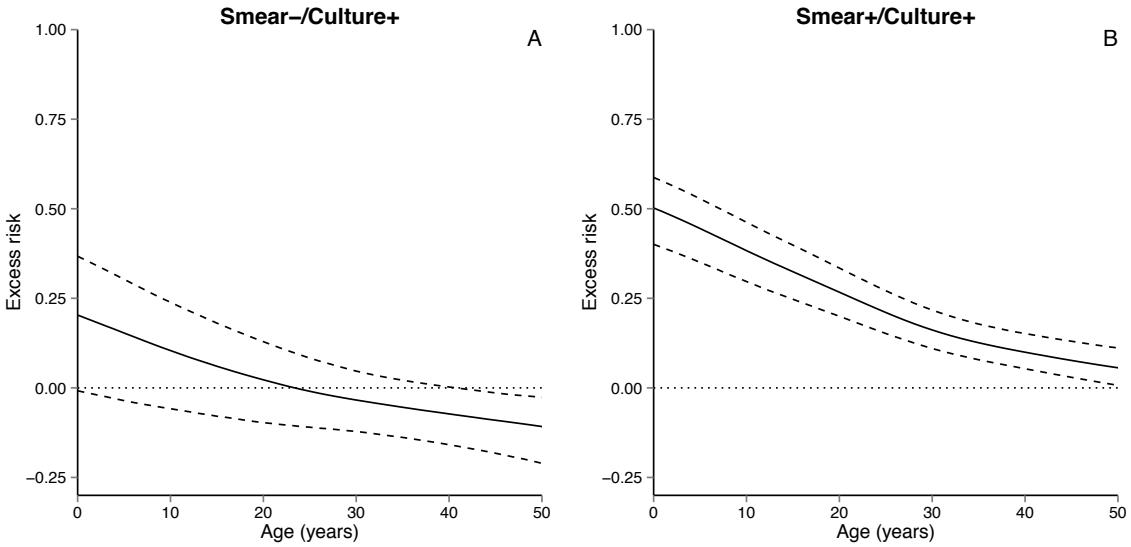
Web Table 1. Predictors of missing baseline tuberculin skin test result. Statistically significant ($p < 0.05$) predictors are highlighted in bold.

	Variable	OR	95% CI
	Intercept	0.41	(0.26, 0.64)
Vaccination	BCG Scar	**0.18**	**(0.15, 0.21)**
Age Groups	0 to 4 Years	**0.09**	**(0.05, 0.16)**
	5 to 9 Years	**0.10**	**(0.06, 0.19)**
	10 to 19 Years	**0.61**	**(0.47, 0.78)**
	20 to 29 Years	REF	-
	30 to 39 Years	1.19	(0.92, 1.53)
	40 to 49 Years	1.09	(0.82, 1.45)
	50+ Years	1.22	(0.96, 1.55)
SES	Overcrowded	0.92	(0.77, 1.10)
	Poor Roof	**0.59**	**(0.42, 0.82)**
	Large Household	1.13	(0.95, 1.34)
Exposure	CPI	**1.38**	**(1.08, 1.76)**
	SCPI	1.18	(0.96, 1.44)
TB	LTBI	1.05	(0.89, 1.24)

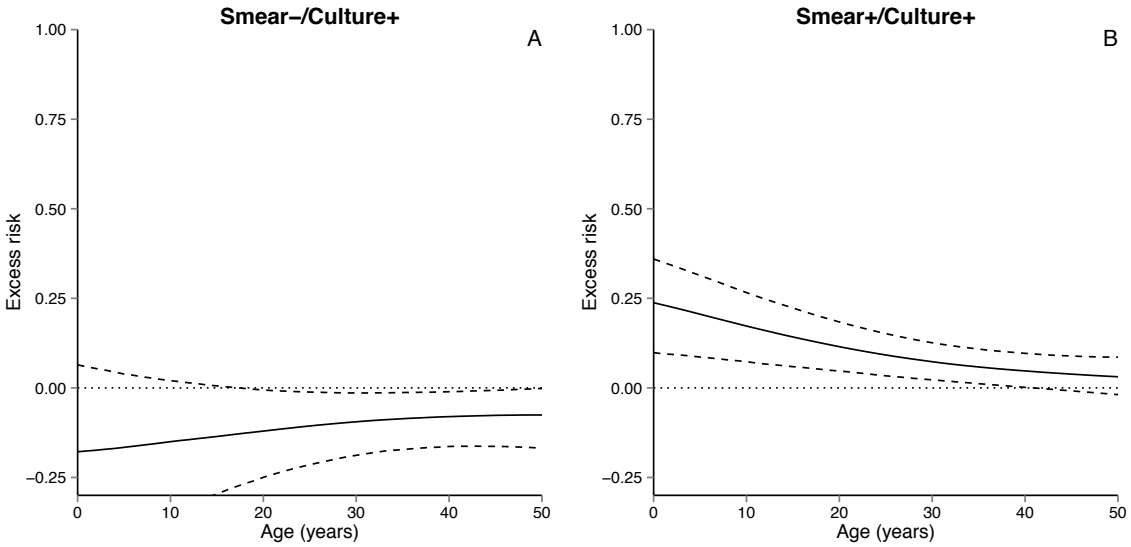
Web Table 2. Predictors of missing BCG vaccination status. Statistically significant ($p < 0.05$) predictors are highlighted in bold.



Web Figure 1. Parameter estimates from sensitivity analysis where NI-exposed individuals without TST results are twice as likely to be TST+ than otherwise equivalent individuals. Solid lines in each panel are point estimates; dashed lines are 95% credible intervals (CIs). CIs are interpreted over the entire age range, i.e. the upper CI is the upper limit in the trend by age. Panel *A* illustrates the age-specific intercept, represented by the change in log-odds of LTBI with age among individuals exposed to an NI. Panel *B* shows age-specific change in the log-odds of TST-positivity associated with BCG vaccination. Panel *C* & *D* show the effects of CPI and SCPI exposure by age, relative to the age-specific intercept for NI exposure. Dotted lines in panels A-D are provided as a guide for assessing statistical significance. Significance for panels B-D is assessed in terms of the difference from the age-specific intercept.



Web Figure 2. Excess risk associated with household smear-positive/culture-positive (SCPI) and culture-positive (CPI) exposure from sensitivity analysis where cases with missing data are twice as likely to be TST+ than otherwise equivalent individuals. Excess risk is measured as the proportion reduction in prevalence between the condition in which all individuals have SCPI exposure and the condition in which no individuals are exposed. The solid line shows the point-estimate as a function of age and the dashed lines show 95% credible intervals. The dotted line is a guide for assessing statistical significance.



Web Figure 3. Excess risk associated with household smear-positive/culture-positive (SCPI) and culture-positive (CPI) exposure from sensitivity analysis where cases with missing data are four times as likely to be TST+ than otherwise equivalent individuals. Excess risk is measured as the proportion reduction in prevalence between the condition in which all individuals have SCPI exposure and the condition in which no individuals are exposed. The solid line shows the point-estimate as a function of age and the dashed lines show 95% credible intervals. The dotted line is a guide for assessing statistical significance.

Web Appendix 2. Socioeconomic status and cumulative exposure

To ensure that our results showing that younger SCPI-exposed individuals experience an elevated excess risk of infection as compared to older individuals with the same exposure are not confounded by the potential relationship between exposure to an SCPI and household socioeconomic status (SES), we conducted an analysis to determine how SCPIs and NIs in our sample differ. We then adjusted the analyses presented in the text for these factors to ensure that our results are not confounded by differences between SCPIs and NIs.

Because self-reported household income is unlikely to be a reliable indicator of overall socioeconomic status in our sample, we used the quality of the roof on the individual's home as a proxy for the overall SES of the household. This is more likely to reflect the overall material well-being of household members than household income, which may fluctuate seasonally with the availability of work. We divided households into those with tiled, shingled or metal roofs, which we denoted as *improved roof* households, and those with mud or thatch roofs, which we denoted as *unimproved roof* households. We also divided households into those with three or more persons per bedroom, and those with 2 or fewer per bedroom, which we denote as *overcrowded* and *uncrowded*, respectively.

We also account for the possibility that the average risk to household contacts may decrease with the size of the household, i.e. that household transmission may exhibit some type of frequency dependence or transmission elasticity (1) in which the per-capita transmission rate scales with household size. To do this, we include two indicator variables, one for large households (defined as households larger than 6 individuals; the 75^{th} percentile of household size distribution in our sample) and exposed to an SCPI, and

another for large CPI-exposed households. Web Table 3 shows the results of a logistic regression model predicting the probability that an individual household contact was exposed to an SCPI as opposed to an NI. This analysis shows a small increase in the probability of living in an overcrowded household for SCPIs relative to NIs (OR = 1.19, 95%CI = 1.10, 1.28), and in a large household (OR = 1.11, 95% CI = 1.03, 1.19). There was also a small difference in roof quality, with individuals in poor roof households less likely to be exposed to an SCPI (OR = 0.82, 95% CI = 0.69, 0.98).

In addition, we also assessed whether there were statistically significant age differences between SCPIs and NIs. The age of an index case may contain information about the cumulative force of infection experienced by that individual and her household contacts. The complicated nature of TB infection means that the relationship between age and risk of infection is unlikely to be as straightforward as in strongly immunizing infections, such as measles, where younger age at infection indicates a greater force of infection, but is still likely to contain information on cumulative exposure in the community. The plot in Web Figure 4 shows that there is not a statistically significant difference in the ages of SCPIs and NIs for the lower 75% of the distribution, but that in the upper quartile, the age distributions diverge slightly, with the oldest 25% of SCPIs slightly younger on average than the oldest 25% of NIs.

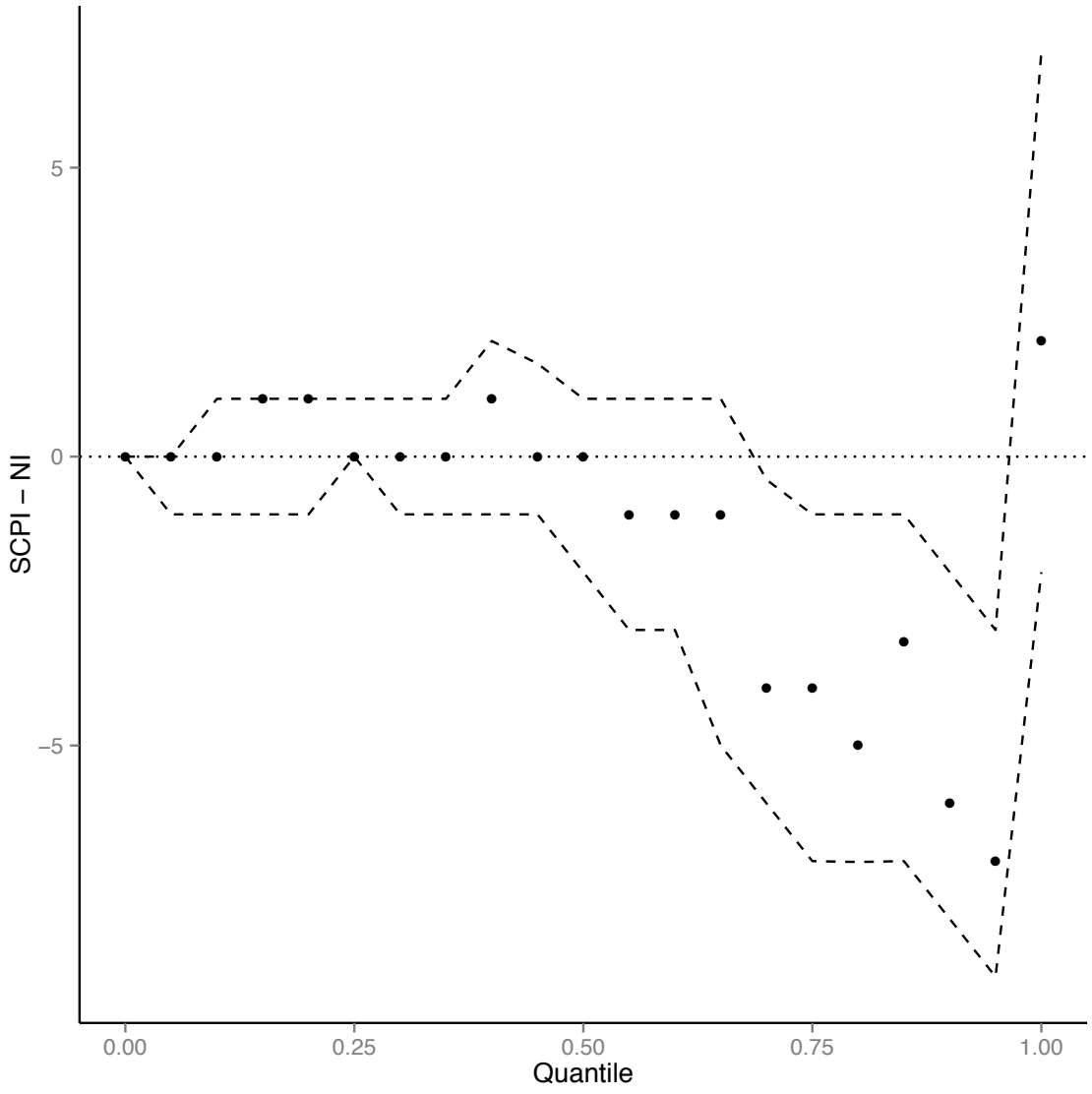
When estimating the generalized additive mixed models (GAMMs) presented in the text, we adjust for the age-specific effects of household overcrowding, roof quality, and index case age. We include age-invariant effects for household size to represent the idea that household size should scale contact rates roughly uniformly across age groups.

We divide index age into a categorical variable with levels for indexes aged 20-29 years, 30-39 years, 40-49 years and 50+ years. Because only households with index cases 14 years or older were enrolled in the study, we keep indexes aged 14-19 years as the reference category. We find that the effects of overcrowding and roof quality are non-significant across all age groups, as illustrated in Web Figure 5. Web Figure 6 illustrates the age-specific effects of exposure to index cases of different ages. This shows that the effects of exposure to an index case aged 30-49 are significant for children and young adults. But the effects of exposure to 20-29 year-old and 50+ year-old index cases are non-significant across all age-groups. These age-specific effects may reflect differential cumulative community exposure. But they are also likely to reflect differences in household contact rates with age. For example, the risk to young children associated with contact between indexes from 20-39 years old may reflect increased rates of contact between parents and young children.

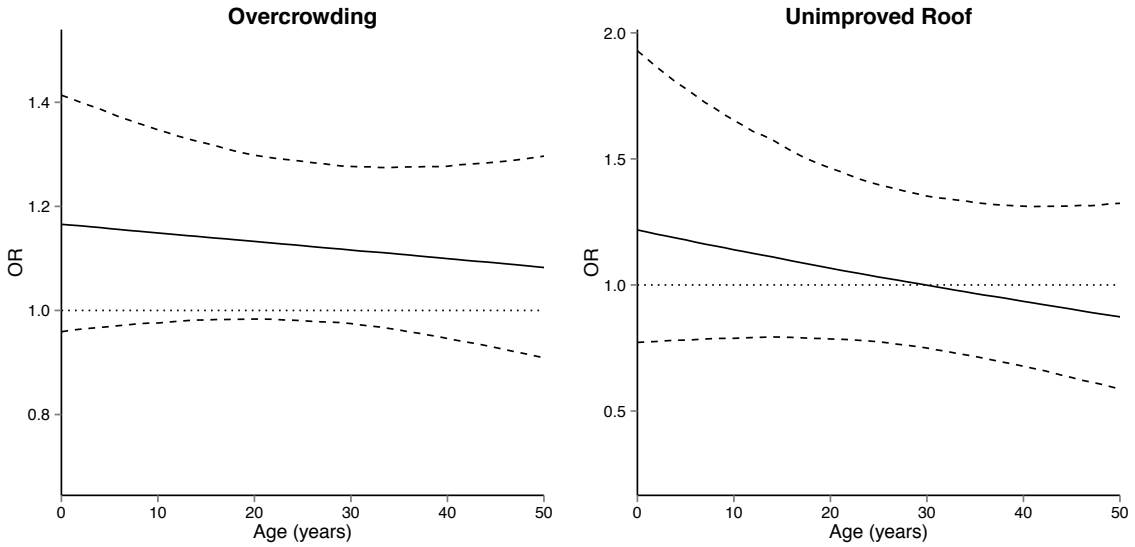
We also find that SCPI-exposed individuals living in a large household are slightly less likely to be TST-positive (OR = 0.72, 95%CI = 0.62, 0.85), while the association between household size and TST positivity is non-significant for CPI-exposed individuals (OR = 0.95, 95%CI = 0.72, 1.25). Sensitivity analysis including age-specific effects of household size resulted in a larger estimate of the effect of SCPI exposure with only marginal improvement to model fit, so we present the more conservative estimates here.

	Variable	OR	95% CI
	Intercept	1.47	(1.19, 1.82)
Vaccination	BCG	0.95	(0.86, 1.05)
Age Groups	0 to 4 Years	0.99	(0.87, 1.12)
	5 to 9 Years	1.00	(0.88, 1.14)
	10 to 19 Year	0.98	(0.88, 1.09)
	20 to 29 Years	REF	-
	30 to 39 Years	0.96	(0.85, 1.09)
	40 to 49 Years	**0.87**	**(0.76, 0.99)**
	50+ Years	0.97	(0.86, 1.09)
SES	Overcrowded	**1.19**	**(1.10, 1.28)**
	Poor Roof	**0.82**	**(0.69, 0.98)**
	Large Household	**1.11**	**(1.03, 1.19)**

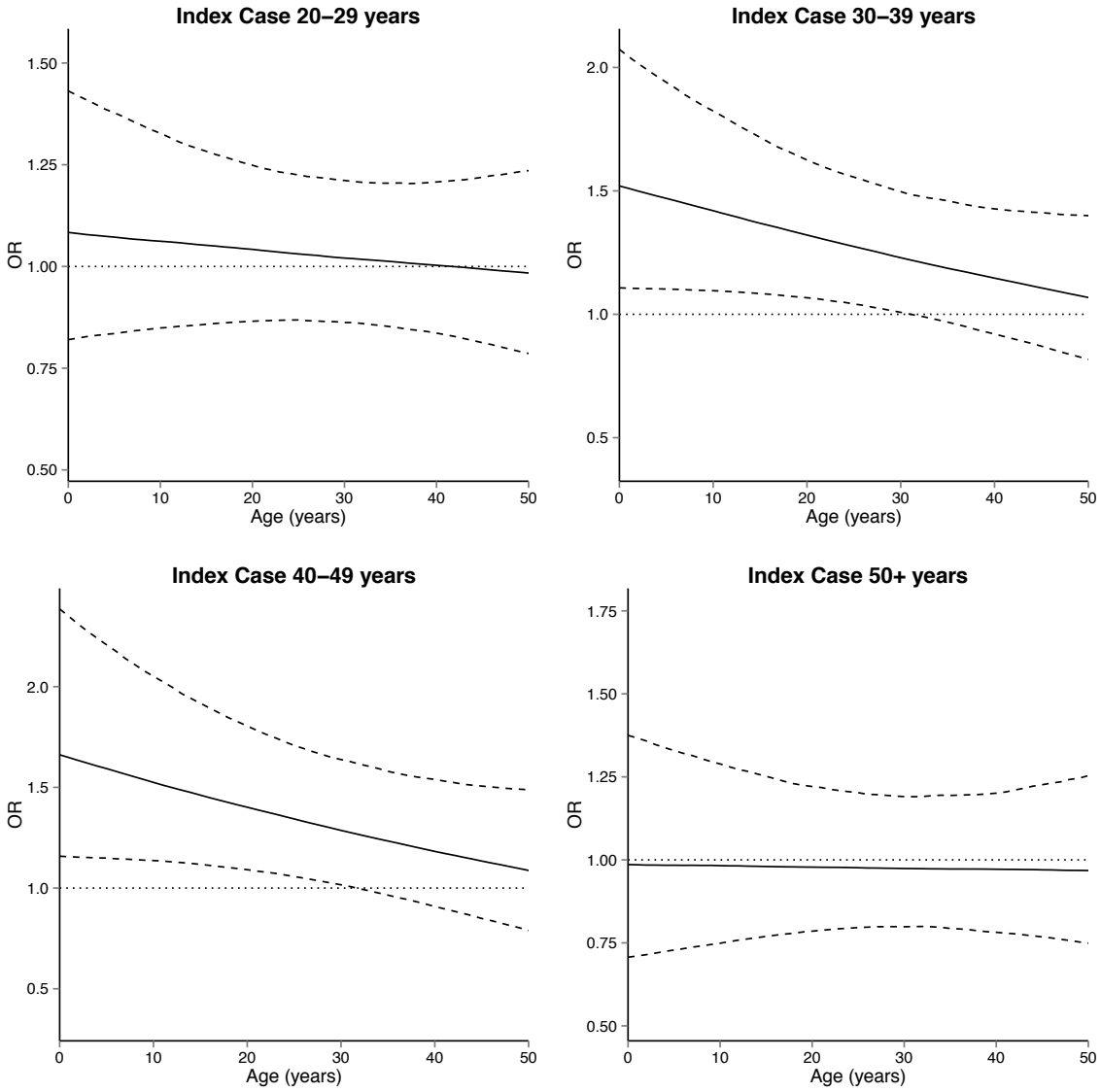
Web Table 3. Predictors of exposure to a Smear-positive/Cuture-positive Index (SCPI) vs. Smear-negative/Culture-negative index (NI). Statistically significant predictors ($p < 0.05$) are in bold.



Web Figure 4. Differences in the age distribution for smear-positive/culture-positive index cases (SCPIs) and smear-negative/culture-negative index cases (NIs). Each filled circle in the plot indicates the difference in the value of the empirical age distribution for SCPIs and NIs for the corresponding quantile specified on the x-axis. The dashed lines are 95% confidence intervals obtained via bootstrap resampling.



Web Figure 5. Age-specific effects of household overcrowding and material conditions on LTBI risk. The figure illustrates the effects of index case age as a function of the age of a household contact. Solid lines indicate age-specific odds-ratios, dashed lines are 95% confidence intervals and the dotted line is provided as a guide to assess statistical significance. All age-specific results are presented relative to the age-specific intercept presented in Figure 4A in the main text.



Web Figure 6. Effects of index case age on risk of LTBI among household contacts. The figure illustrates the effects of index case age as a function of the age of a household contact. Solid lines indicate age-specific odds-ratios, dashed lines are 95% confidence intervals and the dotted line is provided as a guide to assess statistical significance. All age-specific results are presented relative to the age-specific intercept presented in Figure 4A in the main text.

Web Appendix 3. Finite mixture models for identifying LTBI individuals

Finite mixture models are commonly used to infer infection status from quantitative serological data, such as antibody titers (2). This approach has also been applied to the estimation of LTBI prevalence (3,4). In our mixture model, a Gaussian distribution with mean μ_E and standard deviation σ_E represents TST responses of non-infected individuals with TST > 0mm. Responses of individuals with LTBI are represented by a lognormal distribution whose logarithm is a Gaussian distribution with mean μ_L and standard deviation σ_L. We use a lognormal distribution to model LTBI responses to account for large TST responses among some infected individuals, without symmetrically extending the left tail into the uninfected population. The mixing parameter, π, is the proportion of the population drawn from the first mixture component, while the remaining $1-\pi$ are drawn from the second component. We denote the size of individual *i*'s TST response to be s_i and *S* to be the random variable underlying variation in TST responses. If we denote $\mathcal{N}$ to be the PDF of a normal distribution and *g(x)* to be the probability density function (PDF) of a lognormal distribution, we can write the PDF of the mixture model as:

$$Pr(s_i) = \pi\mathcal{N}(s_i; \mu_E, \sigma_E^2) + (1-\pi)g(s_i; \mu_L, \sigma_L^2)$$

To ensure that we are capturing a pattern of TST response representative of the community rather than households with intense exposure, we fit this mixture to the TST responses of the 1221 contacts with $s_i > 0$ mm who had a smear-negative/culture-negative household index case (NI). Because we fit this mixture model to a subset of cases, we ignore the value of the mixing parameter π. For a histogram of the TST sizes of individuals exposed to an NI, see Figure 3 in the main text.

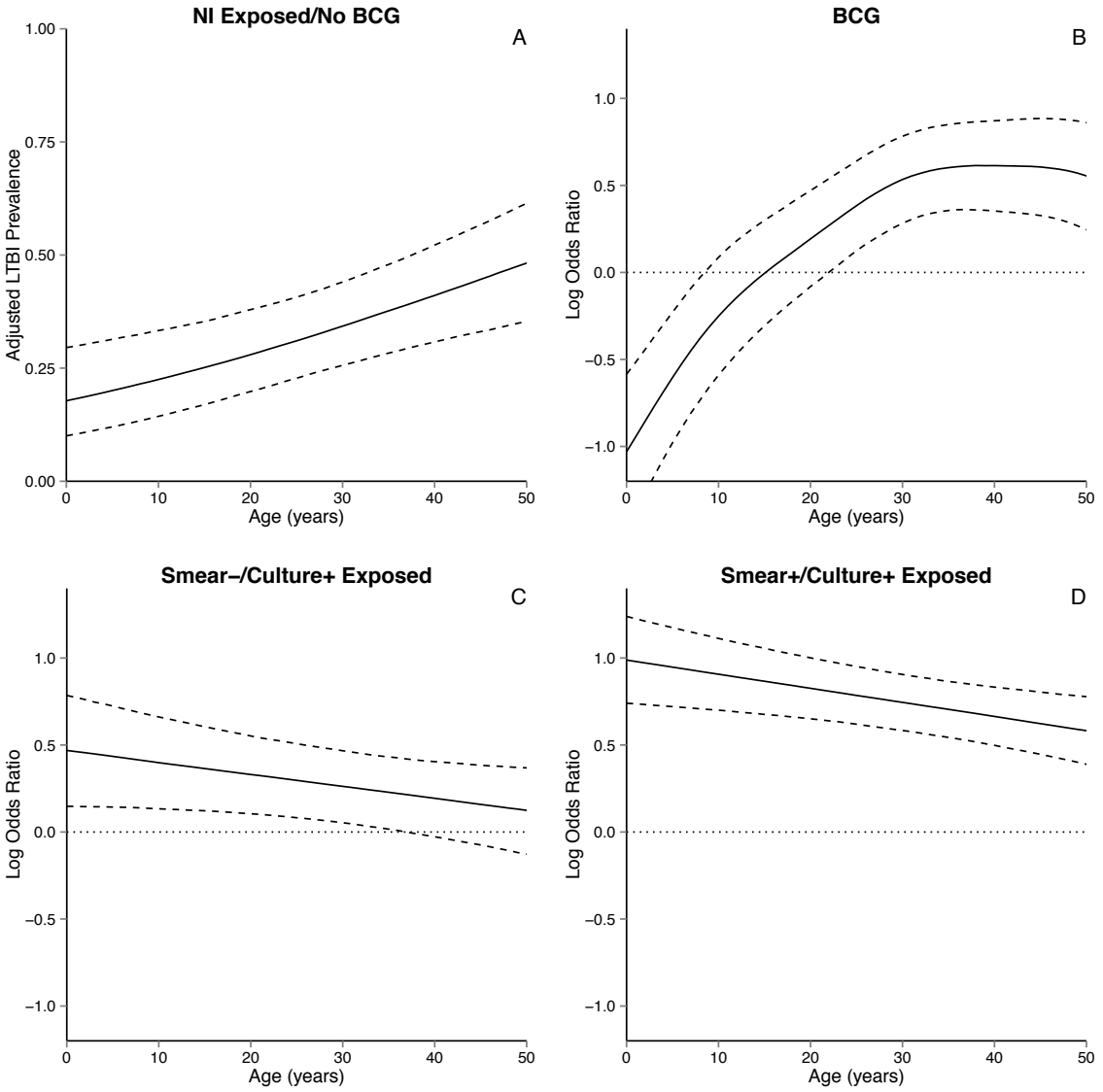
We then use the fitted mixture model to assign infection statuses (i.e. LTBI or no LTBI) to the 6718 household contacts in our analysis with $s_i > 0$mm (an additional 5032 household contacts had $s_i = 0$). We denote the random variable representing an individual's unobserved infection status as Y_i and her realized infection status as y_i. Those individuals with no response (0mm) to the TST are assumed for this analysis to be uninfected, so $\Pr(y_i = 0 | x_i = 0) = 1$. For those with TST > 0mm, we treat y_i as a latent variable to be estimated via Markov chain Monte Carlo (MCMC). We denote the likelihood of an individual's observed TST response conditional on her infection status as $\ell(s_i; y_i)$. Using the values from the fitted mixture model, we can calculate the likelihood of an individual's quantitative TST response given her underlying infection status as:

$$\ell(s_i; y_i) = \begin{cases} \mathcal{N}(s_i; \mu_E, \sigma_E^2), & \text{if } y_i = 0 \\ g(s_i; \mu_L, \sigma_L^2), & \text{if } y_i = 1 \end{cases}$$

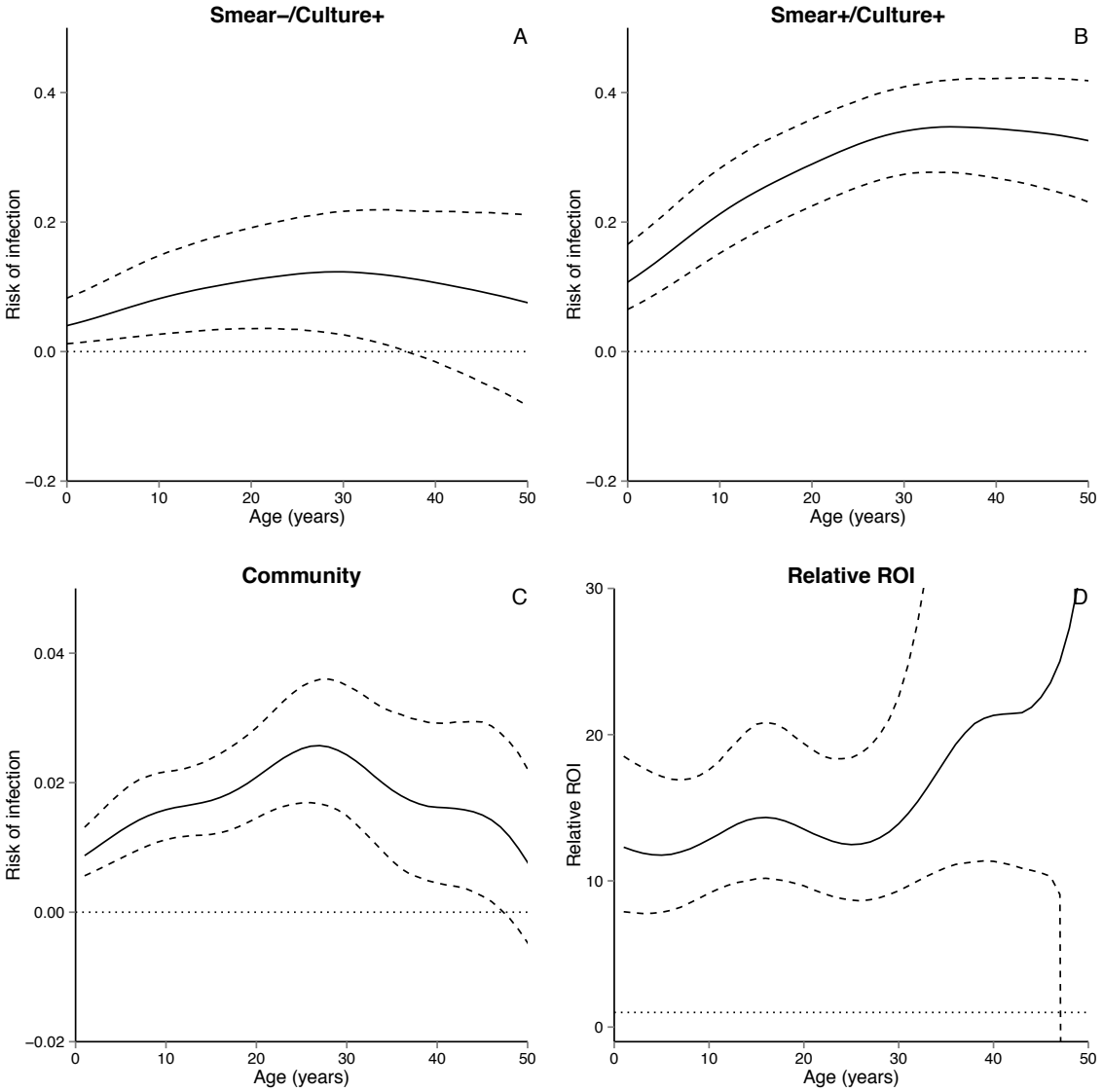
We use the Metropolis-Hastings algorithm (5) to sample from the posterior distribution of y_i for each individual. We take the median of this distribution to be the individual's estimated infection status (2).

Web Appendix 4. Infection Risk Factor Cutoff Model

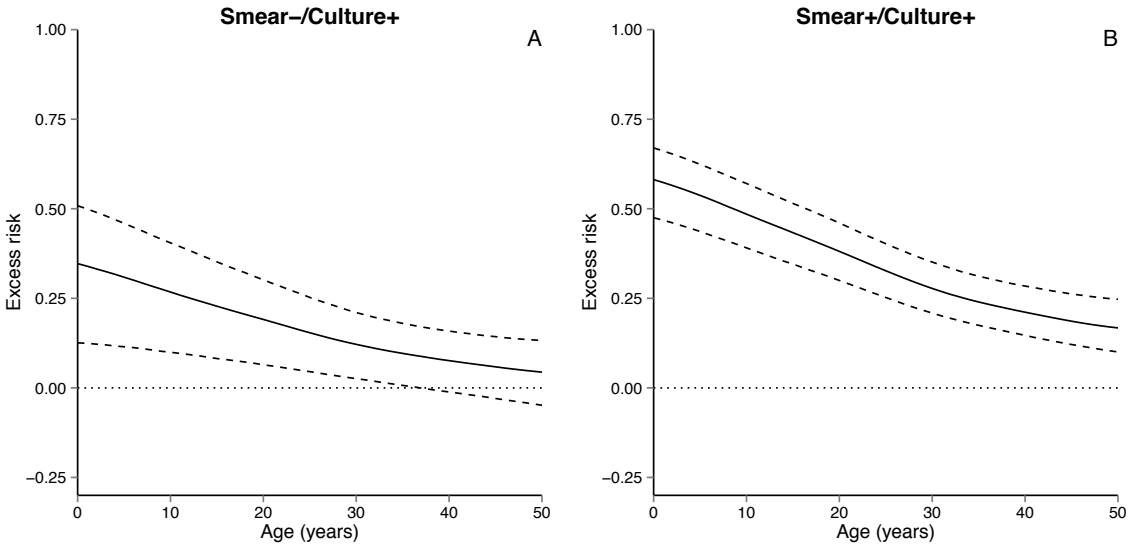
Web Figures 7-9 show parameter estimates, risk of infection (ROI) estimates, and excess risk estimates using the 10mm cutoff. The standard deviation of household-level random intercepts for this model was estimated to be 1.0, and the SD for health-center level intercepts was estimated to be 0.39.



Web Figure 7. Age-specific intercept and age effects for BCG vaccination and SCPI/CPI exposure. Solid lines in each panel are point estimates; dashed lines are 95% credible intervals (CIs). CIs are interpreted over the entire age range, i.e. the upper CI is the upper limit in the trend by age. Panel *A* illustrates the age-specific intercept, represented by the change in log-odds of LTBI with age among individuals exposed to an NI. Panel *B* shows age-specific change in the log-odds of TST-positivity associated with BCG vaccination. Panel *C* & *D* show the effects of CPI and SCPI exposure by age, relative to the age-specific intercept for NI exposure. Dotted lines in panels A-D are provided as a guide for assessing statistical significance. Significance for panels B-D is assessed in terms of the difference from the age-specific intercept.



Web Figure 8. Household and community forces of infection (FOI). Panels *A* and *B* illustrate the age-specific FOI exerted on a household contact by a culture-positive index (A) or smear-positive/culture-positive index (B). Panel *C* illustrates change with age in FOI from community contact. Panel *D* shows the FOI associated with exposure to a household SCPI divided by the community FOI for an unexposed individual of the same age. This may be interpreted as the number of years of community exposure that would be equivalent to the infection pressure exerted by the household index case.



Web Figure 9. Excess risk associated with household smear-positive/culture-positive (SCPI) and culture-positive (CPI) exposure. Excess risk is measured as the proportion reduction in prevalence between the condition in which all individuals have SCPI exposure and the condition in which no individuals are exposed. The solid line shows the point-estimate as a function of age and the dashed lines show 95% credible intervals. The dotted line is a guide for assessing statistical significance.

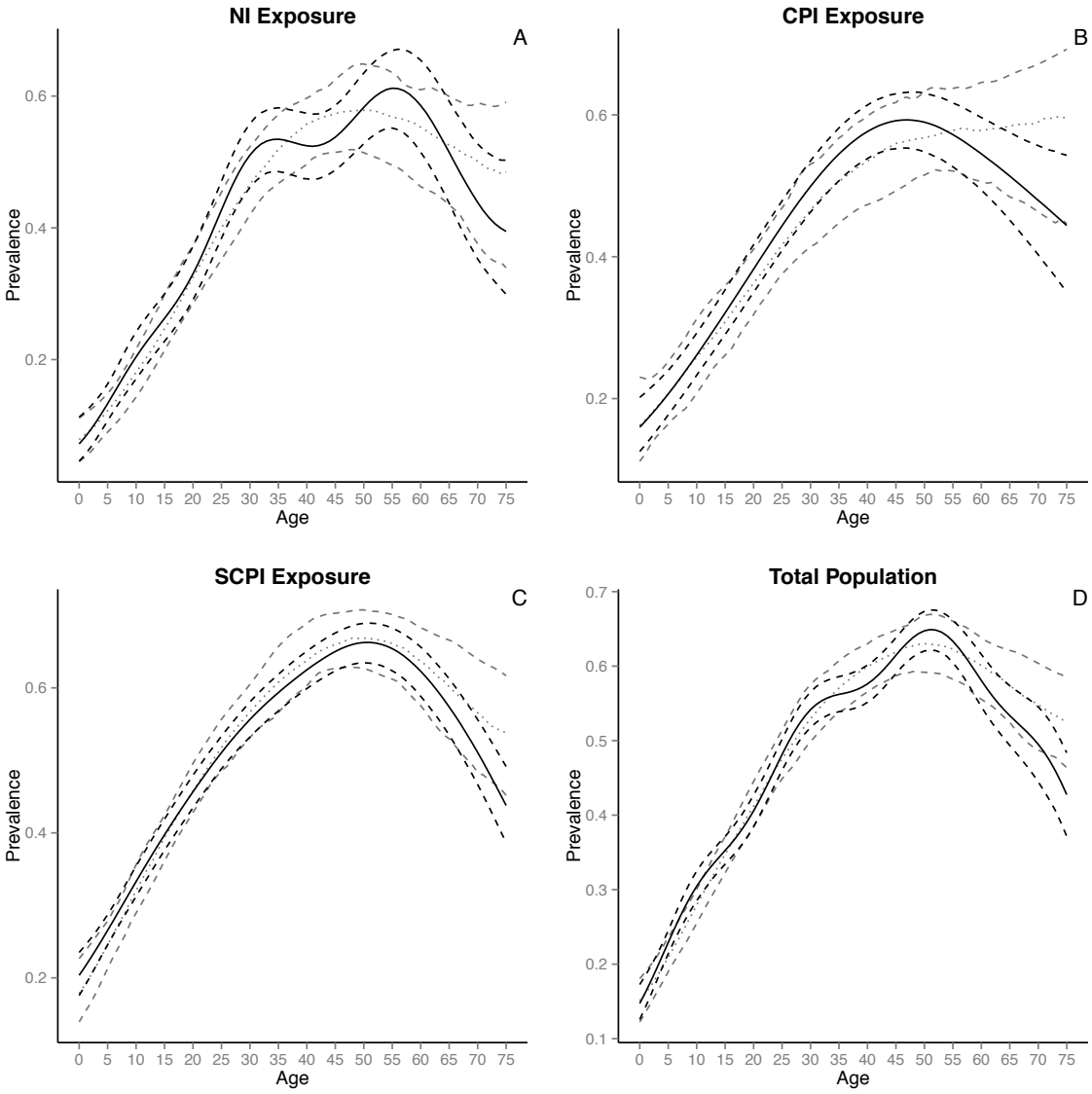
Web Appendix 5. Posterior Predictive Check

To ensure that the risk model adequately describes the data, we employed a posterior predictive check in which we compared the age-prevalence relationship in the household data to age-prevalence curves simulated from the fitted model. Because our results regarding household and community FOI rely upon comparisons between exposure groups, we performed these checks for each risk group (NI, CPI, SCPI) separately as well as for the entire population.

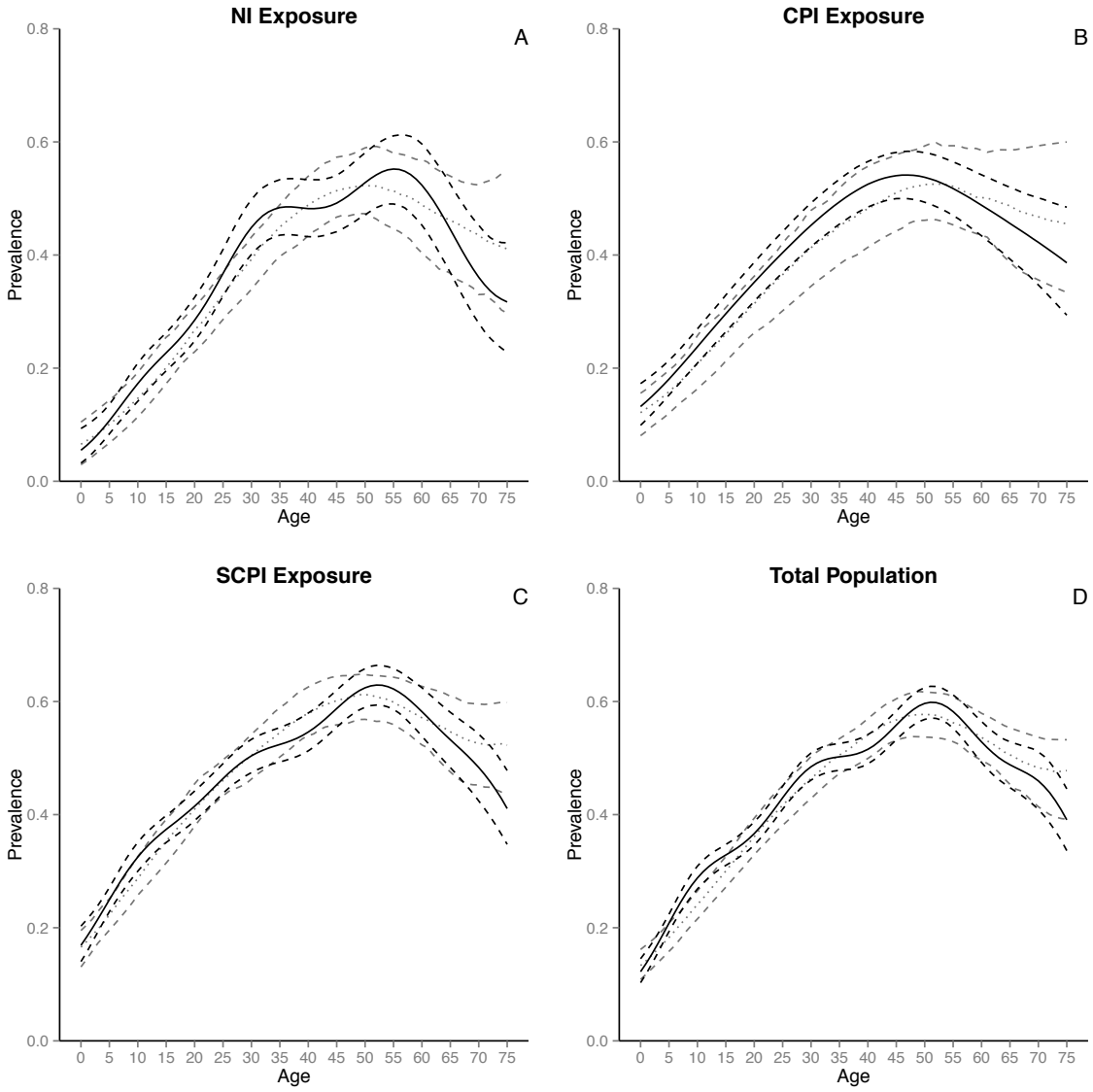
Age-prevalence curves used in the posterior predictive analysis were generated by simulating from the posterior variance-covariance matrix of the fitted generalized additive mixed model (GAMM). Because the *gamm4* fitting procedure does not retain estimates of random effects for individual units (i.e., health centers and households) only the marginal distributions of the parameters are available. Consequently, these plots are most useful for inspecting the qualitative goodness-of-fit and the impact of random effects on prevalence estimates for different age groups.

Web Figure 10 illustrates smoothed prevalence estimates and posterior predictive intervals for the fitted risk factor model based on the mixture model presented in the main text. Web Figure 11 illustrates posterior predictions for the 10mm cutoff model. In all plots, the median of the age-specific posterior predictive distribution is indicated by a dotted line. 95% posterior predictive intervals (PPIs) are illustrated by dashed lines. Solid lines show the smoothed prevalence by age for the specified exposure group. Both sets of models indicate good qualitative fit to the data, suggesting that the inclusion of random

intercepts does not have a strong qualitative impact on estimates. Increasing errors in prediction for older ages likely result from having fewer individuals in older age groups.



Web Figure 10. Unadjusted age-prevalence and posterior predictions from the fitted mixture model. Dark solid and dashed lines indicate unadjusted age-specific smoothed TST prevalence and 95% confidence intervals, respectively. Gray lines indicate the median and 95% posterior predictive intervals of the marginal predictions of the fitted GAMM model.



Web Figure 11. Unadjusted age-prevalence and posterior predictions from the fitted 10mm cutoff model. Dark solid and dashed lines indicate unadjusted age-specific smoothed TST prevalence and 95% confidence intervals, respectively. Gray lines indicate the median and 95% posterior predictive intervals of the marginal predictions of the fitted GAMM model.

Web Appendix 6. Risk of Infection & Excess Risk Calculations

The risk of infection (ROI) quantifies the probability that a particular exposure will result in infection. This is related to the force of infection (FOI), which quantifies the rate at which susceptible individuals are infected for every unit of time. Because the duration of each episode of household exposure is not observed in our data, we estimate discrete risks of infection associated with 1) an average episode of household exposure and 2) a year of exposure in community. This ensures comparability between these two quantities. However, the continuous time FOI can be related to the risk of infection over the period from t to $t+\Delta t$ using the cumulative distribution function of the exponential distribution, so that $ROI(\Delta t) = 1 - \mathrm{e}^{-FOI \times \Delta t}$ and $FOI = \frac{-\log(1-ROI(\Delta t))}{\Delta t}$.

Estimating the Risk Of Infection in the household

We denote ROI associated with household exposure to a smear-positive/culture-positive index (SCPI) as a function $\lambda^{SC+}$and the ROI associated with household exposure to a culture-positive index (CPI) as λ^{C+}. Since the exact duration of exposure to each household index case is unknown, our estimates reflect the average probability that an uninfected individual who undergoes a single episode of exposure to an infectious household index case will become latently infected.

When estimating household ROI, we assume that household SCPI exposure (i.e., $E_{SC+} = 1$) is a transient state, so that the prevalence among unexposed individuals also reflects prevalence within SCPI households immediately prior to the onset of infectiousness of the household index. To calculate the ROI associated with SCPI

exposure, we divide the difference in prevalence for exposed versus unexposed individuals by the probability that the individual will be uninfected in the absence of any exposure to obtain λ_i^{SC+}:

$$\lambda^{SC+}(\eta_i, \omega_i) = \frac{P(Y = 1|E_{SC+} = 1, \eta_i, \omega_i) - P(Y = 1|E_{NI} = 1, \eta_i, \omega_i)}{1 - P(Y = 1|E_{NI} = 1, \eta_i, \omega_i)}$$

Where η_i is the BCG status and η_i is the age of the contact. To calculate the ROI associated with CPI exposure, we employ the same approach, substituting E_{C+} for E_{SC+}.

Estimating Annual Risk Of Infection in the community

We adopt the approach described by Middelkoop et al. (6) to estimate the age-specific annual risk of infection (ARI) in the community (ψ) using the TST responses of individuals exposed to an NI ($E_{NI} = 1$). We define the community ROI to be the difference in the prevalence of LTBI among individuals of age ω and $\omega - 1$, divided by the proportion of individuals exposed to an NI who remain in the non-LTBI state at age ω. We can then write the community ARI as a function of age and vaccination status as:

$$\psi(\eta_i, \omega_i) = \frac{P(Y = 1|E_{NI} = 1, \eta_i, \omega_i) - P(Y = 1|E_{NI} = 1, \eta_i, \omega_i - 1)}{1 - P(Y = 1|E_{NI} = 1, \eta_i, \omega_i - 1)}$$

Our estimate of this quantity explicitly excludes household exposure, so this may be better thought of as a community-only ARI. Consequently, it is expected to be lower than figures obtained from community surveys in which age-specific incidence is estimated without any attempt to exclude individuals with known household exposure, e.g. (6) .

Estimating excess risk associated with household exposure

The excess risk, K, associated with exposure to an SCPI is the expected change in latent infection prevalence among individuals of a particular age if all individuals of that age in a totally exposed population were unexposed, expressed as a fraction of the age-specific prevalence in the fully-exposed condition (7). This quantity allows us to understand the contribution of household exposure to the overall burden of infection at each age. Importantly, it also reflects the amount of information on household infection communicated by a positive skin test of an individual of a particular age who has been exposed to an SCPI or CPI. Using the approach described by Suzuki et al. (7), we can calculate the excess risk associated with SCPI exposure as:

$$K(\eta_i, \omega_i) = \frac{P(Y = 1|E_{SC+} = 1, \eta_i, \omega_i) - P(Y = 1|E_{NI} = 1, \eta_i, \omega_i)}{P(Y = 1|E_{SC+} = 1, \eta_i, \omega_i)}$$

As above, substituting E_{C+} for E_{SC+} allows us to obtain the excess risk associated with CPI exposure.

It is important to note that excess risks for infectious diseases associated with a particular exposure have to be measured with care, as secondary transmission among individuals with a common exposure may artificially inflate the excess risk associated with the exposure (8). However, because we measure excess risks for the acquisition of latent infections that are unlikely to be infectious at or before the time of the survey, our results are unlikely to be impacted by this bias.

References

1. Uys P, Marais BJ, Johnstone-Robertson S, et al. Transmission elasticity in communities hyperendemic for tuberculosis. *Clinical Infectious Diseases*. 2011;52(12):1399–1404.

2. Hens N, Shkedy Z, Aerts M. Modeling Infectious Disease Parameters Based on Serological and Social Contact Data. Springer; 2012.

3. Neuenschwander BE, Zwahlen M, Kim SJ, et al. Determination of the prevalence of infection with Mycobacterium tuberculosis among persons vaccinated against Bacillus Calmette-Guérin in South Korea. *American Journal of Epidemiology*. 2002;155(7):654–663.

4. Villate JI, Ibáñez B, Cabriada V, et al. Analysis of latent tuberculosis and mycobacterium avium infection data using mixture models. *BMC Public Health*. 2006;6(1):240.

5. Robert C, Casella G. Monte Carlo Statistical Methods. *Springer Verlag*. 2004;

6. Middelkoop K, Bekker L-G, Liang H, et al. Force of tuberculosis infection among adolescents in a high HIV and TB prevalence community: a cross-sectional observation study. *BMC Infect Dis*. 2011;11:156.

7. Suzuki E, Yamamoto E, Tsuda T. On the Relations Between Excess Fraction, Attributable Fraction, and Etiologic Fraction. *American Journal of Epidemiology*. 2012;175(6):567–575.

8. Eisenberg JNS, Lewis BL, Porco TC, et al. Bias due to Secondary Transmission in Estimation of Attributable Risk From Intervention Trials. *Epidemiology*. 2003;14(4):442–450.