



Published in final edited form as:

Epidemiology. 2018 January ; 29(1): 117–125. doi:10.1097/EDE.0000000000000751.

Determinants of short-term movement in a developing region and implications for disease transmission

Alicia N. M. Kraay¹, James Trostle^{2,a}, Andrew F. Brouwer^{1,a}, William C. Trujillo³, and Joseph N. S. Eisenberg^{1,*}

¹Department of Epidemiology, University of Michigan, Ann Arbor, MI

²Trinity College, Hartford, CT

³Universidad San Francisco de Quito, Quito, Ecuador

Abstract

Background—Human mobility is important for infectious disease spread. However, little is known about how travel varies by demographic groups and how this heterogeneity influences infectious disease risk.

Methods—We analyzed ten years of survey data from 15 communities in a remote but rapidly changing region in rural Ecuador where road development in the past 15–20 years has dramatically changed travel. We identify determinants of travel and incorporate them into an infection transmission model.

Results—Individuals living in communities more remote at baseline had lower travel rates compared with less remote villages (adjusted odds ratio [OR]=0.51, 95% CI 0.38, 0.67). Our model predicts that less remote villages are, therefore, at increased disease risk. Though road building and travel increased for all communities, this risk differential remained over 10 years of observation. Our transmission model also suggests that travelers and non-travelers have different roles in disease transmission. Adults travel more than children (adjusted OR=1.73, 95% CI 1.30, 2.31) and therefore disseminate infection from population centers to rural communities. Children are more likely than adults to be infected locally (attributable fraction=0.24 and 0.09, respectively) and were indirectly affected by adult travel patterns.

Conclusions—These results reinforce the importance of large population centers for regional transmission and show that children and adults may play different roles in disease spread. Changing transportation infrastructure and subsequent economic and social transitions are occurring worldwide, potentially causing increased regional risk of disease.

Keywords

disease transmission; travel; mobility; movement; road construction; remoteness

*Corresponding Author: Department of Epidemiology, University of Michigan, 1415 Washington Heights, Ann Arbor, MI 48104 (jnse@umich.edu).

^aJ.T. and A.F.B. contributed equally to this work

Conflicts of Interest We have no competing interests.

Data and code used for this analysis is available upon request from the authors.

Introduction

Human mobility contributes to the spread of infectious disease at multiple spatial and temporal levels [1–5]. Human migration—the long-term process by which individuals permanently or semi-permanently relocate—has been implicated in the introduction of pathogens to new locations. For example, movement along the Silk Road introduced the Black Death to Europe in the 14th century, and the forced transatlantic migration of African slaves brought malaria to the US [1]. Short-term movement—travel to another location with subsequent return—can also introduce pathogens when it occurs over large distances [2]. Examples of this process include the spread of SARS [6], pandemic influenza [7, 8], and cholera [9]. Short-term movement also occurs over smaller distances more critical for pathogen circulation, analysis of which can help reveal spatial patterns of infection spread. Short-term movement has been implicated in the 2010 cholera epidemic in Haiti [10] as well as seasonal influenza [11] and dengue transmission [5, 12].

Road availability can influence both frequency and determinants of mobility. For example, as remote towns gain road access, travel frequency and average distance may increase, and towns' sociodemographic environment may change [13]. Easier access to roads also makes urban markets more accessible to rural farmers, which may influence their economic activities within the village as well as reasons for their travel [14]. Finally, road access may also facilitate migrant labor, leading to a more transient population and changing social network structure [14–16].

Therefore, in addition to the benefits of road access, construction of new roads might also lead to increased regional risk of disease [16]. Other work has shown that environmental change occurring in tandem with road construction projects may lead to increased local risk of malaria and dengue [17,18]. As previously remote regions in parts of Asia, Africa, and Latin America gain access to new roads, their changing travel patterns may facilitate both pathogen introduction and pathogen circulation, leading to convergence of regional risk.

The relevance of both short- and long-term movement for transmission has been widely recognized by transmission modelers, who regularly incorporate movement into their models [10,11,19–21]. For example, travel can increase transmission and allow for pathogen invasion even when there would not otherwise be an outbreak [22–24]. Most of this work has focused on developing models that adequately capture the average rate of travel and distribution of destinations rather than the variability in these patterns by demographic subgroups [19, 25, 26]. A separate body of work has looked at the role of population heterogeneity in disease transmission and has shown that outbreaks can occur when the average basic reproduction number (R_0) is less than one if a subgroup of super-spreaders is at higher risk [27, 28]. We aim to combine these two lines of research, investigating how population heterogeneity in both disease susceptibility and travel frequency may influence outbreaks. We use a transmission model, parameterized to reflect rotavirus, as a case study because it is epidemiologically important in our study region and globally and because its attack rate differs by age [29–31].

This paper contributes to a more developed understanding of short-term movement. Using data from a region in rural Ecuador that has been experiencing continuous road development, we evaluate changes in movement patterns and village demographics over a 10-year period (2003–2013) following the construction of a new road. We next identify determinants of short-term travel outside the study region and evaluate the stability of these determinants over time. Finally, we present a transmission model for rotavirus in our study region that accounts for heterogeneity in both travel and disease transmission and illustrates how this heterogeneity can lead to both increased risk of disease and changes in disease etiology.

Methods

Data

Data for this project were collected in rural Ecuador in conjunction with a larger study [16]. In 1996, the government of Ecuador began a road construction project to link the Andes with the Ecuadorian Pacific coast. The highway was completed in 2001, but new roads continue to be built, increasing access for previously more remote villages. Our EcoDess project was initiated in 2003 as a natural experiment to document longitudinal changes that occurred in villages surrounding the local metropolitan center of Borbón due to road construction. The present analysis focused on 15 study villages that were followed from 2003–2013. See reference [16] for more details about the study region. All data collection protocols were approved by institutional review boards at all institutions involved in the project.

In this paper, we use data from movement surveys conducted in 2003, 2007, 2010, and 2013 and census data collected by our study at about the same times. We use stool sample positivity for rotavirus from our ongoing case–control study from 2007—before the vaccine was introduced [32]—to parameterize our transmission model. The methods used to identify rotavirus infection are described in detail elsewhere [29]. Our analyses considered 15 villages, 1,347 households, 5,443 individuals, and 1–4 time points per individual, for a total of 10,725 observations and a median of 2 observations per individual.

Software

Time trend analyses were conducted in R (v. 3.1.2) using the package gee, and other multivariate analyses were conducted in SAS (v. 9.4). Transmission modeling was conducted in R (v. 3.1.2) using the package deSolve.

Defined measures

Remoteness—Remoteness reflects a number of interrelated factors that together capture the relative isolation of a given community. Here, we assess remoteness using two measures: 1) travel frequency, which varies over time, and 2) the feasibility of travel at baseline, based on cost and travel time to reach the nearest large town. For the duration of this paper we refer to the latter metric (cost and travel time) as ‘baseline remoteness.’

Travel frequency: The main outcome of interest in this study was whether an individual had traveled outside of the study region in the past seven days.

Baseline cost and travel time: Project staff used data collected at the start of the study to develop a metric for the remoteness of each village based on the cost and travel time required to reach Borbón, the largest town in the region [16]. This measure was subsequently categorized into three groups: *close*, *medium*, and *far*.

While this metric reflects the feasibility of travel, actual travel frequency may not necessarily be shaped only by road accessibility; thus, the two metrics may measure different aspects of remoteness. Even if the absolute cost and travel time to Borbón changes over time as roads develop, if the *relative* isolation of these communities remains constant, remoteness would remain an important predictor of travel. If, however, road development homogenizes travel frequency across the region, any differences in travel by baseline remoteness should diminish over time. To determine whether or not the regional risk appeared to be converging, we considered baseline remoteness as a covariate in all multivariate models and investigated whether the association between baseline remoteness and travel frequency was stable over time.

Covariates—Occupation was classified as *none*, *domestic*, *student*, *agriculture*, *salaried*, or *other*. Age was calculated using date of birth and was categorized into three groups: <5, 5–13, and >13 years. Each individual’s duration of residence in the village was analyzed using z-scores calculated within age brackets. Highest household education was the highest number of years of schooling reported by any individual of that household. Village-level availability of secondary schools was included in the models as a binary variable and was treated as time-invariant. The mean duration of residence for each village for all individuals over the age of 13 was also included as a proxy for population stability. All of these variables were taken from our census data.

Regression models

To investigate time trends and identify determinants of travel, we used a generalized estimating equation (GEE) framework, which allowed us to account for community clustering but retain a population average interpretation [33]. Because GEE models are not estimated using maximum likelihood, we compared model fit using quasi-information criteria [34]. Change-over-time was first evaluated by building regression models containing only one covariate and years since 2003. These time trends were conceptualized as providing descriptive information about the region and are therefore unadjusted. To identify determinants of travel, forward regression was used. Because all variables except baseline remoteness and secondary school availability were allowed to change over time, our model also adjusts for time-dependent confounding. The stability of associations over time was evaluated by comparing, for each study year, the model containing only cost and travel time, the model containing all variables except cost and travel time, and the full model.

Transmission model

We developed a disease transmission model with a susceptible, infectious, recovered framework (parameterized to reflect rotavirus) to investigate the impact of travel to an urban center (here modeled as analogous to Borbón) on a small village. Because the geography and transportation network of the region causes most travel to traverse Borbón, almost all

travel outside the region is to another urban setting, and travel to the village by city residents is assumed to be negligible, this relatively simple two-community model suffices to investigate the impact of travel by members of the village on their community as a whole. To capture heterogeneity in risk of infection and travel, we stratified the modeled populations into two age categories (under and over 5 years), with children under five being more likely to become infected and less likely to travel (see Figure 1 and the eAppendix for additional details on model structure). The appendix also presents the equations, parameterization, a sensitivity analysis of the transmission parameters, and calculations of the basic reproductive number for each community in isolation, which we call R_0^* . In our study region, the Global R_0 is always greater than 1 for rotavirus infection and thus not relevant (see eAppendix, Transmission Rate section for the calculations necessary to numerically calculate the Global R_0). To understand how disease risk (measured by cumulative incidence) might change with time, we modeled an average close, medium, and far village, using travel frequency and population size in the data for each of the four study years (2003, 2007, 2010, 2013) for a total of 12 models (three villages and four time points). To examine the separate roles of demographic heterogeneity in travel and the level of travel, we conducted model analyses to investigate changes in cumulative disease incidence as i) we increase the average travel frequency for all village residents (assuming children and adults travel the same) and ii) we fix the travel frequency of children but increase the travel frequency of adults. These analyses used the population size and structure from 2013 and considered rotavirus transmission rates corresponding to the maximum and minimum observed in the study region (corresponding to R_0^* of 0.79 and 1.44 respectively).

Results

Evaluating demographic change

Baseline characteristics of the participants are shown in Table 1. Over time, the fraction of individuals traveling outside the region increased for all remoteness levels, with a 9.1%, 8.2% and 4.4% increase in travel for close, medium, and far villages respectively (see Table 1 and Figure 2). In 2013, far villages had similar travel patterns to the close villages in 2003. Far villages consistently had lower levels of travel than the medium or close villages with an average 6.0% and 3.2% difference comparing far with close and medium villages respectively, suggesting that travel patterns for the communities had not yet converged during the study. All far villages that showed an increase in travel patterns did not begin their increases until 2010, suggesting a time lag in the effect of changing infrastructure for more distant communities.

In addition to increased travel, there was also a tendency toward increased socioeconomic status. The proportion of individuals having no occupation decreased over time. The fastest growing occupation was salaried worker, and this change was especially pronounced in the most remote villages. The proportion of individuals working in agriculture decreased slightly for close and far villages.

Identifying predictors of travel

Gender was not associated with travel frequency and was therefore not included in the regression models. Based on prior knowledge, we conceptualized remoteness as being an 'exposure' whose effects were mediated by various demographic factors. However, to estimate the effect of these mediators on travel, we adjusted our demographic models for remoteness because it was a common cause of the various demographic characteristics and travel and therefore a confounder of those associations. See eFigure 1 for causal diagram.

Because the relationship between age and travel patterns was non-linear, we used a categorical age variable (<5, 5–13, >13). The results of model building are shown in Table 2. Although all variables except for age were associated with both remoteness and out-of-region travel (Table 2), only secondary school availability appeared to be markedly confounded by remoteness (comparing the Unadjusted and Remoteness Adjusted models). Secondary school availability and mean duration of residence appeared to be the strongest mediators of the remoteness–travel estimate, as indicated by the reduction in the point estimate (comparing the Unadjusted Model with Model 1). In general, occupation and education appear to have independent effects. However, the point estimate for salaried occupation attenuated after adjusting for education, most likely because of the much higher required education for salaried workers. In other words, there was no association between education and occupation except for salaried workers. The effect of occupation also appeared to be confounded by age (comparing Models 3 and 4).

After adjusting for all variables, travelers were more likely to come from households with higher education (OR=1.10, 95% CI 1.08, 1.12) and to work either in domestic (OR=1.64, 95% CI: 1.14, 2.36), agriculture (OR=1.50, 95% CI: 1.13, 1.98), or salaried occupations (OR=3.27, 95% CI: 2.15, 4.96). Remoteness remained a strong predictor of travel (OR=0.51, 95% CI: 0.38, 0.67), with travel being more common among individuals living in less remote regions. Individuals living in areas with a secondary school were also less likely to travel than people without secondary schools (OR=0.53, 95% CI: 0.37, 0.76). The effect of age was non-linear; individuals aged 5–13 were less likely to travel than children under five (OR=0.65, 95% CI: 0.50, 0.85) and adults over the age of 13 were more likely to travel than children under five (OR=1.73, 95% CI: 1.30, 2.31) (Table 2).

Although the effect estimate for increased travel by household level of education is small for each year increase in education (OR=1.10), the cumulative effect is substantial. For instance, a five-year difference in household education (roughly the difference between completing primary and secondary school) would correspond to an odds ratio of 1.61 for traveling.

Overall, the full model, including both baseline remoteness (cost and travel time) and all demographic variables was the best model for all four study years (Table 3). The quasi-information criteria values including demographic factors were consistently lower than those with remoteness only, suggesting that these variables combined provided more information than the baseline remoteness metric alone. Although the cost and travel time metric provided less information than the demographic variables, adding it to the model improved model fit for each study year, supporting its continued use as a measure of remoteness in our region.

Travel and disease transmission

Component effects—To disaggregate the impact of travel from the impact of heterogeneity in travel by age, we present simulated cumulative disease incidence that consider change in i) the overall travel rate for all ages (homogeneous travel) and ii) a proportional increase in travel for adults only (heterogeneous travel by age).

Homogeneity in travel: As travel increased from levels seen in far villages to those in close villages, the predicted disease incidence rate doubled (15, 27, and 31 per 1,000 people for close, medium, and far villages respectively, where $R_0^*=0.79$ for all village types). While higher transmissibility within the village (R_0^*) was associated with increased disease risk, the estimated effect of travel was much stronger. In general, predicted incidence at the maximum transmission rate ($R_0^*=1.44$) was only 1–2 cases per 1,000 people higher than risk at the minimum transmission rate ($R_0^*=0.79$), less than a 7% increase (details in eAppendix).

Heterogeneity in travel: As the relative travel rates of adults increased compared to children, the risk of infection for the village increased, suggesting that the adults disseminate the infection into local communities (see Table 4). Although this increase in risk was greatest for adults, risk also increased for children. Only results from close villages are shown; results were qualitatively similar for close and medium villages and are shown in the eAppendix. Based on our survey data (Table 2), adults had 1.73 times the travel of children, so we would expect our study region to most closely resemble the medium travel heterogeneity scenario (scenarios defined in Table 4).

Further, as heterogeneity in travel increased, the attributable fraction of infections acquired locally increased for children, but decreased for adults (Figure 3). The magnitude of local transmission depended on transmissibility within the village. Under conditions of high heterogeneity in travel and high transmission, 24% of all transmission to children originated within the village, compared to 6% for no heterogeneity in travel (i.e., adults traveled the same as children) and low transmission. However, regardless of whether R_0^* was greater than 1, most transmission for both age groups occurred in the city. This result held over the transmission parameter sensitivity analysis, but was modified by village population size. Increasing the population of the village relative to the city was associated with proportional increases in the attributable fraction for local transmission but slight decreases in the overall incidence (see eTable 6).

Predicted Incidence—Over time, modeled infection risk increased for all three types of communities (Table 5). In 2013, the risk in far villages was similar to that seen in close villages in 2003. At the end of the study, the predicted risks for close and medium villages were similar, but the risk for far villages remained lower. Thus, although all villages had predicted increases in risk over the study period, the gradient of risk by remoteness was preserved for the most remote villages. Comparison with our component effects model suggests that this increase in risk was almost entirely driven by increased travel rather than changing demographics.

Discussion

Areas with changing transportation infrastructure undergo demographic change that can impact community structure and increase infection risk. In our analysis, this increased risk is largely driven by the city, highlighting the likely importance of population centers for transmission into remote regions. Furthermore, we show that increasing adult travel frequency can indirectly increase risk for those that travel less (e.g., children) and affect the fraction of infections that are acquired locally for both age groups.

Demographic change

Medium and close villages had higher rates of travel than far villages in 2003. Over time, travel rates increased across the region. Similar increases in travel in our transmission model led to increased risk of infection. Although we parameterize our model to reflect rotavirus, our results extend to other pathogens and demonstrate how road construction can lead to increased regional risk of enteric infection, highlighting the need to consider the broader health impacts of constructing roads and other projects like railroads that can increase opportunities for human movement.

Several demographic variables were found to be important predictors of travel; the effect of these variables did not appreciably change over time. Travelers were more likely to come from households with higher education and to be salaried or domestic workers. Individuals with salaried positions were the most likely to engage in out-of-region travel. The increased travel seen among domestic workers may reflect fewer time constraints placed on women who work inside the home. The tendency for young children under the age of 5 to travel more than older children (aged 5–13) may be because the youngest children are too young to be left behind in the villages when the mother is traveling.

Travel and disease transmission

Our transmission model results suggested increased travel could lead to increased infection; most of this risk to the village appears to be driven by the city. Although the city had a higher transmission rate than the villages, our sensitivity analysis revealed that the city's population size was the primary reason that it dominated transmission. Adults experienced the greatest increases in infection risk due to their more frequent travel, but children were also indirectly affected by the travel of adults. This result highlights the role of travelers in dissemination of disease to remote communities and shows that population centers can be important for driving risk in remote regions.

The increased travel observed across all villages over time indicates potentially increased disease transmission for all types of communities—including those in remote regions—as a function of changing mobility. In general, more remote communities tend to benefit from less infection pressure due to their relative isolation [36]. In our study, remoteness remained an important predictor of travel and infection throughout the study, despite ongoing development. Therefore, at least in our study region, the erosion of a remoteness advantage occurs over a time scale longer than 10 years.

Because of demographic differences in travel rates, local interventions may have a greater impact on children than adults but targeting interventions in cities is likely to be the most effective way to lower regional risk. Future studies comparing prevalence of infection and diarrheal disease both by remoteness and over time in this region may provide further insights into how changing travel patterns affect risk. Thus, while population-level proxies (i.e., gravity models, radiation models, etc.) for travel may be inadequate to capture demographic variability in travel, they still may be reasonable predictors of average travel behavior as development proceeds.

Strengths and Limitations

There are a few caveats to our study. First, our remoteness variable was defined at baseline. If community remoteness confounds the association between the covariates and travel, there may be some residual confounding by remoteness in later years that is not captured by the baseline measure. Second, because of the lack of preexisting theory on determinants of remoteness, we relied on the change in estimate criteria to evaluate confounding and mediation.

Additionally, our estimate of R_0^* required an estimate of age at first infection. We used age-specific prevalence of infection as a proxy for age at first infection. Because people can be re-infected with rotavirus, this assumption may have led us to overestimate the average age at first infection and thus underestimate both transmission rates and R_0^* . However, sensitivity analysis demonstrated that i) our results were similar regardless of the actual value of R_0^* and ii) relative infectivity (city vs. community and adults vs. children) was more important than absolute infectivity for determining the source of cases. Because this overestimation of β is non-differential by remoteness and thus relative infectivity is unaffected, our main results should be unbiased. Furthermore, people are rarely infected more than twice, and subsequent infections are less likely to be symptomatic and thus less likely to transmit [31, 37, 38]. As such, any bias in the absolute infectivity is likely to be minimal.

These longitudinal data allow us to assess change over time and show that these covariates have stable relationships with travel. The timing of this study relative to the completion of a road construction project is also a strength because the data provided a useful natural experiment with a unique opportunity to answer these questions in a population with new road development.

Overall, these findings provide useful data and a theoretical framework for future studies investigating the determinants of movement in other populations, particularly areas that are rapidly urbanizing. Given that rural–urban connections are increasingly more common, the relevance of this work continues to increase [39].

Conclusions

Current methods for incorporating travel into disease transmission models generally rely on proxies for travel (e.g., gravity and radiation models) that are based on distance and population size, without any consideration of demographic heterogeneity [10, 11, 19–21]. Our results suggest that while these models may accurately estimate average travel,

accounting for travel heterogeneity in transmission models may both improve model predictions and identify opportunities to target public health interventions.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We would like to thank the EcoDess field staff for their valuable contribution collecting the data used in this work.

Funding This work was funded by the National Institutes of Health (grant R01- AI050038), NIH MIDAS (grant U01GM110712) and the National Science Foundation Water Sustainability and Climate Program (grant 1360330).

References

1. Prothero RM. Disease and mobility: a neglected factor in epidemiology. *International journal of epidemiology*. 1977; 6:259–267. [PubMed: 591173]
2. Tatem AJ, Rogers DJ, Hay SI. Global transport networks and infectious disease spread. *Advances in parasitology*. 2006; 62:293–343. [PubMed: 16647974]
3. Cliff A. Time, travel and infection. *British Medical Bulletin*. 2004; 69:87–99. [PubMed: 15226199]
4. Martens P, Hall L. Malaria on the move : human population movement and malaria transmission. *Emerging Infectious Diseases*. 2000; 103:1–7.
5. Stoddard ST, Morrison AC, Vazquez-Prokopec GM, Soldan VP, Kochel TJ, Kitron U, Elder JP, Scott TW. The role of human movement in the transmission of vector-borne pathogens. *PLoS Neglected Tropical Diseases*. 2009; 3
6. Ruan S, Wang W, Levin SA. The effect of global travel on the spread of sars. *Mathematical Biosciences and Engineering*. 2005; 3:205–218.
7. Cooper BS, Pitman RJ, Edmunds WJ, Gay NJ. Delaying the international spread of pandemic influenza. *PLoS Medicine*. 2006; 3:e212. [PubMed: 16640458]
8. Grais RF, Ellis JH, Glass GE. Assessing the impact of airline travel on the geographic spread of pandemic influenza. *European Journal of Epidemiology*. 2003; 18:1065–1072. [PubMed: 14620941]
9. Lee K, Dodgson R. Globalization and cholera: Implications for global governance. *Global Governance*. 2000; 6:213–236.
10. Tuite AR, Tien J, Eisenberg M, Earn DJD, Ma J, Fisman DN. Cholera epidemic in haiti, 2010: Using a transmission model to explain spatial spread of disease and identify optimal control interventions. *Annals of Internal Medicine*. 2011; 154:593–601. [PubMed: 21383314]
11. Viboud C, Bjornstad ON, Smith DL, Simonsen L, Miller Ma, Grenfell BT. Synchrony, waves, and spatial hierarchies in the spread of influenza. *Science (New York, N.Y.)*. 2006; 312:447–451.
12. Stoddard ST, Forshey BM, Morrison AC, Paz-Soldan VA, Vazquez-Prokopec GM, Astete H, Reiner RC, Vilcarromero S, Elder JP, Halsey ES, et al. House-to-house human movement drives dengue virus transmission. *Proceedings of the National Academy of Sciences of the United States of America*. 2013; 110:994–9. [PubMed: 23277539]
13. Windle J, Cramb RA. Remoteness and rural development: economic impacts of rural roads on upland farmers in sarawak, malaysia. *Asia Pacific Viewpoint*. 1997; 38:37–53.
14. Rudel TK, Richards S. Urbanization, roads, and rural population change in the ecuadorian andes. *Studies in Comparative International Development*. 1990; 25:73–89. [PubMed: 12283566]
15. Zelnor JL, Trostle J, Goldstick JE, Cevallos W, House JS, Eisenberg JNS. Social connectedness and disease transmission: social organization, cohesion, village context, and infection risk in rural ecuador. *American journal of public health*. 2012; 102:2233–9. [PubMed: 23078481]
16. Eisenberg JNS, Cevallos W, Ponce K, Levy K, Bates SJ, Scott JC, Hubbard A, Vieira N, Endara P, Espinel M, et al. Environmental change and infectious disease: how new roads affect the

- transmission of diarrheal pathogens in rural ecuador. *Proceedings of the National Academy of Sciences of the United States of America*. 2006; 103:19460–5. [PubMed: 17158216]
17. Vittor AY, Gilman RH, Tielsch J, Glass G, Shields T, Lozano WS, Pinedo-Cancino V, Patz JA. The effect of deforestation on the human-biting rate of *Anopheles darlingi*, the primary vector of falciparum malaria in the peruvian amazon. *American Journal of Tropical Medicine and Hygiene*. 2006; 74:3–11. [PubMed: 16407338]
 18. Dutta P, Khan S, Sharma C, Doloi P, Hazarika N, Mahanta J. Distribution of potential dengue vectors in townships along the national highways and trunk roads of northeast india. *Southeast Asian Journal of Tropical Medicine and Public Health*. 1998; 29:173–176. [PubMed: 9740295]
 19. Riley S. Large-scale spatial-transmission models of infectious disease. *Science (New York, N.Y.)*. 2007; 316:1298–1301.
 20. Crépey P, Barthélemy M. Detecting robust patterns in the spread of epidemics: A case study of influenza in the united states and france. *American Journal of Epidemiology*. 2007; 166:1244–1251. [PubMed: 17938424]
 21. Balcan D, Colizza V, Gonçalves B, Hu H, Ramasco JJ, Vespignani A. Multiscale mobility networks and the spatial spreading of infectious diseases. *Proceedings of the National Academy of Sciences of the United States of America*. 2009; 106:21484–9. [PubMed: 20018697]
 22. Gatto M, Mari L, Bertuzzo E, Casagrandi R, Righetto L, Rodriguez-Iturbe I, Rinaldo A. Spatially explicit conditions for waterborne pathogen invasion. *Am Nat*. 2013; 182:328–46. [PubMed: 23933724]
 23. Gatto M, Mari L, Bertuzzo E, Casagrandi R, Righetto L, Rodriguez-Iturbe I, Rinaldo A. Generalized reproduction numbers and the prediction of patterns in waterborne disease. *Proceedings of the National Academy of Sciences*. 2012; 109:19703–19708.
 24. Knipl D. A new approach for designing disease intervention strategies in metapopulation models. *J Biol Dyn*. 2016; 10:71–94. [PubMed: 26561360]
 25. Simini F, Gonzalez MC, Maritan A, Barabasi AL. A universal model for mobility and migration patterns. *Nature*. 2012; 484:96–100. [PubMed: 22367540]
 26. Gonzalez MC, Hidalgo CA, Barabási A-L. Understanding individual human mobility patterns. *Nature*. 2008; 453:779–782. [PubMed: 18528393]
 27. Lloyd-Smith JO, Schreiber SJ, Kopp PE, Getz WM. Superspreading and the effect of individual variation on disease emergence. *Nature*. 2005; 438:355–9. [PubMed: 16292310]
 28. Stein RA. Super-spreaders in infectious diseases. *International journal of infectious diseases : IJID : official publication of the International Society for Infectious Diseases*. 2011; 15:e510–3. [PubMed: 21737332]
 29. Bhavnani D, Goldstick JE, Cevallos W, Trueba G, Eisenberg JN. Synergistic effects between rotavirus and coinfecting pathogens on diarrheal disease: evidence from a community-based study in northwestern ecuador. *Am J Epidemiol*. 2012; 176:387–95. [PubMed: 22842722]
 30. Kotloff KL, Nataro JP, Blackwelder WC, Nasrin D, Farag TH, Panchalingam S, Wu Y, Sow SO, Sur D, Breiman RF, et al. Burden and aetiology of diarrhoeal disease in infants and young children in developing countries (the global enteric multicenter study, gems): a prospective, case-control study. *The Lancet*. 2013; 382:209–222.
 31. de Blasio BF, Kasymbekova K, Flem E. Dynamic model of rotavirus transmission and the impact of rotavirus vaccination in kyrgyzstan. *Vaccine*. 2010; 28:7923–32. [PubMed: 20933563]
 32. Widdowson MA, Steele D, Vojdani J, Wecker J, Parashar U. Global rotavirus surveillance: determining the need and measuring the impact of rotavirus vaccines. *J Infect Dis*. 2009; 200(Suppl 1):S1–8. [PubMed: 19817589]
 33. Hanley JA. Statistical analysis of correlated data using generalized estimating equations: An orientation. *American Journal of Epidemiology*. 2003; 157:364–375. [PubMed: 12578807]
 34. Pan W. Akaike's information criterion in generalized estimating equations. *Biometrics*. 2001; 57:120–125. [PubMed: 11252586]
 35. Dietz K. The estimation of the basic reproductive number for infectious diseases. *Statistical Methods in Medical Research*. 1993; 2:23–41. [PubMed: 8261248]

36. Gushulak BD, MacPherson DW. Globalization of infectious diseases: the impact of migration. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America*. 2004; 38:1742–8. [PubMed: 15227621]
37. Glass RI, Parashar UD, Bresee JS, Turcios R, Fischer TK, Widdowson M, Jiang B, R GJ. Rotavirus vaccines: current prospects and future challenges. *The Lancet*. 2006; 368:323–332.
38. Bishop RF, Barnes GL, Cipriani E, Lund JS. Clinical immunity after neonatal rotavirus infection. *New England Journal of Medicine*. 1983; 309:72–76. DOI: 10.1056/NEJM198307143090203 [PubMed: 6304516]
39. Montgomery MR. The urban transformation of the developing world. *Science*. 2008; 319:361–364.

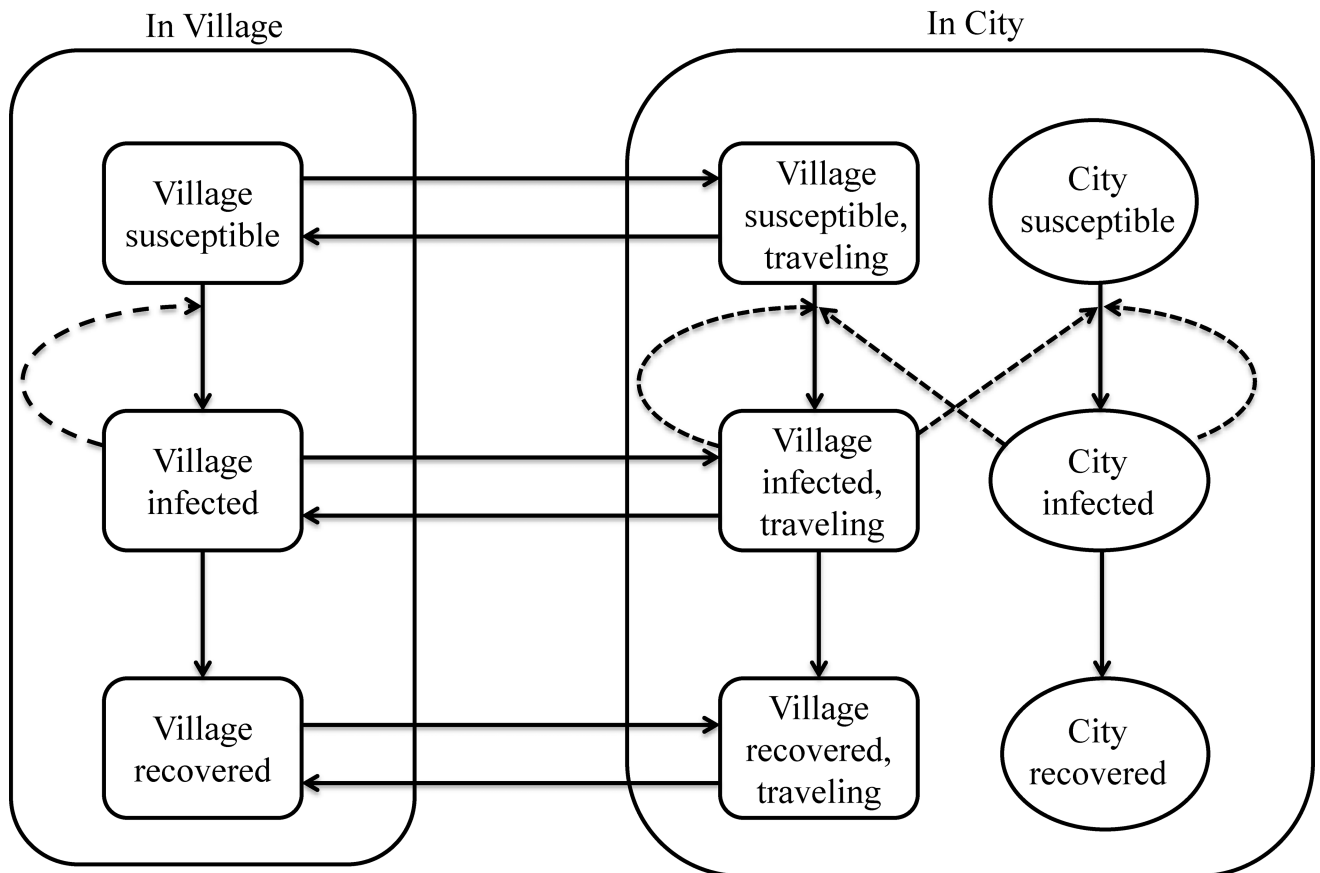


Figure 1.

Rotavirus transmission model diagram. Dashed arrows represent transmission events and solid arrows represent movement of people. Villagers and city residents are classified as susceptible (S), infected (I) and Recovered (R), with separate compartments for individuals under age 5 and ages 5 and older (not shown). Susceptible people can become infected by direct transmission within their own community. Villagers can infect and be infected by the city during travel. We ignore travel by city dwellers and between smaller villages. Village susceptible infectious recovered (SIR) model is stratified by baseline remoteness (close, medium and far) and both village and city SIR models are stratified by year (2004, 2007, 2010, and 2013) resulting in 12 models. Each model is stratified by age (less than or greater than 5 yrs.). This stratification results in 8 transmission parameters (6 village-level due to 2 age-groups and three community groups, and 2 city level due to two age groups), and 24 travel rate parameters to and from the city (2 age groups, 4 years, and 3 community types, where). Population size also varies by age, community type, and year. The recovery rate parameter is the same for all SIR models. See eAppendix for model equations and details.

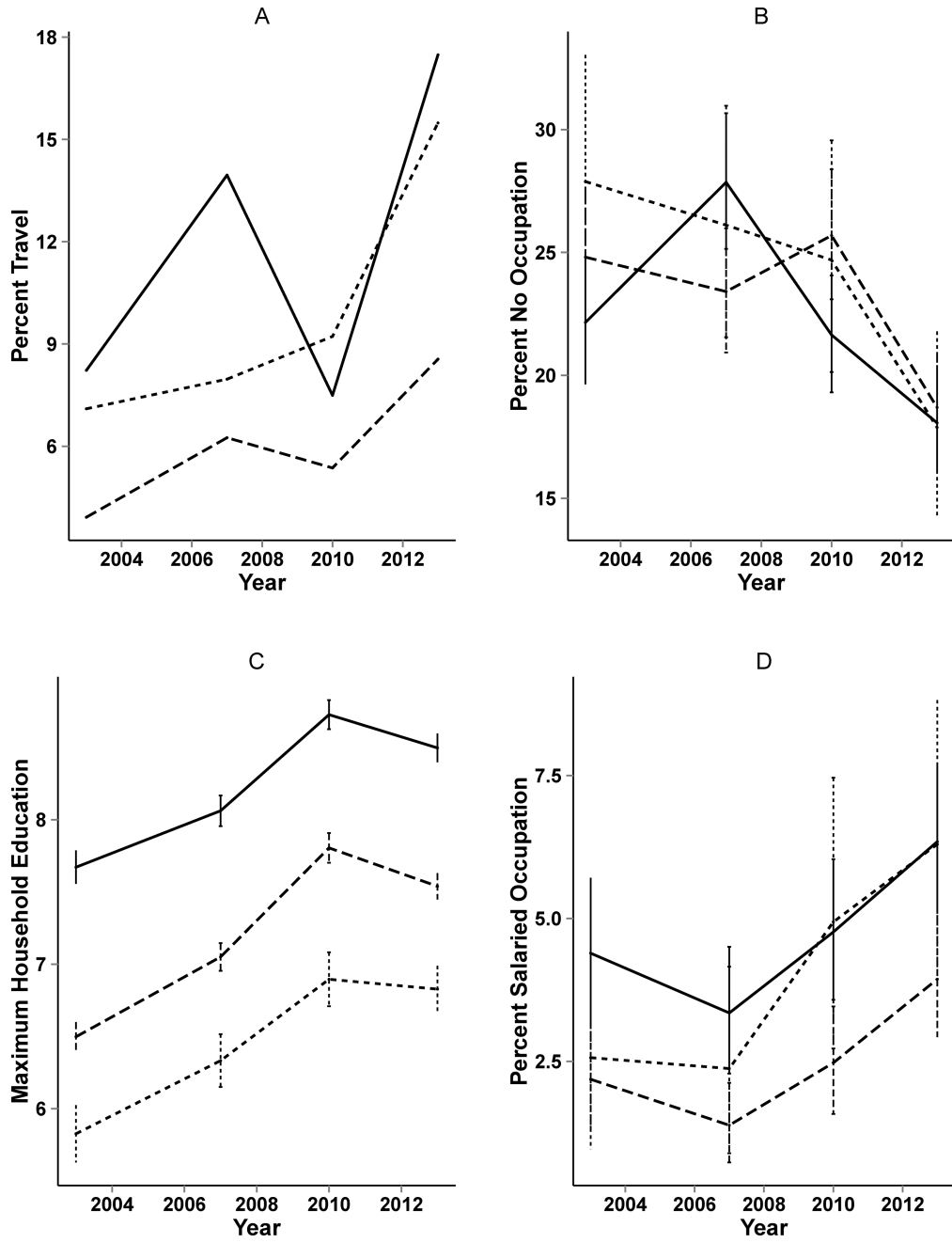


Figure 2. Demographic changes over time by study year and remoteness strata from the survey data. Close, medium, and far villages are shown as solid line, small dashed line, and longer dashed line respectively. A) Percent out-of-region travel (weighted by population size), B) Percent reporting no occupation, C) Maximum household education (years), D) Percent reporting salaried occupation.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

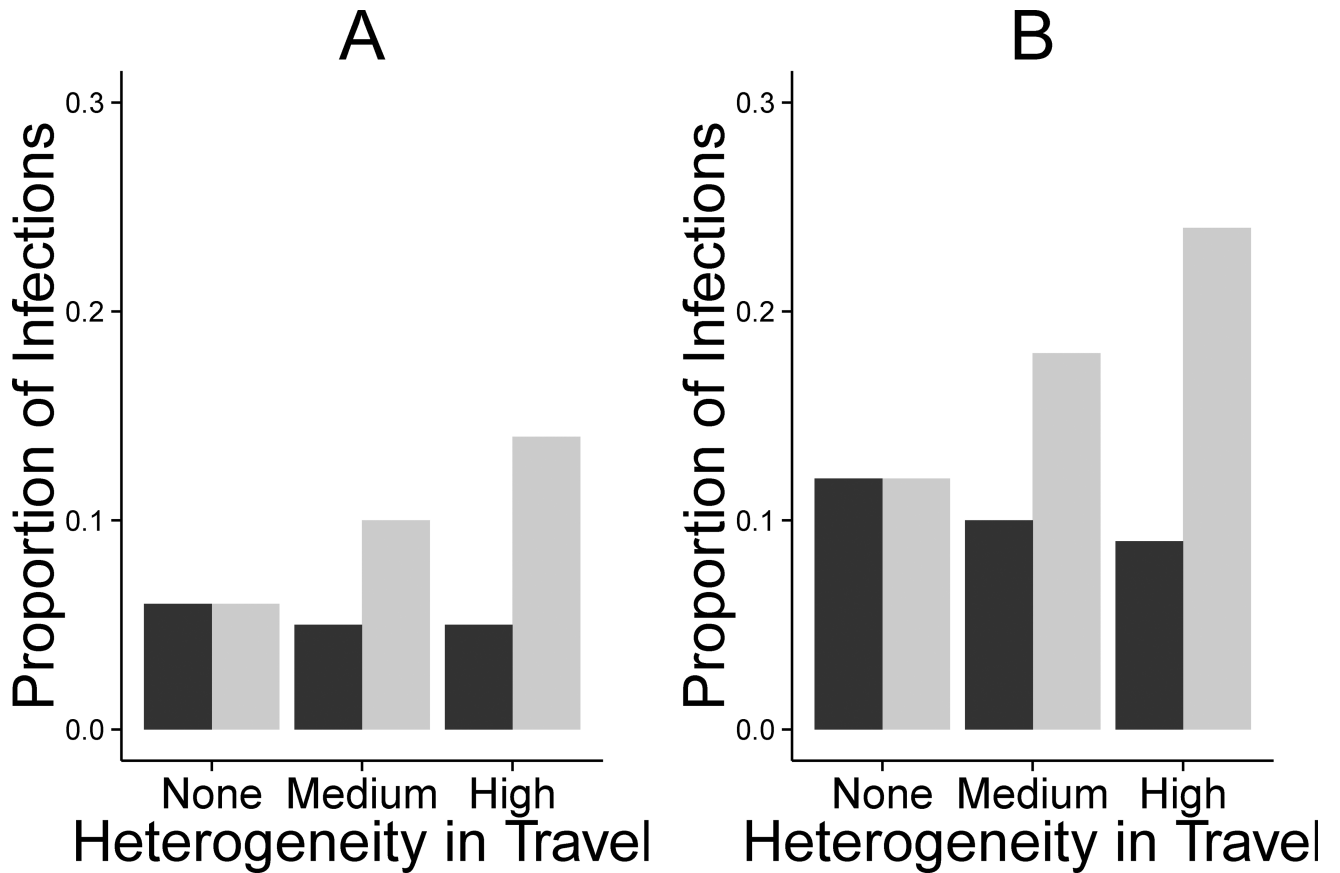


Figure 3.

Attributable fraction for local transmission by heterogeneity in travel (adults vs. children) from the transmission model for A) low within-village transmission (Baseline $R_0=0.79$) and B) high within-village transmission (Baseline $R_0=1.43$). R_0 increases slightly with increasing heterogeneity. Increasing heterogeneity was done using a proportionality constant where $travel_{adult} = c \cdot travel_{child}$ and c takes values of 1 (none), 2 (medium), or 3 (high); $c=1.73$ in the study data. Adults are shown in black and children are shown in gray.

Table 1

Sample characteristics by stratum of remoteness. For baseline values, continuous variables are presented as mean (standard deviation) and categorical variables are presented as percent (n). Time trend values represent the correlation between time and the covariate accounting for community level clustering using a GEE with an independence correlation structure.

	Remoteness: Close		Remoteness: Medium		Remoteness: Far	
	N=1004 ^a Baseline	N=4,748 ^a Time trend β (SE) ^c	N=309 ^a Baseline	N=1,440 ^a Time trend β (SE) ^c	N=927 ^a Baseline	N=4,534 ^a Time trend β (SE) ^c
Village Variables						
Cost/travel time metric	0.018	--	0.07	--	0.17	--
Mean duration (yrs) of residence; mean (SD) ^b	18.2(6.5)	0.215(0.096)	24.4 (3.0)	0.433(0.200)	28.1(3.4)	-0.037(0.187)
Community age (yrs); mean (SD) ^b	23.3(1.6)	0.085(0.056)	25.2(3.2)	0.434(0.278)	23.6(3.1)	0.223(0.088)
Individual and household variables						
Highest household education (yrs); mean (SD) ^b	7.0(3.8)	0.080(0.051)	5.6(3.6)	0.203(0.026)	6.2(3.2)	0.139(0.038)
Housing index; mean (SD) ^b	5.01(0.93)	0.019(0.007)	4.60(0.94)	0.039(0.031)	4.37(1.20)	0.034(0.031)
Duration of residence (z-score); mean (SD) ^b	-0.18(0.99)	0.001(0.006)	0.01(0.98)	-0.012(0.012)	0.13(0.93)	-0.008(0.005)
Age (yrs); mean (SD) ^b	37.9(18.4)	0.155(0.065)	41.8(20.4)	0.826(0.693)	38.2(18.8)	0.111(0.283)
Age; % (No.)						
<5	21(216)	-0.029(0.010)	21(70)	-0.045(0.085)	20(195)	-0.055(0.025)
5-13	26(261)	0.003(0.004)	23(74)	-0.115(0.057)	29(287)	0.027(0.019)
>13	53(540)	0.015(0.008)	56(184)	0.124(0.010)	52(514)	0.011(0.026)
Male; % (No.)	52(523)	-0.005(0.007)	51(165)	-0.005(0.009)	50(491)	-0.003(0.008)
Occupation; % (No.)						
None	22(223)	-0.022(0.008)	28.1(88)	-0.058(0.071)	25(230)	-0.042(0.018)
Domestic	17(169)	0.012(0.006)	18.5(58)	0.039(0.049)	19(174)	0.003(0.015)
Student	33(328)	0.012(0.012)	21(66)	-0.031(0.027)	31(291)	0.053(0.018)
Agriculture	20(199)	-0.022(0.011)	25(80)	0.025(0.040)	20(186)	-0.064(0.024)
Salaried	4(44)	0.048(0.014)	3(8)	0.191(0.040)	2(21)	0.065(0.034)
Other	4(41)	0.019(0.019)	3(9)	0.004(0.064)	3(25)	0.055(0.039)
Outside travel; % (No.)	8(83)	0.051(0.028)	7(23)	0.189(0.016)	4(38)	0.090(0.030)

\hat{p} : For Baseline values the total number of observations come from only individuals surveyed in 2003. For time trend analysis total number of observations come from individuals surveyed at all time points (2003, 2007, 2010, and 2013), including individuals with missing data on community of residence (3 people). These individuals are not included in Tables 2 or 3.

σ : SD is the abbreviation for standard deviation

c : SE is the abbreviation for standard error

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 2

Determinants of travel patterns. We used forward generalized estimating equation (GEE) regression with compound symmetry covariance structure, and all terms are odds ratios (95% CI). First, the unadjusted associations with each variable and the outcome were considered. To evaluate potential confounding by remoteness, each unadjusted association was adjusted for remoteness only. A series of models was used to investigate the relationships between the remaining variables. Model 1 contained only village-level variables. Model 2 was further adjusted for education. Model 3 further adjusted for occupation. Model 4 included all variables in model 3 as well as duration of residence and age.

	Unadjusted	Remoteness Adjusted	Model 1 Village	Model 2 Model 1+ Education	Model 3 Model 2+ Occupation	Model 4 Fully Adjusted
Village variables						
Remoteness	0.69(0.59,0.82)	0.69(0.59,0.82)	0.48(0.35, 0.66)	0.46(0.34, 0.61)	0.51(0.38,0.68)	0.51(0.38,0.67)
Mean duration of residence (yr)	1.06(1.02, 1.10)	1.06(1.02,1.10)	1.05(1.01,1.08)	1.05(1.02,1.08)	1.04(1.01,1.08)	1.04(1.01,1.07)
Secondary school availability (y/n)	1.47(1.11,1.95)	0.76(0.58,1.00)	0.61(0.40,0.94)	0.50(0.35,0.70)	0.56(0.40,0.78)	0.53(0.37,0.76)
Individual and household variables						
Highest household education (yr)	1.11(1.09,1.13)	1.10(1.08,1.13)	--	1.11(1.09,1.13)	1.10(1.08,1.13)	1.10(1.08,1.12)
Occupation			--	--		
None	1.00(Ref)	1.00(Ref)	--	--	1.00(Ref)	1.00(Ref)
Domestic vs. None	2.77(1.99,3.87)	2.70(1.96,3.73)	--	--	2.89(2.10,3.99)	1.64(1.14,2.36)
Agriculture vs. None	2.54(1.81,3.57)	2.47(1.78,3.43)	--	--	2.71(1.97,3.71)	1.50(1.13,1.98)
Salaried vs. None	7.82(5.27,11.58)	7.37(4.91,11.07)	--	--	5.69(3.95,8.20)	3.27(2.15,4.96)
Other vs. None	4.51(2.89,7.03)	4.36(2.78,6.84)	--	--	4.21(2.57,6.90)	2.43(1.51,3.92)
Duration of residence (yr)	0.99(0.90,1.10)	1.00(0.91,1.09)	--	--	--	1.04(0.95,1.13)
Age						
<5	1.00(Ref)	1.00(Ref)	--	--	--	1.00(Ref)
5-13	0.75(0.59,0.95)	0.77(0.62,0.95)	--	--	--	0.65(0.50,0.85)
>13	2.94(2.25,3.85)	2.83(2.15,3.73)	--	--	--	1.73(1.30,2.31)

Quasi-information criteria values for goodness of fit for the regression model by study year and characteristics considered. The lowest quasi information criterion (preferred model) for each year is in bold.

Table 3

	2003	2007	2010	2013	Overall
Full Model	815.42	1343.52	1233.10	2220.80	5626.71
Remoteness-only model	1074.89	1669.96	1331.05	2387.69	6537.63
Demographics model	829.78	1347.97	1252.63	2275.60	5815.40

Table 4

Percent difference in cumulative incidence predicted by the transmission model for adults, children, and the whole community for increasing heterogeneity of travel by age group. Increasing heterogeneity was done using a proportionality constant where $\text{travel}_{\text{adult}} = c \text{travel}_{\text{child}}$ and c takes values of 1 (low), 2 (medium), or 3 (high). Low-within village transmission corresponds to a baseline R_0 of 0.79 and high within village transmission corresponds to a baseline R_0 of 1.43 (R_0 increases slightly with increasing heterogeneity)

Travel heterogeneity	Adults	Children	Community
Low within-village transmission			
Medium vs. None	1.6%	0.2%	1.4%
High vs. None	3.1%	0.3%	2.8%
High within-village transmission			
Medium vs. None	1.7%	0.3%	1.4%
High vs. None	3.3%	0.6%	2.8%

Table 5

Predicted cumulative incidence per 1,000 people by remoteness and study year. Models were parameterized to reflect data collected in our field site over time and from the literature regarding the three disease processes: travel to the city (by community, age group, and study year), transmission rate (by community and age group, but not time), and recovery rate (estimated from literature at five days). Community population size and the fraction of children under 5 years of age were also estimated for each study year using survey data and were incorporated into our model. The population of the city was fixed at 5,000 for all study years. All parameter values are shown in the supplement along with relevant derivations.

Remoteness	Study Year			
	2003	2007	2010	2013
Close Villages	13	24	13	28
Medium Villages	12	16	16	31
Far Villages	7	10	10	14

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript