Original Contribution

Relating Diarrheal Disease to Social Networks and the Geographic Configuration of Communities in Rural Ecuador

Sarah J. Bates¹, James Trostle², William T. Cevallos³, Alan Hubbard¹, and Joseph N. S. Eisenberg⁴

¹ School of Public Health, University of California, Berkeley, CA.
² Trinity College, Hartford, CT.
³ Universidad San Francisco de Quito, Quito, Ecuador.
⁴ Department of Epidemiology, School of Public Health, University of Michigan, Ann Arbor, MI.

Received for publication December 21, 2006; accepted for publication May 22, 2007.

Social networks and geographic structures of communities are important predictors of infectious disease transmission. To examine their joint effects on diarrheal disease and how these effects might develop, the authors analyzed social network and geographic data from northern coastal Ecuador and examined associations with diarrhea prevalence. Between July 2003 and May 2005, 113 cases of diarrhea were identified in nine communities. Concurrently, sociometric surveys were conducted, and households were mapped with geographic information systems. Spatial distribution metrics of households within communities and of communities with respect to roads were developed that predict social network degree in casual contact (“contact”) and food-sharing (“food”) networks. The mean degree is 25–40% lower in communities with versus without road access and 66–94% lower in communities with lowest versus highest housing density. Associations with diarrheal disease were found for housing density (comparing dense with dispersed communities: risk ratio = 3.3, 95% confidence interval (CI): 1.1, 10.0) and social connectedness (comparing lowest with highest degree communities: risk ratio = 3.4, 95% CI: 1.1, 10.1 in the contact network and risk ratio = 4.9, 95% CI: 1.1, 21.9 in the food network). Some of these differences may be related to more new residents, lower housing density, and less social connectedness in road communities.

diarrhea; Ecuador; environment and public health; geography; social support

Abbreviation: CI, confidence interval.

Social networks and geographic structure of communities and households shape the transmission of enteric pathogens that cause diarrheal disease. Understanding the extent of overlap between geographic and social spaces can therefore help determine the role of each exposure within the transmission cycle. We present here social network and geographic data from nine villages in northern coastal Ecuador and examine their associations with diarrheal disease prevalence.

Geographic and social contact patterns have been studied concurrently in the epidemiology of human immunodeficiency virus (1), syphilis (2), and gonorrhea (3). These analyses, however, focus mainly on meeting places of sexual partners, concentrating on sexual networks and transmission, rather than on how geography shapes the networks. Klovdahl et al. (4) present social networks defined by leisure locations in a study of tuberculosis, specifically including geographic locations as actors in social networks. This
approach acknowledges that geography plays a role in not only transmitting pathogens via social networks but also shaping the social networks themselves.

The geographic location of susceptible individuals is critical when determining exposure to pathogens (5) and chemical agents (6), as well as access to health care (7) or clean water (8). Two equally important geographic scales are the distribution of communities with respect to regional-level factors, such as access to roads and transportation networks (9–12), and the distribution of households within communities (6). In the study of enteric pathogens, the environment acts as both source and target of contamination. The geography of these environmental features is important in defining the transmission dynamics of diarrheal disease. Social processes also play a major role in how water is accessed, stored, and used; how food is obtained, distributed, and prepared; and how household dynamics affect hygiene and sanitation behavior. Social networks have been used to study transmission of pathogens via sexual (13–16), respiratory (4, 17), and blood-borne routes (18, 19). These approaches focus primarily on specifically defined social contacts and on pathogens with single-pathway transmission cycles. In contrast, enteric pathogens are transmitted through casual person-to-person contact, as well as via contaminated food, water, and fomites, emphasizing the importance of diverse social contacts.

Studying diarrheal disease patterns requires understanding broadly defined social networks and considering environmental exposures and contamination. We recognize the importance of jointly addressing geographic and social contacts when investigating enteric pathogen transmission. The following analysis quantifies the extent to which geography influences social network structure and formation and the contribution of geographic and social factors to diarrheal disease in rural Ecuador.

MATERIALS AND METHODS

Study area

Our study was situated in Esmeraldas, the northernmost province on the Ecuadorian coast, and was conducted in conjunction with a larger study of diarrheal disease that involves 21 communities in the region (10). Details on the region can be found elsewhere (20–24). In 1996, the Ecuadorian government’s effort to join this northern region with the coast and the Andes principally involved a two-lane asphalt highway, which was completed in 2001. Secondary roads continue to be built, linking the region’s communities to urban centers.

Each of the 21 study communities was visited three times between July 2003 and May 2005. During each visit, a 15-day case-control study was conducted in which all cases of self-reported diarrhea were identified by the field staff, based on a case definition of three or more loose stools in the prior 24 hours. The present analysis focuses on data from nine of the 21 communities, where sociometric interviews were conducted during 2003. Within the larger study, demographic data were collected on all individuals in 2003, including age, gender, race, years of education, and years lived in the community (table 1).

Social network data

All individuals older than 12 years were sought for interview regarding their social interactions. Information about younger participants was obtained from parents or caretakers. The survey design was therefore sociometric, involving all available network members, rather than egocentric, involving a targeted subset of the population. The characteristics measured were expressive, defined by various communications between individuals, or activity specific, defined by participation in a particular activity that increases the risk of exposure to, in this case, enteric pathogens. To measure social interactions between households within each community, interviewees were asked to name people outside their household but in their community with whom they had participated in various activities.

For this analysis, the casual contact (“contact”) and food-sharing (“food”) networks were chosen from seven that were elicited. The contact network (“In general, with whom do you spend time?”) describes loosely defined expressive behaviors involving casual contact between individuals. Of the activity-specific networks, the food network (“In the past week, with whom did you share food?”) was the densest and relates directly to the transmission potential of enteric pathogens.

The network measure, degree, is the number of direct (one-step) connections that a participant has in a network; we use degree here because the number of contacts is likely to be significant in disease transmission potential. Networks were analyzed for degree rather than normalized degree, a measure of connectedness normalized by the total number of connections available. Average household degree is uncorrelated with the size of communities ($R^2 = 0.01$ and 0.00 for contact and food networks, respectively; data not shown); the use of normalized degree would introduce correlation with community size.

The contact and food networks reported here are nondirected, meaning all connections are assumed to be two way; an individual’s degree includes all self-reported connections as well as connections to that individual named by others.

Because global positioning system data were collected for each house, degree is collapsed at the level of the household. Connections between houses are binary and do not reflect multiple relations between multiple actors in two households.

Characterizing geographic space

Latitude and longitude were collected for all houses using a Trimble GeoExplorer 3 mobile unit and organized with Trimble GPS Pathfinder (Trimble Navigation, Limited, Sunnyvale, California) 2002 and ArcGIS (ESRI, Inc., Redlands, California) 2001 software packages. Approximately 60 points were collected and averaged for each house over a period of 1 minute. Most positions were accurate within 3 m. The distance between houses was calculated by use of latitude and longitude data.

At the regional level, communities were classified as road accessible or nonaccessible.
Within communities, two methods were used to characterize geographic space. The first was an index of the spatial proximity (called “spatial index” throughout) of each house, defined as the harmonic mean of distances between house $i$ and all other houses in a community. The spatial index, $SI_i$, for house $i$, was computed as follows:

$$SI_i = \log\frac{1}{n-1} \sum_{j=1}^{n-1} \frac{1}{D_{ij}},$$

where $n$ is the number of houses in the community, and $D_{ij}$ is the distance between houses $i$ and $j$. The harmonic mean prevented a few large distances from exerting a disproportionately large influence. Houses with large indices of spatial proximity had, on average, greater distances to all other houses than those with low spatial indices. We define the mean of this index as a measure of housing density within a community. Because study communities had ambiguous boundaries and many continue to grow outward, we argue that spatial proximity is a good surrogate measure for housing density.

The second method to characterize geographic space was a pairwise distance measure. Geographic distance was plotted against the proportion of houses that shared a connection at that distance. This relation was determined using the locally weighted regression scatterplot smooth function, which smooths binomial data by performing least-squares regressions at each observation.

### Data handling and statistical analysis

Crude and multivariate linear models were used to examine the associations between diarrheal disease prevalence and both housing density (spatial index) and social connectedness (network degree). Data were analyzed with STATA, version 8.0, software (StataCorp LP, College Station, Texas) and UCINET, version 6, a program to analyze sociometric data (25). Graphs were produced with STATA software.

### Human subjects approval

Study protocols were approved by institutional review boards at the University of California, Berkeley; Trinity College; and Universidad San Francisco de Quito. Participants or parents of minor participants gave verbal, informed consent prior to enrollment in the study and collection of data.

### RESULTS

#### Social networks

A total of 2,053 individuals were interviewed in nine communities. The response rate was 81 percent (2,053 of 2,540 study participants); this response rate is an underestimate due to migration not reflected in census records. Those who were missed for interviews were 1.5 times (95 percent confidence interval (CI): 1.1, 2.2) more likely to be minorities and to have had 1.7 additional years of education (95 percent CI: 1.2, 2.2) (table 1). This disparity in education is likely due to the higher educational level of individuals who temporarily left the region. Age, gender, and years in the community did not affect the likelihood of being interviewed.

The mean network degree of households in each community ranged from 4.8 to 18.7 for the contact network and from 0.4 to 10.4 for the food network. The mean degrees in the contact and food networks of a community were highly correlated (adjusted $R^2 = 0.88$), although the degree was
2.2 times greater in the contact network than in the food network. Households in communities with road access had, on average, 3.8 (95 percent CI: 2.2, 5.3) fewer connections in the contact network and 3.0 (95 percent CI: 2.3, 3.8) fewer connections in the food network than did households in communities without road access (table 2).

**Geographic characteristics**

Within the study region, the mean of the spatial index was 4.3 (range: 2.6–6.7), indicating that houses were located an average of 75 (range: 14–774) m from all other houses in the community. The nine communities fell into one of two road categories (table 2). Houses in road communities (6–9) had a mean distance to all other houses of 110 m (mean spatial index = 4.7), and houses in roadless communities (1–5) had a mean distance to all other houses of 38 m (pooled mean spatial index = 3.6). Road communities comprised houses that were, on average, 72 m (95 percent CI: 65, 79) farther apart from each other than in roadless communities.

Community size also varied depending on road access (mean geographic boundaries = 1,397 m for road and 205 m for roadless communities), as did residence duration (households in roadless communities had been part of the communities for an average of 11.0 years (95 percent CI: 1.9, 28.0) longer than those in road communities) (table 2).

Within communities, houses were geographically arranged according to the household members’ years of residence. Communities 5, 7, 8, and 9 were characterized by a negative correlation between the duration of household residence and the spatial index of the house. In these communities, newer residents tended to live on the periphery; that is, the community grew outward. Communities 2 and 3 exhibited positive correlations between spatial index and duration of household residence. In contrast, newcomers in these communities tended to live in the center of a community. Communities 1, 4, and 6 did not exhibit strong correlations ($R^2 < 0.10$) between spatial index and duration of household residence (data not shown).

The mean distance between households (mean spatial index) was strongly negatively correlated with the mean number of social contacts between households. Each unit increase in the spatial index denotes an order-of-magnitude increase in the mean distance between households and was associated with a mean drop of 3.7 contacts (95 percent CI: 0.4, 7.0) in the contact network and 2.7 contacts (95 percent CI: 0.9, 4.5) in the food network (figure 1). Communities with lower housing density had fewer connections in their social networks. The relation between housing density and social network degree did not change after controlling for the number of households or number of occupants in a community.

Just as housing density was related to mean social connectedness and the age of the community, geographic distance between two specific households predicted the probability of a network connection between them. The probability of

<table>
<thead>
<tr>
<th>Community</th>
<th>No. of houses</th>
<th>Contact network size</th>
<th>Food network size</th>
<th>Residence duration (years)*</th>
<th>Contact degree†</th>
<th>Food degree†</th>
<th>Spatial index‡</th>
<th>Geographic boundaries (m)§</th>
</tr>
</thead>
<tbody>
<tr>
<td>Roadless communities</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>29</td>
<td>192</td>
<td>105</td>
<td>37.3</td>
<td>13.2</td>
<td>7.2</td>
<td>3.4</td>
<td>172</td>
</tr>
<tr>
<td>2</td>
<td>32</td>
<td>229</td>
<td>109</td>
<td>33.0</td>
<td>14.3</td>
<td>6.8</td>
<td>3.6</td>
<td>235</td>
</tr>
<tr>
<td>3</td>
<td>21</td>
<td>84</td>
<td>35</td>
<td>35.7</td>
<td>9.3</td>
<td>3.9</td>
<td>3.7</td>
<td>241</td>
</tr>
<tr>
<td>4</td>
<td>49</td>
<td>457</td>
<td>255</td>
<td>37.8</td>
<td>18.7</td>
<td>10.4</td>
<td>3.7</td>
<td>179</td>
</tr>
<tr>
<td>5</td>
<td>69</td>
<td>504</td>
<td>234</td>
<td>46.5</td>
<td>14.8</td>
<td>6.9</td>
<td>3.8</td>
<td>198</td>
</tr>
<tr>
<td>Road communities</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>109</td>
<td>716</td>
<td>263</td>
<td>33.5</td>
<td>13.4</td>
<td>4.9</td>
<td>4.6</td>
<td>868</td>
</tr>
<tr>
<td>7</td>
<td>64</td>
<td>170</td>
<td>90</td>
<td>17.9</td>
<td>5.5</td>
<td>2.9</td>
<td>4.7</td>
<td>1,261</td>
</tr>
<tr>
<td>8</td>
<td>171</td>
<td>1,020</td>
<td>420</td>
<td>31.0</td>
<td>12.3</td>
<td>5.0</td>
<td>4.7</td>
<td>902</td>
</tr>
<tr>
<td>9</td>
<td>11</td>
<td>24</td>
<td>2</td>
<td>9.9</td>
<td>4.8</td>
<td>0.4</td>
<td>6.0</td>
<td>2,557</td>
</tr>
<tr>
<td>Pooled</td>
<td>555</td>
<td>32.7</td>
<td>12.5</td>
<td>5.6</td>
<td>4.3</td>
<td>735</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Roadless</td>
<td>355</td>
<td>39.7</td>
<td>14.9</td>
<td>7.5</td>
<td>3.6</td>
<td>205</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Road</td>
<td>200</td>
<td>28.7</td>
<td>11.2</td>
<td>4.5</td>
<td>4.7</td>
<td>1,397</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*p value (t-test).
Given that degree is negatively associated with both spatial index and both measures of degree seen in figure 1. Clusters had a higher probability of social connection. and nonlinear networks, houses separated by smaller distance dependent. (In the multivariate model, regression coefficients for both contact degree and spatial index increased more than threefold with the crude model, regression coefficients for both contact degree and spatial index were significant for both covariates. Compared to the crude model, the multivariate model including both social network degree and spatial index showed no association with disease prevalence. The analogous case holds for spatial index. The food and contact networks displayed similar relations with disease prevalence. There was no evidence of confounding by other demographic (community population size, educational level, mean residence duration, or socioeconomic status) or risk factor (water source or sanitation level) variables.

The multivariate log-linear model predicts diarrheal incidence approximately 3.3 times (95 percent CI: 1.1, 10.0) greater in the highest housing density than in the lowest housing density communities, when holding the mean degree constant. Similarly, diarrheal incidence is approximately 3.4 times (95 percent CI: 1.1, 10.1) greater in the highest mean contact degree than in the lowest mean contact degree communities, when holding the mean spatial index constant. Diarrheal incidence is 4.9 times (95 percent CI: 1.1, 21.9) greater in the high mean food degree than in the low mean food degree communities (figure 3).

**DISCUSSION**

Diarrheal disease is transmitted through multiple environmental pathways, including person-to-person contact and exposure to contaminated water, food, and fomites. Numerous observational studies and intervention trials attest to the strong association between environmental risk factors and diarrheal disease. These risk factor studies, however, do not capture how pathogens flow through individuals and communities. For example, diarrheal disease pathogens can move between households because of geographic proximity or social ties.

In our analysis of nine communities in rural Ecuador, social network and geographic measures predict diarrheal disease prevalence at the community level. The two measures are highly correlated, and both are strongly associated with diarrheal disease prevalence. The relation between social and physical space holds when using the spatial index, a measure of housing density based upon the mean distance to all other households in the community (figure 1), or a pairwise distance metric, in which the unit of analysis is the distance between two households (figure 2).

Negative associations between physical and social space metrics are evidence that both increased crowding and decreased social connectedness are risk factors for diarrheal disease (figure 3). The association between increased crowding and diarrheal disease is well documented in the literature (26, 27). Less well documented is the relation between social measures of connectedness and diarrheal disease risk. The associations observed in this study suggest that the social networks we measured have a positive effect on disease risk; that is, increased connectedness decreases disease risk. This finding suggests that being socially connected in our study is a measure of social cohesion rather than of pathogen transmission. This result contrasts with the sexually transmitted disease literature showing that more network connections increase transmission (13, 28, 29).

**FIGURE 1.** Mean spatial index versus mean network degree in the community for contact (filled circles and solid line) and food (unfilled circles and dashed line) networks, Ecuador, 2003–2005. There were nine communities. The mean spatial index is the geometric mean distance of one household to all others in the community, and the mean network degree is the average number of community members nominated by the interviewee. Refer to Materials and Methods for more details. Data points under “Roadless communities” represent those communities without road access; those under “Road communities” represent communities that reside on the road.

a social connection between any two houses was negatively correlated with the distance between them, although the exact nature differed by network (figure 2). Four of these relations were linear (communities 1, 2, and 4 in the contact network and community 3 in the food network); the majority, however, attenuated nonlinearly (communities 5, 6, 7, and 8 in the contact network and communities 1, 2, 4, 5, 6, 7, and 8 in the food network). That is, the polynomial test for second-order terms is significant ($p < 0.01$). For these linear and nonlinear networks, houses separated by smaller distances had a higher probability of social connection.

**Disease outcomes**

Field staff identified 113 cases of diarrheal disease during three 15-day case-control studies in the nine communities. The identified prevalence of all-cause diarrheal disease, $D$, was 0.45 cases/person-year. Diarrheal incidence is highly age dependent. ($D = 1.9$ cases/person-year for <5 years of age and 0.16 cases/person-year for ≥5 years of age.)

Crude linear regression models including either social network degree or spatial index showed no association with diarrheal disease prevalence in these communities. A multivariate model including both social network degree and spatial index was significant for both covariates. Compared with the crude model, regression coefficients for both contact degree and spatial index increased more than threefold in the multivariate model ($-0.0018$ vs. $-0.0005$ and $-0.0093$ vs. $-0.0028$, respectively), evidence of confounding. Further evidence is the negative correlation between spatial index and both measures of degree seen in figure 1. Given that degree is negatively associated with both spatial index and diarrhea, an increase in degree is associated with both a decrease in disease and a decrease in spatial index. This decrease in spatial index, however, is also associated with an increase in disease, resulting in confounding toward the null in the unadjusted model of degree versus diarrhea rates. The analogous case holds for spatial index. The food and contact networks displayed similar relations with disease prevalence. There was no evidence of confounding by other demographic (community population size, educational level, mean residence duration, or socioeconomic status) or risk factor (water source or sanitation level) variables.
Spatial metrics are derived from global positioning system readings of latitude and longitude and require minimal assumptions. Social network and disease metrics, however, have important assumptions that should be made explicit. Social network data were collected in a single survey and, therefore, lack variability information. Previous studies, however, have suggested that, although individual connections are not stable, overall network properties such as those presented here are stable over time (30). Because of the case-control nature of the study, community prevalence estimates require that all cases of diarrhea be identified. We addressed this concern by conducting ethnographic interviews to integrate local definitions of diarrheal disease and by using local field staff to identify cases more accurately.

The particularities of our observed relations between social and geographic space can be discussed in terms of the effects of geographic configuration within communities on social networks, the effects of road access on social networks and on the geographic configuration of houses within communities, and the joint effects of geographic configurations and social networks on transmission dynamics.

**Effects of geographic configuration within communities on social networks**

Our analysis supports the concept that the physical space in which social networks occur has a crucial role in network structure and formation. Geographically dispersed communities, those with a high mean spatial index, contain social networks with lower network degree. High dispersion of households contributes to low social network degree simply by increasing the effort required to create and maintain social connections.

The formation of specific contacts in social networks is shaped by geography as well as by the overall network structure. Geographic distance between households predicts the probability of social connection in both the contact and the food networks (figure 2). This variation likely depends on community contexts; differences may be rooted in topographic factors, remoteness, overall social cohesion, land ownership, inheritance or residence rules, or other factors beyond the scope of this analysis. A similar trend was observed by Rothenberg et al. (31), where socially connected pairs in sexual and injection networks were, on average,
Effects of road access on social networks and geographic configuration within communities

The recently constructed road contributes to observed variation in social networks and geographic configurations by promoting differential growth patterns in the region. Communities with road access are more geographically dispersed, have shorter duration of household residence, and are less socially connected. Within these communities, housing density predicts the social connectedness of a community, and geographic distance between households predicts the probability of specific social connections, suggesting the importance of including both when examining enteric pathogen transmission.

The average residence duration illustrates a general picture of regional growth patterns with respect to roads. Two of the roadless study communities (communities 2 and 3) exhibit the expected positive correlation between spatial index and residence duration, indicating that newcomers enter the interior of the community. This growth pattern contributes to a higher mean spatial index.

The lower social connectedness of road villages could be explained by the age of the community; that is, newly introduced individuals have fewer social connections. As a community becomes more established and its members have longer residence duration, social connections increase. Our data show, however, that both overall social network structure (figure 1) and connections between individual actors (figure 2) are more strongly associated with the geographic configuration of houses in a community than they are with the residence duration.

Transmission dynamics

The significance of geography in relation to social networks has been recognized in sexually transmitted disease research (1–3, 31) and in tuberculosis transmission (4). In contrast to sexually transmitted infections, the study of enteric pathogen transmission within populations requires an accurate estimate of the overall probability of contact between individuals, rather than detailed contact tracing of particular individuals.

Social and geographic processes affect the transmission dynamics of various pathogens to different extents. For example, pathogens such as *Vibrio cholera*, which are transmitted primarily via the contaminated environment, depend largely on geographic configurations (32). Conversely, pathogens such as *Shigella* spp., which do not have long survival times in the environment (33), rely more heavily on person-to-person transmission, and infection patterns

FIGURE 3. Point estimates of diarrheal incidence as predicted by a multivariate model including both geographic and social factors, Ecuador, 2003–2005. Estimates are shown for the lowest and highest observed spatial indices, contact degrees, and food degrees. Risk ratios (RRs) are ratios between the lowest and the highest as predicted by a log-linear multivariate model. Numbers in parentheses, 95% confidence interval.
will reflect the social patterns in a given population. Most enteric pathogens, however, have significant components of both person-to-person and environmental transmission and reflect both social and geographic patterns of populations. How best to target interventions should therefore be informed by the specific relation between social and geographic configurations (whether a complex joint effect or one approaching collinearity), as well as by the transmission patterns of pathogens.

ACKNOWLEDGMENTS

This study was supported by grant RO1-AI050038 from the National Institute of Allergy and Infectious Diseases. The authors of this paper would like to thank the Ecologia, Desarrollo, Salud, y Sociedad (EcoDeSS) field team for their invaluable contribution collecting the data. Conflict of interest: none declared.

REFERENCES