Supporting Information

Dowdy et al. 10.1073/pnas.1203517109

SI Materials and Methods

Population Selection. The city of Rio de Janeiro is divided into 160 geographic regions called *bairros*, each of which has a population between 180 and 202,000. These *bairros* are aggregated into 10 larger planning areas. The population and incidence of tuberculosis (TB) of each *bairro* is reported by the city's Department of Health and Civil Defense on an annual basis. The last year for which data are available is 2009. In that year, city-wide TB incidence was 95.3 per 100,000.

We identified hotspots as those areas fulfilling three criteria: (*i*) geographically and administratively contiguous; (*ii*) TB incidence at least twice the city-wide average; and (*iii*) at least 120 incidence cases of TB in 2009 (i.e., 2% of the city's total). Three areas fulfilled these criteria (Fig. S1).

Planning Area 1 (AP 1.0, red) consists of the city center, with a population of 248,908 and TB incidence of 219 per 100,000 (545 cases). Manguinhos and Bonsucesso (blue) are geographically contiguous with AP 1.0, but administratively distinct; they are also separated from AP 1.0 by a large highway. Manguinhos is largely a slum (*favela*), and Bonsucesso contains an aging, industrial population. These areas had a combined 2009 population of 55,647 (Manguinhos 37,544; Bonsucesso 18,103), with a TB incidence of 313 per 100,000 (174 cases). Rocinha (green) is Rio de Janeiro's largest *favela*, with a 2009 population of 70,600 and TB incidence of 382 per 100,000 (270 cases). Taken together, these three populations contain 6.0% of Rio de Janeiro's population (6.29 million), but 16.5% of its notified incident TB cases in 2009 (5,993). Because reporting of TB cases relies on a passive surveillance system, the relative contribution of these hotspots to TB incidence in the city is likely underestimated (as passive case reporting is likely less complete in these areas than in the city as a whole).

Model Selection. Because our aim was to demonstrate the potential importance of heterogeneity on TB transmission as a proof of concept—not to evaluate specific interventions—we sought to create the most parsimonious model that would incorporate (*i*) geographic heterogeneity in recent TB transmission and (*ii*) existing epidemiological data from surveillance systems in Rio de Janeiro and/or Brazil. We identified nine potentially relevant data points (also listed in Table 1):

TB incidence, city-wide
TB incidence, hotspot areas (see above)
TB prevalence
TB mortality
Proportion of incident TB represented by retreatment cases
HIV prevalence
HIV mortality
TB/HIV incidence
TB/HIV mortality

The most parsimonious model that could incorporate all of these parameters, as well as geographic heterogeneity in transmission (vs. reactivation of latent disease), would use four TB compartments (susceptible, latent infection, active disease, and recovered/cured), two HIV compartments (HIV-negative and HIV-positive), and two geographic compartments (hotspot and general population). We expanded the number of TB compartments to five (splitting latent infection into a recent and remote compartment), as this fundamentally requires no more assumptions than a four-compartment model that allows both fast and slow progression, and likely represents the risk of progression to active disease more appropriately.

Certain assumptions governing model equations (e.g., that mean time spent in a compartment is equal to 1/rate) are formally valid only at equilibrium/steady-state (i.e., rate of flow into each compartment is equal to the rate of flow out of that compartments). Furthermore, neither HIV nor TB incidence has varied by more than 10% over the past 4 y. Thus, we modeled the baseline population as being at equilibrium. Our primary results do not vary to a substantive degree when small variations from equilibrium (e.g., 5% decline in TB incidence year over year) are introduced into the model.

Estimating the Relative Rate of Hotspot-to-Hotspot vs. Hotspot-to-Community TB Transmission. Empirical data on this relative transmission rate are sparse, owing to the limitations of existing methods (e.g., molecular epidemiology) in estimating the proportion of TB due to recent transmission (1, 2) as well as the expense of conducting population-based molecular epidemiological studies. To inform an estimate of the relative transmission probability within vs. across geographical hotspots, we conducted a focused review of the literature in PubMed on April 13, 2012, using the terms (tuberculosis AND geograph* AND ["molecular epidemiology" OR fingerprint*]). We identified 69 titles initially, and searched the reference lists of relevant articles for further citations. Results from population-based molecular epidemiological studies with appropriate geographical data to inform this estimate are summarized here.

Only one study was identified that overlapped with the present population (3). This study evaluated 10 transmission clusters (mostly of two patients each) within the district of Manguinhos, one of the hotspots in this analysis. The authors found that three clusters linked patients in distant *favelas*, whereas seven clusters linked patients in the same or nearby *favelas*. If one assumes that the relative rate of hotspot-to-community vs. hotspot-to-hotspot transmission follows this same ratio, then 3/7 = 0.43 hotspot-to-hotspot transmissions would occur for every hotspot-to-community transmission. However, the authors did not evaluate transmission outside this district; thus, the relative rate of transmission outside the hotspot could not be assessed.

We identified three other population-based molecular epidemiological studies carried out at the city level and presenting geographical information (4–6). In Baltimore, from 1994 to 1996, 16 of 28 patients (57%) in documented transmission clusters with known epidemiological links showed significant geographic aggregation; among those in molecularly similar transmission clusters without a known epidemiological link, only 11 of 58 patients (19%) showed such clustering (4). The mean intracluster distance was 1.72 km in the first group and 3.10 km in the second group; the largest of our three hotspots has a square area of 8.4 km^2 (http://portalgeo.rio.rj.gov.br/bairroscariocas/index_ra.htm), corresponding to a radial distance of 1.64 km using a circular approximation. Thus, the mean presumed distance of transmission in Baltimore was larger than the approximated radius of the largest hotspot in Rio de Janeiro. Taken together, these data suggest that in Baltimore, at least half of all TB transmission occurred outside of geographically defined regions the same size as hotspots in the present study.

In New York City from 1989 to 1992, 39 patients were identified in 12 documented transmission clusters. Of the 34 with available geographic data for mapping, 16 (47%) occurred within a single geographical region approximately the same size (1.5-mile diameter) as our largest hotspot. This again suggests that a substantial proportion of recent transmission occurs across geographical regions,

but the lack of cluster-specific geographic data precludes an estimate of within- vs. across-hotspot transmission.

An analysis of 187 patients in Harare, Zimbabwe, revealed a remarkably high rate of transmission clustering: 147 patients were confirmed to be part of a transmission cluster (6). Of these, only 36% of patients with a shared spoligotype (and fewer with shared tandem-repeat pattern) were geographically aggregated at the level of the city district, again suggesting that over half of all transmission occurs across geographical boundaries of similar size as those of the hotspots in the present study.

These findings are corroborated by data from larger geographical catchment areas, which showed that 47% of all transmission clusters in England spanned geographical regions the size of London (7), and that the largest transmission cluster in Thailand ($n = 13$ patients) spanned five full provinces, with no one province containing more than six (46%) patients (8).

Taken together, these studies suggest that only 30–50% of recent TB transmission, as detected by population-based molecular epidemiology, is geographically aggregated at a level similar to that of the hotspots in this model. Nevertheless, the one study performed in a Rio de Janeiro *favela* suggests that geographic aggregation may be higher (up to 70%) in this setting. Thus, for our baseline scenario, we estimated that a case of active TB in the hotspot would generate 0.5 secondary transmission events outside the hotspot for every transmission event occurring within the hotspot—a scenario that would lead to 67% geographic aggregation among secondary cases if linked to the index case. However, we performed sensitivity analysis across a wider range of parameter values, as shown in the text, including a scenario in which one hotspot-to-community transmission would occur for every hotspot-to-hotspot transmission (i.e., 50% geographic aggregation).

Model Description and Equations. The model consists of five TB compartments: (*i*) S_{hg}, susceptible; (*ii*) $L1_{hg}$, latently infected, recent; (*iii*) $L2_{hg}$, latently infected, remote; (*iv*) A_{hg}, active TB; and (*v*) R_{hg}, recovered/treated. Each of these five compartments is subdivided into two HIV compartments (subscript $h = 0$ if HIV uninfected, 1 if infected), and two geographic compartments (subscript $g = 0$ if general population, 1 if resident of hotspot).

Rates of flow between compartments are governed by the following differential equations. The first five equations are the primary model equations; the subsequent three equations describe parameters used in the primary equations. The values of all parameters are given in Table 1 with their symbolic representation in Table S1. The model was programmed in R (R Project for Statistical Computing), and the source code for the model is available from D.W.D. on request.

Susceptible (S):

$$dS_{hg}/dt = (mort_{gh}) - (\lambda_g + \mu_0 + \mu_h) * S_{hg} + (hiv_{hg} * S_{hg}), \quad \textbf{[S1]}$$

where $mort_{gh}$ (Eq. S7) is the sum of all mortality in geographic region g (to keep the population constant), λ_g (Eq. S6) is the force of infection in geographic region g, μ_0 represents the background mortality rate, μ_h represents the HIV-related mortality rate (0 if $h = 0$ and μ_1 if $h = 1$), and hiv_{hg} (Eq. S8) represents HIV incidence.

Susceptible individuals enter through birth (held equal to all mortality in the corresponding geographic region, to maintain a constant population in each subpopulation) and exit through TB infection and mortality. HIV infection causes transfer of individuals from the HIV-negative to the HIV-positive compartment. Migration between the subpopulations does not occur.

Latent, recently infected (L1):

$$\frac{dL1_{hg}}{dt} = \lambda_g * [S_{hg} + L2_{hg} * (1-p) + R_{hg}] - (\eta + \zeta_h + \mu_0 + \mu_h) * L1_{hg} + (hiv_{hg} * L1_{hg}), \quad \textbf{[S2]}$$

where λ_g (Eq. S6) is the force of infection in geographic region g, p is the degree of partial immunity to reinfection if latently infected, η = [1/(duration of recent infection phase)], ζ_h is the rate of primary progression per year, μ_0 represents the background mortality rate, μ_h represents the HIV-related mortality rate (0 if $h = 0$ and μ_1 if $h = 1$), and hiv_{hg} (Eq. S8) represents HIV incidence.

Latent, recently infected individuals enter through infection of the susceptible, latent (remote), and recovered populations. The latent compartment exhibits partial immunity from rapid progression after reinfection, which is modeled as a reduction in the rate of flow from the remotely infected to recently infected compartment. Latent (recent) infections exit through progression to remote infection via passage of time, progression to active TB (rapid/primary progression), and mortality (at the same rate as the TB-uninfected population). HIV infection causes transfer of individuals from the HIV-negative to the HIV-positive compartment.

Latent, remotely infected (L2):

$$dL2_{hg}/dt = \eta * L1_{hg} - [\lambda_g * (1-p) + \upsilon_h + \mu_0 + \mu_h] * L2_{hg} + (hiv_{hg} * L2_{hg}), \quad \textbf{[S3]}$$

where η = [1/(duration of recent infection phase)]; λ_g (Eq. S6) is the force of infection in geographic region g, p is the degree of partial immunity to reinfection if latently infected, υ_h is the rate of reactivation to active TB per year, μ_0 represents the background mortality rate, μ_h represents the HIV-related mortality rate (0 if $h = 0$ and μ_1 if $h = 1$), and hiv_{hg} (Eq. S8) represents HIV incidence.

Latent, remotely infected individuals enter through progression of latent (recent) TB infection through passage of time. These individuals exit through reactivation, reinfection (after consideration of partial immunity to rapid progression after reinfection), and mortality (at the same rate as the TB-uninfected population). HIV infection causes transfer of individuals from the HIV-negative to the HIV-positive compartment.

Active TB (A):

$$dA_{hg}/dt = (\zeta_h * L1_{hg}) + (\upsilon_h * L2_{hg}) + (\psi_h * R_{hg}) - [\rho_h + \mu_0 + \mu_h + \mu_{TBh}] * A_{hg} + (hiv_{hg} * A_{hg}), \quad \textbf{[S4]}$$

where ζ_h is the rate of primary progression per year, υ_h is the rate of reactivation to active TB per year, ψ is the rate of relapse per year, ρ_h is the rate of diagnosis and treatment (including spontaneous recovery) per year, μ_0 represents the background mortality rate, μ_h represents the HIV-related mortality rate (0 if $h = 0$ and μ_1 if $h = 1$), μ_{TBh} represents the TB-related mortality rate, and hiv_{hg} (Eq. S8) represents HIV incidence.

Individuals with active TB enter through progression of latent TB infection (both recent and remote), as well as relapse after treatment/recovery. These individuals exit through diagnosis and treatment (or spontaneous recovery), as well as mortality. HIV infection causes transfer of individuals from the HIV-negative to the HIV-positive compartment.

Recovered/treated (R):

$$dR_{hg}/dt = (\rho_h * A_{hg}) - [\lambda_g + \psi_h + \mu_0 + \mu_h] * R_{hg} + (hiv_{hg} * R_{hg}), \quad \textbf{[S5]}$$

where ρ_h is the rate of diagnosis and treatment (including spontaneous recovery) per year, λ_g (Eq. S6) is the force of infection in geographic region g, ψ is the rate of relapse per year,

μ_0 represents the background mortality rate, μ_h represents the HIV-related mortality rate (0 if $h = 0$ and μ_1 if $h = 1$), and and hiv_{hg} (Eq. S8) represents HIV incidence.

Recovered/treated individuals enter through diagnosis and treatment (or spontaneous recovery) of active TB and exit through relapse, reinfection, and mortality. HIV infection causes transfer of individuals from the HIV-negative to the HIV-positive compartment.

The following equations are used to calculate quantities in the above primary model equations:

Force of infection (λ):

$$\lambda_g = \left(\beta_g / z_g\right) * \left(A_{g0} + ri * A_{g1}\right) + \left(\beta_{g'} / z_{g'}\right) * rt * \left(A_{g'0} + ri * A_{g'1}\right), \quad \textbf{[S6]}$$

where β_g is the transmission parameter for geographic region g, z_g represents the proportion of the total population living in region g (i.e., density-dependent transmission), ri represents the relative infectiousness of HIV-infected (vs. HIV-uninfected) TB cases (e.g., lower proportion of smear-positive disease), and rt represents the relative transmission of TB across (vs. within) geographic regions. In this equation, g' represents the opposite geographic region. Thus, in evaluating the force of infection for the hotspot (λ_1), $g = 1$ and $g' = 0$.

Summed mortality:

where μ_0 is the baseline mortality, μ_1 is the HIV-associated mortality, and μ_{TBh} is the mortality associated with TB according to HIV status.

$$\begin{aligned} mort_{gh} = {} & \mu_0 * \left(S_{g0} + S_{g1} + L1_{g0} + L1_{g1} + L2_{g0} + L2_{g1} + A_{g0} + A_{g1} + R_{g0} + R_{g1}\right) \\ & + \mu_1 * \left(S_{g1} + L1_{g1} + L2_{g1} + A_{g1} + R_{g1}\right) + \mu_{TB0} * A_{g0} + \mu_{TB1} * A_{g1}, \end{aligned} \quad \textbf{[S7]}$$

HIV infection

$$\begin{aligned} hiv_{gh} &= \varphi \; if \; g = 0 \; and \; h = 0; \\ &= -\varphi \; if \; g = 0 \; and \; h = 1; \\ &= rh * \varphi \; if \; g = 1 \; and \; h = 0; \\ &= -rh * \varphi \; if \; g = 1 \; and \; h = 1, \end{aligned} \quad \textbf{[8]}$$

where φ is the HIV incidence rate and rh is the relative HIV incidence in the hotspot vs. the general population.

1. Glynn JR, Vynnycky E, Fine PE (1999) Influence of sampling on estimates of clustering and recent transmission of Mycobacterium tuberculosis derived from DNA fingerprinting techniques. *Am J Epidemiol* 149:366–371.
2. Murray M, Alland D (2002) Methodological problems in the molecular epidemiology of tuberculosis. *Am J Epidemiol* 155:565–571.
3. Mendes JM, et al. (2008) Molecular diversity of Mycobacterium tuberculosis strains in a slum area of Rio de Janeiro, Brazil. *J Bras Pneumol* 34:1063–1068.
4. Bishai WR, et al. (1998) Molecular and geographic patterns of tuberculosis transmission after 15 years of directly observed therapy. *JAMA* 280:1679–1684.
5. Alland D, et al. (1994) Transmission of tuberculosis in New York City. An analysis by DNA fingerprinting and conventional epidemiologic methods. *N Engl J Med* 330:1710–1716.
6. Easterbrook PJ, et al. (2004) High rates of clustering of strains causing tuberculosis in Harare, Zimbabwe: A molecular epidemiological study. *J Clin Microbiol* 42:4536–4544.
7. Love J, et al. (2009) Molecular epidemiology of tuberculosis in England, 1998. *Int J Tuberc Lung Dis* 13:201–207.
8. Rienthong D, et al. (2005) Restriction fragment length polymorphism study of nationwide samples of Mycobacterium tuberculosis in Thailand, 1997-1998. *Int J Tuberc Lung Dis* 9:576–581.

Fig. S1. Map of Rio de Janeiro *bairros*. This map shows all 159 *bairros* of Rio de Janeiro. The *bairros* corresponding to the three TB transmission hotspots are shown in color, as described in *SI Materials and Methods*. Map courtesy of the government of the City of Rio de Janeiro.

Table S1. Model parameters and their symbolic representation

Parameter	Description	Value
β_0	Number of transmissions per active TB per year, community	3.71
β_1	Number of transmissions per active TB per year, hotspot	9.74
rt	Relative rate of transmission, hotspot-to-hotspot vs. hotspot-to-community	0.03
z_1	Proportion of total population residing in the hotspot	0.06
ζ_1	Rate of rapid progression after recent infection, HIV-positive, per year	0.31
υ_1	Rate of slow progression after remote infection, HIV-positive, per year	0.08
μ_{TB0}	TB mortality rate, HIV-negative, per year	0.031
μ_{TB1}	TB mortality rate, HIV-positive, per year	0.074
ρ_0	TB detection/treatment rate, HIV-negative, per year	0.87
ρ_1	TB detection/treatment rate, HIV-positive, per year	1.74
φ	HIV incidence, per year	1.5×10^{-4}
μ_1	HIV mortality rate (non-TB), per year	0.026
ψ	TB relapse rate, per year	0.0083
ri	Relative infectivity of HIV/TB cases	0.68
p	Partial immunity to reinfection if latently infected	0.56
$1/\eta$	Duration of "recent infection" phase	5 y
ζ_0	Rate of rapid progression during this phase, HIV-negative, per year	0.03
υ_0	Rate of slow progression of remote TB infection, HIV-negative, per year	0.0005
$1/\mu_0$	Life expectancy	73 y
rh	Relative HIV incidence in hotspot vs. community	2.13