The Science of Human Connection: A Study of the Effect of Social Networks on Acute Gastrointestinal Illness in Rural Ecuadorian Communities

by

Sonia Tara Hegde

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Epidemiological Science)
in The University of Michigan
2018

Doctoral Committee:

Professor Joseph N. Eisenberg, Chair
Associate Professor Marisa C. Eisenberg
Professor Bhramar Mukherjee
Professor James A. Trostle, Trinity College

Sonia Tara Hegde

hegdes@umich.edu

ORCID iD: 0000-0001-6426-0096

Acknowledgements

I would first like to thank my dissertation committee, Drs. Joseph Eisenberg, Marisa Eisenberg, Bhramar Mukherjee, and James Trostle. I thank all of them for continually challenging my thinking and providing me a good theoretical and methodological foundation. Under their guidance, I've grown as an independent scientist, feeling more assured in my ability to defend my ideas and explore new possibilities. I would also like to thank Dr. Philippa Clarke for her relentless support and encouragement to seek deeper truths about the data and delve into social epidemiology. I am incredibly grateful for the mentors I've had at the University of Michigan.

Second, I would like to thank my family. I am the daughter of two immigrants, who came to this country in search of a better life where the pursuit of knowledge, access to healthcare and basic necessities was considered a fundamental right. I am the niece of nine women who were not permitted to go to school beyond the 7th grade, who were married at the age of 16, and some of whom contracted polio and tuberculosis but did not have access to healthcare. I am the granddaughter of two women who were not educated beyond the 2nd grade, who were married at the age of 14, and who were told not to have opinions beyond needing to birth and raise children. And I am the sister to two incredible women who have shown me what strength of character, resilience, and education can achieve in the face of adversity.

I dedicate this dissertation to all the women who came before me who made this degree possible – my mother, aunts, grandmothers, great grandmothers, and my sisters. I dedicate this dissertation to the women in the study for sharing their stories with me so openly and to the women that came before them. I dedicate this dissertation to the band of women I call dear friends without whom this degree would not have been possible. And I dedicate this dissertation to the young girls that will come after me. To all of you, your unwavering encouragement and support gives me the strong foundation on which I stand and your hopes and dreams for a brighter future gives me the motivation to look ahead.

An Urdu poet, Muhammad Iqbal, wrote,

“Tu shaheen hai kar baser phado ki chattano main. Parvwaaz hai kaam tera. Tere saamne aasmaan aur bhi hain."

“You are an eagle, living in the high mountains. Your passion is flight. You have more skies ahead of you yet to soar.”

I will always believe our best days have yet to come as a society and in the power of education. Thank you for this opportunity.

Table of Contents

List of Tables

List of Figures

List of Appendices

Abstract

Background

Diarrheal disease is an important cause of childhood mortality and is spread by two main mechanisms: human contact and contamination of the environment. Though individual- and household-level Water, Sanitation, and Hygiene (WASH) interventions are primarily used to intersect these transmission pathways, seldom are community-level factors considered to ensure both intervention adoption and sustainability. Social constructs like social cohesion are believed to influence the quality and effectiveness of interventions, especially those based on action at the community-level. Few studies, however, identify a causal framework for how social constructs impact WASH interventions and diarrheal disease occurrence, and fewer use social network data. Previous studies in coastal Ecuador showed diarrheal disease spreads more slowly to and in rural villages that have a greater density of social ties, suggesting a greater spread of individual and collective water practices that help reduce transmission of diarrheal disease.

Objective

This dissertation research aims to extend previous work by methodically defining social cohesion as an important social construct using different types of social network data, examining temporal variability of the effect of social cohesion on diarrheal disease, whether this relationship is mediated by WASH, and the role that gender plays in social cohesion and WASH in rural, coastal Ecuador.

Methods

Using longitudinal sociometric data from villages in rural, coastal Ecuador, we identify important network determinants of social cohesion and in turn the temporal effect of social cohesion on WASH interventions and diarrheal disease incidence. We use statistics for the analysis of network graph data and a novel two-stage Bayesian hierarchical model. We importantly theorize a causal framework for the observed phenomena through use of qualitative methods.

Results

Different types of social networks illustrate the multidimensionality of social processes at the household- and community-levels that influence diarrheal disease incidence. While a network comprised of individuals who pass time together becomes a stronger measure of risk over time, due to density of people and increased travel, having a network of core discussants with whom to discuss important matters is a consistent measure of protection. Having a strong community network of core discussants results in 0.87 (0.71, 1.06) fewer odds of diarrheal disease in 2007 and 0.34 (0.26, 0.45) fewer odds of diarrheal disease by 2013. This protective effect is partially mediated by WASH related factors like community sanitation and improved water use over time, suggesting the importance of social constructs at the community-level for intervention implementation and in turn the reduction of diarrheal disease. Qualitative data collected in the same communities, however, revealed the contributions of infrastructural development and an increasing wage economy to the increasing importance of community. Qualitative data also revealed the importance of gender equity for both community social cohesion and adoption of WASH practices. Analysis of social network data shows communities that are more assortative by gender (i.e. that have less gender equity) are less likely to engage in WASH practices at the household-level over time.

Significance

By understanding how community correlates of social networks affect intervention practices and diarrheal disease transmission, we can leverage social networks to influence positive behavior change and WASH infrastructure. This research objective is in line with target 5 and 6b of the United Nations Sustainable Development Goals, which aim to achieve gender equality and support and strengthen participation of local communities in improving WASH.

Chapter I

Background

"We don't accomplish anything in this world alone…whatever happens is the result of the whole tapestry of one's life and all the weavings of individual threads from one to another that creates something." - Sandra Day O'Connnor, the first woman on the United States Supreme Court.

Fundamental to the human experience is human connection. In 1933, Dr. J. L. Moreno displayed the first sociogram at a meeting of the Medical Society of the state of New York. He examined the effects of social networks on disease propagation. He was so enthralled with his work, he took his study of social connection to a classroom of fourth graders to study attraction between boys and girls. Of course, this was prior to the days of parental consent. Nevertheless, his research endeavor was met with enthusiasm and flourished.

Social influences play a critical role in public health [1,2], and have been well studied in the field of social epidemiology. Such factors include occupation, gender dynamics, infrastructural environment, and social connectivity measured by social networks. The application of social networks, in particular, has grown in number in the last decade.[3] In 1979 Berkman and Syme [4] demonstrated those with weaker social ties had significantly higher mortality rates. A robust body of literature across disciplines built on this early result has since indicated that being socially integrated or feeling socially connected reduces mortality and various disease morbidities in the U.S. compared to other leading health indicators.[2] Social connectedness is a construct that represents the different ways an individual can connect to others socially: physical, behavioral, social-cognitive, and emotional.[2] We can measure social connectedness through both social networks and other self-report measures like trust and participation in institutional organization. Such social constructs importantly have a hypothesized role in information spread of interventions for the prevention of different infections, including diarrhea.

Diarrheal infections result in 700,000 deaths in children under five years of age annually, with 72% of deaths occurring in the first two years of life.[5] Despite dramatic reductions in childhood mortality in the past decade, diarrhea remains a major cause of preventable childhood deaths worldwide.[6,7] In low-income settings, diarrheal disease transmission occurs through multiple pathways, largely human contact and the contaminated environment, including water, food, sanitation, and lack of hygiene.[8] Aside from vaccination, well known measures of diarrheal disease prevention include implementation of safe water, sanitation, and hygiene (WASH) practices.[8]

Adoption of and adherence to these behavioral practices, however, are influenced by a multitude of factors, including seldom considered societal level constructs.[9,10] While interventions are often administered at both the individual- and household-levels, the community plays a critical role in ensuring intervention dissemination and sustainability.[11] Indeed, though community effects have been implicated in the causation of diarrhea [12], few studies have shown the protective effects of community on diarrheal disease occurrence. One such study conducted in rural, coastal Ecuador, demonstrated that more remote communities have more dense social networks and less diarrheal disease, which was mediated by improved water quality and sanitation practices. [13] These insights though were not methodically incorporated into a causal framework or mechanistic model, and the study neither investigated temporal trends of this relationship nor sought to further explore the construct of social cohesion.

1.1 Interventions

Early intervention trials dating back to the 1970's assumed water quality as the critical source of diarrheal infection and primarily investigated the effect of expanding public water services for reducing diarrheal disease incidence.[14] This single intervention mantra continued through the next decade when studies either focused on water supply [15,16] or sanitation alone [17], a result of the United Nations designating the decade as the "International Drinking Water Supply and Sanitation Decade" to bring attention and support for clean water and sanitation worldwide.[18] Though water systems were implemented at the community-level, few studies highlighted the utility of community engagement in intervention implementation.[19]

By the 1990's, hygiene interventions like hand-washing emerged as an important tool for diarrheal disease reduction and studies began to investigate intervening on multiple transmission pathways. [20–23] There was a shift from focusing on public interventions to private (household- and individual-level).[24] In the following decade, household water interventions, like household water treatment (HWT) that focused on contamination between water supply and point-of-use, escalated.[25] Studies that investigated the use of more than one intervention, such as hand-washing and HWT, demonstrated that implementation of more than one intervention in the household had no greater benefit than only implementing one [26], except when environments contain many sources of contamination.[8,27]

In this evolution of diarrheal disease interventions, however, randomized trials followed and the issue of individual-level compliance surfaced.[28–30] In order for an intervention to be sustained, individual compliance is needed, but studies only examined the efficacy and not effectiveness of interventions.[31] While investigations should focus on the long-term sustainability of WASH interventions [32–34], by researching both issues of intervention compliance and the relative applications of compliance and sustainability to each intervention, a synthesis of the literature suggests community-level infrastructure should be re-examined as a necessary component of intervention effectiveness. Understanding community-level structures and how to leverage them could help overcome issues of compliance and sustainability of behavioral interventions at the community-, household- and individual-levels.

Compliance and sustainability are not mutually exclusive, but prove important in different ways for different types of interventions. For sanitation, compliance is critical. Even when coverage levels for large scale sanitation interventions are high, uptake and reduction of pathogen exposures do not necessarily follow.[35] The development and assessment of sanitation intervention requires significant change to defecation behavior and environmental contamination.[36] For water, on the other hand, sustainability may require more critical attention as recent studies have shown that regardless of compliance, all improved water source interventions resulted in reduced disease, especially at the household-level.[37] One study, using simulations to model the effect of different levels of child compliance to HWT on childhood diarrhea, demonstrated that even incomplete compliance in communities results in disease reduction.[38] However, as compliance is often poorly measured and defined in the field as it requires observation of behavior, this study did not use observational data to define compliance, but rather assumed incomplete compliance. For drinking water in infrastructure constrained settings, there often is no decision making between types of water sources as there are seldom options to begin with. Thus, if access to an improved water source exists nearby, individuals tend to comply.

Research on the underlying factors, like social or community constructs, that strengthen intervention implementation and increase acceptance, adoption, and sustained use is rare but should be invested in.[39] Few theoretical frameworks for behavior change that are used to precede intervention implementation take a broader ecological approach and place individual- and household-level interventions within a multi-level causal framework.[10] One such model for water treatment and safe storage, by Figuerora and Kincaid, suggests interventions influence behavior outcomes by individual-, household-, and community-level factors, including community action and cohesion.[40] Another model, by Dreibelbis et al., suggests an integrated behavioral framework for all WASH interventions that extends beyond the individual- and household-levels, and theorizes that interventions that act at the structural level have the capacity to reach large sections of a population and be highly cost-effective.[10,41] Studying WASH behaviors at each of these levels is important for sustaining behavior change, but research is especially lacking on the influence and significance of community-level social factors.

1.2 Social cohesion

While diarrheal disease prevalence has been attributed to low uptake of varying WASH interventions [42,43], few studies have considered the social determinants beyond examining poverty that contribute to whether a child gets a diarrheal infection or not. In social epidemiology literature, there is wide appreciation that neighborhood environments play a role in determining health outcomes.[44,45] Within the past two decades, literature on community variation of health has significantly advanced although it has primarily focused on health afflictions in higher-income settings, and thus chronic health disease.[46] Though such upstream social factors as education and income are not similarly applicable in rural, low-income settings, as populations are more socioeconomically homogeneous [47], in this dissertation we aim to understand the contextual social environment that affects health outcomes downstream, and whether it results in infection or prevention.

In discussions of multi-level causal frameworks for WASH interventions, the community-level represents both the physical and social environment where individuals and households are nested. Changes in the social environment produce changes in individuals, and the support of individuals in the community is essential for implementing environmental changes.[48] Collective efficacy, social capital, and social cohesion are all latent constructs of the social environment believed to influence the quality, effectiveness, and sustainability of interventions, especially those based on action at the community-level.

In the literature, social scientists discuss various conceptualizations of these social constructs, and there is much debate about how these constructs relate to each other. General conceptualizations of collective efficacy assume it is a latent construct comprised of a combination of the structural and cognitive components that facilitate a community's shared belief in its ability to come together and execute actions related to a common goal.[49] Social capital, on the other hand, is often conceptualized as features of social structures, such as trust, norms of reciprocity, and mutual aid, that act as resources for individuals to facilitate collective action.[50–52] Social capital is also commonly conceptualized as a component of social cohesion.[53,54] Some conceptualizations of social cohesion conceive it as a bottom-up process with a foundation in local social capital, where the social capital of a community takes on a strong sense of local space, albeit with ambiguous and fluid boundaries.[54] Therefore, social cohesion refers to two broader features of society, 1) the absence of latent social conflict (i.e. presence of social homogeneity); and 2) the presence of strong social bonds.[53]

The influence that collective efficacy more broadly, and social cohesion more specifically, has on intervention effectiveness may be explained in part by the theory of diffusion of innovations. This theory suggests that innovative behaviors diffuse much more rapidly in communities that are cohesive and in which members know and trust each other.[55] Such theoretical conceptualizations are supported by an empirical evidence base that suggests communities high in social constructs, have higher uptake of WASH interventions and substantial health benefits.[11,13] However, standard metrics or analytical approaches to investigating the relationship between social cohesion and intervention effectiveness, and related recommendations, do not exist in the WASH sector. As such, in this dissertation, we draw attention to the role that social constructs, specifically social cohesion, play in WASH programming and research.

1.3 The role of gender

Women have a critical role to play in community driven programs and efforts to establish clean water sources. Though women in low-resource settings are often excluded from key decision-making roles in their communities, they bear the burden of collecting, storing, and protecting water sources, and as such experience higher psychosocial stress.[56] Time spent walking in search of clean water or treating water often results in safety issues as women are afraid of encountering men alone and in dark spaces, which sometimes results in acts of violence. This time could be spent on education, work, or improving the overall health and nutrition of the household. Including women in key decision-making roles, to improve access to clean, nearby water sources, empowers women to improve their futures and bring families and communities out of poverty, in addition to improving downstream health effects.[56,57] Furthermore, prior research has demonstrated effects of chronic psychosocial stress beyond inhibiting behavioral uptake of safe WASH practices. As chronic stress can induce structural changes in the gut microbiota[58,59], maternal stress can alter child neuroendocrine-immune function thereby increasing disease risk.[60]

Empowering women as economic, political, and social actors in a community that can change key decisions and make choices is a critical component of development. Though reducing global income inequality is important, so is closing the gap in wellbeing between males and females to improve economic efficiency and improve other developmental outcomes. Giving women more voice at the local level can lead to a greater provision of public goods, including WASH resources.[61]

1.4 Social networks

Social networks are a tool for understanding the social environment of a community. Ties in social networks can correspond to casual acquaintance, close friendship and trust, or economic exchange, and can even measure differences between genders. In public health, social networks are commonly used to study the transmission of infectious diseases as they provide a map of direct contact for person-to-person transmission.[62,63] As a result, relationships between individuals are associated with greater individual-level risk.[64] However, imperative to the functioning of communities, networks also embody social cohesion and organization.[65–67] For example, if an individual has many ties to other individuals in her community social network and also belongs to a community organization aimed at improving water quality, exposure to pathogens may be reduced in the entire community.

While many studies have articulated the infectious nature of social networks, few studies demonstrate the protective nature of networks on community-level infectious disease risk.[68] Social networks are conduits of both social change and disease risk, and allow for the study of both phenomena. Indeed, examining social cohesion using social network data allows for a look at not only comprehensive community metrics but the effect that individuals and households within a community have on one another and on community structure as a whole. Using social network data allows for a hierarchical systems approach: a better understanding of the interdependent relationship between community level factors, households, and the individual in the WASH sector.[8]

Using social network data from rural, coastal Ecuador, previous studies have demonstrated the infective and protective effects of social connectedness on diarrheal disease risk. Trostle et al. examined social networks related to food sharing in these rural communities and showed how ties amongst community members differed across remoteness and resulted in different risk estimates.[69] Another study demonstrated the difference between social networks and geographic structures of communities, as spatial proximity between individuals and households is an important component of social networks in rural areas.[70] Only one study in the same Ecuadorian population, however, demonstrated the protective effect of social connectedness, as determined by network density, on diarrheal disease risk.[13] We extend this work by methodically defining social cohesion as an important social construct, examining temporal trends, and illustrating the importance of core discussion networks and gender for disease reduction and intervention implementation using both qualitative and quantitative methods.

1.5 Specific aims

We sought to examine the role of community-level social constructs in the multi-level causal framework of WASH interventions for the reduction of diarrheal disease. Using longitudinal social network data and qualitative data collected during the course of this degree from the same communities is rural Ecuador, we 1) examined a two-step penalization and shrinkage approach for binary response data that is jointly separated and correlated to study the effects of social networks on diarrheal disease, 2) examined the multifactorial effects of two different types of social networks at the household- and community-levels on diarrheal disease over time, 3) examined whether the observed effects of the community-level cohesion on diarrheal disease was mediated by WASH practices, and 4) examined the role of gender in community social structures and water insecurity and WASH practices.

Chapter II

A two-step penalization and shrinkage approach for binary response data that is jointly separated and correlated: The effects of social networks on diarrheal disease

2.1 Abstract

Epidemiologic data often violate common modeling assumptions of independence between subjects due to study design. Statistical separation is also common, particularly in the study of rare binary outcomes. Statistical separation for binary outcomes occurs when regions of the covariate space have no variation in the outcome, and separation can negatively impact the validity of logistic regression model parameters. When data are correlated, we generally use multi-level modeling for parameter estimation, and statistical approached have also been developed for handling statistical separation. Approaches for analyzing data with *both* separation and complex correlation, however, are not well-known. Extending prior work, we demonstrate a two-stage Bayesian modeling approach to account for both separated and highly correlated data through a motivating example examining the effect of social ties on Acute Gastrointestinal Illness (AGI) in rural Ecuador. The two-stage approach involves fitting a Bayesian hierarchical model to account for correlation using priors derived from parameter estimates from a Firth-corrected logistic regression model to account for separation. We compare estimates from the two-stage approach to standard regression methods that only account for either separation or correlation. Our results demonstrate that correctly accounting for separation and correlation when both are present can potentially provide better inference.

2.2 Introduction

Diarrhea is an important disease globally, resulting in approximately 1.3 million deaths annually [71]. Despite significant reductions in disease burden in the last decade, diarrheal disease continues to persist in low-resource settings, primarily through human contact and contaminated environments, including water, food, sanitation, and lack of hygiene [8]. Aside from vaccination, well known measures of diarrheal disease prevention include implementation of safe water, sanitation, and hygiene (WASH) practices, which are commonly spread by word of mouth [8]. Though social network data is more commonly used to study disease transmission in public health, previous studies from northern coastal Ecuador have shown a greater density of social ties between individuals may lead to the spread of sanitation practices, both individual and collective, thereby reducing the transmission of diarrheal disease [13]. This phenomenon, however, has yet to be examined over time using longitudinal data.

We collected social network data from a cohort in coastal Ecuador at three cross-sectional time-points (2007, 2010, and 2013). We asked individuals to name people within their community with whom they discuss important matters and collected data on self-reported diarrhea and fever at each time-point. Though most individuals listed friends in their network, few individuals reported having diarrhea or fever (approximately 10% per time-point), indicating the diarrheal disease outcome is a rare event. We collected data across multiple communities and for multiple households within each community at multiple time-points (Figure 2.1). This leads to a hierarchical/multi-level data structure (individual responses nested within households and households nested within communities) that is longitudinal. This data structure is commonly seen in epidemiological research, particularly in the study of infectious diseases. Resulting challenges include dealing with repeated measures within an individual, accounting for cluster-level correlation, and statistical separation of the outcome and predictors, a phenomenon often seen for rare binary outcomes and described in detail below. In this paper, we present a statistical approach used to analyze the study data that accounts for both complex multi-level correlation and separation in serially measured binary outcome data.

In epidemiology, we generally use logistic regression for binary outcomes. However, the uniqueness, existence, and consistency of maximum likelihood estimates for the logistic regression model depend on the configuration of data in the outcome-covariate space [72,73]. Separation for binary outcomes occurs when regions of the covariate space have no variation in outcome (all one or all zero). This condition is driven by factors including sample size, the number of covariates, the joint distribution of covariates, the strength of outcome-covariate association and whether the response variable is unbalanced/rare [74]. When there is separation in the data, numerical algorithms searching for the maximum likelihood estimate and its variance may lead to poor results [72,75]. A finite solution may not be reached, since one or more parameters in the model become theoretically infinite when data are separated [76]. When the likelihood for one or more parameters is maximized at very large but not infinite parameter values, the model experiences quasi-complete separation [72]. In studies of rare outcomes, separation may exist even when sample sizes are sufficiently large due to the unbalanced outcome distribution. Unbalanced and rare outcomes are particularly common in epidemiological research, leading to a potential need to address statistical separation in analyses.

When observations are independent but separation exists, we can obtain parameter estimates using Firth-corrected logistic regression. Firth correction is a penalized likelihood method originally introduced to eliminate small-sample bias but which can also be used to address issues of statistical separation [77]. Firth correction introduces a penalty term to the logistic regression likelihood involving the square root of the information matrix. This penalty is negligible when sample size increases. A comprehensive review of how Firth correction works in a binary logit model with a single dichotomous covariate can be found in a paper by Heinze and Schemper [74]. Firth correction can also be viewed in a Bayesian framework as using Jeffrey's prior on all regression parameters. Firth correction has proven useful for addressing complete or quasi-complete separation in binary response models, providing a better approach to separation than omitting problematic covariates [74,78].

In addition to issues of separation, epidemiologic data often violate common modeling assumptions of independence between subjects due to study design. In our study, the data are clustered (individuals nested within households and households nested within communities). Moreover, within-subject longitudinal observations are correlated. The general approach for analysis of correlated binary data is to use a Generalized Estimating Equation (GEE) or Generalized Linear Mixed Model (GLMM) (Figure 2.2). GLMM is preferable for our data structure for ease of handling nested clustering, unequal or small sized clusters, and missing-at-random data. GEE does not allow for multiple cluster-specific variance component estimates and, currently, accessible software does not handle multiple levels of clustering with computational ease. Unlike GLMM, GEE does not require distributional assumptions on the random effects, since estimation of the population average model is based on specifying the first two moments and not the entire joint distribution of observed data and random effects [79]. Due to lack of collapsibility in the logistic link function, the estimated odds ratio (OR) from a GEE model is often closer to the null value of 1 than the corresponding marginal OR in a simple random intercept GLMM model [80].

Although there are standard and widely-used approaches to address situations with either separation or a violation of independence, approaches for analyzing data with complex correlation *and* separation are not extensively well-known. Typical choices regarding which analytic approach to use are often ad hoc and based on ignoring one of the issues (Figure 2.2). Though a Firth-penalized likelihood could be used for random effects logistic regression, extending Firth correction to GLMM is difficult [78,81]. Given our interest in learning about the hierarchical structure of the data and the associated variance components, we explore existing methodology in the Bayesian paradigm.

Gelman (2006) and Gelman et al. (2008) explore Bayesian methods for fitting hierarchical logistic regression models with separation. These methods involve specification of weakly informative Cauchy priors for model parameters [81,82]. Abrahantes and Aerts (2012) proposed an alternative approach to dealing with separated and clustered binary data [78]. Their method uses a penalized likelihood approach to obtain data-driven priors for the regression coefficients that account for the separation. These prior distributions are then used under a Bayesian hierarchical model for inference. Abrahantes and Aerts examine one random effect in their hierarchical model and focus on defining weakly informed priors for only those covariates with separation issues. With separation issues, uninformative priors generally lead to convergence problems, and strong informative priors lead to results depending heavily on the mean and variance of the prior distribution. Therefore, their recommendation is to elicit and use weakly informative priors [83].

To better address the needs of our data, which are both correlated and separated, we extend the two-step approach in (15) to a multi-level model by replacing the second step with a hierarchical Bayes GLM with weakly informative priors on covariates with separation issues. Using these methods, we examine how social ties, individually and collectively, affect diarrheal disease over time. In this motivating example, we explore longitudinal clustered data, allowing for multiple random effects and accounting for separation in all covariates. We additionally explore different regression approaches and demonstrate how effect estimates and standard errors differ when we do not account for both separation and correlation.

2.3 Methods

DATA STRUCTURE

We collected sociometric data from 20 villages during three cross-sectional waves (2007, 2010, 2013) in northern, coastal Ecuador to examine the effect of social ties, derived from social network data, on Acute Gastrointestinal Illness (AGI). All community members ≥ 13 years of age were asked to participate. We surveyed all study participants who provided informed consent (approximately 80% each wave). All data collection protocols were approved by institutional review boards at the University of Michigan and Universidad de San Francisco de Quito.

Our outcome of interest was based on self-reported diarrhea and fever data collected in the sociometric survey. Participants were asked if they had a fever in the last week and if they had three or more liquid stools in one day in the last week. We combined these two measures to assess an individual's risk of having Acute Gastrointestinal Illness (AGI) to achieve more specificity in the context of enteric disease than just diarrhea. Investigators have used different terms for gastrointestinal illness, including Intestinal Infectious Disease [84,85] and Highly Credible Gastrointestinal Illness (HCGI) [86]. We define AGI as having diarrhea or fever, similar to other studies [87,88].

Our exposure of interest is a set of four covariates that define different aspects of social cohesion at the individual-, household-, and community-levels. Often ascertained by collecting survey data, social cohesion is a complex concept that is hierarchical by nature; individuals are influenced by their social environment in multiple dimensions [89]. Here, we assess social cohesion by use of both social network data and self-reported measures.

Social network data was collected by asking survey participants to identify members of their village outside their household with whom they discuss important matters, an indicator of an individual's core discussion network [90]. From this, we assessed the number of social ties an individual has to other individuals in the same community network. We then extended this measure to the household level and measured an highest number of ties in an individual's household (called their household degree) and how large the household degree is relative to other households within the same village. We refer to this relative degree as the household degree deviance. Continuous average community degree was measured by averaging the number of social ties across individuals in each community. Other measures of social cohesion examined are whether an individual has trust in her/his community and the number of organizations an individual belongs to (treated as a continuous measure).

We also examined remoteness, age, and sex as possible confounders. Remoteness is a function of time and cost to the nearest township from each village and is an indicator of infrastructural development [91], which may influence how individuals interact with each other. To avoid computational issues due to scale differences between covariates, remoteness was normalized by rescaling each community's remoteness score to be between zero and one, with the most remote village having a remoteness of one. Additionally, in our longitudinal model, we assumed a linear rate of change by time, coded ordinally as 0,1,2. We restricted analyses to individuals that were surveyed at all three time-points.

ANALYSIS

We assessed whether separation exists among covariates in the dataset by examining skewness, distributional plots, and defining prevalence estimates of covariates. We determined that separation was present (e.g. Supplementary Figure 2.1).

We consider four modeling strategies, each of which attempts to address the non-independence (GLMM and GEE methods), the separation (Firth-corrected logistic regression), or both (2-stage Bayesian GLM). Below, we describe the four analytical methods.

Two-Stage Bayesian GLMM

Due to the binary nature of our outcome, we used the following general multi-level hierarchical model structure, where we considered random effects at the individual- and household-levels, and subjects who share the same index (*i, j, or k*) are correlated. Including an additional random effect variable for community did not change the results of the full model, so we decided not to include it for parsimony and to limit computational complexity.

Level 1 regression equation:

$$logit(p_{ijk}) = log\left[\frac{p_{ijk}}{1-p_{ijk}}\right] = \beta_{0jk} + \beta_1 sex_{ijk} + \beta_2 age_{ijk} + \beta_3 trust_{ijk} + \beta_4 organizations_{ijk} + \beta_5 time_{ijk} + \beta_6 time_{ijk}\ x\ trust_{ijk} + \beta_7 time_{ijk}\ x\ organizations_{ijk} + \beta_8 time_{ijk}\ x\ household\ degree_{jk} + \beta_9 time_{ijk}\ x\ remoteness_k + \beta_{10} time_{ijk}\ x\ average\ community\ degree_k + \mu_i$$

Level 2 regression equation:

$$\beta_{0jk} = \gamma_{00k} + \gamma_1 household\ degree_{jk}\delta_{0j0}$$

Level 3 regression equation:

$$\gamma_{00k} = \pi_0 + \pi_1 remoteness_k + \pi_2 average\ community\ degree_k$$

where $\mu_i \sim N(0,\sigma_\mu^2)$ is the random effect associated with repeated measures within individuals, $\delta_j \sim N(0,\sigma_\delta^2)$ is random effect associated with multiple individuals within a household. Here, i indexes individual (i = 1,…,N), j indexes household (j = 1,…,n_j), and k indexes community (k = 1,…, n_k).

We extended the Abrahantes and Aerts approach by (1) adding multiple random effects and (2) using weakly informative, normally distributed priors obtained from a Firth-corrected logistic regression without any adjustment for the nested structure. We fit this regression using the following model structure:

$$\begin{aligned} logit(p_{ijk}) = log\left[\frac{p_{ijk}}{1-p_{ijk}}\right] &= \beta_0 + \beta_1 sex_{ijk} + \beta_2 age_{ijk} + \beta_3 trust_{ijk} + \beta_4 organizations_{ijk} \\ &+ \beta_5 time_{ijk} \\ &+ \beta_6 household\ degree_{jk} + \beta_7 remoteness_k \\ &+ \beta_8 average\ community\ degree_k + \beta_9 time_{ijk}\ x\ trust_{ijk} \\ &+ \beta_{10} time_{ijk}\ x\ organizations_{ijk} + \beta_{11} time_{ijk}\ x\ household\ degree_{jk} \\ &+ \beta_{12} time_{ijk}\ x\ remoteness_k + \beta_{13} time_{ijk}\ x\ average\ community\ degree_k \end{aligned}$$

where parameter $\boldsymbol{\beta}$ is a 14 x 1 vector. The log-likelihood is penalized with a Firth correction as follows, where $L(\boldsymbol{\beta})$ is the unpenalized log-likelihood and $I(\boldsymbol{\beta})$ is the corresponding information matrix:

$$\boldsymbol{ln}L^P(\boldsymbol{\beta}) = \boldsymbol{ln}L(\boldsymbol{\beta}) + \tfrac{1}{2}\cdot \boldsymbol{ln}|I(\boldsymbol{\beta})|.$$

We then used a Bayesian hierarchical model described previously to obtain inference using prior distribution $\mathbf{N}(\bar{\boldsymbol{\beta}}, \bar{\sigma}_\beta{}^2)$ for each fixed effect, where $\bar{\boldsymbol{\beta}}$ is the maximum likelihood estimate from the Firth corrected logistic regression and $\bar{\sigma}_\beta{}^2$ is the estimated variance. We assumed a noninformative **Inv-Gamma**(10^{-3}, 10^{-3}) prior for the random effect variances σ_μ^2 and σ_δ^2. We also examined a less informative "large variance" prior distribution of $\mathbf{N}(\bar{\boldsymbol{\beta}}, 2\cdot\bar{\sigma}_\beta{}^2)$) for the fixed effects to allow for extreme values of beta and compared this to the more informative default $\mathbf{N}(\bar{\boldsymbol{\beta}}, \bar{\sigma}_{\beta}{}^{2})$ prior.

We fit the Bayesian hierarchical model using *Stan*, which uses a Hamiltonian Monte Carlo No-U-Turn Sampler algorithm. This algorithm avoids random walk behavior and sensitivity to correlated parameters but is sensitive to step size and desired number of steps [92]. We ran each model for 5 chains of 10,000 iterations with a thinning of 5, after a 2,000 iteration burn-in. Convergence was assessed by the Geweke test [93], and we reported the 2.5% and 97.5% quantiles of the posterior distribution for the credible intervals.

GLMM, GEE, and Firth-Corrected Logistic Regression

As a comparison, we analyzed our data either ignoring the separation or certain elements of the clustering using statistical models routinely available in standard *glm* or *glmm* packages. For the GLMM method, we fit the Bayesian hierarchical model structure described for the 2-stage modeling using frequentist methods. For the GEE method, we fit a model using the logistic regression model mean structure used in the Firth corrected regression described earlier and accounting for correlation between the repeated measures but ignoring the clustering within households and communities. Both methods ignore the issue of covariate separation. We also compare these methods to the Firth-corrected logistic regression fit in the 2-stage procedure, which does not account for the clustering.

SOFTWARE

Social network degrees were calculated in R (v. 3.4.2, R Foundation for Statistical Computing, Vienna, Austria) using the package *igraph*. Regression analyses were conducted in R (v. 3.4.2) using packages *lme4*, *geepack*, *logistf*, *brms* and *rstan*. The codes for analysis are available on github (https://github.com/hegdesonia/two-stage-bayes).

2.4 Results

There was a total of 944 individuals observed across three-time points in the longitudinal dataset (Table 2.1). We noted that AGI increased 18% over time and was rare, with approximately 10% prevalence at each time point. Our outcome was unbalanced, and we noted separation issues across all covariates. For example, few subjects had trust and experienced AGI (Supplementary Figure 2.1).

By not accounting for separation, the GLMM model failed to converge due to excessive zeros in the parameter space because of the rare outcome (i.e. separation), resulting in larger parameter estimates and standard errors (Figure 2.3). We report the effect estimates computed in R software at the maximum likelihood value evaluated in tables and figures though the GLMM model did not converge. Using the GLMM model, valid inference could not be made though we were accounting for correlation.

Results from the GEE model had smaller confidence intervals than GLMM (Figure 2.3). Although the estimated effect stayed relatively similar to GLMM, the effect of trust in 2007 had a markedly smaller confidence interval in the longitudinal model (GEE 1.40, 95% CI: 1.12, 1.78 vs. GLMM 1.50, 95% CI: 0.96, 2.34) (Table 2.2). The odds ratio of AGI in 2007 for every one unit increase in the average number of social ties in the community was 0.89 (95% CI: 0.67, 1.19). As in the GLMM model, community-level network ties were not significantly associated with AGI. The GEE model also showed that for every one unit increase in household social ties away from the community mean, an individual's odds ratio of AGI in 2007 is 0.89 (95% CI: 0.79, 1.00). Though the mean effects were similar, the GLMM model demonstrated insignificant effects. By changing the model type to account for clustering differently, the significance of certain covariates changed, and our inference changed.

The regression with the Firth corrected likelihood accounts for covariate separation but not correlation between and within individuals, resulting in differences in the fixed effect estimates compared to both the GLMM and GEE results. Because we are not accounting for clustering and assume independence, we see larger standard errors due to positive intra-cluster correlation (i.e. how large the variance of the random effect is). The large difference in fixed effect point estimates in the Firth corrected models compared to the GEE and GLMM models suggest an impact of accounting for separation (Figure 2.3). The odds ratio of AGI in 2007 for those who trust the community (1.84, 95% CI: 1.23, 2.75) was greater than both the GEE and GLMM estimate (Table 2.2). The odds ratio of AGI in 2007 for every one unit increase in household social ties away from the community mean was significantly protective (0.82, 95% CI: 0.68, 0.98). The effect direction and significance of covariate effects changed by examining separation alone and ignoring correlation. Due to these marked changes in the fixed effects and our marginal exploration into separation for these data (e.g. Supplementary Figure 2.1), we ultimately determined that separation needed to be accounted for in addition to correlation.

Comparing this 2-stage approach that accounts for both highly separated and correlated data to the other regression methods used, we noted fixed effect estimates and standard errors that reflect the separation and correlation found in the data structure (Figure 2.3). The odds ratio of AGI in 2007 given a one unit increase in average number of social ties in the community is 1.15 (95% credible interval: 0.77, 1.70,) and the odds ratio of AGI in 2007 given a one unit increase in household social ties away from the community mean is 0.83 (95% CI: 0.68, 1.00). Compared to the estimates for GEE and GLMM, these estimates reflect wider 95% intervals and different point estimates. While our inference from the GEE and GLMM models hint that having a higher average number of community social ties may be protective against AGI (not significant), the 2-stage model suggests that having a higher average number of community social ties may be a risk while having a greater number of household ties compared to other community members is protective. We also found living in a remote community has a markedly stronger protective effect in the 2-stage model compared to both GEE and GLMM (Table 2.2). Having trust in the community remained a significant risk for AGI (OR 1.70, 95% CI: 1.16, 2.48) like the GEE and Firth corrected model showed in 2007. Gender had limited effects in all models. Importantly, in this analysis, we also illustrated there is little difference in standard errors in the Bayesian models when we allow for extreme values by scaling the prior variance by 2 (Table 2.2). We note the changes in effect estimates for each covariate over time in Supplementary Figure 2.2.

2.5 Discussion

Correctly identifying a model type to handle both separated and correlated data can result in markedly different inference. We demonstrate this by comparing different model types that only account for correlation, that only account for separation, and that account for both. By using the 2-stage Bayesian GLM approach, we were able to illustrate the protective effects of social ties at the household-level against Acute Gastrointestinal Illness (AGI) over time.

GLMM and GEE estimation account for correlation, but there is currently no software available that runs a GLMM or GEE accounting for separation. As we noted when comparing GEE and a Firth corrected logistic regression, the fixed effects did change markedly when either separation or correlation was ignored, changing our inference. Since we were interested in examining both individual and collective effects of social ties on AGI and how one level influences the other, GEE was also limited in terms of interpretation compared to GLMM.

The 2-stage Bayesian approach for analyzing highly separated and correlated data proved to be a useful alternative to ignoring correlation and/or separation in analyses. We recommend this approach be adopted more widely, especially for rare, clustered binary response data. Though ideally we would like to conduct a full Bayesian model with Bayesian sampling, we are limited by software. The proposed method is an efficient approach for epidemiologists as it uses existing functions and software. Though there is concern about using data twice for both the prior distribution and model fitting, both Gelman and Abrahantes demonstrate by simulation that this is not an issue [78,81]. Also, the approach taken in this paper reduces bias when both separation and clustering are present. Additionally, it attains particularly good results for small sample sizes (N < 100) and when there are greater than 100 clusters compared to methods ignoring separation and/or clustering [78].

A limitation of this approach is the use of weakly informed priors without heavier tails that allow for more extreme values as suggested by Gelman's Cauchy prior. Our approach, as previously explained, produces smaller posterior standard deviations with smaller tails of the prior densities used. It's also possible that in the context of modeling rare conditions [94], a weaker prior distribution (e.g. a Cauchy with mean zero and scale 10) might lead to more realistic results [82]. We try to control for this by comparing a prior variance of different scales to allow for extreme values and find there is not much difference in the credible intervals between a scaled variance prior by two and non-scaled in our setting, though this might differ for other datasets. However, we did not compare this to credible intervals and point estimates obtained from a weaker informative prior as suggested by Gelman, as that method is more difficult and time-intensive to implement. Furthermore, our method assumes independence of observations to construct the weakly informative priors for all covariates when the binary outcome presents separation issues. Additionally, though we standardized continuous covariates in our model, we did not standardize the binary variables to be symmetric as Gelman previously suggested to handle separation. Our assumption is that using this 2-stage approach does not require the standardization of binary covariates, which can be hard to do.

In the motivating example, we assume covariate effects change linearly by time and present these results in Supplementary Figure 2.2. We would likely obtain more intuitive time trend results if we did not make this linearity assumption, which might avoid some of the direction switching that we note in Supplementary Figure 2.2. We would expect the main effect differences between model types (GEE, GLMM, etc) to be similar.

Overall, this approach proves useful and results in minimal statistical bias, assuming we have specified the correct model. As the data is longitudinal and the number of clusters and sample size are sufficiently large, this method provides fixed effect estimates that better reflect the data separation and narrower credible intervals. Typically, in global health we predict effect estimates, like prevalence and incidence, of rare outcomes. However, rare outcomes generally result in separated data, and we often ignore separation and only account for correlation through the common use of GEE and GLMM models. For infectious diseases, which often have low prevalence in study data, accounting for separation is especially important for minimizing statistical bias and for making inference. By adapting the Abrahantes and Aerts method [78], we can account for multi-level data structures and provide a good solution for handling correlated data and making inference across nested clusters. This approach allows us to account for both highly separated and highly correlated data, leading to more accurate results and predictions.

Figure 2.1. Hierarchical model structure

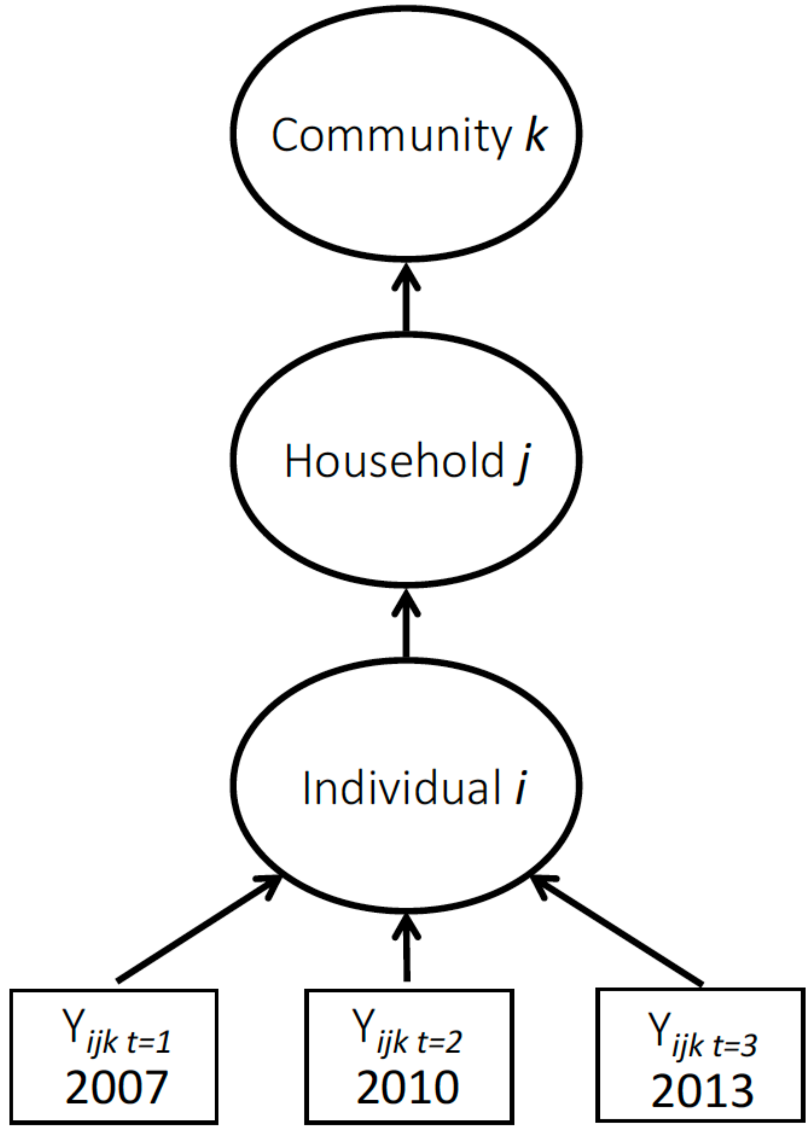


Figure 2.2. Analytical options for separated and correlated data

Correlation across repeated binary outcomes and within clusters	Separation: Yes	Separation: No
Yes	?	Generalized estimating equation (GEE) & Generalized linear mixed model (GLMM)
No	Logistic regression with Firth correction (logistf)	Logistic regression

Figure 2.3. Model results from different analytical methods

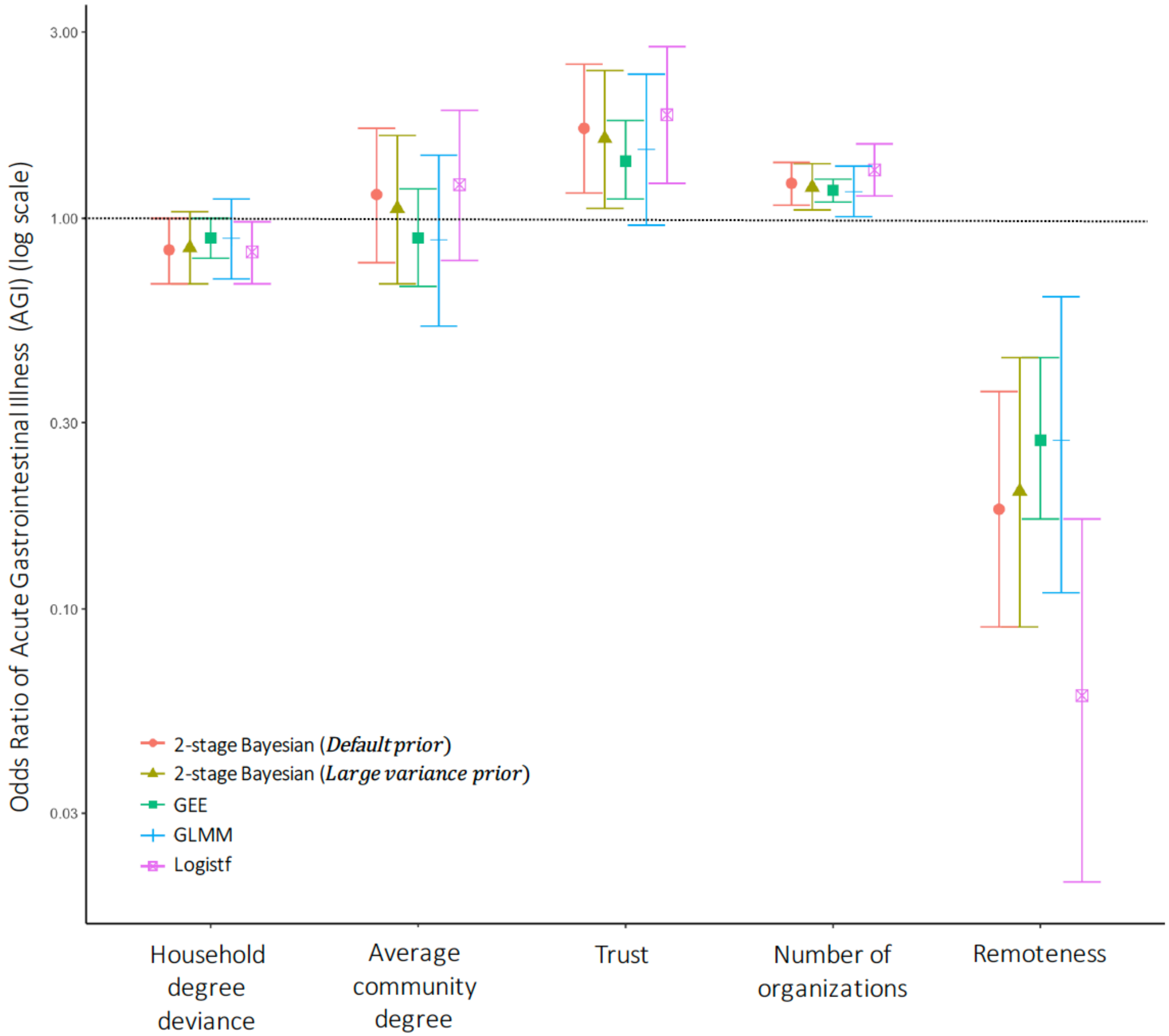


Table 2.1. Descriptive statistics of longitudinal data

	2007 N=944 Households=681 Communities = 20	**2010** N=944 Households=681 Communities = 20	**2013** N=944 Households=681 Communities = 20	**Variable description**
Acute gastrointestinal illness (AGI)*	10.1% (94)	10.0% (94)	12.6% (118)	The proportion (N) of individuals with AGI; having diarrhea or fever.
Household degree	3.7 (0-22)	3.0 (0-22)	3.5 (0-20)	The median (range) across all individuals of the maximum number of degrees (social ties) in their households.
Average community degree	2.2 (0.8-3.3)	1.9 (0.5-3.0)	2.4 (1.1-4.1)	The median (range) across all communities of the average number of degrees (social ties) within each communities.
Age	41.0 (13.0-86)	44.2 (14.0-91)	47.1 (17.0-94)	The median (range) of an individual's age in the dataset.
Sex (Female)	58.7%	58.7%	58.7%	The proportion of women in the dataset.
Remoteness	0.464 (0.06-1.00)	0.464 (0.06-1.00)	0.464 (0.06-1.00)	The median (range) across all communities of the normalized remoteness score based on time and cost to the nearest township.
Trust (Yes)	52.2%	48.5%	33.8%	The proportion of individuals who have trust in their community in the dataset.
Number of organizations	1.9 (0-7)	1.6 (0-8)	1.00 (0-7)	The median (range) of the number of organizations an individual belongs to.

Table 2.2. Model results from different analytical methods

	Bayes OR (Credible Interval)* **		**GEE OR (95% CI)**	**GLMM OR (95% CI)**	**Logistf OR (95% CI)**
Variable	*Default prior*	*Large variance prior*			
Household degree deviance*	0.83 (0.68, 1.00)	0.84 (0.68, 1.04)	0.89 (0.79, 1.00)	0.89 (0.70, 1.12)	0.82 (0.68, 0.98)
Average community degree	1.15 (0.77, 1.70)	1.06 (0.68, 1.63)	0.89 (0.67, 1.19)	0.88 (0.53, 1.45)	1.22 (0.78, 1.89)
Age	1.00 (0.99, 1.00)	1.00 (0.99, 1.00)	1.00 (0.99, 1.00)	1.00 (0.99, 1.01)	0.99 (0.98, 1.00)
Sex (Male)**	0.98 (0.77, 1.25)	0.99 (0.76, 1.28)	0.99 (0.87, 1.12)	0.93 (0.71, 1.22)	0.95 (0.71, 1.27)
Remoteness*	0.18 (0.09, 0.36)	0.20 (0.09, 0.44)	0.27 (0.17, 0.44)	0.27 (0.11, 0.63)	0.06 (0.02, 0.17)
Trust*	1.70 (1.16, 2.48)	1.60 (1.06, 2.39)	1.40 (1.12, 1.78)	1.50 (0.96, 2.34)	1.84 (1.23, 2.75)
Number of organizations	1.23 (1.08, 1.39)	1.20 (1.05, 1.38)	1.18 (1.10, 1.26)	1.17 (1.01, 1.36)	1.33 (1.14, 1.55)
Time	3.03 (1.68, 5.53)	2.48 (1.32, 4.76)	1.68 (1.14, 2.48)	1.68 (0.84, 3.39)	28.9 (10.2, 81.9)
Time X Household degree deviance	1.17 (1.02, 1.34)	1.15 (0.99, 1.34)	1.12 (1.02, 1.22)	1.11 (0.94, 1.31)	1.40 (1.15, 1.71)
Time X Average community degree	0.70 (0.52, 0.93)	0.76 (0.54, 1.04)	0.90 (0.72, 1.11)	0.91 (0.63, 1.32)	0.48 (0.34, 0.69)
Time X Remoteness	2.32 (1.39, 3.90)	2.14 (1.23, 3.74)	1.74 (1.25, 2.42)	1.70 (0.92, 3.15)	0.87 (0.37, 2.04)
Time X Trust	0.63 (0.47, 0.84)	0.66 (0.49, 0.90)	0.73 (0.62, 0.87)	0.70 (0.50, 0.98)	0.41 (0.28, 0.59)
Time X Number of organizations	0.85 (0.77, 0.94)	0.87 (0.78, 0.97)	0.89 (0.83, 0.94)	0.88 (0.78, 1.00)	0.77 (0.68, 0.87)
σ_δ^2 *(Standard deviation)*	0.21 (0.24)	0.22 (0.25)	-	0.199 (0.446)	-
σ_μ^2 *(Standard deviation)*	0.36 (0.31)	0.32 (0.30)		0.180 (0.425)	
Correlation parameter (Standard error)	-	-	0.0763 (0.0273)	-	-

*Standardized continuous variables
**Binary variables
*** We report the 2.5% and 97.5% quantiles of the posterior distribution for the credible intervals.

Effect estimates of Acute Gastrointestinal Illness (AGI) using different regression methods, 2007-2013. Household- and individual-level clustering is taken into account for GLMM and Bayes models. Only individual-level clustering is taken into account for GEE. Time is modeled as ordinal variable (0,1,2). We report OR and 95% CI.

Supplementary Figure 2.1. Example of separated variable

2007
N=944

AGI	Trust: No	Trust: Yes
No	373	409
Yes	42	46

2010
N=944

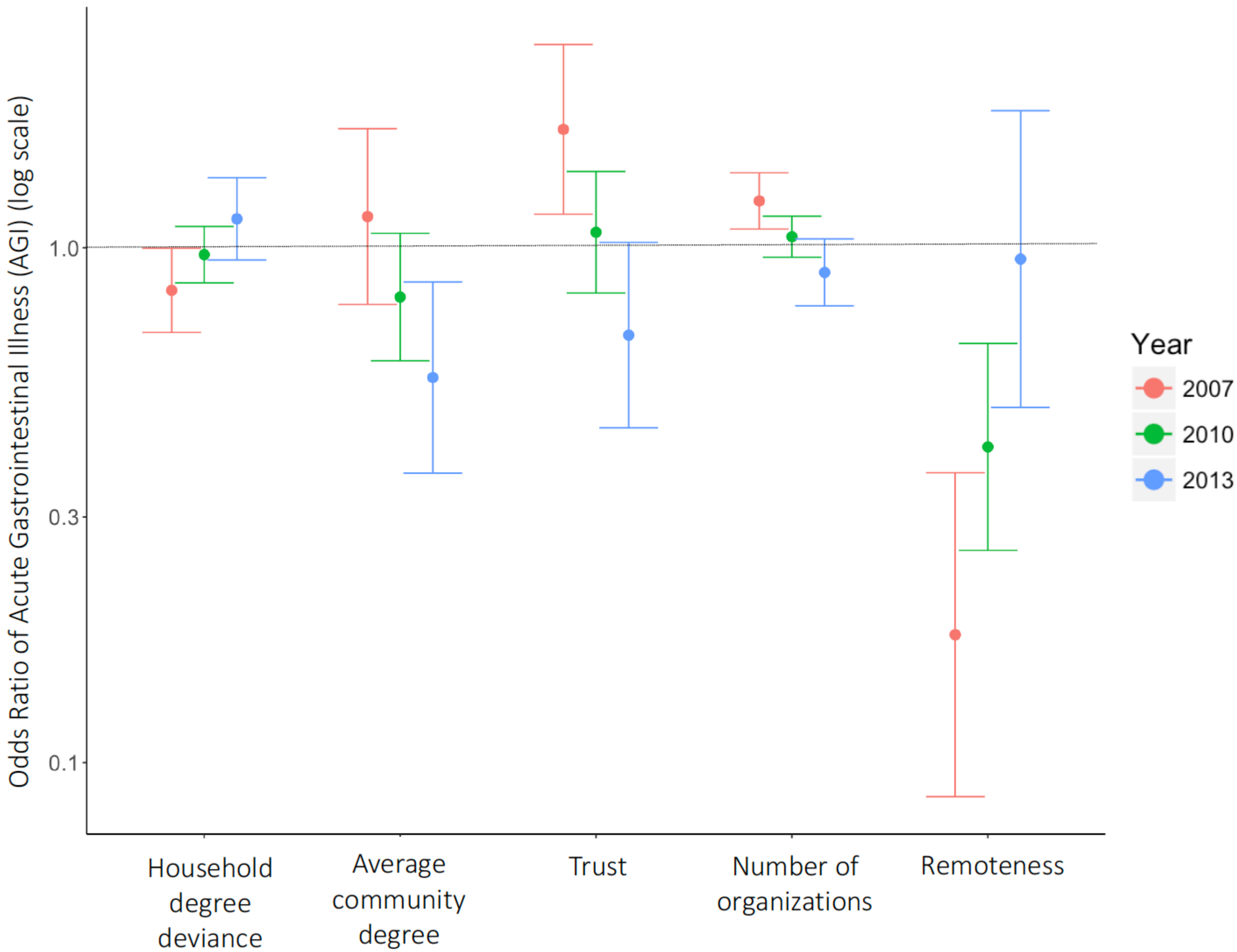
AGI	Trust: No	Trust: Yes
No	412	387
Yes	46	43

2013
N=944

AGI	Trust: No	Trust: Yes
No	505	271
Yes	85	31

Separated binary indicator with binary response variable, 2007-2013.

Supplementary Figure 2.2. Two-stage Bayesian model results



Forrest plot of 2007-2013 effect estimates using the two-stage Bayesian regression method with the *default prior*. The error bars represent the credible intervals for the effect estimates.

Chapter III

Multifactorial effect of social networks on acute gastrointestinal illness in rural Ecuadorian communities

3.1 Abstract

Social constructs play a critical role in public health, and represent different ways an individual connects to others and is influenced by her social environment. Social networks are part of an individual's social construct and for infectious diseases have largely been employed to describe social connection as conduits of transmission and not disease reduction. Generally, social network research in public health has focused on analyzing one social network criterion relation at a time, viewing the single network as an indicator of a social process that is realistically more complex. Prior cross-sectional analyses using social network data from northern coastal Ecuador suggested that a greater density of social ties between individuals may lead to the spread of water and sanitation practices, both individual and collective, that help reduce the transmission of acute gastrointestinal illness (AGI). This study showed social networks are potentially useful for intervention implementation and the prevention of AGI. Here we extend our earlier findings to examine how social connectedness, embodied through the joint effects of multiple social networks and other measures of an individual's social construct, influences AGI from 2007-2013 in rural Ecuador. We use a two-stage Bayesian hierarchical model to estimate effects, accounting for the statistically separated and correlated data structure. Having a larger community network of people to discuss important matters with, having trust in one's community, and participating in institutional organization becomes more protective against AGI over time. Effect modification of networks occurs within households. While a passing time network becomes a stronger measure of risk over time, due to density of people and increased travel, having a larger network of individuals to discuss important matters with is a consistent measure of protection. By 2013, the household networks become a greater risk for AGI and we observe synergistic effects as the people an individual passes time with becomes the people they

go to for important matters. Different network types contribute to the multidimensionality of social processes that occur at the household-level and that in turn influence individual health. Having a strong community network of ties to individuals to discuss important matters with is importantly protective against AGI.

3.2 Introduction

Social influences play a critical role in public health [1,2], and have been well studied in the field of social epidemiology. Such factors include occupation, gender dynamics, infrastructural environment, and social connectivity measured by social networks. The application of social networks, in particular, has grown in number in the last decade [3]. In 1979 Berkman and Syme [4] demonstrated those with weaker social ties had significantly higher mortality rates. A robust body of literature across disciplines built on this early result has since indicated that being socially integrated or feeling socially connected reduces mortality and various disease morbidities in the U.S. compared to other leading health indicators [2]. Social connectedness, as we'll refer to here as *sociality*, is a construct that represents the different ways an individual can connect to others socially: physical, behavioral, social-cognitive, and emotional [2], including through both social networks and other self-report measures like trust and participation in institutional organization.

Much of the work demonstrating the protective effects of *sociality* on health has focused on chronic disease [46]. For infectious diseases, social networks, specifically contact networks, have largely been employed to describe social connection as conduits of transmission and not disease reduction [13]. In both cases, however, social network research in public health has focused on analyzing one social network at a time, viewing the single network as an indicator of a social process that is realistically more complex. Indeed, many networks play a role in both risk and protection. Here we focus on how the joint effects of two social networks potentially influence risk of diarrheal disease.

Though different types of social networks may share similar features, studies have shown they are in fact distinct. Among the different types of egocentric networks, the important matters name generator (generated by asking the question "who do you go to for important matters?") first appeared in the 1985 U.S. General Social Survey and has been hypothesized to establish a network of social influence, called a core discussion network (CDN) [90]. Given a strong historical base in social science research, the important matters network has been particularly well studied. Using the CDN elicits important connections to an individual who she may not necessarily feel close to, but who are context-dependent social supporters [95], likely chosen intentionally and selectively [96]. The benefits of *sociality*, including information transfer, influence, and solidarity, are sought through verbal exchange with such context-dependent social supporters. Thus, CDNs map access to ideas or resources that an individual might activate in forming attitudes or in pursuing goals [96]. On the other hand, the passing time name generator (generated by asking the question “in the past week who have you spent time with?”) has been used to establish contact networks and therefore has proved useful in studying the transmission of infections like influenza and SARS [70,97].

Though some social processes may lead to infection while others to protection from disease, as highlighted when comparing the important matters and passing time networks, social network analyses seldom consider the multidimensionality of social processes. Therefore, distinguishing their effects in analyses is key. In public health, it is especially important to consider criterion relations that may contribute to disease transmission more than protection or vice versa; *sociality* is multidimensional and prevention can spread at the same time as disease. Here we extend earlier findings [13] to examine how *sociality*, embodied through the joint effects of two social networks (important matters and passing time) and other measures of social influence, affects diarrheal disease, an infection mitigated by infrastructure and social capital [12], which are influenced by social connection [98].

Diarrheal disease results in over 500,000 deaths in children under five years of age and 1.3 million deaths across all age groups annually [71]. Despite dramatic reductions in childhood mortality in the past decade, diarrhea remains a major cause of preventable childhood deaths worldwide [6,7] and is a leading cause of DALYs (71.6 million DALYs) [71]. In low-resource settings, diarrheal disease transmission occurs through multiple pathways, largely human contact and the contaminated environment, including water, food, sanitation, and lack of hygiene [8]. Aside from vaccination, diarrheal disease is controlled through implementation of safe water, sanitation, and hygiene (WASH) infrastructure, which is commonly spread by word of mouth [8], relevant for the application of social networks.

Prior cross-sectional analyses using social network data from northern coastal Ecuador suggested that a greater density of social ties between individuals in remote communities may lead to the spread of WASH practices, both individual and collective, that help reduce the transmission of diarrheal disease [13]. Though this study examined only one network criterion relation at a time, we learned that social networks are potentially useful for WASH intervention implementation and the prevention of acute gastrointestinal illness (AGI). We therefore expand on this study methodologically to gain a more nuanced understanding of how underlying social structures influence an individual's health behavior in the context of diarrheal disease and WASH. This is critical for targeting interventions among socially relevant groups where social norms develop [99], and is particularly important for achieving sustainable behavior change and interventions.

With no ascribed method in public health for examining *sociality* as having a multidimensional effect on disease, here we illustrate one such approach to examine the effect of *sociality* on AGI over time in Ecuador (i.e. the relative importance of different network types for AGI reduction and risk). In this paper, we examined two types of egocentric networks simultaneously using data collected through structured network surveys as well as unstructured qualitative interviews and focus groups. We additionally show these effects alongside other measures of social influence, including community trust and organizational belongingness. The analysis of these networks compliment those data that largely exist in the U.S. and other high-resource settings where infrastructure is different.

3.3 Methods

Quantitative data

We collected sociometric and census data from 20 villages in northern coastal Ecuador, in the province of Esmeraldas. Sociometric data was collected during three cross-sectional waves in 2007, 2010, and 2013 from all consenting community members ≥13 years of age. Census data was collected from all communities just prior to each sociometric survey. Compared to village censuses, the average sociometric response rate across communities was approximately 80% each wave. The study population consists of primarily Afro-Ecuadorians, Mestizos, and Chachis, an indigenous group of the Cayapas River in the region. All study participants provided informed consent and all data collection protocols were approved by institutional review boards at the University of Michigan and the University of San Francisco of Quito.

Outcome

Self-report diarrhea and fever data was collected in the sociometric survey. Participants were asked if they had a fever in the last week and if they had three or more liquid stools in one day in the last week. We combined these two measures to assess an individual's risk of having Acute Gastrointestinal Illness (AGI) to achieve more specificity in the context of enteric disease than just diarrhea. Investigators have used different terms for gastrointestinal illness, including Intestinal Infectious Disease [84,85] and Highly Credible Gastrointestinal Illness (HCGI) [86]. We define AGI as having diarrhea or fever, similar to other studies [87,88], and consider AGI as a binary outcome.

Exposure

We measured our primary exposure, *sociality*, using several variables: household and community network measures, trust, and number of organizations an individual belongs to. Social network data was collected using two different name generator questions on the sociometric survey. Study participants were asked to identify members of their village outside their household with whom they have spent time with in the previous week. We refer to this as a *passing time* network. Participants interviewed, or *egos*, were also asked to identify members of their village outside their household with whom they can discuss important matters. We refer to this as a core discussion network (CDN). The names generated created an "ego perceived friend" network, where approximately 63% of *alter* names generated were also interviewed. *Egos* were asked to generate names without a cap on the number of listed alters, which has been proven to be a sufficient number of names [100].

We measured *sociality* at the individual, household, and community-levels. To measure *sociality* at the individual-level we calculated the number of ties to another individual, per person in each network (called the degree). We aggregated individual degree to the household-level because AGI interventions like hygiene, water, and sanitation occur at the household-level. To accomplish this we took the individual with the highest degree in an ego's household and defined this as the ego's household degree. To measure the effect of an ego's household social connectedness relative to other households in the same village, we then standardized this measure within each village to have mean zero and unit variance. *Household degree deviance* from the village mean was measured in standard deviation units from the mean village household degree. We also aggregated individual degree to the community-level, calculating the *average community degree* per person in a village. Here we assume that the number of social connections per person positively affects overall community social connectedness [69,70]. Average community degree was measured in 1-unit increments. Average community degree and household degree deviance, the two social network variables included in our models, was measured for each type of network.

We additionally examined *trust* and *number of organizations* an individual belongs to as measures of *sociality*. We assessed trust, an indicator of individual perception, by asking if people generally trust one another in the whole community. Trust was used as a binary indicator. We assessed organizational belongingness, an indicator of institutional organization, by asking participants if they participated in community groups or local organizations in the last 12 months in their village. We specifically asked about 9 groups/organizations. The variable used in the model was the number of organizations a specific individual belonged to and was measured in 1-unit increments. Both trust and organizational belongingness are factors of social influence not accounted for in structural, network measures

Covariates

The study villages exist along three river basins: Cayapas, Santiago, Ónzole and vary by remoteness, which is a function of time and travel cost to the nearest township, Borbón [101]. Since 1996, paved roads have been built connecting this township to the coast and Andes. Smaller roads continue to be built linking villages to the main road. Remoteness may affect both *sociality* and AGI and was therefore used as a continuous variable in study models. Remoteness was normalized by rescaling each community's remoteness value to be between zero and one, with the most remote community having a remoteness of one. For more details see Eisenberg et al. 2006 [101].

From the census, we also examined an individual's age (continuous), gender (binary), household size (continuous), highest household education (categorical), highest household education of women (categorical), and community-level percent asset deprivation (continuous) as possible confounders. To measure community asset deprivation, we first measured household asset deprivation using the Multidimensional Poverty Index standard of living indices.

Deprivation was indicated by not having at least one asset related to information (e.g. TV, stereo, cell), or having at least one asset related to information but not having at least one asset related to mobility (e.g. canoe, motorbike, bicycle, motor), or not having at least one asset related to livelihood (e.g. fridge, arable land, livestock) [102]. After measuring this at the household-level, we then calculated community-level asset deprivation by summing the total number of households with asset deprivation and dividing by the total community population at each time-point.

Regression analysis

Since individuals entered and left the study continuously between 2007 and 2013 we chose not to limit our sample to those only included in each of the three surveys (2007, 2010, and 2013). To increase sample size and our ability to detect effects, therefore, we developed regression models using cross-sectional data. We selected a model of best fit based on known confounders from prior analyses [13] and the Akaike Information Criteria (AIC).

Our data exhibit both separation, which occurs for binary outcomes when regions of the covariate space have no variation in outcome (all one or all zero) [74], and correlation, as individuals are nested within households and households nested within communities. Standard data analyses dealing with separation generally use a penalized likelihood regression [74], while standard approaches for dealing with correlation consist of hierarchical linear models. We therefore used a two-stage Bayesian analysis approach with a binomial distribution to account for: 1) separation in the data, when maximum likelihood estimates are not proven to exist and are not unique [103], and 2) the nested structure of the data resulting in correlation. For more details of this method and the model equation, please see Supplementary Figure 3.1.

We applied this structure to two types of models: 1) where passing time network and CDN measures are examined separately to avoid collinearity; and 2) where passing time network and CDN measures are examined together with an interaction effect to examine effect modification as there might be confounding by two network features (e.g., the CDN may be related to the passing time network which in turn is causally associated with the outcome). We chose to examine effect modification, using an interaction term between household degree deviance in a passing time network and household degree deviance in a CDN. There was no significant interaction between average community degree in a passing time network and average community degree in a CDN, and there was no significant interaction between the household-level and community-level network measures.

We illustrate the interaction between the household degree deviance variables by estimating the predicted probability of AGI given a dichotomization (high vs. low) of each household degree deviance measure. This results in four categorizations of individuals (LL = low passing time network & low CDN, LH = low passing time network & high CDN, HL = high passing time network & low CDN, HH = high passing time network & high CDN). We used the 75^{th} percentile as the cut off for having a high versus low household degree deviance. We also report these estimates by remoteness level, categorized as either high or low based on the midpoint. We additionally illustrate the interaction effects by estimating the marginal effect of a household degree deviance in a passing time network on AGI as the household degree deviance in a CDN increases over time by 1) high versus low remoteness and 2) those that have trust in their community compared to those who do not have trust in their community when all other model covariates are controlled for. In supplemental material, we table demographics of the most connected individuals in households in a passing time network and CDN, and show the percentage overlap in ties between the networks (i.e. how often is the person visited for important matters also the person an individual spends time with). With this information, we then show the marginal effect of the odds ratio of having AGI for every one unit increase in household degree deviance in a passing time network as household degree deviance in a CDN increases, given the joint effects model and predicted for individuals where there is any overlap in ties between the passing time and important matters network and where there is no overlap over time.

Qualitative data & analysis

In 2016, we collected qualitative data from 15 of the 20 study villages to help interpret our quantitative results. We conducted a focus group with 6-8 community leaders and in-depth interviews with one key informant in each community. In total, we interviewed 15 different key informants and conducted 15 focus groups with community leaders in the study villages (3 communities situated by a road, 5 at a medium distance from a road, and 7 only accessible by canoe). We developed the same discussion guide for the focus group and in-depth interview, focusing on community problem solving, social organization, and kinship. We also focused on relationships individuals have with persons outside the community to complement our social network data that was based on community-centric ties. Specific questions under each theme underwent an iterative process of change, whereby we continually updated our discussion guide, to help fully understand topics and reach saturation (i.e. no longer attain diversity of opinions) in data collection. Key informants were either community health promoters or leaders of community organizations (i.e. persons of influence). Community leaders for the focus groups were purposively sampled. All focus groups and key informant interviews were recorded on a voice recorder and afterwards transcribed. Transcriptions were then analyzed in Spanish for consistent themes across communities based on underlying codes identified from our quantitative analysis and from social theory, and was iteratively compared to an emergent conceptual framework. All study participants provided written consent.

Data visualization

To note general differences in social connectedness in passing time networks and CDNs by remoteness levels, we employed the Kamada-Kawai algorithm to visualize networks of ties between individuals in a single remote community compared to a single roadside community over time [104]. Communities were chosen based on the qualitative data as self-declared unified (remote) versus disorganized (roadside) communities. This algorithm iteratively repositions nodes to reduce the number of ties that cross each other; the fundamental pattern of ties or topology of the social network image is fixed [104,105].

We also examined how the same selected remote and roadside community compared in CDNs specifically by visualizing network modularity, algorithmically defined sub-communities based on ties between individuals, over time. The detection and characterization of sub-community structures in the networks illustrates densely connected groups of vertices, with only sparser connections between groups. The approach we used optimizes modularity over the possible divisions of a network expressed in terms of the eigenvectors of a characteristic matrix for the network, the modularity matrix. This approach, developed by Mark Newman, leads to a spectral algorithm for community detection that returns results of higher quality than competing methods in shorter running times [106].

Software

Network analyses were conducted in R (v. 3.4.2, R Foundation for Statistical Computing, Vienna, Austria) using the package igraph. Regression analyses were conducted in R (v. 3.4.2) using packages brms and rstan.

3.4 Results

Regression analysis

Our total study population consisted of 2,204 individuals in 2007, 2,371 in 2010, and 2,326 in 2013. On average, an ego's household had 6.3 ties with other village households in a passing time network and 3.8 ties in a CDN in 2007 (Table 3.1). The number of ties per household decreased over time in both networks. The average number of ties at the community-level was less than the average number of ties at the household-level for both types of networks (Table 3.1). Trust decreased markedly over the entire six-year period (51% to 33%; Table 3.1), and was consistently higher in more remote communities across time (64% vs. 37% in 2007 & 45% vs. 21% in 2013). The number of organizations an individual belongs to decreased from a mean of 2 in 2007 to 1 in 2013. There also was a tendency toward increased socioeconomic status, where both the proportion of households with asset deprivation and households with a primary education decreased over time (68% to 57% and 27% to 20%, respectively).

Based on prior models conducted in this study population and AIC, age, sex, and remoteness were left in our full regression models as covariates (Supplementary Figure 3.1). The socioeconomic indicators of household education and household asset deprivation had no significant effect on AGI and were excluded from the final model. When we examined the network features in separate models, in a passing time network the household degree deviance went from being protective to a risk over time (Table 3.2). In 2013, for every 1 SD increase in an ego's household degree from the village mean, the odds of getting AGI increased by 1.20 (1.11, 1.30). Average community degree, on the other hand, had no significant effect in 2007 or 2013. In contrast, in a CDN, household degree deviance had no significant effect, while average community degree was protective against AGI over time (Table 3.2); in 2013, for every 1 unit increase in the village average degree in a CDN, the odds of getting AGI decreased by 0.74 (0.63, 0.87). Both networks also shared numerous attributes: remoteness remained protective over time; trust and being male became significantly protective over time; and the number of organizations an individual belongs to had limited effects on AGI (Table 3.2).

In the second set of models, where we examined the joint effects of the household-level passing time network and CDN, both household degree deviance and average community degree in a passing time network trended toward becoming risks from 2007 to 2013 (Table 3.3). By 2013, for every 1 degree increase in an ego's community average passing time degree, the odds of getting AGI increased by 1.55 (1.38, 1.75). On the contrary, household degree deviance and average community degree in a CDN trended toward becoming protective from 2007 to 2013. By 2013, for every 1 degree increase in an ego's community average CDN degree, the odds of getting AGI decreased by 0.34 (0.26, 0.45). The number of organizations an individual belongs to had limited effects on AGI and age had no effect on AGI. Trust and being male, like the first set of models, became significantly protective over time. Remoteness was significantly protective in 2007 (OR: 0.36 (0.25, 0.53)), but became less protective by 2013 (OR: 0.95 (0.73, 1.23)).

The interaction term between household degree deviance in a passing time network and CDN remained significant at all three time points. After dichotomizing household degree deviance for both network types into high vs. low to examine the interaction effects, we found the joint effects of the co-network features were different than the marginal effects originally observed. Individuals with a low passing time network and high CDN household degree experienced the lowest predicted probability of AGI by 2013 compared to the other subgroups (Figure 3.1). The effect of reduced predicted probability of illness was stronger in more remote communities (Figure 3.2B). The joint effect of having a low passing time network and high CDN resulted in reduced odds of AGI over time (2007: OR 1.16 (0.94, 1.44), 2013: OR 0.56 (0.45, 0.71)) (Supplementary Table 3.1). On the other hand, having a high passing time network and low CDN resulted in increased odds of AGI over time. Furthermore, we identified synergistic effects among individuals with a high passing time network and high CDN household degree, resulting in increasing predicted probability of AGI over time. There was no difference in these synergistic effects by remoteness level (Figure 3.2D). In the interaction model, the marginal effect of household degree deviance in a passing time network on the odds of AGI decreases in 2007 and 2010 as an individual's household degree deviance in a CDN increases. This changes effect direction in 2013 (Figure 3.3). This effect increasingly differs for those who have trust in their community compared to those who do not have trust in their community from 2007 to 2013, with trust becoming more protective over time. By 2013, there is no difference in risk for high versus low remoteness communities (Figure 3.3). The marginal effect of household degree deviance in a passing time network on AGI consistently decreases over time as household degree deviance in a CDN increases for households that have ties to persons they discuss important matters with independent of the persons they pass time with (Supplementary Table 3.3 & Supplementary Figure 3.5).

Qualitative analysis

Community problem solving & social organization

Both the key informant interviews and community leader focus groups characterized remoteness, and more generally the physical environment, as an important influence on community social cohesion, particularly a community's resolve to solve problems and self-perception of success and organizational leadership. Across all levels of remoteness, when asked to describe community problems, key informants and leaders mentioned potable water, contaminated water due to mining activities, lack of jobs, and trash disposal. Only roadside communities, however, mentioned disorganized youth, lack of internal collaboration, and chronic illness, like hypertension, drug addiction, and cancer, as issues. More remote communities mentioned lack of a pharmacy, medic, teachers, and school structure as issues. When asked if issues among residents were solved in the last year, all communities except those at a medium level of remoteness said at least one issue was resolved.

Remote communities had a tendency of resolving issues more and with social structures, like community leadership groups, designed specifically to solve problems. Persons visited to help solve important matters were community leaders, sometimes president of the community, elders, or leaders of organized collective labor groups called *mingas*. The most remote communities stated they had attained more success in their communities compared to other communities. These successes included: building a community house, soccer field, church, tourist hotel, sidewalks, staircase from the riverbed, piped water, energy system, a night club, and pooling community money to send a sick child to a hospital. All remote communities attributed their successes to similar concepts: an ability to solve problems through organization and unity, considering themselves as a single unit or commune.

On the contrary, roadside communities claimed they had less success compared to remote communities due to lack of participation and poor relationships between community members. Though both roadside and very remote communities commented on having external influences, like non-governmental organizations (NGOs), roadside communities all commented on the negative disruption that resulted from NGOs visiting their easily reachable households. Examples of negative disruption included one-time distribution of interventions, like water filters or bed nets or building of wells, without oversight or continuity. In contrast, remote communities cited more positive involvement from governmental ministries due to economically inspired pursuits like agriculture (i.e. gold, cacao) and tourism. Importantly, communities at a medium level of remoteness stated being continually neglected by external influences.

Kinship

Key informants discussed the communities forming based on families expanding and needing more space. The oldest community, for example, is a remote community formed at the time Afro-Ecuadorians were brought as slaves by the Spanish from West Africa to do mining nearly 400 years ago. Communities downstream since formed based on families expanding and needing new territory. The more remote communities, as a result, discussed having greater kinship and connectedness, unlike roadside communities where there is heavier migration. Though kinship is a unifying force in remote communities, members also cite kinship as an issue due to incest and lack of exposure to other individuals or sex education (e.g. one community said two siblings married each other). Across all remoteness levels, leaders mentioned having relationships with members in neighboring communities for school, work, and sport and farming organizations. For more information on the qualitative findings, please see Supplementary Table 3.2.

Data visualization

Remote communities overall were more visually cohesive and had more highly connected individuals in a passing time network compared to a CDN (Supplementary Figure 3.2). When we examined modularity in CDNs, we noted a different number of sub-communities in remote versus roadside villages. The roadside community visualized had more separated or disjointed communities with no ties between them compared to the remote community, which had on average fewer sub-communities (Supplementary Figure 3.3 5 & Supplementary Figure

3.4); remote communities were more cohesive compared with roadside communities. However, these differences decrease over time as the remote community described in the visual heuristics goes from having 10 sub-communities in 2007 to 7 in 2013, while the roadside community goes from having 17 sub-communities in 2007 to 9 in 2013.

3.5 Discussion

This paper demonstrates how social constructs can both influence protection against and risk of acute gastrointestinal illness (AGI) in rural Ecuadorian communities. The results highlight the importance of examining the joint effects of two types of networks within households to elucidate the multidimensionality of different social processes on health. At the household-level a passing time network becomes more protective against AGI as households increase the number of individuals they visit for important matters (CDN) indicating the importance of community leaders (Supplementary Figure 3.5), but this effect changes over time as the economy and infrastructure change and wage earnings and travel increase. As a passing time network becomes a stronger measure of risk over time due to density of people and increased travel, however, having a larger community-level CDN consistently leads to protection against AGI over time likely because of increased spread of intervention awareness and safe WASH practices. Though at a smaller magnitude, having trust in one's community and participating in institutional organization leads to reduced odds of AGI over time as well. Below we expand on these key results of the analysis, provide a discussion of the mechanisms of sociality, limitations, and future implications.

The effect of sociality on AGI

For AGI, social influences play an important role in mitigating risk and have public health significance for reducing disease burden. Understanding who individuals go to for discussing important matters versus who they pass time with is important for understanding prevention and risk of AGI. Our findings first suggest the importance of having individuals of influence for disease reduction (i.e. a strong CDN) at the household-level; who individuals go to for discussions of important matters is critical for change in this study population independent of who they pass time with. Importantly, social processes at the household-level influence one another and are multidimensional; there was no significant interaction between community-level network covariates or community- and household-level network covariates.

From 2007 to 2010, a passing time network is increasingly protective at the household-level as a CDN increases, but by 2013, we note increased risk of a passing time network as a CDN increases (Figure 3.2D & Figure 3.3). Over time, the people individuals go to for important matters become the people they pass time with (Supplementary Table 3.3) and vice versa, reaching 40% overlap by 2013. Across years, the more ties an individual has indicates more overlap in ties between a passing time network and CDN (i.e. the more social a person is the more likely he or she is to pass time with persons in their CDN). We show, however, that in 2013, it's those individuals that have overlap in their ties that experience increased risk of AGI while individuals who have an independent CDN are protected against AGI (Supplementary Figure 3.5). Therefore, having many people with whom to pass time with may indicate more possible interactions for disease transmission to occur in 2013 as oppose to 2007 and 2010 and a passing time network is more a conduit of transmission for AGI at the household level.

Given prior analyses in this study population and our qualitative data, in the study period, regional risk of diarrhea can also be attributed to movement patterns [107], as there was an increase in travel over time for both non-remote and remote communities [108] as infrastructure changed and the economy shifted toward wage-earning. With the travel, individuals were seeking more wage employment and steering away from local farming, which introduces individuals to more environmental contamination and disease risk. Furthermore, overlap in individuals listed in the two networks may occur as we ask *egos* to generate a list of *alters* they visit for important matters first and then generate a list of *alters* they pass time with, there may be question order effects and redundancy of names generated [109]. "Satisficing" could also occur if *egos* think interviewers expect more names for the passing time network than the previous important matters question and so list extra people that are not necessarily close, thus making the passing time network larger [109].

Second, our findings suggest having a large network of individuals who can discuss important matters at the community-level is critical for AGI reduction over time. Independent of everything else and when we control for the multidimensional effects at the household-level, we note stronger effects of the community-level CDN becoming more protective over time (Table 3.2 vs. Table 3.3). Indeed, as movement increases, communities become dislocated, and exposure to disease risk increases, so what happens at the community-level becomes more important for prevention and having a CDN becomes a stronger indicator of protection. As shown in our qualitative data, study communities with a stronger CDN presented an ability to overcome disruption and conflict over time. Thus, the data indicate that sociality at the community-level is more an indicator of cohesion while sociality at the household-level may be more an indicator of direct risk exposure.

Third, our findings suggest CDNs are more stable and stronger as community remoteness increases, as there is decreased mobility and less transient populations. Remoteness becoming less protective over time is an indication of infrastructure development and the introduction of more environmental contamination through travel; for AGI, access to roads and the changing wage economy has introduced greater disease risk in this study population. The effect of remoteness being protective against AGI, however, reduces over time though CDNs become increasingly protective, indicating that community-level cohesion is not just an attribute of remoteness but is an important attribute of society for prevention as development reaches rural areas.

Fourth, our findings suggest other measures of social influence at the individual-level, like trust and participation in institutional organization, similarly show increased protection over time against AGI, but at smaller magnitudes than the CDN (Table 3.3). There is a multifactorial effect of social influences on AGI over time; we note reduced risk of AGI for those who have trust in their community compared to those who do not trust their community by 2013. Importantly, the measures of sociality derived from the social network data impact individual-level AGI risk more strongly than trust or participation in institutional organization; identifying social networks is an important method for disentangling social processes.

Lastly, our findings suggest that when we examine networks singularly, we fail to capture the multidimensional nature of social processes and incorrectly approximate their impact on health at the individual-, household-, and community-levels; by not considering the joint effects, we would fail to see that the effects of a passing time network at the household-level are modified by a CDN. For AGI, we avoid confounding by co-network features with the joint effects model as passing time confounds the effects of an important matters network. Having different social processes, like trust and organizational belongingness, in the same model and controlling for confounding of co-network features at the household-level allows us to tease apart these effects and elucidate relationships critical for protection and risk of AGI.

Mechanisms of sociality

In this paper, we examine different components of sociality through social network data and other self-report measures like trust and organizational belongingness. We focus on two types of social networks: a passing time network and CDN. While a passing time network becomes a stronger measure of risk over time in our study population, due to density of people and increased travel, the CDN is a consistent measure of protection. Here, we expand on our qualitative results and the literature to assess the mechanisms behind a CDN, and the other measures of social influence, with respect to health and risk of AGI.

Individuals of influence

Our data indicate that the social process is indeed different between individuals we pass time with and individuals sought to discuss important matters with. There is a difference in influence between who we go to for important matters and who we simply pass time with, and being close to a person is not necessarily indicative of their ability to influence us as highlighted in 2013 when individuals of influence in a CDN are protective at the household-level when they are independent of the individuals people spend time with (Supplementary Figure 3.5). Over time, we observe this effect is nuanced: a passing time network can be protective if households have a strong core discussion network. Theoretically, however, it remains true that the social processes are different each year for the different network types.

Compared to a passing time network, the CDN has a stronger foundation in social theory and therefore consistent meaning. Studies have shown that when participants are asked to name persons they seek advice from (i.e. important matters), participants are more likely to nominate *alters* of higher social status, while persons they talk to (i.e. pass time with) are more likely egalitarian relationships [110]. Though CDNs may not indicate independent strong ties or people who are important to an *ego*, core networks indicate those who are core for change. For support, research has shown that people seek potential helpers based on some deliberation whereby the following are assessed: trustworthiness, skill, intimacy, and accessibility [111]. Thus, CDNs are innately hierarchical and reveal locally specific networks of leadership and influence. If an individual has more resources or access to people of influence, she can lead a healthier life and know who to contact when in need, particularly in the global context. Community leaders, not just people you pass time with, are critical players and important for social change and disease reduction.

While research on CDNs could benefit from having a clear definition of the underlying mechanism behind the effects we see in social networks [95,96,112], our findings support existing research showing that people do activate *alters* with distinct characteristics for different kinds of discussion topics or functions. Though the problems listed by our study communities are largely similar, the remote communities are accessed by government ministries, like tourism and mining, and NGOs more. While the roadside communities also experience NGO involvement, the internal critical relationships are not as many or as strong. Despite having similar external influencers, remote communities have less negative disruption because of amenities provided by the government due to key economic products like gold and cacao. Though all communities share problems of water and trash contamination, they differ in kinship, with remote communities being largely descendants of the same families while roadside communities have much more transient, migratory populations [108].

As the qualitative data illustrated, a strong CDN indicates resolve to overcome disruption from outside and internal forces (i.e. resistance to disruption of cohesion), and thus represents more critical relationships than a passing time network. These communities that are resistant to disruption and that are strong in important matters have communal societal structures where they share both their problems and money and so resolve problems together. These communities have strong structures of social organization, all individuals join the commune, and leaders are easily identifiable and interested in making change. Of note, those communities that have strong CDNs have female leadership and social groups, in both remote and roadside communities. As noted by roadside communities reporting solving their issues despite disorganization, people of influence are important in all communities. Thus, irrespective of remoteness, leadership is critical for resolving issues and is critical for change.

Collectivism

Importantly, individuals of influence have an effect at the community-level in addition to the household-level. The community average CDN degree is protective consistently over time in both the marginal and joint effects models, indicating there is a difference in how the collective influences an individual compared to an individual discussant.

As lack of social integration is a major risk factor for both morbidity and mortality at the individual-level [113], the emotional sharing that occurs in social integration at the individual-level may have community-level effects as *sociality* is multilevel by nature. Per psychology literature, listening to an individual share information elicits emotions, which in turn results in the dissemination of information with others [114,115], and when similar emotions are shared through dissemination, group cohesion increases [116]. A sense of community is created between the narrator and audience, and it's this repeated sharing of emotional information that enriches social beliefs, enhancing the collective and social integration of group members. Participation in such collective emotional events enhances social identity [117], ethnic identification [118], identity fusion with others [119], social cohesion [120,121], perceived social support [122], and solidarity [123]. Importantly, life in our study communities recurrently involves collective emotional gatherings like demonstrations, collective feasts, sport, musical, or religious events. This social sharing of emotions illustrates the continuous transactions linking the individual experience, interpersonal relationships, and the collective.

Also, intentionality is central to the occupations or activities individuals, communities, and societies engage in [124]; there is something that drives collective human engagement whether it be toward social cohesion or dysfunction, or advancement of or aversion to a common good. Central to some communities in our study is the idea that they are part of a collective; there is a strong belief in the interconnectedness between the individual and the collective and that one does not exist without the other. Therefore, much of the intention is focused on how to further the advancement of the community collectively. In fact, historically communities often engaged in *mingas*, a type of collective labor to accomplish tasks needed in the community and meant to further the collective. For example, this would include tending to local farms and picking cacao that would then be exported, and lead to profit for the entire community. Given much of our study population are Afro-Ecuadorians, originally West Africans who have integrated minimally with other ethnic groups since arriving nearly 400 years ago, there are stronger collective ideals and strong CDNs in the more remote communities.

However, as infrastructure has developed and more remote communities have become accessible by road and water taxis more available, occupations and the role of the collective have changed. For instance, *mingas* are no longer as prominent as a wage economy has been introduced and labor relations have shifted from within the community to outside the community. In the roadside communities, there was a shift from working on your own land to working on other's land for a secure wage income, and a wage economy is more singularly-based compared to an economy sustained by collective farming in more remote areas. Thus, the CDN became more protective over time at the community-level as the sources of employment and outside engagement increased; shorter term economic pursuits reduce collectivism in communities as jobs move outside the community. Such employment industries include palm oil and plantations in the roadside communities, and cacao farming and gold mining, supported by the government, in far communities. Importantly, the employment companies, government, and NGOs that do not complete development projects, act as negative external influences, rendering the CDN and community cohesion more relevant for social sustainability and disease reduction.

Like the community-level CDN, trust and organizational belongingness also resulted in increased protection over time. However, these measures, though commonly used in studies of social capital and social cohesion [53], did not have as strong of an effect as the social network measures. Trust is generally viewed as a level of interpersonal engagement, a strong social bond, which acts as a resource for individuals and facilitates collection action [53]. Institutional organization similarly acts as a measure of social bonds that generates confidence and has the ability to support social reform [125]. We see trust and participation in institutional organization as important proxies of the CDN.

<u>Remoteness</u>

As noted throughout this paper, remoteness plays a critical role in protection against AGI. In the modularity analysis, more remote communities had fewer sub-communities and stronger, more stable CDNs, indicating a proclivity for more resilience to disruption (i.e. the breaking apart of the larger component of the network). Indeed, in more remote communities, there is decreased mobility and less transient populations [112]. With stronger CDNs and greater collectivism, these communities therefore experience less disease unlike roadside communities that have more dynamic CDNs. Though as movement has increased over time in remote communities [108], we see a closing gap in disease reduction between low and high remoteness levels (Supplementary Table 3.1, 2007: 0.52 (0.37, 0.74); 2013: 0.96 (0.73, 1.27). Modern lifestyles, like cell phones and computers and road access, likely contribute more to a lack of social relationships in roadside communities in addition to the changing wage economy. New technologies will change how individuals feel isolation and lack social integration in present day, and in low-resource settings that translates to infrastructural development [1].

As learned from the qualitative data, more remoteness, a proxy of a community's physical environment, indicates more economical investment from the government because of agricultural products. Historically, human beings formed social structures (i.e. trust, group belongingness, connectivity, cohesion, social organization) settling near physical environments of interest like water and precious metals or stones (Figure 3.4). In turn, the social environment affected the physical environment changing its make-up with the pursuit of precious elements, waste creation, contamination of water systems and soil, infrastructural development for education, healthcare, and transportation. In this evolving dynamic process, the physical environment in turn influenced *sociality* by establishing issues of access and forming transient, migratory populations. This feedback loop between the social and physical environment emphasizes the impactful role of outside forces, like NGO or government visitors induced by the physical environment, that influence a community.

As communities with a strong CDN are qualitatively better at resolving issues, and are more remote, we suggest the internal forces or influences are stronger in such communities. For example, though NGOs built pharmacies and changed access to medication in roadside communities, a supposed positive influence, these communities have weaker resolve to solve issues; roadside communities have strong external but weak internal forces. Importantly, roadside communities also have many NGOs visit with short term, ineffective projects (i.e. negative disruptions). On the contrary, remote communities have strong internal and strong external forces but have less negative disruption as the government provides benefits due to gold mining and cacao production. Medium communities, however, have weak internal and weak external forces of influence and little to no disruption. Understanding how communities deal with conflict is just as important as understanding positive attachment and solidarity; an ability to deal with conflict is an indication of group rationality.

Limitations

One limitation of the study is that the network data is made of "ego perceived friends" or unidirectional ties. The implications of this on our results, however, is that if an individual has a large perceived network of persons they can go to for important matters, there is disease reduction regardless of whether the friendship is reciprocated. Indeed, the person an individual visits to discuss important matters is not necessarily someone close, but rather someone of influence [95]. Our use of name generator questions to elicit network data is supported by literature that shows name generator network questions elicit specific and accurate *alters* based on frequency and recent timing of contact [109]. We use the same instrument to generate two separate lists or samples of names and analyze the two subsets of names generated. A strength of our approach is that we use qualitative data to further describe and validate our quantitative analysis. Other limitations of the data are that additional relationships are not accounted for in the data like connections within the household, kinship among network ties, ties with members of other communities, and geographic space or movement within communities. As these all play a role in social connectedness, we aimed to address these through the qualitative data analysis. While we did not account for spatial autocorrelation, by accounting for the hierarchical structure of the data there may be no residual error due to space that's not already accounted for. Another study from our study site demonstrated the difference between social networks and geographic structures of communities, as spatial proximity between individuals and households is an important component of social networks in rural areas [70].

Though examining the network features separately allows us to avoid collinearity and confounding by co-network features, with network data we can never assume independent and identical distribution (IID) because there is an underlying network that induces dependence (i.e. individuals are dependent on one another via social ties). Nonetheless, effect modification allows us to examine the independent effects of each network feature by the other. Additionally, in using a CDN we assume individuals identify persons in their social support network by means of deliberation and not a random process. Though we believe individual decisions affect the composition of and resources gained from networks, we do not actually know whether the CDN consists of intentionally chosen or random contacts. Of note, our analysis is focused on the most connected individuals within households and does not consider that different individuals in the household may have different behavioral dynamics or may not interact with the same individuals the most connected person in the household interacts with. However, as our analysis demonstrates strong effects on individual risk of AGI given the household-level measures of sociality, this suggests individuals within a household are similar.

Of note, we use network measures at the household- and community-levels based on degree in this analysis though we have network data that allows us to estimate structural network measures at the community-level. Importantly, we wanted to examine the multi-level effects of sociality at the individual-, household-, and community-levels and thus chose to examine degree at each level. We also use household-level degree deviance as a measure and not simply household-level average degree to get a sense of how individuals and household exist relative to their community as out network data is restricted to within community. Deviance has been shown in other studies to be an effective measure of sociality [13]. As this analysis only demonstrates one way of estimating the network effects, new studies should explore other methods.

Future implications

In this paper, we present a method for examining the multidimensional nature of *sociality* whereby different social ties result in different effects on AGI jointly, and generate new hypotheses for how these social processes interact. We demonstrate an analytical approach that allows us to examine dependency between two network covariates without having to model a dynamical process. We also demonstrate the utility of a mixed methods approach, the use of both quantitative and qualitative methods, important for providing causal reasoning of the mechanism by which these effects occur. Though observational studies are intended to estimate causal processes, they may result in spurious associations. As such, correctly identifying the assumptions of the model and collecting qualitative data provide evidence of causation and not just correlation. Here, we have the advantage of having longitudinal network data and thus can demonstrate temporality in addition to a strength of association. We can demonstrate consistency of the results across studies in different disciplines and coherency within our own study population between our different methodological approaches.

Additionally, social networks are a useful tool for studying *sociality* as a health determinant in community structures globally and its impact on infectious diseases. Currently, network studies have forayed into using electronic sensors to collect data and define an *ego's* social ties. Based on our study, it will be important to decipher an individual's CDN from an individual's passing time network (i.e. the criterion relation sensors identify) to identify relationships of influence and points of social change, particularly in rural community and adult settings. Sensors may not elicit relationships of influence as individuals may visit them infrequently. Thus, further research should focus on elucidating the effects of the multidimensional social processes elicited by different network types at the household-level.

Furthermore, it's important to understand the utility of such data for AGI intervention implementation and sustainability. We hypothesize the protective effect of having a strong CDN leads to WASH behavior change and therefore disease reduction over time. Thus, focusing on individuals of influence and community leaders for information dissemination may be key for intervention adoption and behavior change. Indeed, social networks and community-based participatory research can be useful tools for designing intervention implementation by identifying persons in a community with leadership roles who are particularly influential discussants and therefore for disease reduction. Future studies should also examine the multiplex nature of social ties through simulation to describe how information diffuses in a CDN relative to risk of infection, which may provide a road map for sustainability of preventive behaviors.

Table 3.1. Descriptive statistics of cross-sectional data

	2007 N=2204 Households=1005	2010 N=2371 Households=1121	2013 N=2326 Households=1100
Outcome			
Acute gastrointestinal illness (AGI)*	10.6%	10.1%	12.5%
Covariates			
Household Passing Time Degree	6.3 (0-25)	3.6 (0-31)	5.0 (0-33)
Average Community Passing Time Degree	4.2 (2.2-6.6)	2.1 (0.8-3.0)	3.5 (1.8-6.2)
Household Important Matters Degree	3.8 (0-22)	3.0 (0-29)	3.4 (0-20)
Average Community Important Matters Degree	2.1 (0.8-3.3)	1.9 (0.5-3.0)	2.3 (1.1-4.1)
Age	36.9 (13.0-90.5)	37.7 (13.0-91.7)	37.9 (13.0-95)
Sex (Female)	48.8%	52.4%	52.7%
Remoteness	0.451 (0.06-1.00)	0.437 (0.06-1.00)	0.442 (0.06-1.00)
Trust (Yes)	51.0%	44.0%	32.8%
Number of organizations	1.8 (0-8)	1.58 (0-8)	1.00 (0-7)
Household size	2.9 (1-8)	2.8 (1-10)	2.7 (1-8)
Number infected in Household	0.28 (0-3)	0.27 (0-4)	0.32 (0-3)
Household asset deprivation	68.3%	62.8%	56.7%
Households highest education is primary (Yes)	26.9%	22.5%	19.5%

* Having diarrhea or fever

Reporting proportions and means(ranges) of each variable by year. At the household-level, we report degree and household degree deviance from the community average. Remoteness is a continuous measure that is a function of time and cost from the community to the nearest township.

Table 3.2. Model 1 results of network measures examined separately .

	2007	2010	2013
Passing time			
Household degree deviance	0.71 (0.66, 0.77)	0.91 (0.84, 0.97)	1.20 (1.11, 1.30)
Average community degree	1.04 (0.94, 1.15)	0.68 (0.59, 0.80)	1.06 (0.98, 1.15)
Trust	1.15 (0.98, 1.35)	1.26 (1.09, 1.45)	0.64 (0.54, 0.75)
Organizations	1.10 (1.04, 1.17)	1.03 (0.99, 1.07)	0.94 (0.88, 1.01)
Remoteness	0.28 (0.21, 0.38)	0.66 (0.53, 0.82)	0.53 (0.42, 0.67)
Age	1.01 (1.01,1.01)	1.00 (1.00, 1.00)	1.00 (1.00, 1.00)
Sex (Male)	1.09 (0.94, 1.27)	0.57 (0.50, 0.66)	0.74 (0.64, 0.85)
Important matters			
Household degree deviance	1.09 (1.01, 1.19)	1.00 (0.93, 1.07)	1.04 (0.97, 1.12)
Average community degree	0.89 (0.74, 1.06)	0.61 (0.50, 0.73)	0.74 (0.63, 0.87)
Trust	1.03 (0.88, 1.21)	1.27 (1.09, 1.48)	0.61 (0.52, 0.71)
Organizations	1.07 (1.01,1.14)	1.04 (1.00, 1.08)	0.96 (0.91, 1.02)
Remoteness	0.40 (0.28, 0.57)	0.73 (0.58, 0.92)	0.78 (0.61, 1.00)
Age	1.01 (1.01, 1.01)	1.00 (1.00, 1.00)	1.00 (1.00, 1.00)
Sex (Male)	1.05 (0.91, 1.22)	0.57 (0.49, 0.65)	0.77 (0.68, 0.88)

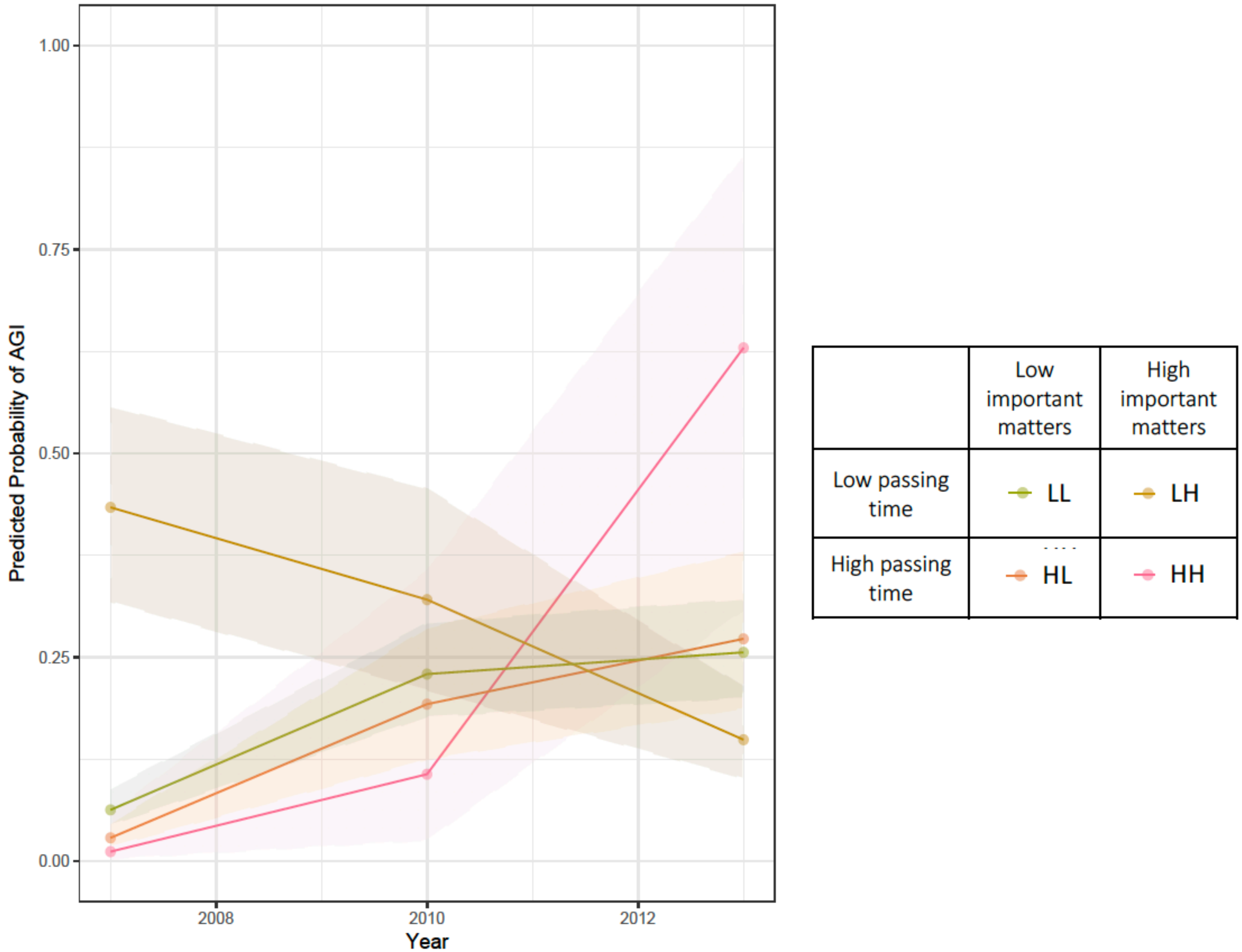
Passing time and important matters network measures examined separately.

Table 3.3. Model 2 results of network measures examined jointly

	2007	2010	2013
Passing time			
Household degree deviance	0.64 (0.56, 0.71)	0.91 (0.85, 0.98)	1.13 (1.04, 1.22)
Average community degree	1.04 (0.93, 1.16)	0.79 (0.65, 0.95)	1.55 (1.38, 1.75)
Important matters			
Household degree deviance	1.44 (1.31, 1.60)	1.08 (1.01, 1.16)	0.94 (0.87, 1.02)
Average community degree	0.87 (0.71, 1.06)	0.79 (0.63, 0.99)	0.34 (0.26, 0.45)
Interaction			
Household degree deviance passing time x Household degree deviance important matters	0.77 (0.71, 0.84)	0.94 (0.91, 0.98)	1.20 (1.13, 1.27)
Other covariates			
Trust	1.11 (0.94, 1.30)	1.26 (1.09, 1.45)	0.67 (0.57, 0.79)
Organizations	1.09 (1.03, 1.16)	1.03 (0.99, 1.07)	0.91 (0.86, 0.97)
Remoteness	0.36 (0.25, 0.53)	0.73 (0.58, 0.92)	0.95 (0.73, 1.23)
Age	1.01 (1.01, 1.01)	1.00 (1.00, 1.00)	1.00 (1.00, 1.00)
Sex (Male)	1.07 (0.92, 1.25)	0.57 (0.50, 0.66)	0.75 (0.66, 0.85)

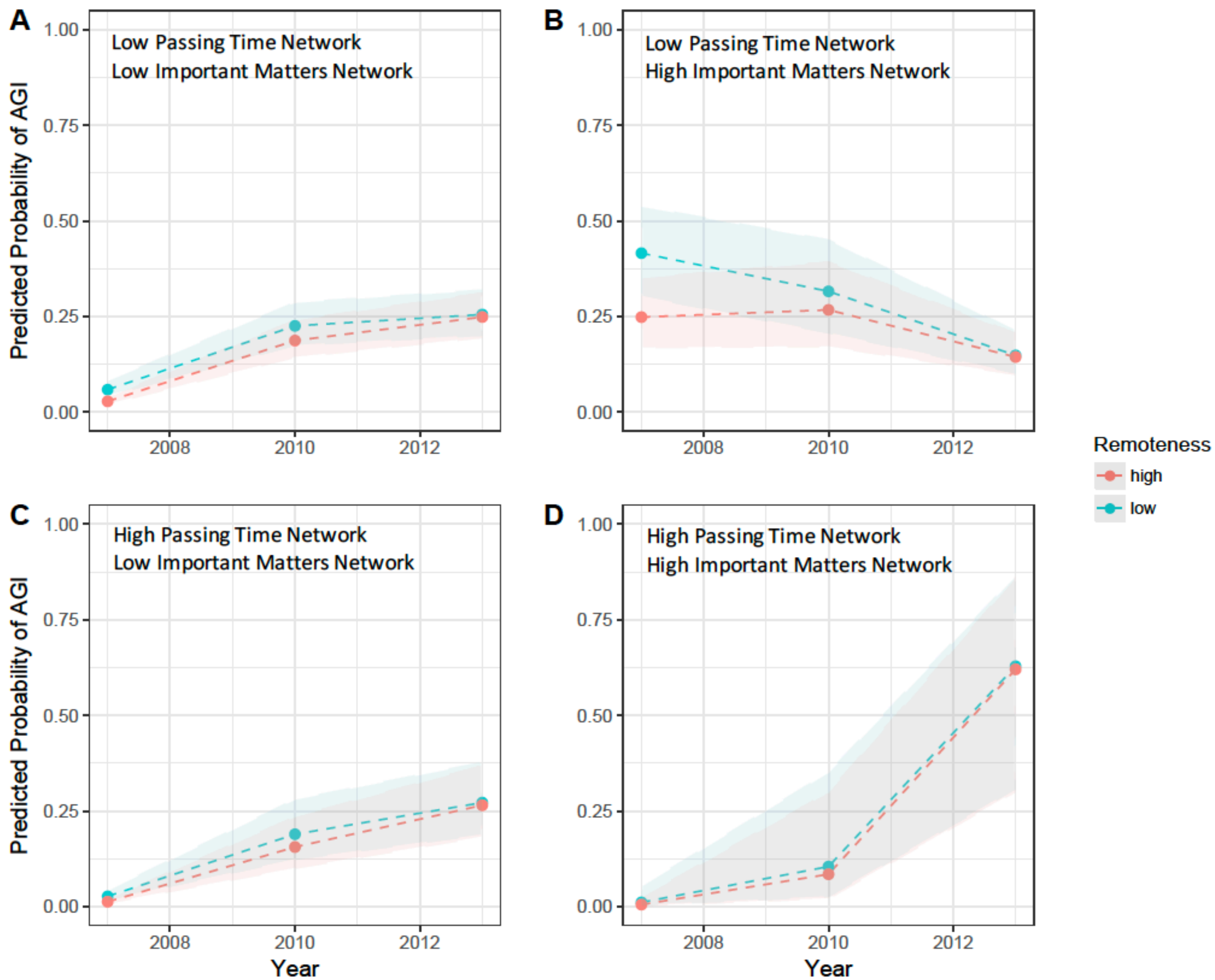
Passing time and important matters network measures examined together with an interaction effect.

Figure 3.1. Model 2 household-level interaction effects



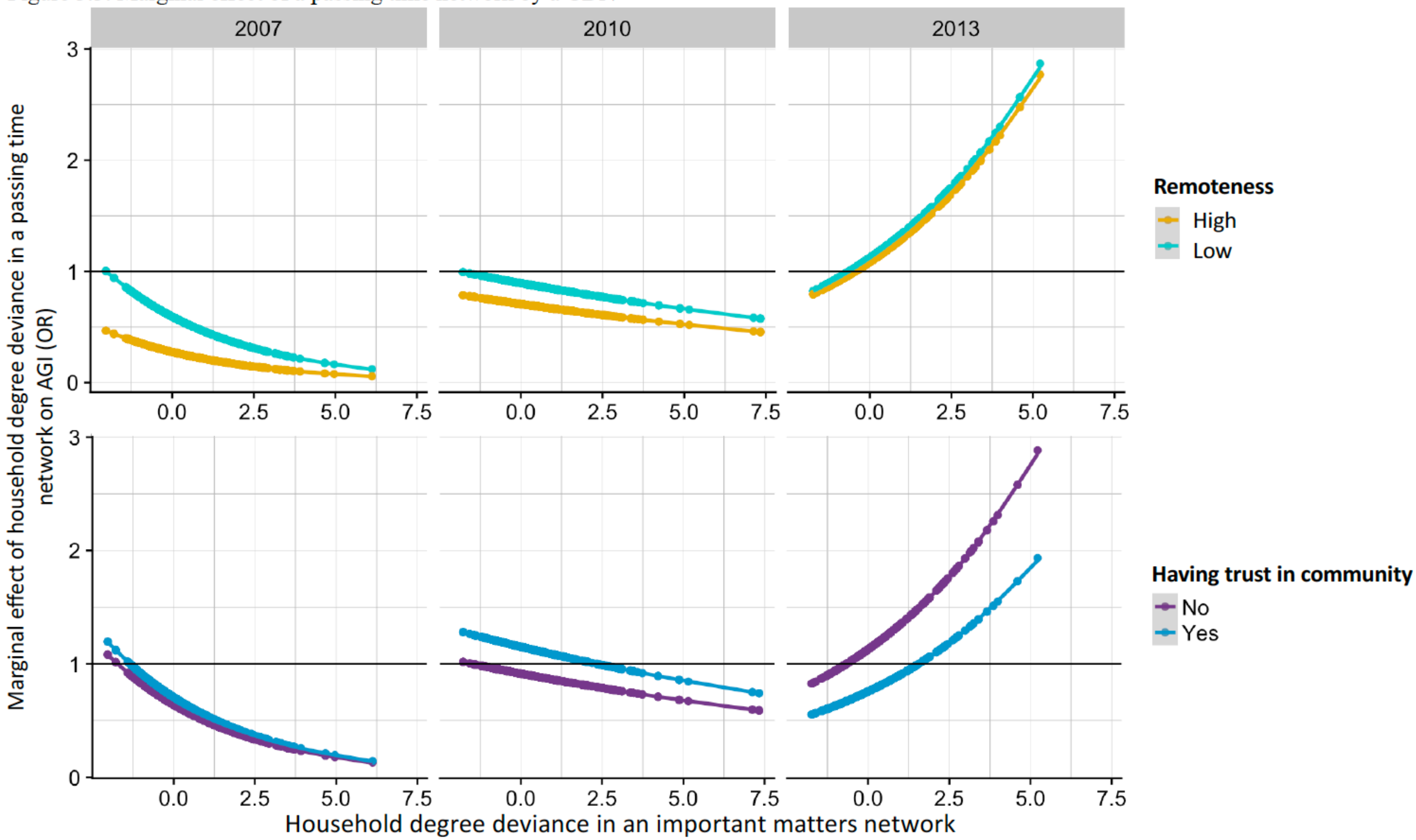
Model 2 interaction effects by high vs. low household degree deviance, 2007-2013.

Figure 3.2. Model 2 household-level interactions effects by remoteness



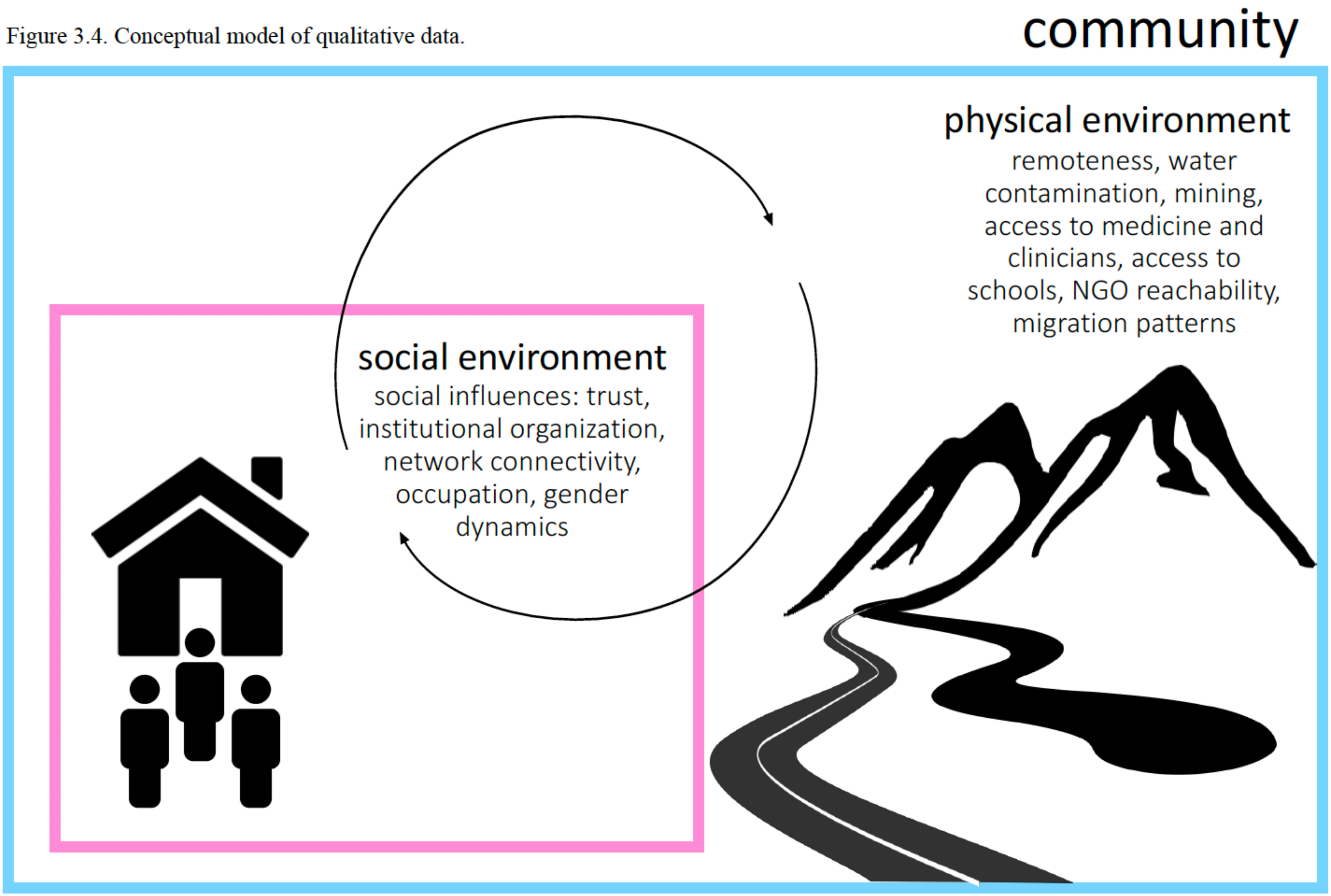
Model 2 interaction effects by high vs. low household degree deviance and remoteness, 2007-2013.

Figure 3.3. Marginal effect of a passing time network by a CDN



Marginal effect from the joint effects model of the odds ratio of having AGI for every one unit increase in household degree deviance in a passing time network as household degree deviance in an important matters network increases by remoteness and trust in one's community, 2007-2013.

Figure 3.4. Conceptual model of qualitative data.



Supplementary Figure 3.1. Two-stage Bayesian hierarchical model.

The two-stage approach consisted of: 1) conducting a firth penalized likelihood logistic regression to account for data separation, and 2) using these fixed effects as normally distributed weakly informative priors in a Bayesian hierarchical model to account for the nested data structure. We assumed a non-informative Inv-Gamma(10^{-3}, 10^{-3}) prior for the random effect variance at the household-level. Including an additional random effect variable for the community in did not yield significantly different results. Using Stan, our model uses a Hamiltonian Monte Carlo and No-U-Turn Sampler (HMC-NUTS) (30). We ran each model for 5 chains of 10,000 iterations with a thinning of 5.

Due to the binary nature of our outcome, we used the following general hierarchical model structure, whereby we considered random effects at the household-level for cross-sectional analyses for each year's data and anyone who shares the same index (*i, j, k*) are correlated

Level 1 individual regression equation:

$$logit(p_{ijk}) = log\left[\frac{p_{ijk}}{1-p_{ijk}}\right] = log(odds) = \beta_{0jk} + \beta_1 sex_{ijk} + \beta_2 age_{ijk} + \beta_3 trust_{ijk} + \beta_4 organizations_{ijk}$$

Level 2 household regression equation:

$$\beta_{0jk} = \gamma_{00k} + \gamma_{01k} household\ degree_{0jk} + \delta_{0j0}$$

Level 3 community regression equation:

$$\gamma_{00k} = \pi_{000} + \pi_{001} remoteness_{00k} + \pi_{002} average\ community\ degree_{00k}$$

where random effect $\delta_{0j0} \sim N(0,\sigma_\delta^2)$ is the j^{th} level adjusted for covariates on the intercept, and i= individual(i = 1,…,N), j= household (j = 1,…,n_j), k= community(k = 1,…, n_k).

Supplementary Table 3.1. Model results of examining joint effects

	2007	2010	2013
Joint effects			
Low passing time and Low important matters	1.00 (Ref)	1.00 (Ref)	1.00 (Ref)
Low passing time and High important matters	1.16 (0.94, 1.44)	0.93 (0.74, 1.15)	0.56 (0.45, 0.71)
High passing time and Low important matters	0.70 (0.55, 0.89)	0.82 (0.66, 1.01)	0.90 (0.72, 1.12)
High passing time and High important matters	1.25 (0.97, 1.60)	0.89 (0.70, 1.14)	1.46 (1.21, 1.77)
Other covariates			
Average Community Passing Time Degree	1.05 (0.95, 1.16)	0.83 (0.67, 1.03)	1.38 (1.21, 1.56)
Average Community Important Matters Degree	0.82 (0.67, 0.99)	0.79 (0.61, 1.02)	0.45 (0.33, 0.61)
Trust	1.04 (0.88, 1.23)	1.19 (1.02, 1.39)	0.75 (0.64, 0.88)
Organizations	1.06 (1.00, 1.11)	1.03 (0.99, 1.07)	0.94 (0.89, 1.00)
Remoteness	0.52 (0.37, 0.74)	0.78 (0.62, 0.99)	0.96 (0.73, 1.27)
Age	1.01 (1.01, 1.01)	1.00 (1.00, 1.00)	1.00 (1.00, 1.00)
Sex (Male)	1.03 (0.89, 1.19)	0.65 (0.56, 0.76)	0.82 (0.72, 0.93)

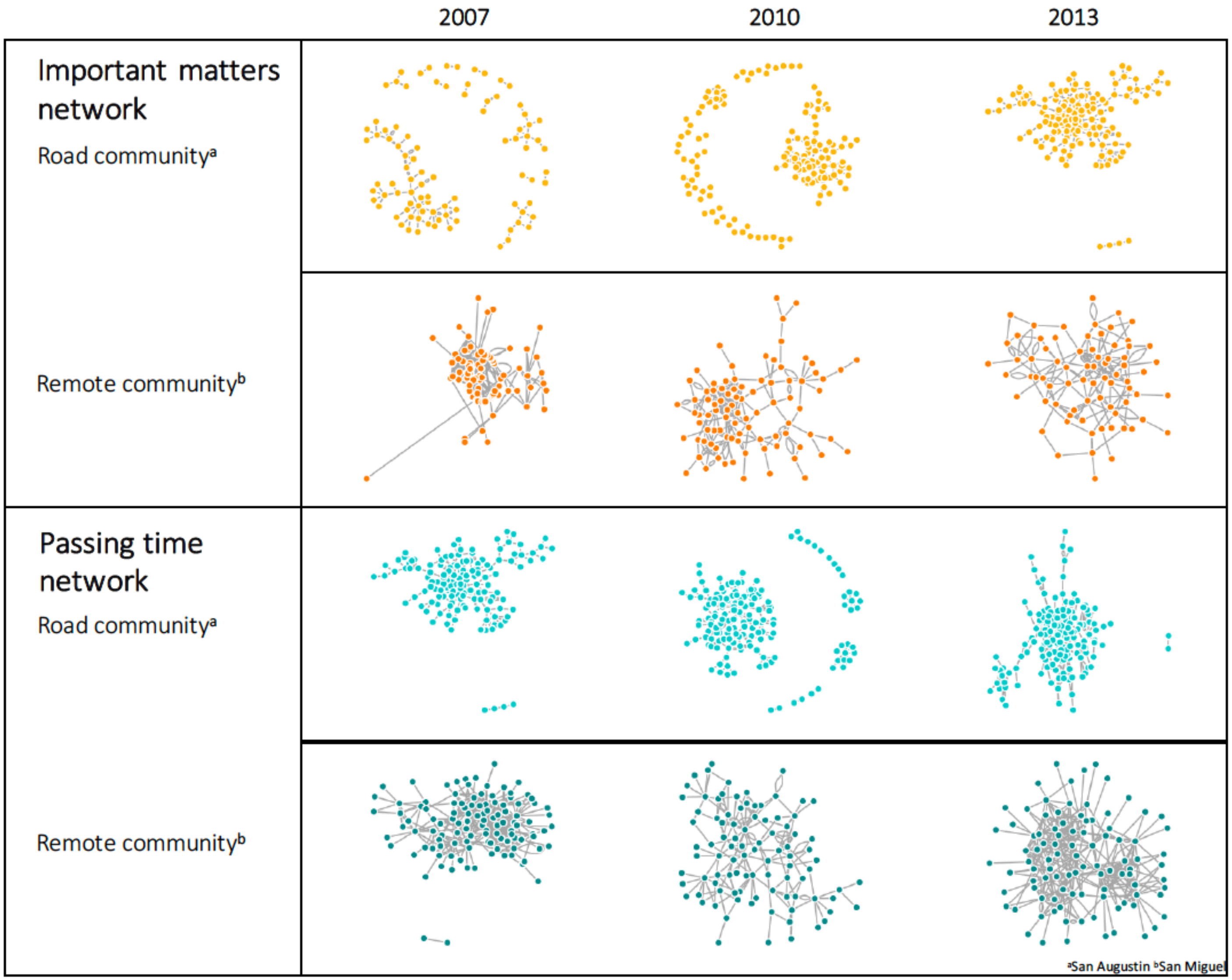
Model results, 2007-2013, examining the joint effects of having either a high or low passing time and important matters network.

Supplementary Table 3.2. Qualitative findings

Remote communities	With respect to water insecurity and problem solving more specifically, members of highly cohesive communities discussed the river nearby having fresh water and many fish in years past, but since being highly contaminated with fewer fish. They discussed not wanting river contamination and coming together to protest and write songs about the issue, with their strength to fight coming from their blood and support from family and friends. These communities felt united and said social organization is the impactful element that is different in their respective communities compared to other communities. In fact, remote communities tend to cite their own community as incomparable and unique, even independent (i.e. "doing everything ourselves"). One community told a story of a young child getting his arm cut off by an accident in the night and the community reuniting the same hour to help and provide funds to the family to take the boy on a canoe to the nearest healthcare facility. People come together and give money if someone needs help in the area for health or construction. Community members are clear about their pride in their community, being proud to be where they're from and to have the community in their blood. Though some remote communities, with high government involvement due to gold mining, stated having a household tube water system from mountain springs, they still experience water shortages when it doesn't rain (minimum of 15 days/year) and competition for water. Social organizations in remote communities included much of the same, but also included a: marimba group, tourism group, education committee, social forest program, provincial government, festival committee, job committee, and cacao group. More remote communities created social organizations to help solve the issues they listed, like education and work.
Medium distance communities	Communities at a medium distance from a road said they have been fighting with the government for 30 years and have felt ignored by the government. The issues they have included contaminated rivers and fish, distance to water for collection, land fights and deaths over lumber, lack of community support, and lack of healthcare and schools (the nearest school is four hours away by canoe and road resulting in most kids not going to school). Though government officials have visited these communities and know of these issues, they have done nothing. Medium distance communities want a road, constant water supply, and electrical power, whereas remote communities prefer there not to be a road. An engineer visited in years past to one community to build a well and tube system to each household from the well, but this system has since broken.
Roadside communities	Communities close to road infrastructure feel their communities are divided, disorganized, with no functional clean water, but that people sometimes help each other, known as *mingas*. They also still have marimba and *arullos*, singing, groups, where members gather to sing their troubles and joys. They say researchers and aid organizations will come and collect data and then leave without truly helping or sharing information. These close communities, like the medium communities, also report not having any assistance from the government saying the government only wants to do what benefits themselves, consistently overlooking the Afro-Ecuadorian and indigenous populations. Though the government put in the roads, there are no walkways or sidewalks which is a big problem. Additionally, most people have health issues that are a direct result of the contaminated water (e.g. vaginal and leg infections, respiratory problems, cancer), as there is mercury and other chemical water contamination across all seasons due to gold mining and the palm oil industry. However, there are no interventions and no medics to help understand the issue. Furthermore, community members said there is no good water system: some neighbors do share water supplies, some treat their water and others do not and some only use the river water to drink, some do have taps in the household but unsure of where the tap water comes from, and others do not have latrines in their households and defecate in the river they collect water from. In the one highly cohesive road community, the water system built in 2003 does not serve the entire community as it requires each household to pay for use and not many can afford it. Thus, community members report using rain, spring, and river water daily. When asked how these communities close to road infrastructure differ from the more remote communities, members said the more remote communities are better organized and have heart – that their motivation comes from unity. Social organizations in road communities included a: commune (i.e. primary community leadership organization), parochial meeting, singing group, sports groups, women's group, health committee, church group, and a union. Different than other road communities, one road community also had a group specific for taking action. Road communities, on the other hand, deal with issues like unfair job practices.
High NGO involvement communities	Overall, communities with high NGO involvement, particularly church groups, define community as an ideation of social connections. These communities claimed foreign churches have input things like day cares, a computer center, social project to build trust, spiritual and emotional health, helped men to have their own income source and not be dependent on others. Religion is an important aspect of social organization.

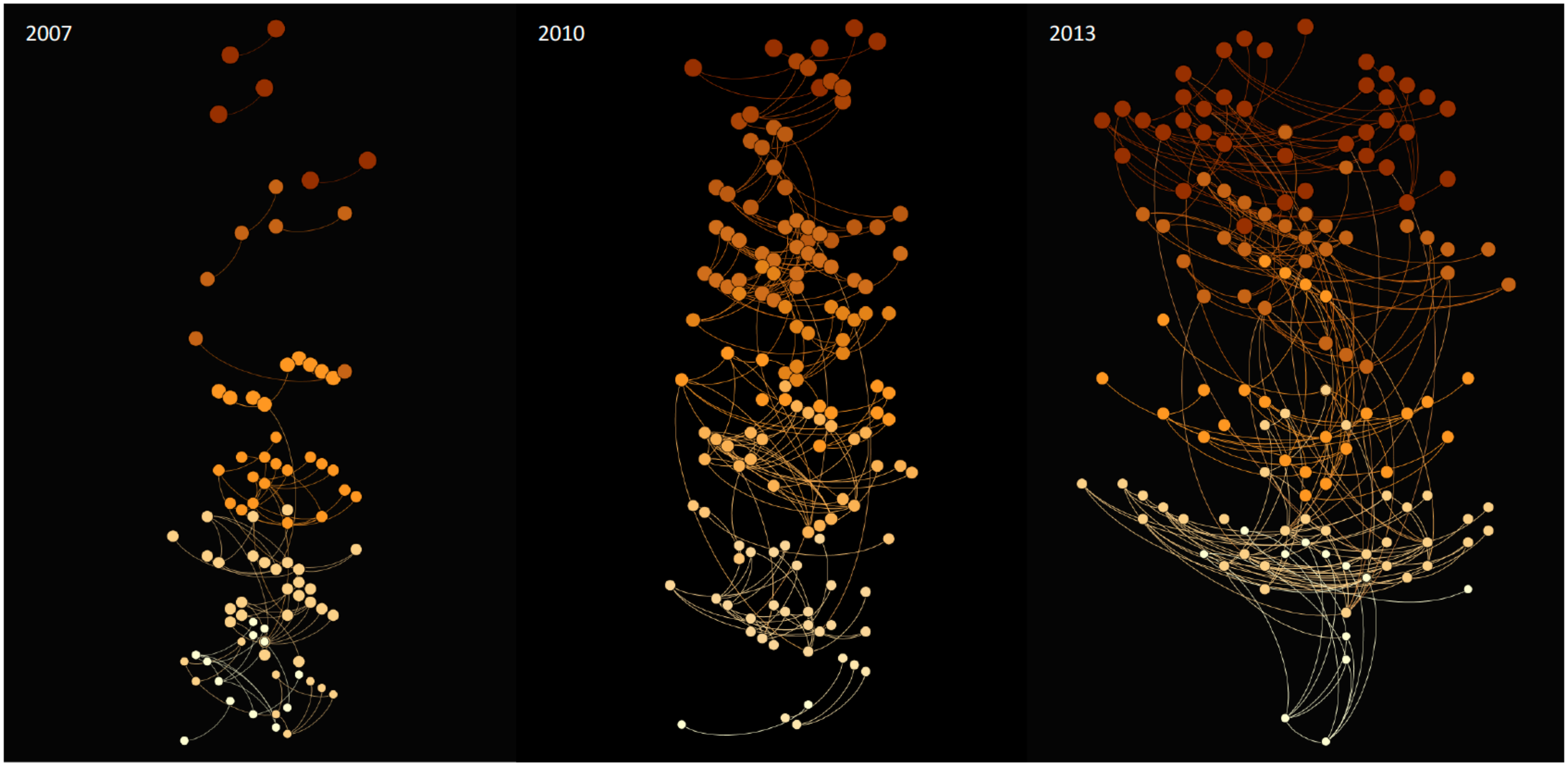
Extension of qualitative findings by remoteness level and external community influence.

Supplementary Figure 3.2. Community network visualizations by remoteness



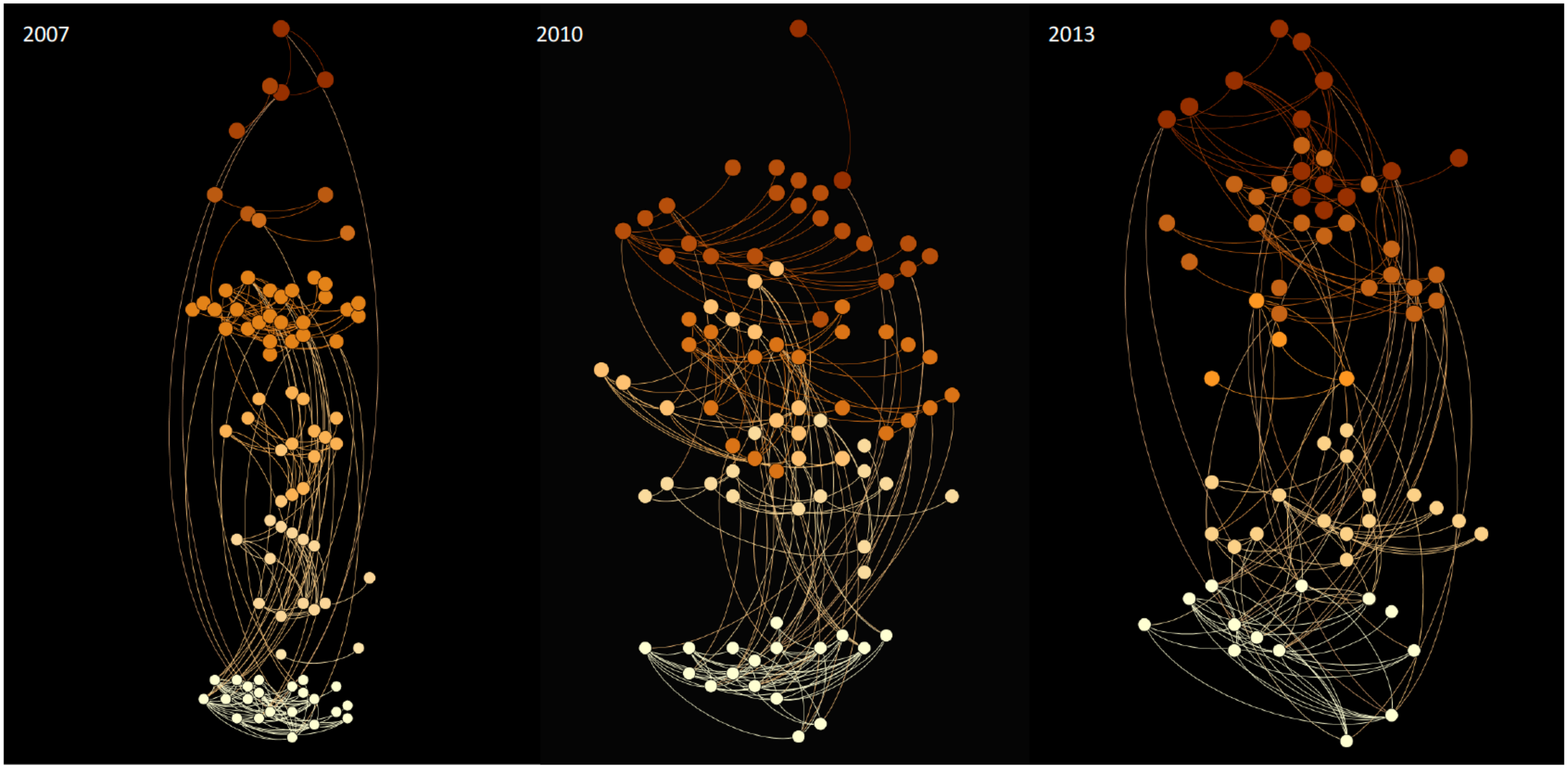
Kamada-Kawai network layout of both an important matters and passing time network in a roadside (San Augustin) and remote community (San Miguel) from 2007-2013.

Supplementary Figure 3.3. Temporal network visualization in roadside community



A network visualization of modularity, indicated by color, of an important matters network in a roadside community (San Augustin) from 2007 to 2013.

Supplementary Figure 3.4. Temporal network visualization in far community

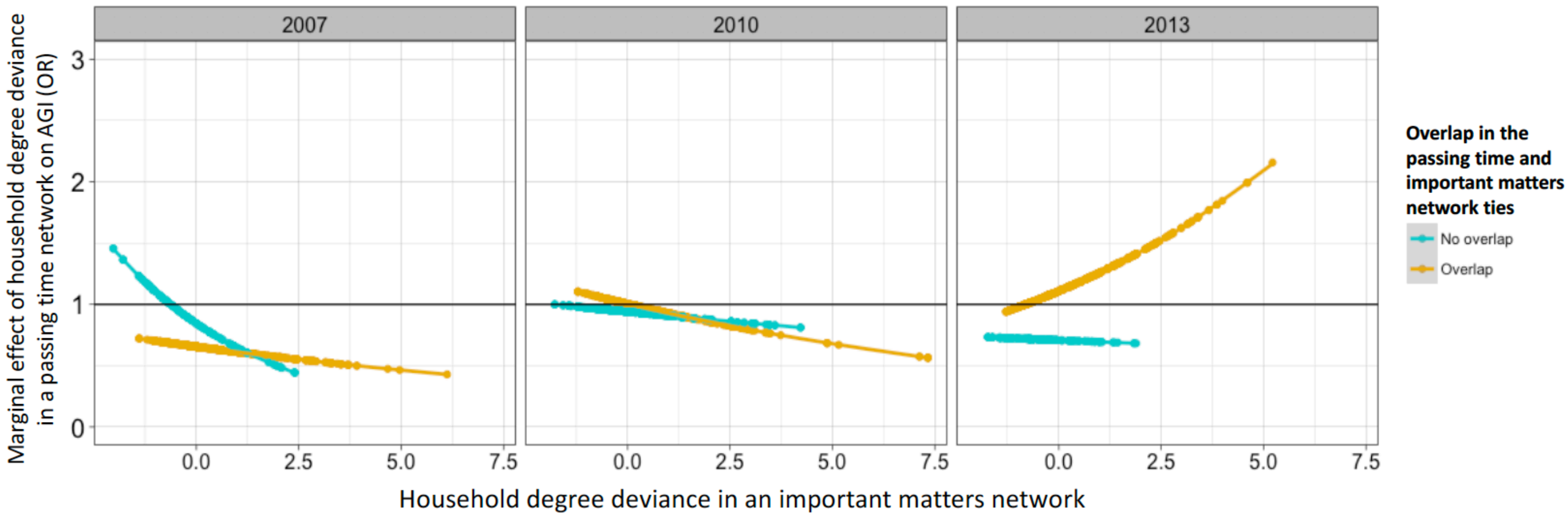


A network figure of modularity, indicated by color, of an important matters network in a remote community (San Miguel) from 2007 to 2013.

Supplementary Table 3.3. Descriptive statistics of the most connected individuals

	2007 N=1005		2010 N=1121		2013 N=1100	
	Passing Time	Important Matters	Passing Time	Important Matters	Passing Time	Important Matters
Outcome						
Acute gastrointestinal illness (AGI)	11.9%	12.0%	11.2%	10.2%	13.2%	12.2%
Covariates						
Household Passing Time Degree	5.8 (0-25)	5.8 (0-25)	3.1 (0-31)	3.1 (0-31)	4.6 (0-33)	4.6 (0-33)
Average Community Passing Time Degree	4.1 (2.2-6.6)	4.1 (2.2-6.6)	2.1 (0.8-3.0)	2.1 (0.8-3.0)	3.5 (1.8-6.2)	3.5 (1.8-6.2)
Household Important Matters Degree	3.1 (0-22)	3.1 (0-22)	2.7 (0-29)	2.7 (0-29)	3.2 (0-20)	3.2 (0-20)
Average Community Important Matters Degree	2.1 0.8-3.3)	2.1 0.8-3.3)	1.8 (0.5-3.0)	1.8 (0.5-3.0)	2.3 (1.1-4.1)	2.3 (1.1-4.1)
Age	39.3 (13.0-87.8)	40.1 (13.0-87.8)	39.8 (13.0-91.7)	41 (13.0-91.2)	41.2 (13.0-92.2)	41.9 (13.1-92.2)
Sex (Female)	44.3%	47.6%	54.3%	54.2%	53.0%	53.6%
Trust (Yes)	54.7%	54.5%	48.5%	48.9%	34.2%	35.2%
Number of organizations	1.8 (0-7)	1.8 (0-7)	1.7 (0-8)	1.6 (0-8)	1.0 (0-7)	1.0 (0-7)
Household size	2.2 (1-8)	2.2 (1-8)	2.1 (1-10)	2.1 (1-10)	2.1 (1-8)	2.1 (1-8)
Number infected in Household	0.22 (0-3)	0.22 (0-3)	0.21 (0-4)	0.21 (0-4)	0.26 (0-3)	0.26 (0-3)
Household asset deprivation	70%	70%	66%	66%	61%	61%
Households highest education is primary	32.3%	32.3%	28.5%	28.5%	26.1%	26.1%
Percentage belonging to a women's group	13.5%	13.8%	14.3%	14.6%	6.5%	7.4%
Percentage overlap of social ties between networks						
Overall	26.2%		23.6%		35.2%	
What percentage of the passing time ties among the most connected household members are also ties to individuals they go to for important matters?	27.5%		25.7%		35.3%	
What percentage of the important matters ties among the most connected household member are also ties to individuals they pass time with?	33.5%		27.5%		40.4%	

Supplementary Figure 3.5. Marginal effect of household networks by network overlap



Marginal effect of the odds ratio of having AGI for every one unit increase in household degree deviance in a passing time network as household degree deviance in an important matters network increases given the joint effects model, predicted for individuals where there is any overlap in ties between the passing time and important matters network and where there is no overlap, 2007-2013.

Chapter IV

The intangible essence of community and its relationship to WASH intervention success in rural Ecuador: a mediation analysis

4.1 Abstract

Diarrheal disease results in 1.3 million deaths annually in the world, primarily in low-resource settings, and requires implementation of safe water, sanitation, and hygiene (WASH) practices to mitigate risk aside from vaccination. Such WASH practices are generally adopted through behavior change and infrastructural development, however, can also be influenced by social constructs at the community-level. While studying WASH behaviors at the individual- and household-levels is important for sustaining behavior change, research is lacking on the influence and significance of community-level social factors, particularly as they pertain to maximizing social influence for behavior change. Previously, we've shown that community cohesion, as defined by a social network of core discussants, impacts the reduction of acute gastrointestinal illness (AGI) in rural Ecuadorian communities. Using data from the same communities, here we estimate whether engaging in safe WASH practices, including hygiene, community sanitation, and improved water source, mediates the relationship of community cohesion on AGI from 2007 to 2013. We do so by modeling 1) the effect of community cohesion on WASH practices, and 2) the effect of WASH practices on AGI when community cohesion is controlled for. Our results demonstrate partial mediation of the protective effect of community cohesion on AGI by community sanitation and improved water use over time, suggesting the importance of social constructs at the community-level for intervention implementation and in turn the reduction of diarrheal disease. Our results also underscore the important role of infrastructure development in changing access to WASH facilities over time.

4.2 Introduction

Diarrhea is an important disease globally, resulting in approximately 1.3 million death annually, with over 500,000 deaths occurring among children under five [71]. Disproportionately affecting young children and low-resource settings, it is a leading cause of disability adjusted life years (DALYs) at 71.6 million DALYs [71]. Despite significant reductions in disease burden in the last decade, diarrheal disease continues to persist in low-resource settings, primarily through human contact and contaminated environment, including water, food, sanitation, and lack of hygiene [8]. Aside from vaccination, well known measures of diarrheal disease prevention include implementation of safe water, sanitation, and hygiene (WASH) practices [8], which are largely adopted through both behavior change and infrastructural development.

To reduce diarrheal disease burden, early intervention trials since the 1970s focused on a single intervention at a time, initially assuming water quality as the critical source of diarrheal infection and investigating the effect of expanding public water services [14], followed by focusing on improving sanitation alone [17]. By the 1990's, however, hygiene interventions like hand-washing emerged as an important tool for diarrheal disease reduction and there was a shift from focusing on public interventions to private individual- and household-level interventions [24]; studies started intervening on multiple transmission pathways [20–23]. Studies that investigated the use of more than one intervention, like hand-washing and household water treatment (HWT), showed implementation of more than one intervention in the household had no greater benefit than only implementing one [26], except when environments contain many sources of contamination [8,27]. Presently, there has a been a return to the early dogma as behavior change at the individual- and household-level has proven difficult, to a more centralized intervention approach and expanding clean public water services [126,127].

Although centralized WASH systems are implemented at the community-level, few studies have highlighted the utility of community engagement in intervention implementation [19]. Though studies have focused on the importance of community health-care workers and emphasized intensive promotion is necessary for sustained uptake [128,129], the attention has not been on community engagement at large, but rather on trained community health-care workers. Indeed, few theoretical frameworks for behavior change used to precede intervention implementation take a broader ecological approach and place individual- and household-level interventions within a multi-level causal framework [10]. One such model for water treatment and safe storage, by Figuerora and Kincaid, suggests interventions influence behavior outcomes by individual-, household-, and community-level factors, including community action and cohesion.[40] Another model, by Dreibelbis et al., suggests an integrated behavioral framework for all WASH interventions that extends beyond the individual- and household-levels, and theorizes that interventions acting at the structural level have the capacity to reach large sections of a population and be highly cost-effective.[10,41] Studying WASH behaviors at each of these levels is important for sustaining behavior change, but research is especially lacking on the influence and significance of community-level social factors, particularly as they pertain to maximizing social influence for behavior change [130].

Thus, observing how community-level social constructs or community cohesion influences the adoption of WASH practices will help inform our understanding of the impact of these multi-level causal frameworks on reducing diarrheal disease. Though community cohesion has been defined a multitude of ways, here we use longitudinal social network data from rural Ecuador to describe the community social environment. We obtained data from a name-generator survey where participants were asked to list individuals they visit to discuss important matters with. In sociology, this is referred to as a core discussion network (CDN). Ties in a CDN correspond to people an individual may not necessarily feel close to, but who are context dependent social supporters [95], likely chosen intentionally and selectively [96]. The benefits of the social connection, including information transfer, influence, and solidarity, are then sought through verbal exchange with such context-dependent social supporters. Thus, CDNs map access to ideas or resources that an individual might activate in forming attitudes or in pursuing goals [96], like improving WASH practices.

While social networks are commonly used to study the transmission of infectious diseases as they provide a map of direct contact for person-to-person transmission [62,63], here we examine the CDN as a conduit of information spread for dissemination of safe WASH practices. In prior studies, we have shown the CDN is protective against acute gastrointestinal illness (AGI) over time at the household- and community-levels in rural Ecuador. In addition to community, however, local infrastructural development importantly plays a role in access to WASH. Sanitation and centralized water systems require construction and are difficult to build by communities alone when access to resources is limited due to poor road infrastructure and political neglect. In Ecuador, our study period extends over six years when infrastructural development like roads connecting the nearest township to the coast and capital and smaller roads connecting some communities to the nearest township were built. Thus, different than other datasets used in WASH studies, we can examine remoteness and infrastructural development as significant modifiers of the effect of community on WASH and diarrheal disease.

Focusing on the role of community cohesion, we first estimate the effect of community cohesion, as measured by a community-level metric of the CDN, on hygiene practices, community sanitation, piped water, and rain water in communities in rural Ecuador from 2007 to 2013. To elucidate why a CDN is increasingly protective, we then estimate whether WASH infrastructure mediates the protective effect we observed of community cohesion on AGI in the same communities and how remoteness modifies this effect. We additionally show the effects of having trust in community and participation in institutional organization, other self-report measures of cohesion, on WASH practices.

4.3 Methods

Data

We collected sociometric and census data from 23 villages in northern coastal Ecuador, in the province of Esmeraldas. Sociometric data was collected during three cross-sectional waves in 2007, 2010, and 2013 from all consenting community members ≥13 years of age in 20 of the communities where the study population consisted of primarily Afro-Ecuadorians, Mestizos, and Chachis, an indigenous group of the Cayapas River in the region. Sociometric data was collected during a single wave in 2010 from all consenting community members ≥13 years of age in 3 majority Chachi communities. Census data was collected from all community households prior to each sociometric survey. Compared to village censuses, the average sociometric response rate across communities was approximately 80% each wave. We also collected case-control data from the same communities to estimate risk factors for diarrheal disease in the study region. Census and case-control data was collected yearly from 2003 - 2013, not just during years when the sociometric survey was administered.

In 2016, we visited 18 of the study communities, including all 3 Chachi communities, and collected qualitative data through focus groups with community leaders, where we inquired more about community social organization and WASH practices. We highlight these results in the discussion to help elucidate our findings and focus on the qualitative findings in the Chachi communities in the results. All study participants provided informed consent and all data collection protocols were approved by institutional review boards at the University of Michigan and the University of San Francisco of Quito.

Outcome

Self-report diarrhea and fever data was collected in the sociometric survey. Participants were asked if they had a fever in the last week and if they had three or more liquid stools in one day in the last week. We combined these two measures to assess an individual's risk of having Acute Gastrointestinal Illness (AGI) to achieve more specificity in the context of enteric disease than just diarrhea. Investigators have used different terms for gastrointestinal illness, including Intestinal Infectious Disease [84,85] and Highly Credible Gastrointestinal Illness (HCGI) [86]. We define AGI as having diarrhea or fever, similar to other studies [87,88], and consider AGI as a binary outcome.

Exposure

We measured our primary exposure, *community cohesion*, using social network data collected from name generator questions on the sociometric survey. Participants interviewed, or *egos*, were asked to identify members of their village outside their household with whom they can discuss important matters. We refer to this as a core discussion network (CDN). The names generated thus created an "ego perceived" network at the individual-level. To measure community cohesion, we averaged the individual-level number of social ties, or *degree*, within each community. The average community degree was measured in 1-unit increments. The number of social connections per person in a CDN positively affects overall community social connectedness [69,70].

We additionally examined *trust* and *organizational belongingness* an individual belongs to as proxy measures of *community cohesion*, but measured at the individual-level. Trust, an indicator of individual perception, was measured by asking whether people generally trust one another in the entire community. Trust was used as a binary indicator. Participation in institutional organization was measured by asking if survey participants were active in community groups or local organizations in the last 12 months in their village. We specifically asked about 9 groups/organizations. The variable used in the model was the total number of organizations a specific individual belonged to and was measured in 1-unit increments. Both trust and organizational belongingness are factors of social influence not accounted for in structural, network measures

WASH measures

We examined four measures of WASH practice at the household-level: hygiene, sanitation, piped water, and rain water. All data was obtained from either the census or case-control surveys, and matched to households in the sociometric dataset by wave of data collection. If a household was missing data for any WASH variable in a wave, we imputed household data from the previous year of data since census or case-control data was collected yearly. We assumed that WASH measures that are more infrastructure dependent are less likely to change in a year. This was primarily done for the 2007 data wave as many of the household WASH observations were collected on the case-control survey where not every household in a community was surveyed. The 2010 and 2013 waves of data were matched to household data obtained from the census.

We measured hygiene as a score based on the proportion of 'Yes' answers to a series of 23 binary response questions related to hygiene conditions inside and outside the household. The full list of questions is in Supplementary Figure 4.1. There was approximately 11% of households with missing hygiene score data across all data waves. As conceptualized here, hygiene score is a marker of individual decision-making within a household.

We measured sanitation as the proportion of households within a 500 meter radius of each household with improved sanitation. We first determined whether each household in a community had improved sanitation using the 2015 World Health Organization (WHO)/UNICEF Joint Monitoring

Program (JMP) definition [131]: a flush or pour flush system to a piped sewage system, septic tank, or pit latrine. This was a binary indicator per household. We then calculated the proportion of households within a 500m radius of each household that has improved sanitation. We refer to this as community sanitation. There was approximately 5% of households with missing sanitation data across all data waves.

We measured a household's water source two ways: by whether a house had piped water and whether a house collected rain water for use. Both measures were used as binary indicators. These measures were asked jointly and originally meant to derive a composite measure of improved drinking water using the 2015 WHO/UNICEF JMP definition, however, given recent updates to the improved drinking water definition [126], which now includes having access to water on the premise, availability when needed, and contamination free water, we have chosen to leave rain water and piped water separate. Notably, it is possible that both types of water sources can be contaminated and so should be estimated separately. There was approximately 13% of households with missing water data in 2007, 8% in 2010, and 1% in 2013.

Covariates

Other measures accounted for in the models are: age, sex, and remoteness. The study villages exist along three river basins: Cayapas, Santiago, Ónzole and vary by remoteness, which is a function of time and travel cost to the nearest township, Borbón [101]. Since 1996, paved roads have been built connecting this township to the coast and Andes. Smaller roads continue to be built linking villages to the main road. Remoteness may affect both *community cohesion* and AGI and was therefore used as a continuous variable in study models. Remoteness was normalized by rescaling each community's remoteness value to be between zero and one, with the most remote community having a remoteness of one. For more details see Eisenberg et al. 2006 [101]. We obtained individual age (continuous) and gender (binary) data from the census.

From the census, we additionally examined household size (continuous), highest household education (categorical), highest household education of women (categorical), and community-level percent asset deprivation (continuous) as possible confounders. To measure community asset deprivation, we first measured household asset deprivation using the Multidimensional Poverty Index standard of living indices. Deprivation was indicated by not having at least one asset related to information (e.g. TV, stereo, cell), or having at least one asset related to information but not having at least one asset related to mobility (e.g. canoe, motorbike, bicycle, motor), or not having at least one asset related to livelihood (e.g. fridge, arable land, livestock) [102]. After measuring this at the household-level, we then calculated community-level asset deprivation by summing the total number of households with asset deprivation and dividing by the total community population at each time-point. None of these covariates were significant in our model and are not shown in this analysis.

Mediation analysis

As our data was not originally designed to estimate causal mediation effects, our study was not powered to examine three-way interactions and thus exposure-mediator, exposure-outcome, or mediator-outcome confounding as is done in the causal mediation literature [132]. As a result and to avoid complexity, we use the product method, which is used more frequently in social science, to examine mediation in each wave of data cross-sectionally [132], and assume the mediators do not affect one another (i.e. one WASH practice doesn't influence the occurrence of another WASH practice). For each wave, we thus examine 1) a mediator model: the effect of community cohesion on the mediator controlling for other covariates, and 2) an outcome model: the effect of community cohesion and the mediator on AGI controlling for other covariates, where the mediator is each WASH practice modeled separately.

For the outcome models, our data exhibit both separation, which occurs for binary outcomes when regions of the covariate space have no variation in outcome (all one or all zero) [74], and correlation, as individuals are nested within households and households nested within communities. Standard data analyses dealing with separation generally use a penalized likelihood regression [74], while standard approaches for dealing with correlation consist of hierarchical linear models. We therefore used a two-stage Bayesian analysis approach with a binomial distribution that: 1) conducted a firth penalized likelihood logistic regression to account for data separation in our rare outcome [103], and 2) used these fixed effects as normally distributed weakly informative priors in a Bayesian hierarchical model to account for the nested data structure. We assumed a non-informative Inv-Gamma(10^{-3}, 10^{-3}) prior for the random effect variance at the household-level. Including an additional random effect variable for the community did not yield significantly different results. Using Stan, our model uses a Hamiltonian Monte Carlo and No-U-Turn Sampler (HMC-NUTS) (30). We ran each model for 5 chains of 10,000 iterations with a thinning of 5.

Due to the binary nature of AGI, we used the following general hierarchical model structure, whereby we considered a random effect at the household-level for each year's data. Anyone who shares the same index (*i, j, k*) are correlated:

Level 1 individual regression equation:

$$logit(p_{ijk}) = log\left[\frac{p_{ijk}}{1-p_{ijk}}\right] = log(odds) = \beta_{0jk} + \beta_1 sex_{ijk} + \beta_2 age_{ijk} + \beta_3 trust_{ijk} + \beta_4 organizations_{ijk}$$

Level 2 household regression equation:

$$\beta_{0jk} = \gamma_{00k} + \gamma_1 WASH\ practice_{jk} + \delta_{0j0}$$

Level 3 community regression equation:

$$\gamma_{00k} = \pi_{000} + \pi_1 remoteness_k + \pi_2 average\ community\ degree_k$$

where random effect $\delta_{0j0} \sim N(0,\sigma_\delta^2)$ is the j^{th} level adjusted for covariates on the intercept, and i= individual(i = 1,…,N), j= household (j = 1,…,n_j), k= community(k = 1,…, n_k). The household WASH practice includes either hygiene score, community sanitation, having piped water, or rain water.

For the mediator models, we used the two-stage Bayesian hierarchical model for modeling both piped water and rain water as outcomes, as these variables exhibited statistical separation. Due to the binary nature of these WASH practices, we used the following general hierarchical model structure, whereby we considered a random effect at the household-level as our outcome was now household-level WASH practices and assumed a binomial distribution. Anyone who shares the same index (*i, j, k*) are correlated

Level 1 individual regression equation:

$$logit(p_{jk}) = log\left[\frac{p_{jk}}{1-p_{jk}}\right] = log(odds) = \beta_{0jk} + \beta_1 sex_{ijk} + \beta_2 age_{ijk} + \beta_3 trust_{ijk} + \beta_4 organizations_{ijk}$$

Level 2 household regression equation:

$$\beta_{0jk} = \gamma_{00k} + \delta_{0j0}$$

Level 3 community regression equation:

$$\gamma_{00k} = \pi_{000} + \pi_1 remoteness_k + \pi_2 average\ community\ degree_k$$

where random effect $\delta_{0j0} \sim N(0,\sigma_\delta^2)$ is the j^{th} level adjusted for covariates on the intercept, and i= individual(i = 1,…,N), j= household (j = 1,…,n_j), k= community(k = 1,…, n_k). We use priors generated from a penalized likelihood logistic regression to account for data separation and use these fixed effects as normally distributed weakly informative priors in a Bayesian hierarchical model to account for the nested data structure. We assumed a non-informative Inv-Gamma(10^{-3}, 10^{-3}) prior for the random effect variance at the household-level. Using Stan, our model uses a Hamiltonian Monte Carlo and No-U-Turn Sampler (HMC-NUTS). We ran each model for 5 chains of 10,000 iterations with a thinning of 5. For all the two-stage Bayesian hierarchical models, we report the 2.5% and 97.5% credible intervals.

For the household WASH practices of hygiene score and community sanitation, which had no separation issues, we used a Generalized Linear Mixed Model (GLMM) with a binomial distribution as these mediators are proportions that are frequencies (i.e. a count of binary outcomes of a Bernoulli-distributed random process). Such data can be modeled with the assumption of a binomial error structure using logistic regression [133]. We further chose to model hygiene and community sanitation as binomially distributed in hierarchical models to make calculating the indirect effects with the binomially distributed outcome models more straightforward and comparable. Thus, for data waves, 2007, 2010, 2013, we used the same general hierarchical model structure as above, where random effect $\delta_{0j0} \sim N(0,\sigma_\delta^2)$ is the j^{th} level adjusted for covariates on the intercept, and i= individual(i = 1,…,N), j= household (j = 1,…,n_j), k= community(k = 1,…, n_k). For the GLMM models, we bootstrapped each model, resampling from the data for 1,000 iterations each, to obtain 97.5% confidence intervals for all effect estimates.

To estimate the indirect effects of community cohesion on AGI (i.e. the reduction of the effect of the exposure variable on the outcome when the mediator is present), we then report the following product $\pi_2 \cdot \gamma_1$ for all WASH variables where π_2 is estimated in the mediator model and γ_1 is estimated in the outcome model. For hygiene and sanitation, γ_1 estimates were obtained from bootstrapping the GLMM models and the π_2 estimates obtained from iterating the Bayesian hierarchical model to calculate the confidence intervals of the indirect effects. As shown before, standard errors of the indirect effects can be derived by using bootstrapping [134]. For logistic regression and the odds ratio scale, the product method in mediation analysis does not often yield numerically identical results to the difference method [135], however, in circumstances when the outcome is rare, there is normally distributed error in regression assuming models are correctly specified, and there is no interaction between the exposure of interest and mediator, the product method can approximate the difference method (i.e. the odds ratio approximates the risk ratio) [135]. The natural indirect effect odds ratio thus captures the odds ratio for AGI comparing a particular WASH practice given community cohesion and no community cohesion if the subject had in fact had community cohesion [135]. An indirect effect is significant when the confidence interval of the estimate does not cross 1.

Lastly, though we modeled multiple comparisons, we do not use a Bonferroni correction as it is based on an alpha level with respect to reporting p-values. On the contrary, we bootstrap GEE models and report the 2.5% and 97.5% tails of the distribution for each parameter allowing for less conservative confidence intervals and report credible intervals from the Bayesian hierarchical models, which provide a range of values from the posterior probability distribution derived from multiple iterations. We do not report p-values. We report all estimates in figures on a linearized, log scale.

Software

Network analyses were conducted in R (v. 3.4.2, R Foundation for Statistical Computing, Vienna, Austria) using the package igraph. Regression analyses were conducted in R (v. 3.4.2) using packages brms, rstan, and geepack.

4.4 Results

Our total study population consisted of 2,204 individuals in 2007, 2,371 in 2010, 2,326 in 2013, and 274 individuals in the Chachi communities in 2010. The mean of the average community degree in a CDN was 2.1 ties in 2007 compared to 2.3 ties in 2013, and 1.2 ties in the Chachi communities (Table 4.1). The prevalence of AGI increased from 10.6% to 12.5% in the six-year period, and was 15.2% in the Chachi communities. In the six-year period, the prevalence of household asset deprivation decreased from 68% to 57% along with the highest household education-level being primary school. Household asset deprivation was 48.9% in the Chachi communities.

Overall, rain water usage decreased over six-year period from 41% to 31%. This was differential by remoteness level as the most remote communities maintained ~50% rain water use while communities close to road infrastructure decreased rain water use from 24% to 4% (Supplementary Table 4.1). Presence of piped water was similarly differential by remoteness level, increasing overall from 1.5% to 42%, from 1% to 78% in communities close to road infrastructure and remaining relatively non-existent in communities at a medium distance and increasing only slightly (2.3% to 17%) in the most remote communities. Rain water usage in the Chachi communities was 84% and there was no piped water infrastructure.

The average hygiene score in households increased from 0.62 to 0.74 with a range of 0 to 1 (Table 4.1); the distribution of household hygiene score noticeably shifted toward becoming more hygienic in the six-year period (Figure 4.1). The average hygiene score in Chachi communities was 0.52. The average community sanitation or proportion of households within a 500 meter radius of each household with improved sanitation increased from 0.60 to 0.82 with a range of 0 to 1; the distribution of community sanitation noticeably shifted to increased community sanitation in the six-year period. The average community sanitation in Chachi communities was 0.49 with a wide distribution.

Without considering WASH practices as mediators in the relationship between community cohesion and AGI, having a one-unit increase in average community degree in a CDN resulted in 0.89

(0.74, 1.06) fewer odds of AGI compared to having no increase in average community degree in 2007 (Supplementary Table 4.2). This protective effect became stronger in 2010 and 2013 with a one-unit increase in average community degree in a CDN resulting in 0.74 (0.63, 0.87) of AGI compared to having no increase in average community degree in 2013. In Chachi communities, the odds ratio of community average degree regressed on having AGI was 0.47 (0.13, 1.61).

For hygiene, average community degree consistently predicted increased hygiene over time (2007: 1.99 (CI: 1.53, 2.82), 2013: 1.75 (CI: 1.04, 3.28)) in the mediator model (Supplemental Table 4.3). Trust and organizational belongingness had null effects. In 2007, being more remote resulted in 0.38 (0.21, 0.66) fewer odds of having increased hygiene compared to being less remote. There was no effect of remoteness on hygiene in 2010 or 2013. Community cohesion had a significant effect on hygiene in the Chachi communities (Figure 4.6A). In the outcome model, every one-unit increase in hygiene predicted 0.57 (0.34, 0.94) lower odds of AGI in 2007 (Supplemental Table 4.3). This effect was slightly attenuated by 2013 (Figure 3). For every one-unit increase in average community degree, the odds of AGI was 0.79 (0.65, 0.97) lower in 2007. This effect became attenuated over time (2013: 0.90 (CI: 0.73, 1.10)). Trust, on the other hand, was significantly protective by 2013 (0.72 (0.59, 0.88)), while organizational belongingness had a null effect. Remoteness became more protective against AGI over time when hygiene was also controlled for. In the Chachi communities, neither average community degree nor hygiene had significant effects on the odds of AGI occurring (Figure 4.6B). However, having trust in the community was a significant predictor of AGI (3.31 (1.59, 7.01)). There were no significant indirect effects of community cohesion, as defined by average community degree, on AGI through hygiene (Figure 4.7). Indirect effects were marginal for trust and organizational belongingness in 2007-2010 (Supplementary Figure 4.2).

For community sanitation, average community degree consistently predicted increased community sanitation over time (2007: 1.28 (CI: 1.15, 1.33), 2013: 2.25 (CI: 2.01, 3.10)) in the mediator model (Supplemental Table 4.4). Trust and organizational belongingness had null effects. In 2007, being more remote resulted in 1.87 (1.72, 2.17) greater odds of increased community sanitation compared to being less remote, however, this effect switched directions in 2010 or 2013 (Figure 4.3). For every one-unit increase in average community degree in the Chachi communities, the odds of community sanitation decreased 0.02 (0.00, 0.20). In the outcome model, every one-unit increase in community sanitation predicted 0.59 (0.31, 1.14) lower odds of AGI in 2007 (Supplemental Table 4.4). This effect switched directions in 2010 and 2013 (Figure 4.3). For every one-unit increase in average community degree, the odds of AGI was 0.84 (0.69, 1.01) lower in 2007. This protective effect became stronger over time (2013: 0.79 (CI: 0.64, 0.97)). Trust became significantly protective by 2013 (0.69 (0.57, 0.83)), while organizational belongingness had a null effect. Remoteness became less protective against AGI over time when sanitation was also controlled for. In the Chachi communities, every one-unit increase in average community degree resulted in 0.13 (0.02, 0.74) lower odds of AGI and increased community sanitation resulted in 0.27 (0.06, 1.16) lower odds of AGI. Trust had the opposite effect on AGI (3.27 (1.61, 6.95)). There were significant indirect effects of average community degree on AGI through community sanitation in 2010 and 2013, and for the Chachi communities (Figure 4.7).

For piped water use, average community degree did not have a significant effect on having piped water in 2007 in the mediator model. In 2010, every one-unit increase in average community degree resulted in 2.84 (2.21, 3.66) greater odds of having piped water and in 2013 the effect direction changed (0.68 (0.54, 0.84)) (Supplemental Table 4.5). Trust and organizational belongingness did not have significant effects. Over time, every one-unit increase in remoteness resulted an even lower likelihood of having piped water (Figure 4.4). In the outcome model, having piped water went from resulting in greater odds of AGI in 2007 (1.93 (0.98, 3.82)) to resulting in lower odds of AGI by 2013 (0.76 (0.63, 0.93)). Average community degree was protective against AGI over time when piped water was controlled for, with the effect estimate crossing the null by 2013 (0.86 (0.70, 1.05)). Trust became significantly protective by 2013 (0.73 (0.60, 0.88)), while organizational belongingness had a null effect. Remoteness was significantly protective in 2007 and 2013 when piped water was also controlled for. There were significant indirect effects of average community degree on AGI through piped water in 2010 and 2013 (Figure 4.7).

For rain water use, average community degree did not have a significant effect on rain water in 2007 in the mediator model. In 2010, every one-unit increase in average community degree resulted in 0.38 (0.32, 0.46) lower odds of having rain water and in 2013 the effect direction changed (1.38 (1.11, 1.74)) (Supplemental Table 4.6). Having trust in community resulted in 1.55 (1.27, 1.90) greater odds of having rain water in 2013 while belonging to more organizations resulted in 0.90 (0.83, 0.97) lower odds having rain water. Remoteness was an increasingly strong predictor of having rain water between 2007 to 2013 (Figure 4.5). In the Chachi communities, every one-unit increase in average community degree resulted in 0.10 (0.02. 0.45) lower odds of having rain water. In the outcome model, rain water resulted in greater odds of AGI (1.37 (1.12, 1.69)) in 2007. This effect became attenuated over time and switched direction, however, was not significant. For every one-unit increase in average community degree, the odds of AGI was 0.82 (0.67, 1.00) lower in 2007. This effect became stronger in 2010 (0.66 (CI: 0.53, 0.84)) and attenuated in 2013 (0.86 (0.71, 1.06)). Trust became significantly protective by 2013 (0.73 (0.60, 0.89)), while organizational belongingness had a null effect. Remoteness became less protective against AGI over time when rain water was also controlled for. In the Chachi communities, every one-unit increase in average community degree resulted in 0.25 (0.07, 0.96) lower odds of AGI and having rain water resulted in 0.30 (0.17, 0.54) lower odds of AGI. Trust had the opposite effect on AGI (3.63

(1.75, 7.76)). There were significant indirect effects of average community degree on AGI through rain water use in 2007 and 2010, and for the Chachi communities (Figure 4.7).

Qualitative findings

In the focus groups with leaders, all Chachi communities reported no outside agencies, like government or development agencies, visiting or working in the communities. While all communities reported some form of a collective or social organization, only one community thought they had more success as a community. The other two communities felt they had less success because they lacked organizational direction, unity, and didn't work well together, though one of these communities stated building a soccer field in the last 3 years. The community that felt they had more success than other communities said they meet every two weeks as a community to discuss issues, that they were more organized as they built a staircase from the riverbed to the community dwellings to make accessing canoes and water easier, and that they were better at disposing of trash. The social groups across communities included a central Chachi group across all neighboring Chachi communities, cacao farming, soccer club, Evangelical group, Catholic group, school committee, men's group. One community stated they have a social group with the nearest Afro-Ecuadorian community.

One community reported solving community issues in the year prior, one reported having solved a community issue in the last year but did not list what those problems solved were, and one community said they had not solved any issues in the last year. Overall, the problems listed included the economy and lack of resources like food, lack of education, having land that was disconnected by the river, having land disputes, benefits from cacao farming, lack of toilets, and moral and social concerns including incest.

Only one community reported having a female leader. She was a part of the school. The other communities reported no female leaders. Though one community stated having a women's group, when spoken to separately, the women stated no such group existed, that women do not have a role in community leadership, and spoke of domestic violence issues.

4.5 Discussion

In this paper, we demonstrate partial mediation of the protective effect of community cohesion on acute gastrointestinal illness (AGI) by community sanitation and improved water use over time, suggesting the importance of social constructs at the community-level for intervention implementation and in turn the reduction of diarrheal disease (Figure 4.8). Our results also underscore the important role of infrastructure development in changing access to WASH facilities over time. As shown in prior studies in this region, communities that are closest to a road have gained access to WASH infrastructural development while communities that are farther away and have not provided the government or outside agencies with economic incentives (like cacao or gold) have been consistently ignored. In contrast, in the indigenous communities (the Chachis), having greater community cohesion resulted in smaller likelihood of community sanitation and rain water use, though both community sanitation and rain water were protective against AGI. This indicates that the concept of community cohesion is represented differently in these communities and therefore has the opposite effect. Below we expand on these key results by discussing: the effect of community cohesion on safe WASH practices, the effect of varied WASH practices on AGI, the mediation effect of WASH practices on the relationship between community cohesion and AGI over time, and the effects found in the Chachi communities.

Community cohesion and WASH practices

Higher average community degree in a CDN (a network consisting of relationships with whom people discuss important matters) results in a higher hygiene score and community sanitation over time. This has opposing effects on rain and piped water likely due to infrastructural development within and near the communities over time as rain water increases and piped water decreases in more remote communities. Importantly, for all the WASH practices and communities examined, average community degree was a stronger measure of social influence than both trust and participation in institutional organizations, which had limited effects on increasing any of the WASH practices over time; degree within a CDN is a seemingly more accurate measure of influence with respect to disease.

The social properties, however, assessed through the measure of degree is likely different for each WASH practice or intervention as the environment around a household gets cleaner by different mechanisms of human behavior and social change (Figure 4.9). The CDN, specifically, extends the functional specific hypothesis that individuals seek out conversations with others they believe to be helpful, knowledgeable, or sympathetic about a particular topic [96]. Therefore, the community social construct we describe here is at baseline a measure of how social behavior is influenced by verbal exchange with general and problem-specific discussion networks with other community members. It is this network that then has a differential effect on social processes at the individual-, household-, and community-levels, which in turn affect adoption of different WASH practices and influence AGI.

Increased community cohesion consistently resulted in better hygiene over time. Hygiene, which is more a measure of individual-level decision making, was not affected by remoteness of a community; in 2013, the mean hygiene score was not differential by remoteness category (Close: 0.73, Medium: 0.73, Far: 0.75). Rain water, on the other hand, is more a marker of household-level decision, while community sanitation and piped water access are markers of the ability of a community to construct facilities for use. Hence, in discussing the effects that community cohesion has on these different WASH practice, we are more realistically referring to the different social processes that social ties aggregated at the community-level affect, like behavior versus construction.

The effect of increased community degree in a CDN resulted in increased rain water use in 2007, decreased rain water use in 2010, and significantly increased rain water use in 2013. Rain water also increased dramatically from 2007 to 2013 as remoteness increased. Harvesting rain water, a marker of household decision making as water is shared at the household-level[127], is likely a function of seasonality and water contamination of the rivers as all the remote communities are situated by riverbeds and would otherwise use river water for daily activities. The qualitative data brought to light issues of mercury contamination from gold mining and other chemicals from industries like palm oil. Thus, community members were forced to use more rain water as there was limited access to other types of water systems aside from seasonal drivers (i.e. when there is rain or flooding versus not). As the percentage of rain water use decreased from 41% to 4% in communities with access to a road, rain water use remained at 54% in the more remote communities (Supplementary Table 4.1), where community cohesion is consistently higher.

Remoteness had the opposite effect on having piped water as a water source. In 2007, only 1.5% of households had access to piped water (Supplementary Table 4.1). By 2013, communities with road access had 78% coverage of piped water in households while more remote communities had only 17% coverage; the communities at a medium distance had 0% coverage. As such, as remoteness increased, the likelihood of having piped water decreased over time. Piped water, unlike, hygiene and rain water, requires access to resources and more strongly requires decision-making at the community-level. As shown in our prior qualitative work, communities at a medium level of remoteness are continually neglected by outside agencies that may build infrastructure like piped water. Though the more remote communities are visited more frequently by government agencies due to the cacao, gold, and eco-tourism industries, these communities experience more one-time intervention implementations that don't consider quality control or long term sustainability, so some communities have reported piped water systems working in the past but losing functionality over time. While some communities commented on coming together to fix the installed water system, the majority of communities did not. Increased community cohesion, as a result, was associated with increased piped water use from 2007 to 2010, but decreased piped water use in 2013 as communities with smaller network densities are getting more piped water due to infrastructural development and construction coming from outside resources. Different than rain water, the strength of the community network and social solidarity may be more manifested through piped water.

Sanitation, similar to piped water, also relies on access to resources and construction, but is more focused on household-decision making (Figure 6) as sanitation needs may differ across the community. Increased community cohesion resulted in increased community sanitation, the proportion of households within 500 meters with improved sanitation, over time; the effect slightly decreased from 2010 to 2013, which was likely an indicator of the difficulty of community cohesion or collective action to influence sanitation adoption. Though there have been examples of success of collective action, addressing collective sanitation problems within a community may in fact decrease as proximity to sanitation matters which means that needs differ across the geography of a community (i.e. open defecation in a local area may impact one group, while poorly constructed and maintained pit latrine affects another, and the collective risks are unlikely to overlap) [136]. Nonetheless, studies have shown that the key-decision makers in informal settlements for sanitation interventions are both landlords and tenants, indicative of household heads and community leaders being important in more rural settings [137]. Like piped water, increased remoteness resulted in decreasing likelihood of community sanitation over time, similarly indicative of a lack of access to resources. In 2013, the mean proportion of households with improved sanitation within 500 meters was relatively the same between the least and most remote communities and markedly less for communities at a medium level of remoteness (Close: 0.85, Medium: 0.59, Far: 0.86).

WASH practices and AGI

Having better hygiene is a predictor of reduced AGI in 2007 and 2010. These results are indicative of hygiene being more a measure of individual decision-making, with the effect attenuating in 2013. Though the overall hygiene score is increasing over time, it's possible that long-term utilization of the hygienic practices is faltering and survey responses were biased. Hygienic practices are important measures for mitigating further microbial contamination in the household, but also require adherence [37]. Additionally, though we estimate the effects of WASH practices on AGI, the improved water and sanitation variables do not capture behavior but rather access and thus our results represent the effect of access on AGI when community cohesion is controlled for.

In more remote and low-income settings, improved water sources include having protected groundwater or harvested rainwater, which are alternatives to surface water sources like rivers. However, these sources of water are likely sources of contamination depending on what has been introduced into the harvested water system or how long it's been stagnant and so are not as effective as centralized piped-in water systems [37]. While piped water became increasingly protective over time, rain water shifted from being a risk to having a null effect. Furthermore, as piped water increased drastically over the six-year time-period in communities, the effects of reducing exposure to contaminated water took time to become protective as construction increased and quality improved.

Sanitation, on the other hand, shifted from being protective against AGI in 2007, to being a risk in 2010, and then closer to the null in 2013. This shifting of effect directions is likely an indicator of the decreasing quality of latrines and the subsequent improvement as improved sanitation increased in quantity in 2013 (Figure 1). Improved sanitation is measured by access to a piped sewage system, septic tank, or pit latrine, however, pit latrines have been shown to decrease in quality over time unless maintained, resulting in increased contamination of the local environment and ground water [138] and also increased open defecation [139] increasing diarrheal disease risk through fecal-oral transmission. Multiple studies have highlighted the relationship of unsanitary latrines and a concentration of pathogens in shallow water and the surrounding environment as far as 70 meters away [140,141]. Even without considering the quality of the pit latrine, as the distance between a pit latrine and dug-well decreases, the fecal coliform count increases [142]. Thus, poor quality latrines result in increased local risk of diarrheal disease [107].

Mediation effects

As time increased, we observed an attenuation of the effects of community cohesion on AGI when the WASH measures were controlled for; the mediation effect became stronger over time. This suggests that WASH explains more of the protective effect of cohesion over time as infrastructure development and construction increased, reducing contamination. However, as we only observe partial mediation through community sanitation and piped water in 2010-2013, there are likely other variables not included in our model that account for the remainder of the protective relationship between community cohesion and AGI. The changing directions of the indirect effects are result of the changing directions of the effect of community cohesion on the varied WASH practices. As it is, community cohesion maintains a direct relationship with AGI that is not accounted for through WASH.

This direct relationship that persists between community cohesion and AGI could be explained by other measures of reduced contamination of the environment that result from social gatherings. In the more remote communities, members have engaged in *mingas*, or organized labor, to address community issues like cleaning up trash. Though the social dynamics in more remote communities are shifting as a wage economy was introduced and occupations were transferred from within the community to away from the community, it is possible that the limited social activities communities are engaging in are becoming more effective as the protective effect of community cohesion on AGI increases over time. Other sources of contamination that could be reduced through community cohesion include from livestock, fecal exposure, food, and pathogens in the environment. In addition, increased community cohesion may affect travel to healthcare facilities or pharmacies to reduce infection. Lastly, other measures of socioeconomic status, aside from asset deprivation, could lie in the causal pathway between community cohesion and AGI.

Other measures of cohesion, trust and participation in institutional organization, had limited effects on increasing any of the WASH practices over time, and thus have marginally significant indirect effects through the varied WASH practices, particularly through hygiene. Hygiene is not a community-level measure but is more an indicator of individual-level behavior and so partial mediation was observed with these measures of cohesion.

Chachi communities

In contrast, in the indigenous communities (the Chachis), having greater community cohesion resulted in smaller likelihood of hygiene, community sanitation, and rain water use. There was no piped water in these communities likely due to an even greater lack of access to resources than the more remote majority Afro-Ecuadorian communities. As noted in our qualitative data, the concept of community cohesion is represented differently in the Chachi communities and therefore has an opposite effect. Though the male community leaders stated either having successes as a community or that certain community groups existed, when spoken to separately, women stated that the notion of being socially connected or organized was a façade and that women were not a part of community leadership and didn't have a voice. Women, who are the primary child caretakers and working in and around the households daily, additionally reported cases of intimate partner violence in each community. In any attempt to report such issues to community leaders, they were ignored. Importantly, hygiene practices in the household are generally influenced by women, the primary caretakers, and when their basic rights are abused engaging in such practices likely becomes difficult. In the Chachis communities, both community sanitation and rain water were protective against AGI, which resulted in mediation effects of these two WASH variables on the effect of community cohesion on AGI, making average community degree more protective. The reasons for these effects should be further looked into.

Limitations

Due to our limited sample size and the clustered nature of our data, we were unable to examine three-way interactions and thus multilevel causal mediation. Ideally, we would have been able to examine mediation effects at household-level, where social processes are multidimensional and we can elucidate the effects of different types of social networks on WASH measures and AGI. Indeed, examining social cohesion using social network data allows for a look at not only comprehensive community metrics but the effect that individuals and households within a community have on one another and on community structure as a whole. Using social network data could allow for a hierarchical systems approach: a better understanding of the interdependent relationship between community level factors, households, and the individual in the WASH sector [8]. Still, we likely present some measurement error by not accounting for the multilevel nature of the mediation (i.e. that the exposure is at the community-level, mediator at the household-level, and outcome at the individual-level). Multilevel mediation methods using the counterfactual approach require strict assumptions that are often not met, including that the effect of the exposure on the outcome and mediator on the outcome is unconfounded given covariate history up until the time period [143]. As such, researchers have suggested using marginal structural models or structural equation modeling to estimate the effect of mediation [144]. Multilevel mediation frameworks for binary dependent data should be further looked into including through mechanistic simulations to test mediation. Though our data structure precludes us from fitting a more sophisticated causal model, we have tried to account for the nested structure by using analyzing our data cross-sectionally and using Baysiean hierarchical model and GEE, and then bootstrapping methods for measuring the indirect effects.

Other limitations of our data include possible misclassification of our WASH variables which could lead to non-differential bias. Additionally, harvesting rain water is likely seasonally dependent according to our qualitative data, however, this variable was measure at a single time point in each wave of data. Lastly, the WASH variables of improved water and sanitation in this analysis do not capture usage but rather are a measures of access. Measuring latrine use is especially difficult [145], and as a result, we use the proportion of households within a 500 meter radius with improved sanitation to each household as a proxy measure of improved contamination in the surrounding area and community. WHO/JMP has also recently changed their definition of improved water to include contamination, however, our data did not allow us to measure this.

Conclusion

Social network data provide important insight on how to increase WASH practices. As such, our results have broad implications on intervention strategy through the use of social network data, with an emphasis on the positive effect of community on diarrheal disease reduction through increased WASH practices. In a systematic review done on risk factors for diarrhea mortality in 2012, it was noted that 502,000 deaths were attributable to inadequate drinking water, 280,000 deaths to inadequate sanitation globally, and 297,000 deaths to inadequate hand hygiene, accounting for 58% of the total diarrheal disease burden [146]. These estimates confirm the significance of improving different interventions in the WASH sector. Given our results, engaging with communities and local social groups will be important not only for the implementation of interventions, for both intervention compliance and sustainability as shown through the effectiveness of community healthcare workers [32–34,128], but for the well-being of individuals beyond what WASH accounts for.

Our results highlight that communities with no access to outside resources are continually neglected. Infrastructural development and outside agencies are a large part of how community members can engage in WASH practices, specifically improved water and sanitation. In emerging economies, like Ecuador, there is immense difficulty in gaining the attention of outside resources to improve access in more rural settings as much of the attention is focused in growing urban areas. As shown across multiple studies, having a reliable centralized water system is important for reducing diarrheal disease [147], however, this is largely a byproduct of access to resources and how larger organizational entities like NGOs and governments choose to spend their money. In our study region, attention is primarily focused in areas with economic incentives. Indeed, finding sustainable solutions that involve community engagement and local administrations requires innovation and funding. Though sanitation is less costly than installing a centralized water system, sanitation compliance is more complicated. In contrast, recent studies have shown that regardless of compliance, all improved water source interventions resulted in reduced disease, especially at the household-level.[37]

There are many factors, social, political, economic, and geographical that contribute to increased collective action, the downstream effect of community cohesion. Herd protection, for example, is considered a collective action based on the desire of community members to improve the health of the collective. When women, who are the primary household caretakers, are distressed, however, and suffer from a lack of respect, don't have a role in community, or are inhibited from organizing collectively, the potential for collection action to improve different household WASH measures is likely greatly diminished [136]. As shown in the Chachi communities, this can undermine community cohesion and in turn collective action. Future studies could examine mechanistic models to simulate how networks influence WASH behavior at the household-level and influence individual disease risk, and the probable role of gender.

Lastly, there exists a gradient of social processes for different WASH practices that should be considered during intervention implementation. The protective effect of community cohesion on diarrheal disease is only partially mediated by community sanitation and improved water, but not hygiene as hygiene is a more individual-level measure and not as affected in magnitude by community social structures. Nevertheless, community social structures should be leveraged during intervention implementation through community leaders and healthcare workers at all levels: the individual-, household-, and community-levels. Future studies should examine the effect of community social constructs on behavioral uptake of WASH practices and not just access to WASH.

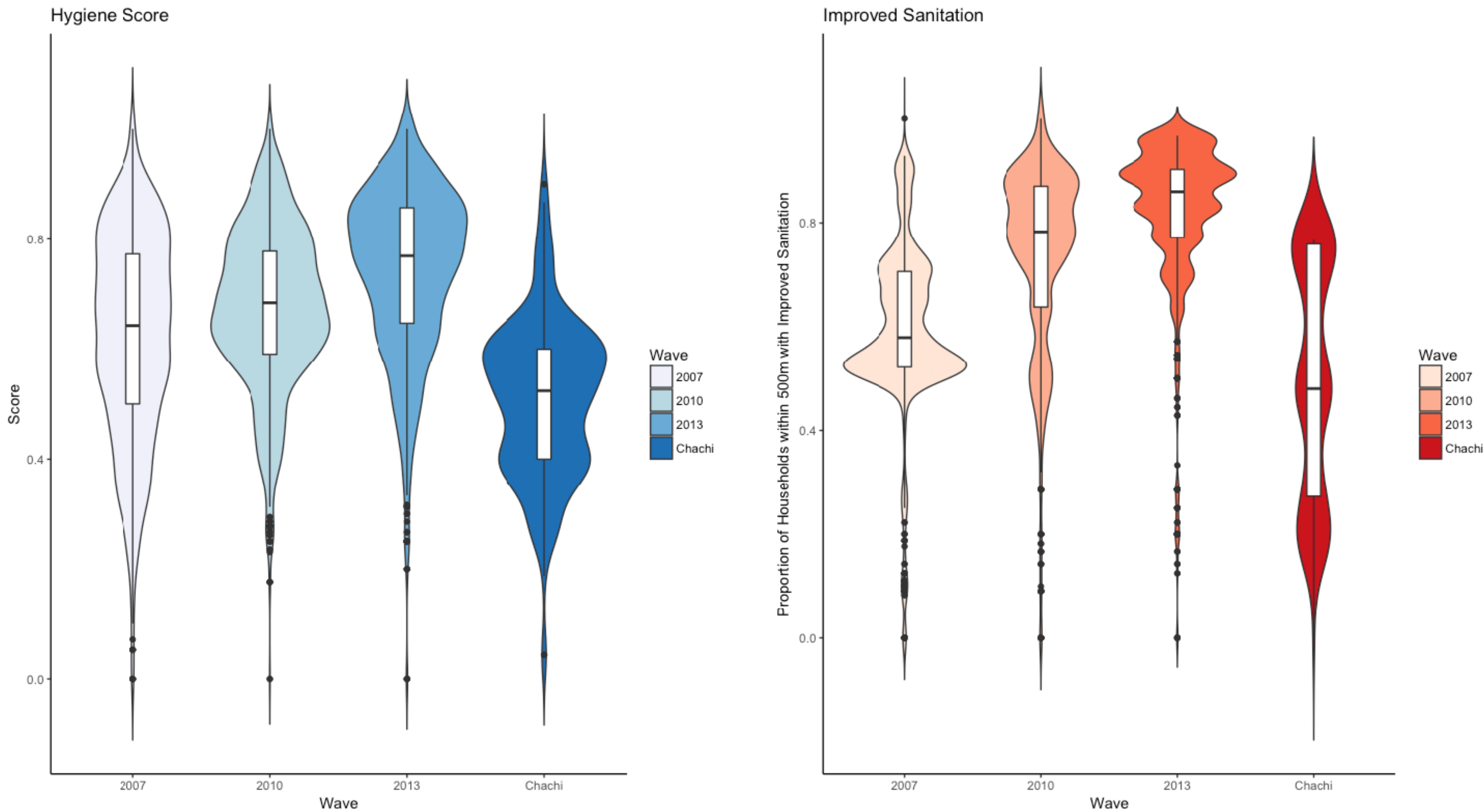
Table 4.1. Descriptive statistics of cross-sectional data and WASH measures

	2007 N=2204 Households=1005	2010 N=2371 Households=1121	2013 N=2326 Households=1100	Chachi communities 2010 N=274 Households=93
Outcome				
Acute gastrointestinal illness (AGI)*	10.6%	10.1%	12.5%	15.2%
Covariates				
Average Community Degree**	2.1 (0.8-3.3)	1.9 (0.5-3.0)	2.3 (1.1-4.1)	1.4 (1.2-1.7)
Age	36.9 (13.0-90.5)	37.7 (13.0-91.7)	37.9 (13.0-95)	34.1 (13.1-75.4)
Sex (Female)	48.8%	52.4%	52.7%	52.0%
Remoteness	0.451 (0.06-1.00)	0.437 (0.06-1.00)	0.442 (0.06-1.00)	0.97 (0.92-1.00)
Trust (Yes)	51.0%	44.0%	32.8%	79.1%
Number of organizations	1.8 (0-8)	1.6 (0-8)	1.0 (0-7)	1.8 (0-8)
Household asset deprivation	68.3%	62.8%	56.7%	48.9%
Households highest education is primary	26.9%	22.5%	19.5%	6.9%
Household WASH practices				
Rain water	40.7%	51.2%	30.6%	84.2%
Piped water	1.5%	23.0%	42.0%	0.0%
Hygiene score (1 indicates having good hygiene)	0.62 (0-1)	0.67 (0-1)	0.74 (0-1)	0.52 (0.04-0.90)
Proportion of households within 500m with improved sanitation	0.60 (0-1)	0.74 (0-1)	0.82 (0-1)	0.49 (0-0.77)

*Having diarrhea or fever
**Using an important matters network

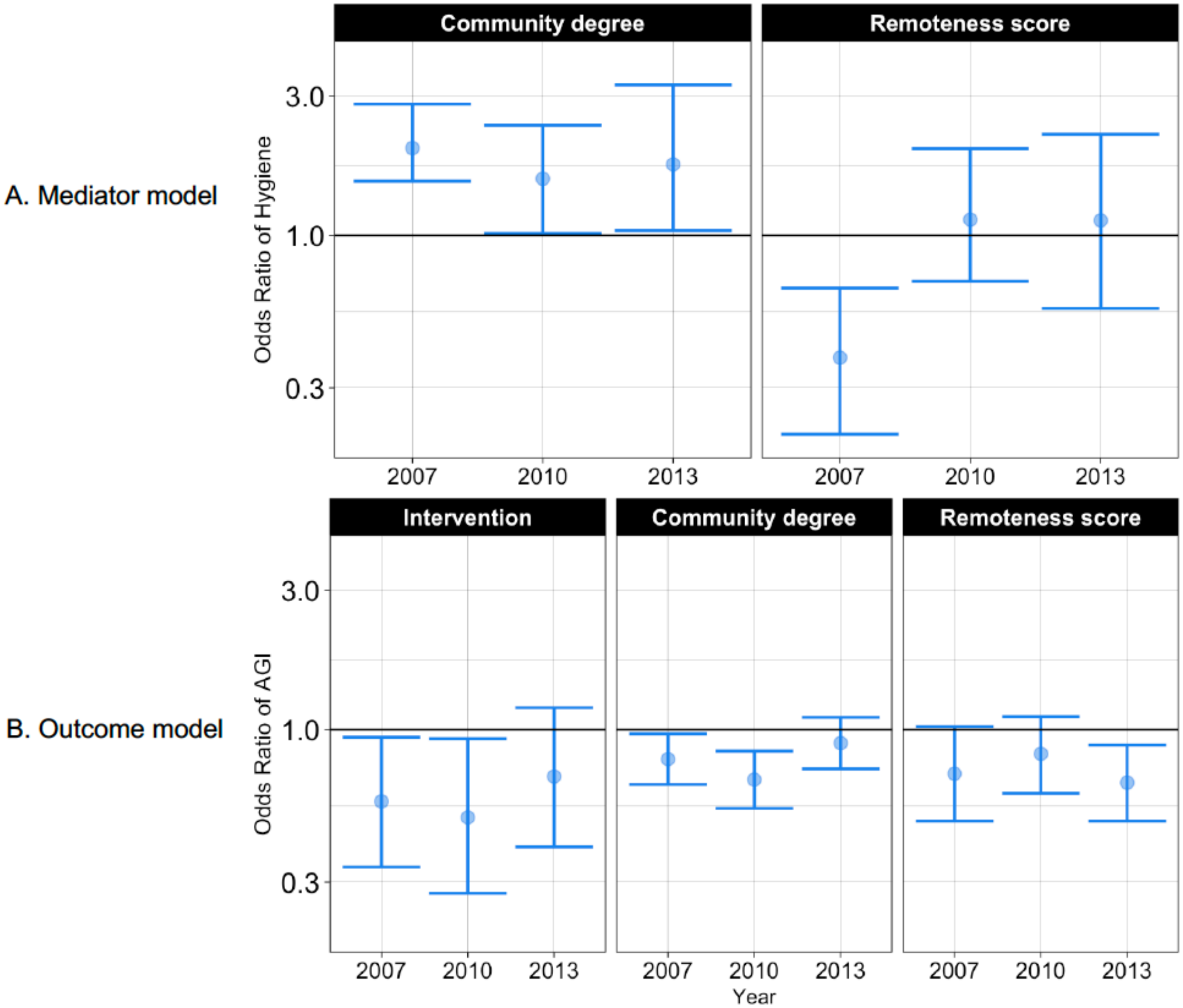
Descriptive statistics of study data, 2007-2013, and for the indigenous population, the Chachis, studied in 2010. Reporting proportions and means and ranges.

Figure 4.1. Violin plots of household WASH



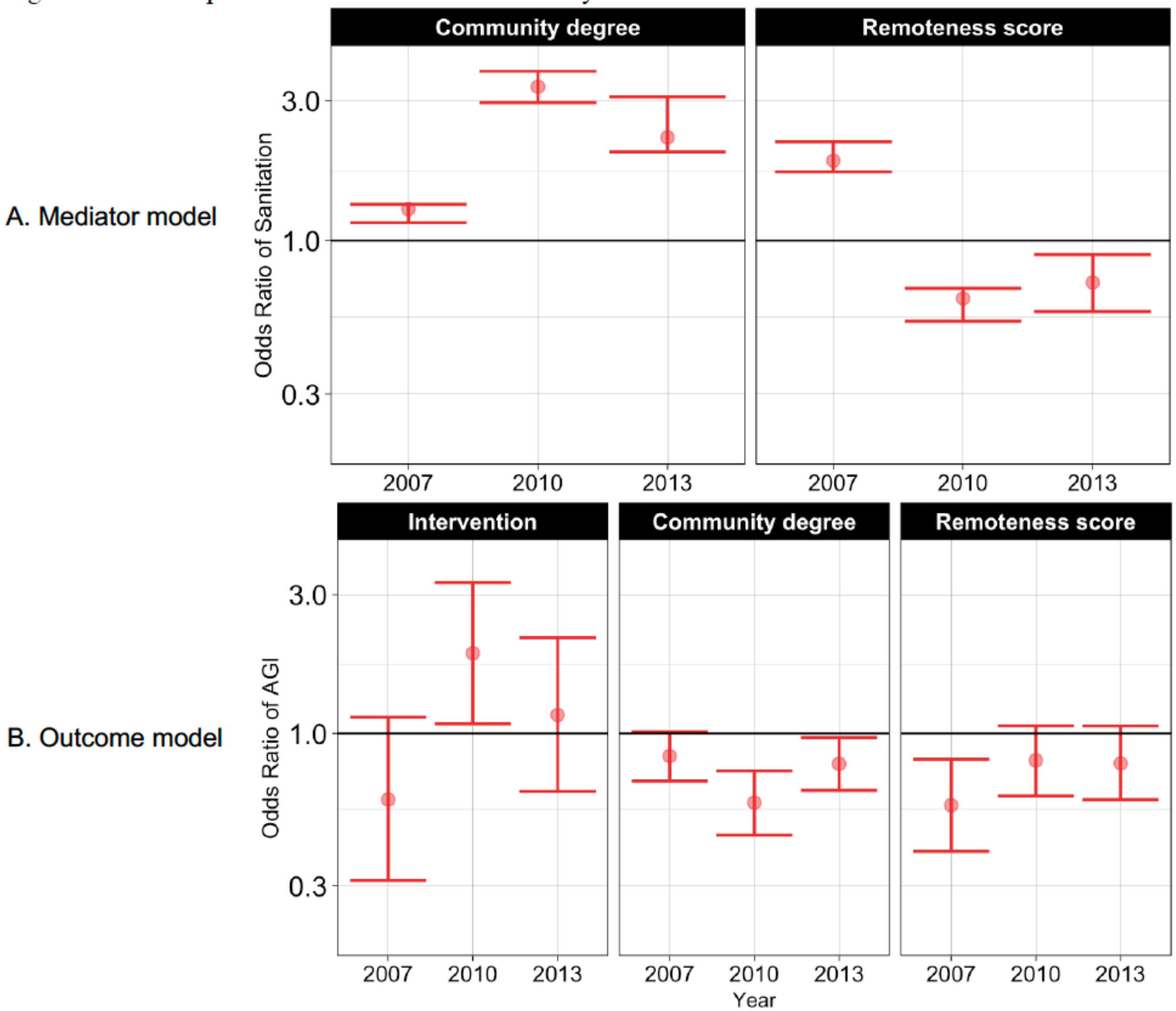
Violin plots of household hygiene score and the proportion of households with improved sanitation within 500m (community sanitation) from 2007-2013, and for the indigenous population, the Chachis, studied in 2010.

Figure 4.2. Forrest plots of model results for hygiene score



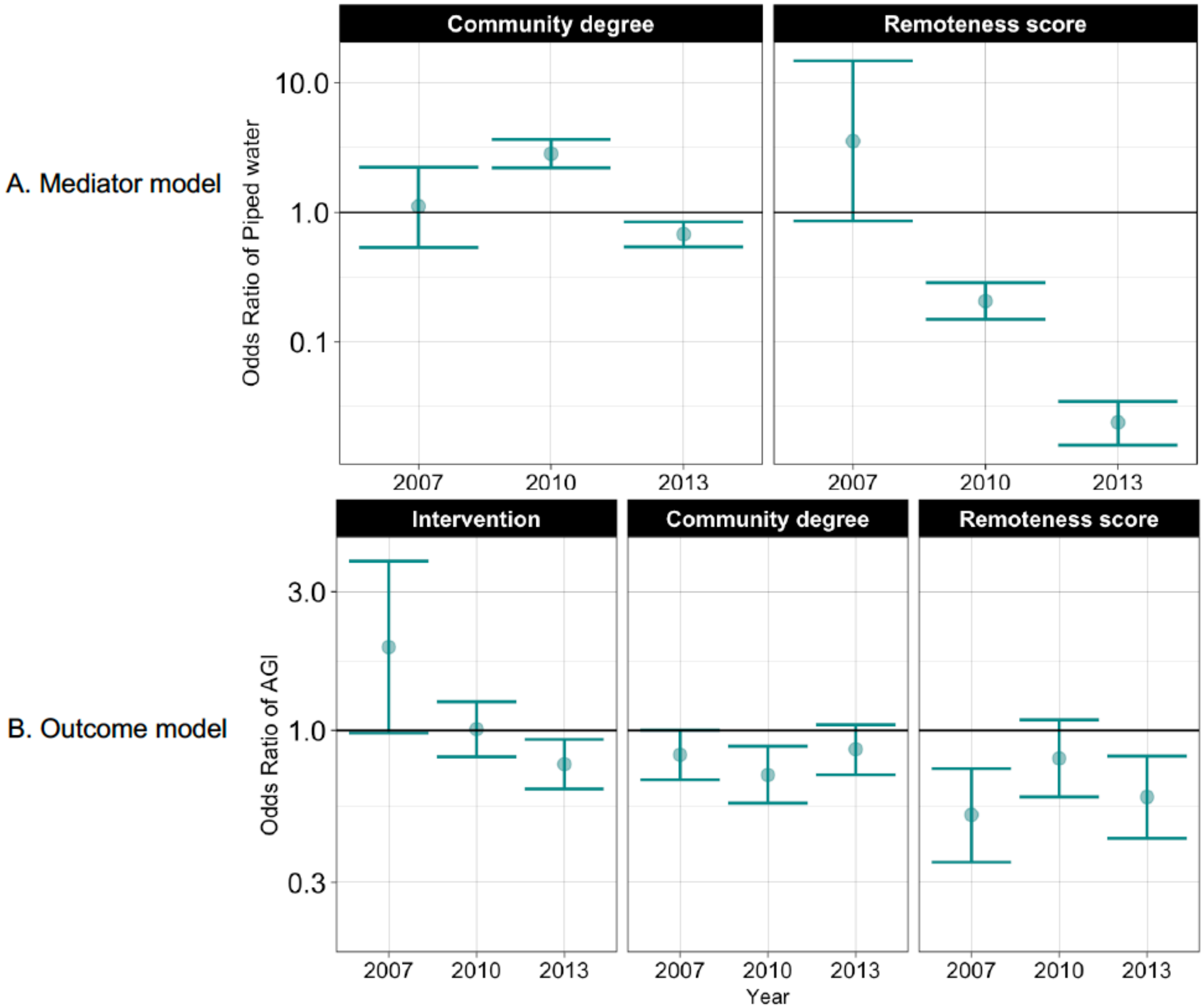
A. Forrest plots of the mediator model results of the effect of community cohesion and remoteness on the odds of engaging in increased hygiene compared to not at the household-level from 2007-2013 using bootstrapped GLMM to attain confidence intervals. B. Forrest plots of the outcome model results of the effect of household-level hygiene on acute gastrointestinal illness (AGI) controlling for community cohesion and remoteness from 2007-2013 using the two-stage Bayesian hierarchical model.

Figure 4.3. Forrest plots of model results for community sanitation



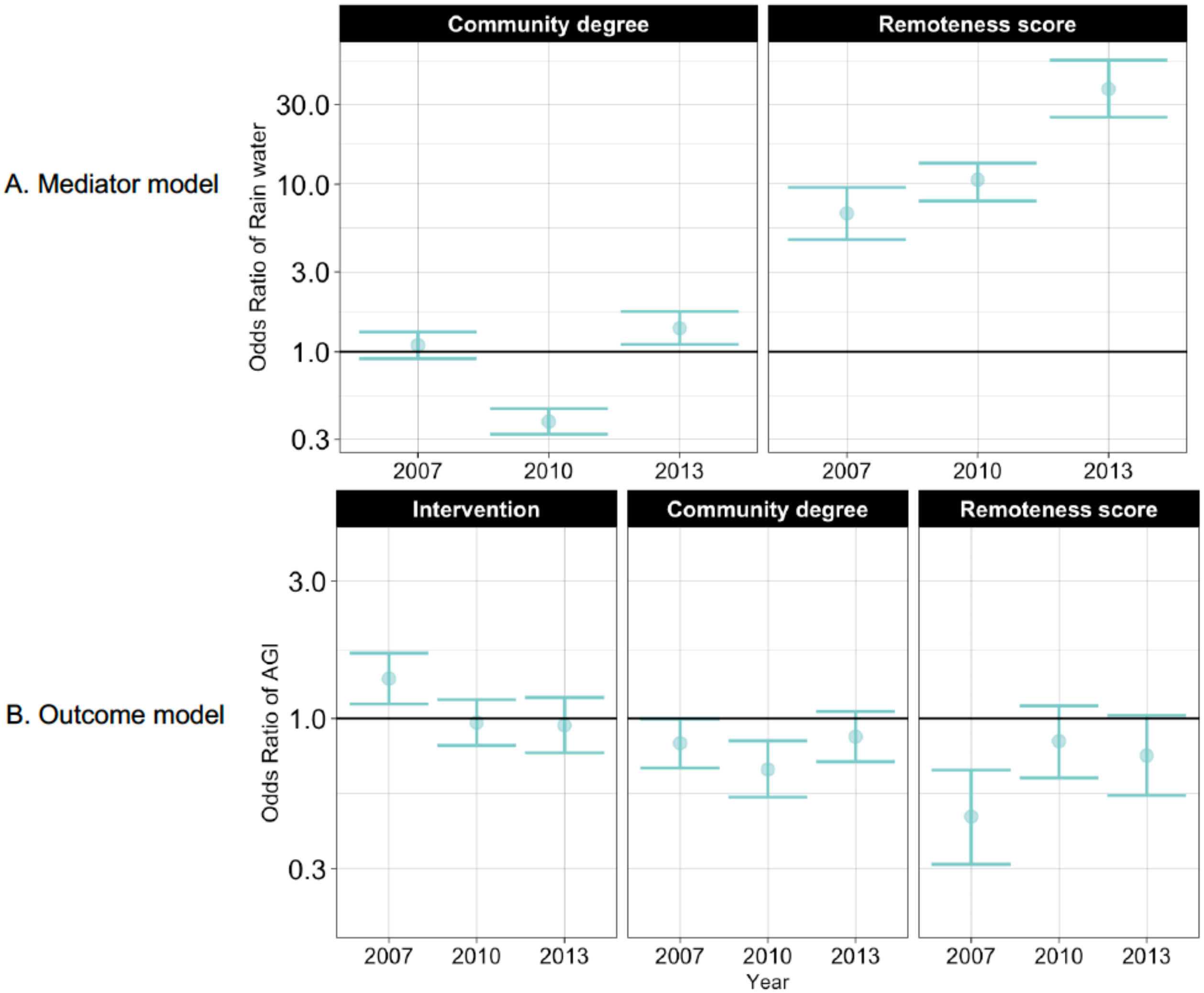
A. Forrest plots of the mediator model results of the effect of community cohesion and remoteness on the odds of having increased community sanitation around the household compared to not from 2007-2013 using bootstrapped GLMM to attain confidence intervals.
B. Forrest plots of the outcome model results of the effect of community sanitation on acute gastrointestinal illness (AGI) controlling for community cohesion and remoteness from 2007-2013 using the two-stage Bayesian hierarchical model.

Figure 4.4. Forrest plots of model results for piped water



A. Forrest plots of the mediator model results of the effect of community cohesion and remoteness on the odds of having piped water at the household-level from 2007-2013 using the two-stage Bayesian hierarchical model. B. Forrest plots of the outcome model results of the effect of household-level piped water on acute gastrointestinal illness (AGI) controlling for community cohesion and remoteness from 2007-2013 using the two-stage Bayesian hierarchical model.

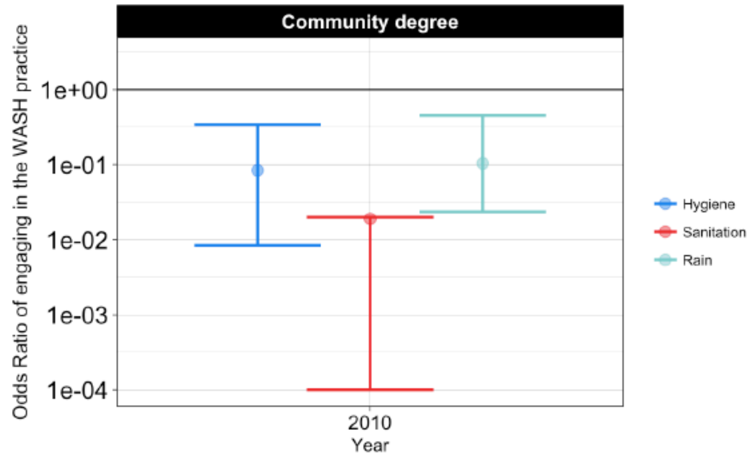
Figure 4.5. Forrest plots of model results for harvesting rain water



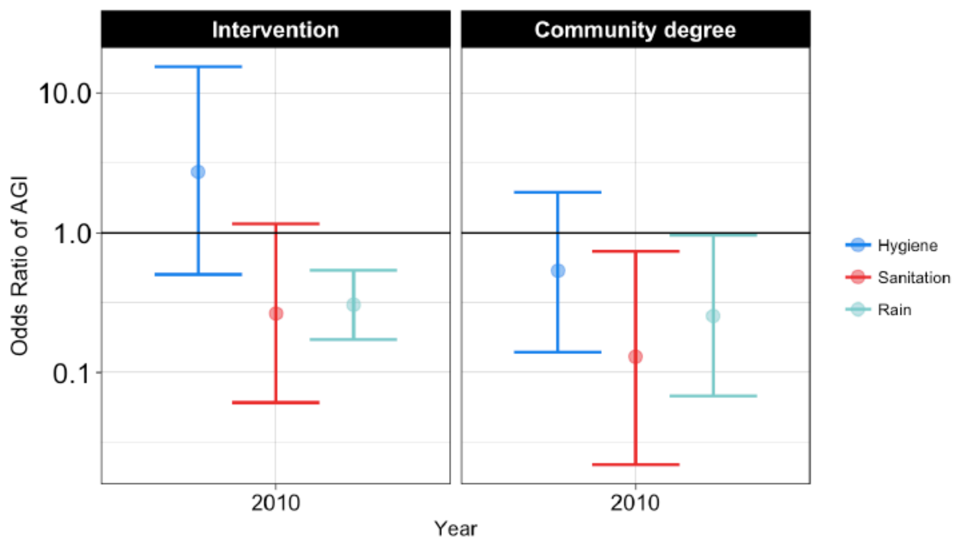
A. Forrest plots of the mediator model results of the effect of community cohesion and remoteness on the odds of harvesting rain water at the household-level from 2007-2013 using the two-stage Bayesian hierarchical model. B. Forrest plots of the outcome model results of the effect of harvesting rain water at the household-level on acute gastrointestinal illness (AGI) controlling for community cohesion and remoteness from 2007-2013 using the two-stage Bayesian hierarchical model.

Figure 4.6. Forrest plots of model results for Chachi communities

A. Mediator model

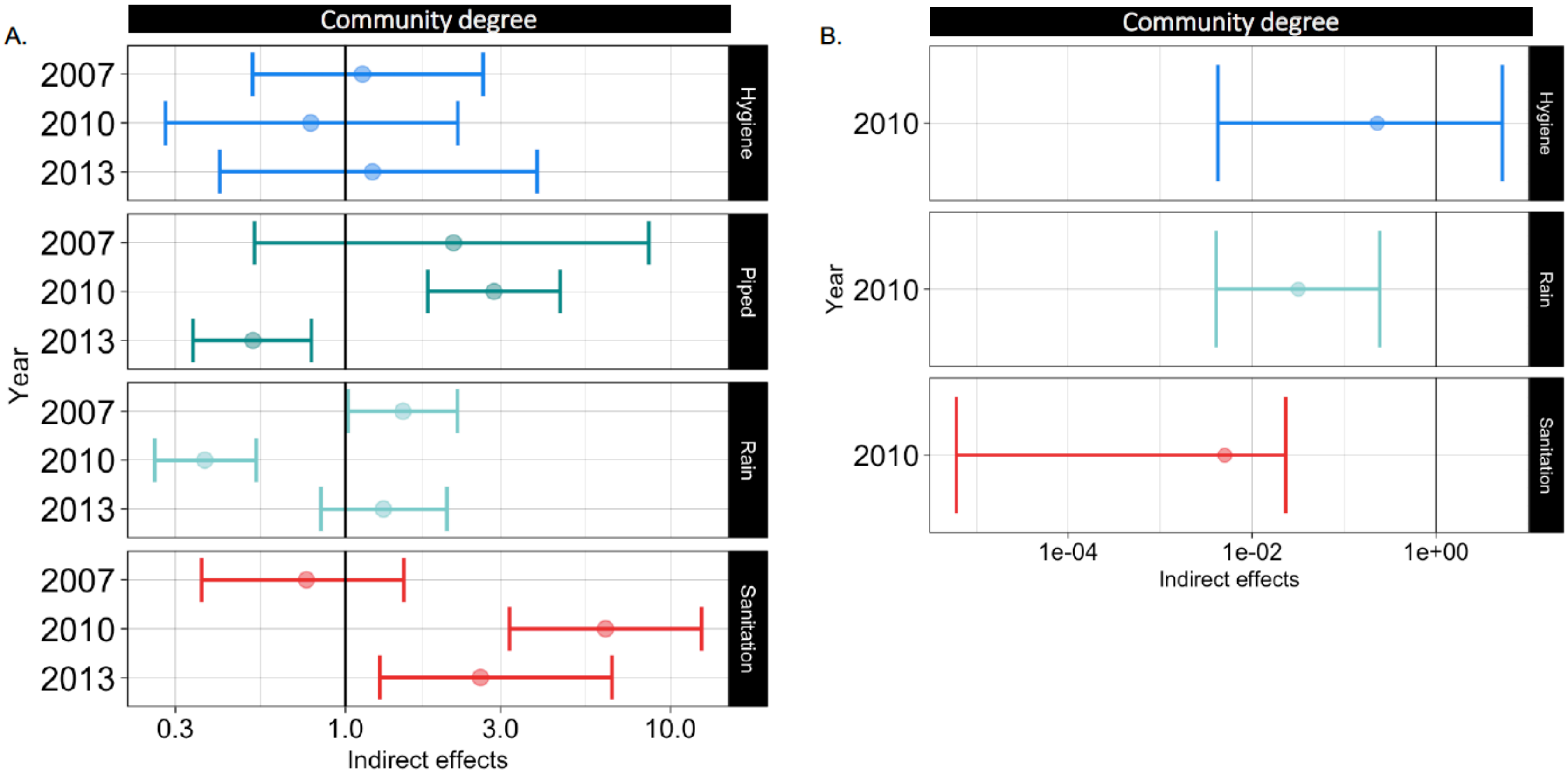


B. Outcome model



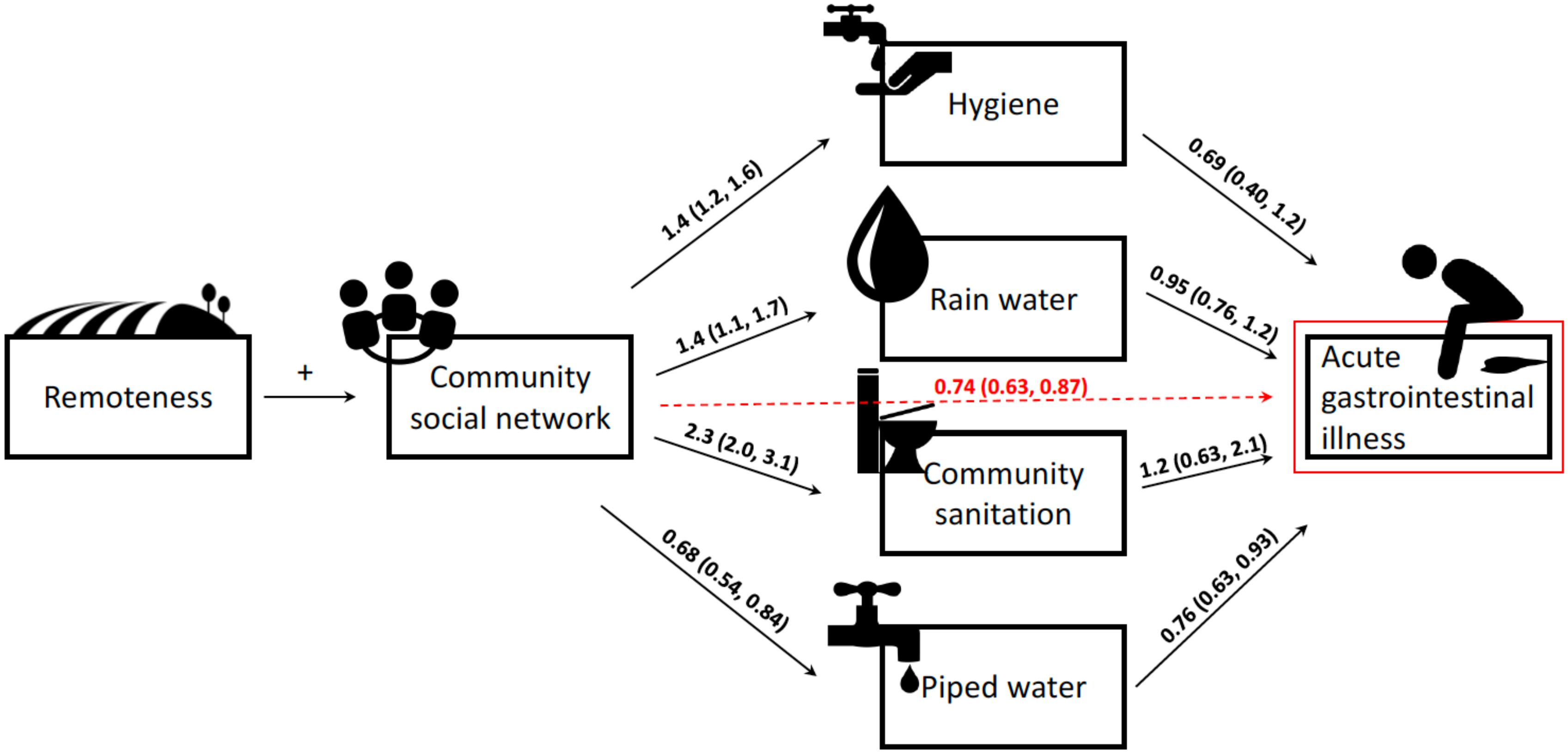
A. Forrest plots of the mediator model results of the effect of community cohesion on the odds of engaging in particular WASH practices compared to not: household-level hygiene score, community sanitation, and harvesting rain water in Chachi communities in 2010. B. Forrest plots of the effect of community cohesion and WASH practices on AGI. The two-stage Bayesian hierarchical model was used for the outcome models.

Figure 4.7. Indirect effects of cohesion on acute gastrointestinal illness



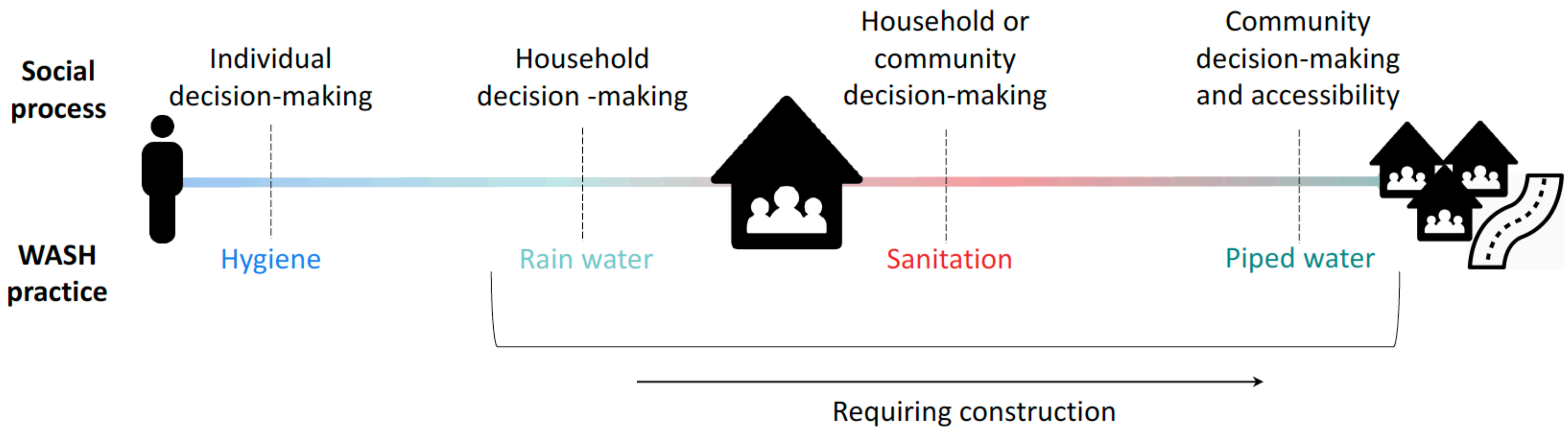
(A) Forrest plots of the indirect effects of community cohesion on AGI mediated through different WASH practices from 2007-2013 and (B) indirect effects of community cohesion on AGI mediated through different WASH practices in Chachi communities in 2010.

Figure 4.8. Causal diagram of mediation effects



Causal diagram of the relationship between the community social network as measured by the average community degree in an important matters network and water, sanitation, hygiene (WASH) practices and acute gastrointestinal illness (AGI) in 2013 for the majority non-indigenous communities. Of note, the relationships vary over the observed study period.

Figure 4.9. Gradient of WASH practices and social processes



Gradient of the observed water, sanitation, hygiene (WASH) practices and the related social processes that occur at different levels of human involvement and decision-making.

Supplementary Figure 4.1. Hygiene survey questions

Observations outside of the house:

Lavatory				
AA.	Is the lavatory clean?	☐ Yes	☐ No	☐ Not observed
BB.	Does the lavatory's hole have enough space (not full of feces?)	☐ Yes	☐ No	☐ Not observed
CC.	Is the lavatory covered?	☐ Yes	☐ No	☐ Not observed
Water				
D.	Is there water stored outside the house?	☐ Yes	☐ No	☐ Not observed
E.	Is the tank where water is stored outside the house covered?	☐ Yes	☐ No	☐ Not observed
F.	Is the water stored outside the house free of debris?	☐ Yes	☐ No	☐ Not observed
Other				
H.	Are the animals corralled or tied up?	☐ Yes	☐ No	☐ Not observed ☐ No animals
I.	Are the children wearing shoes when they are outside of the house?	☐ Yes	☐ No	☐ Not observed ☐ No children

Observations inside or outside the house:

Household Cleanliness				
J.	Is the inside and outside of the house free of human or animal feces?	☐ Yes	☐ No	☐ Not observed
K.	Is the inside and outside of the house free of trash?	☐ Yes	☐ No	☐ Not observed
Children				
L.	Are the faces of the children clean?	☐ Yes	☐ No	☐ Not observed ☐ No children
M.	Are the hands of the people who take care of the children clean?	☐ Yes	☐ No	☐ Not observed ☐ No children
Handwashing				
N.	Is there water for handwashing?	☐ Yes	☐ No	☐ Not observed
O.	Is there soap for handwashing?	☐ Yes	☐ No	☐ Not observed
P.	If there is soap, is it wet?	☐ Yes	☐ No	☐ Not observed

Observations inside the house:

Kitchen				
Q.	Is the kitchen free of flies?	☐ Yes	☐ No	☐ Not observed
R.	Are all the plates, pots, pans, and silverware clean?	☐ Yes	☐ No	☐ Not observed
S.	Is the counter where food is prepared clean?	☐ Yes	☐ No	☐ Not observed
T.	Is the prepared food covered?	☐ Yes	☐ No	☐ Not observed
U.	Are the cooking food supplies covered?	☐ Yes	☐ No	☐ Not observed
Water				
V.	Is there separate drinking water?	☐ Yes	☐ No	☐ Not observed
W.	Is the drinking water covered?	☐ Yes	☐ No	☐ Not observed
Animals				
X.	Is the house free of animals inside?	☐ Yes	☐ No	☐ Not observed

The list of questions about household hygiene practices from the census and case-control surveys administered in study communities.

Supplementary Table 4.1. Descriptive statistics of improved water by remoteness

	Piped water (N)	Rain water (N)	Total Population (N)	Proportion with piped water	Proportion with rain water
2007	30	820	2013	1.5%	40.7%
Close	7	195	798	0.9%	24.4%
Medium	3	150	328	0.9%	45.7%
Far	20	475	887	2.3%	53.6%
2010	470	1047	2046	23.0%	51.2%
Close	257	319	772	33.3%	41.3%
Medium	0	193	323	0.0%	59.8%
Far	213	535	951	22.4%	56.3%
2013	900	655	2141	42.0%	30.6%
Close	745	42	959	77.7%	4.4%
Medium	0	123	263	0.0%	46.8%
Far	155	490	919	16.9%	53.3%
Chachi (2010)	0	229	272	0.0%	84.2%

Descriptive statistics of of the household water infrastructure by remoteness categories of study communities, 2007-2013.

Supplementary Table 4.2. Outcome model results without mediation

		2007			2010			2013			Chachi	
Variable	OR	LL	UL	OR	LL	UL	OR	LL	UL	OR	LL	UL
Original outcome model: No intervention~												
Community important matters degree	0.89	0.74	1.06	0.61	0.50	0.73	0.74	0.63	0.87	0.47	0.13	1.61
Age	1.01	1.01	1.01	1.00	1.00	1.00	1.00	1.00	1.00	1.01	1.00	1.03
Sex (Male)	1.05	0.91	1.22	0.57	0.49	0.65	0.77	0.68	0.88	0.42	0.26	0.68
Remoteness score	0.40	0.28	0.57	0.73	0.58	0.92	0.78	0.61	1.00	-	-	-
Trust	1.03	0.88	1.21	1.27	1.09	1.48	0.61	0.52	0.71	3.26	1.58	6.90
Organizational belongingness	1.07	1.01	1.14	1.04	1.00	1.08	0.96	0.91	1.02	1.02	0.90	1.16

Outcome model results of the effect of social influences on AGI with no WASH practices controlled for from 2007-2013 and for the indigenous population, the Chachis, studied in 2010. All Chachi communities are remote and so remoteness was not included as a covariate in the Chachi model. These models used the two-stage Bayesian hierarchical model approach. (OR = Odds Ratio, LL = Lower Limit, UL = Upper Limit)

Supplementary Table 4.3. Mediation model results for hygiene score

	2007			2010			2013			Chachi		
Variable	**OR**	**LL**	**UL**	**OR**	**LL**	**UL**	**OR**	**LL**	**UL**	**OR**	**LL**	**UL**
Mediator model: Hygiene score ~												
Community important matters degree	1.99	1.53	2.82	1.56	1.02	2.38	1.75	1.04	3.28	0.08	0.01	0.34
Age	1.00	1.00	1.01	1.00	0.99	1.01	1.00	0.99	1.01	1.00	0.98	1.02
Sex (Male)	1.00	0.81	1.28	1.11	0.82	1.53	1.05	0.71	1.52	1.67	0.88	3.45
Remoteness score	0.38	0.21	0.66	1.13	0.70	1.98	1.12	0.56	2.22	-	-	-
Trust	0.89	0.72	1.16	1.26	0.90	1.74	0.82	0.55	1.27	0.68	0.29	1.45
Organizational belongingness	1.13	1.04	1.24	1.09	0.99	1.21	0.98	0.84	1.16	0.92	0.76	1.09
Outcome model: AGI ~												
Community important matters degree	0.79	0.65	0.97	0.67	0.54	0.84	0.90	0.73	1.10	0.54	0.14	1.95
Hygiene score	0.57	0.34	0.94	0.50	0.27	0.93	0.69	0.40	1.19	2.72	0.50	15.36
Age	1.01	1.01	1.02	1.00	1.00	1.01	1.00	0.99	1.00	1.02	1.00	1.03
Sex (Male)	1.03	0.86	1.24	0.66	0.55	0.79	0.84	0.71	0.99	0.38	0.23	0.61
Remoteness score	0.71	0.49	1.02	0.83	0.60	1.11	0.66	0.49	0.89	-	-	-
Trust	0.89	0.73	1.08	1.25	1.03	1.50	0.72	0.59	0.88	3.31	1.59	7.01
Organizational belongingness	1.07	1.00	1.15	1.05	1.00	1.11	1.02	0.94	1.10	1.03	0.91	1.17

Mediation model results of the effect of social influences on AGI with hygiene score modeled as a potential mediator from 2007-2013 and for the indigenous population, the Chachis, studied in 2010. All Chachi communities are remote and so remoteness was not included as a covariate in the Chachi model. GEE models were used to model the WASH practice as the outcome (mediator model) and bootstrapped to attain confidence intervals. (OR = Odds Ratio, LL = Lower Limit, UL = Upper Limit)

Supplementary Table 4.4. Mediation model results for community sanitation

	2007			2010			2013			Chachi		
Variable	**OR**	**LL**	**UL**	**OR**	**LL**	**UL**	**OR**	**LL**	**UL**	**OR**	**LL**	**UL**
Mediator model: Sanitation ~												
Community important matters degree	1.28	1.15	1.33	3.35	2.96	3.78	2.25	2.01	3.10	0.02	0.00	0.02
Age	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.99	1.01
Sex (Male)	1.00	0.92	1.05	0.99	0.91	1.07	1.00	0.90	1.09	1.00	0.78	1.74
Remoteness score	1.87	1.72	2.17	0.64	0.53	0.69	0.72	0.57	0.90	-	-	-
Trust	1.00	0.93	1.07	1.02	0.96	1.15	1.00	0.81	1.02	1.00	0.00	1.32
Organizational belongingness	1.03	1.00	1.07	1.00	1.00	1.04	1.01	0.99	1.10	1.00	0.97	1.00
Outcome model: AGI ~												
Community important matters degree	0.84	0.69	1.01	0.58	0.45	0.74	0.79	0.64	0.97	0.13	0.02	0.74
Sanitation	0.59	0.31	1.14	1.89	1.08	3.30	1.16	0.63	2.13	0.27	0.06	1.16
Age	1.01	1.01	1.02	1.00	1.00	1.00	1.00	1.00	1.00	1.01	1.00	1.03
Sex (Male)	1.01	0.85	1.21	0.65	0.54	0.77	0.82	0.70	0.96	0.42	0.26	0.67
Remoteness score	0.57	0.39	0.82	0.81	0.61	1.06	0.79	0.59	1.06	-	-	-
Trust	1.03	0.85	1.24	1.15	0.96	1.38	0.69	0.57	0.83	3.27	1.61	6.95
Organizational belongingness	1.07	1.01	1.14	1.02	0.97	1.07	0.97	0.90	1.04	1.04	0.91	1.18

Mediation model results of the effect of social influences on AGI with community sanitation (the proportion of improved sanitation within 500m) modeled as a potential mediator from 2007-2013 and for the indigenous population, the Chachis, studied in 2010. All Chachi communities are remote and so remoteness was not included as a covariate in the Chachi model. GEE models were used to model the WASH practice as the outcome (mediator model) and bootstrapped to attain confidence intervals. (OR = Odds Ratio, LL = Lower Limit, UL = Upper Limit)

Supplementary Table 4.5. Mediation model results for piped water

	2007			2010			2013			Chachi		
Variable	**OR**	**LL**	**UL**	**OR**	**LL**	**UL**	**OR**	**LL**	**UL**	**OR**	**LL**	**UL**
Mediator model: Piped water ~												
Community important matters degree	1.11	0.54	2.24	2.84	2.21	3.66	0.68	0.54	0.84	-	-	-
Age	1.00	0.99	1.02	1.00	1.00	1.01	1.00	1.00	1.01	-	-	-
Sex (Male)	0.81	0.45	1.48	0.98	0.82	1.18	1.02	0.85	1.23	-	-	-
Remoteness score	3.55	0.86	14.78	0.21	0.15	0.29	0.02	0.02	0.03	-	-	-
Trust	0.75	0.40	1.40	1.01	0.83	1.22	0.88	0.72	1.08	-	-	-
Organizational belongingness	1.18	0.97	1.44	1.05	1.00	1.11	1.11	1.02	1.20	-	-	-
Outcome model: AGI ~												
Community important matters degree	0.82	0.68	1.00	0.70	0.56	0.88	0.86	0.70	1.05	-	-	-
Piped water	1.93	0.98	3.82	1.01	0.81	1.25	0.76	0.63	0.93	-	-	-
Age	1.01	1.01	1.02	1.00	0.99	1.00	1.00	0.99	1.00	-	-	-
Sex (Male)	1.04	0.86	1.25	0.67	0.56	0.81	0.85	0.72	1.00	-	-	-
Remoteness score	0.51	0.35	0.74	0.80	0.59	1.09	0.59	0.42	0.82	-	-	-
Trust	1.02	0.85	1.24	1.07	0.89	1.30	0.73	0.60	0.88	-	-	-
Organizational belongingness	1.05	0.98	1.12	1.05	0.99	1.10	0.99	0.92	1.07	-	-	-

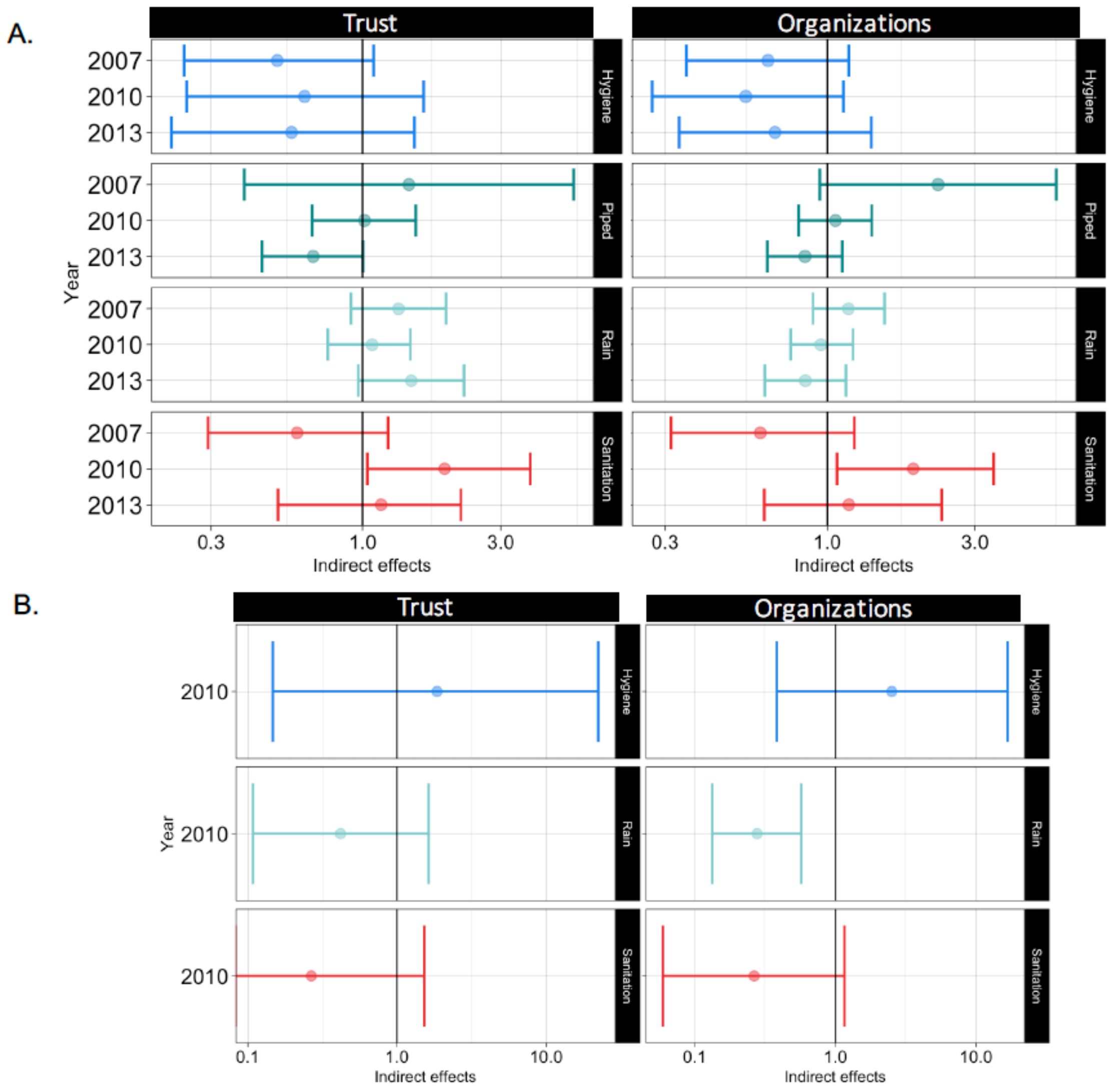
Mediation model results of the effect of social influences on AGI with piped water modeled as a potential mediator from 2007-2013. There was no recorded piped water for the indigenous population, the Chachis. (OR = Odds Ratio, LL = Lower Limit, UL = Upper Limit)

Supplementary Table 4.6. Mediation model results for harvesting rainwater

		2007			2010			2013			Chachi	
Variable	**OR**	**LL**	**UL**	**OR**	**LL**	**UL**	**OR**	**LL**	**UL**	**OR**	**LL**	**UL**
Mediator model: Rain water ~												
Community important matters degree	1.09	0.91	1.31	0.38	0.32	0.46	1.38	1.11	1.74	0.10	0.02	0.45
Age	1.00	1.00	1.01	1.00	1.00	1.00	1.00	1.00	1.01	0.99	0.97	1.01
Sex (Male)	0.98	0.83	1.15	1.00	0.85	1.16	0.96	0.80	1.15	1.06	0.57	2.00
Remoteness score	6.71	4.68	9.58	10.65	7.94	13.38	37.11	25.18	55.20	-	-	-
Trust	0.97	0.81	1.15	1.11	0.94	1.26	1.55	1.27	1.90	1.37	0.63	3.03
Organizational belongingness	0.85	0.80	0.91	0.99	0.94	1.04	0.90	0.83	0.97	0.91	0.78	1.06
Outcome model: AGI ~												
Community important matters degree	0.82	0.67	1.00	0.66	0.53	0.84	0.86	0.71	1.06	0.25	0.07	0.96
Rain water	1.37	1.12	1.69	0.97	0.80	1.16	0.95	0.76	1.18	0.30	0.17	0.54
Age	1.01	1.01	1.02	1.00	1.00	1.01	1.00	0.99	1.00	1.01	1.00	1.03
Sex (Male)	1.04	0.86	1.24	0.63	0.53	0.75	0.85	0.72	1.01	0.42	0.26	0.69
Remoteness score	0.46	0.31	0.66	0.83	0.62	1.11	0.74	0.54	1.03	-	-	-
Trust	1.02	0.84	1.23	1.20	1.00	1.44	0.73	0.60	0.89	3.63	1.75	7.76
Organizational belongingness	1.06	0.99	1.13	1.03	0.98	1.08	0.98	0.91	1.06	1.01	0.89	1.14

Mediation model results of the effect of social influences on AGI with rain water modeled as a potential mediator from 2007-2013 and for the indigenous population, the Chachis, studied in 2010. All Chachi communities are remote and so remoteness was not included as a covariate in the Chachi model . Two-stage Bayesian hierarchical model used for both the mediator and outcome model. (OR = Odds Ratio, LL = Lower Limit, UL = Upper Limit)

Supplementary Figure 4.2. Indirect effects of trust and organizational belongingness



(A) Forrest plots of the indirect effects of trust and organizational belongingness on AGI mediated through different WASH practices from 2007-2013 and (B) indirect effects of trust and organizational belongingness on AGI mediated through different WASH practices in Chachi communities in 2010.

Chapter V

The role of gender in community social structures and water insecurity in rural, coastal Ecuador

5.1 Abstract

Despite dramatic improvements in public health in the past decade, 2.2 million people, mostly children, die globally each year due to unsafe drinking water and poor sanitation. Good water, sanitation, and hygiene (WASH) practices are influenced by a multitude of factors, including community social constructs like social cohesion. Women experience a continual tradeoff in daily tasks, particularly in low-resource settings, and play a unique role in influencing community-level social constructs. Previous studies conducted on coastal Ecuadorian population have identified that a greater density of social ties between individuals in remote communities leads to increased WASH practices at the household-level and reduced diarrheal disease. Here, we extend prior work by examining how gender roles at the individual- and community-level are related to social organization in the context of WASH in communities in rural, coastal Ecuador through qualitative data and social network analysis. In 2016, we conducted in-depth interviews with men and women (5 per gender), and focus groups with men and women in each community. The study team transcribed, coded, and discussed interviews for a thematic analysis by comparing existing theories on power dynamics with themes found in the data. Using longitudinal social network data collected in the same communities from 2004-2013, we then estimated assortativity, cluster modularity, and edge density by gender of the community networks, using bootstrapping methods to determine confidence intervals. Using GEE models, we then assessed the effect of these gender based network measures on household hygiene, improved sanitation, and rainwater collection. The qualitative data showed that women experience a distinct set of stressors that inhibit the role they play in social organization in general, and importantly the collection and treatment of water. Men play a critical role in

creating gender equity and enhancing agency among women. Both men and women, however, identified water insecurity as a primary stress and deterrent to social organization. The intensity of the stressors experienced by women were modified by ethnicity, living environment, and access to a natural water source. The indigenous populations in the region experienced more severe gender inequity than other ethnicities. Communities that are more assortative between genders are less likely to engage in WASH practices at the household-level, with this effect decreasing over time. This study has the potential to inform context-specific and gender-sensitive interventions.

5.2 Introduction

Clean water is essential for life and water resource management is becoming increasingly important. One in ten persons on this planet lacks access to safe water sources, exposing nearly 1 billion people to easily preventable waterborne diseases and rendering millions to live in poverty. Despite dramatic improvements in public health in the past decade, 2.2 million people, mostly children, die globally each year due to unsafe drinking water and poor sanitation [131]. Furthermore, as global population is predicted to grow and climate change persists, per capita water availability will decrease, and as global water consumption increases less water will exist for agriculture, industry, hygiene, and drinking.

Women have a critical role to play in community driven programs and efforts to establish clean water sources. Though women in low-resource settings are often excluded from key decision-making roles in their communities or face gender-based violence (GBV) issues, they bear the burden of collecting, storing, and protecting water sources, and as such experience higher psychosocial stress [56]. Time spent walking in search of clean water or treating water often results in safety issues and could be spent on education, work, or improving the overall health and nutrition of the household. Including women in key decision-making roles, to improve access to clean, nearby water sources, empowers women to improve their futures and bring families and communities out of poverty, in addition to improving downstream health effects [56,57]. Prior research has demonstrated effects of chronic psychosocial stress on inducing structural changes in the gut microbiota [58,59] and maternal stress altering child neuroendocrine-immune function thereby increasing disease risk [60]. Indeed, empowering women as economic, political, and social actors that can change community choices is a critical component of development and likely adoption of interventions, and more specifically water, sanitation, and hygiene (WASH) interventions.

For determining ways for WASH interventions to be successful, researchers have turned to multilevel causal frameworks that outline the importance of studying behavior and social constructs at the individual-, household-, and community-levels [10,40,148]. Importantly, changes in the larger social environment affect changes in individuals and having the support of individuals in community is essential for implementing environmental changes. While understanding community-level structures and how to leverage them could not only help overcome issues of compliance and sustainability of behavioral interventions at multiple levels, research on the underlying factors, like social or community constructs, that strengthen intervention implementation and increase acceptance, adoption, and sustained use is rare [39]. As such, taking a step further, we've integrated a multi-level causal framework with social network theory and theory centered around gender roles in the context of social capital and cohesion.

Collective efficacy, social capital, and social cohesion are all latent constructs of the social environment believed to influence the quality, effectiveness, and sustainability of interventions, especially those based on action at the community-level. The influence that collective efficacy more broadly, and social cohesion more specifically, has on intervention effectiveness may be explained in part by the theory of diffusion of innovations. This theory suggests that innovative behaviors diffuse much more rapidly in communities that are cohesive and in which members know and trust each other [55]. Such theoretical conceptualizations are supported by an empirical evidence base that suggests communities high in social constructs, have higher uptake of WASH interventions and substantial health benefits [11,13]. However, studies examining the role of community social constructs in the WASH programming and research are few, particularly those that also examine the role of women in community and WASH.

Previous research in communities in rural, coastal Ecuador have shown that social constructs, defined by social network data, at the household- and community-levels reduce acute gastrointestinal illness (AGI). Community-level cohesion, defined by community social ties in a network comprised of ties of with individuals visited for important matters, was additionally associated with increased WASH practices like improved water and sanitation, partially mediating the protective effect of the social constructs on AGI. In the indigenous communities in the study region, community-level cohesion had the opposite effect on AGI. Using qualitative data collected in 2016 from the same villages in rural Ecuador and longitudinal social network data from 2004-2013, we extend these prior analyses to examine how gender roles at the individual- and community-level are related to community social structures and water security in this study population. We hypothesize that gender is an important indicator of social cohesion and affects adoption of WASH practices.

5.3 Methods

Qualitative data & analysis

Building off existing data, both qualitative and quantitative, from the same communities in Ecuador, we developed a qualitative survey tool, including probes, of approximately 35 questions, that addressed 4 primary themes: social organization, water, and gender (including general sources of stress). From August to November 2016, we conducted in-depth interviews with men and women (~5 per gender), and 2 focus groups in 18 communities in rural, coastal Ecuador. We conducted one focus group with men and one with women to capture the diversity of responses with approximately 6-8 people. We visited 15 majority Afro-Ecuadorian communities (7 far from an urban center, 5 at a medium distance, 3 near a road) and 3 Chachi communities (all far from an urban center). In total, we conducted 31 focus groups and 196 in-depth interviews. Participants were purposively sampled and interviews were conducted in either Spanish or Chapalachi, the language of the Chachis, through a translator. All study participants provided informed consent and all data collection protocols were approved by institutional review boards at the University of Michigan and the University of San Francisco of Quito.

Given the qualitative nature of the study, the survey tool was subject to change through iterative sampling, adapting the survey tool throughout the data collection process based on emergent findings specific to categories of people or study site. In the preliminary analysis, the study team reviewed the in-depth interview and focus group transcripts to make initial detailed summaries to help inform the iterative sampling. We conducted open-coding on a random sub-set of transcripts to ascertain themes that were covered and that needed to be added to the tool. After data collection was completed, all transcripts were transcribed by study team members, coded, and discussed for a thematic analysis. As social cohesion, gender dynamics, and water security are all themes of social relationships, we analyzed the data for competing forms of power dynamics. We compared existing theories on forms of human relationships that emphasize different components of power to our observations, and assessed the relative strength of these theories by unpacking processes. Inductive theory development [149] creates an ongoing dialogue between pre-existing theory and new insights generated from empirical observation for the development of new explanations or for refining pre-existing theory.

We reached saturation of results, when the results reported are the same, after coding 70% of the in-depth interviews and focus groups in each community. In the results, we use the contextual factors of remoteness and ethnicity to help explain the findings. The study villages exist along three river basins: Cayapas, Santiago, Ónzole and vary by remoteness, which is a function of time and travel cost to the nearest township, Borbón [101]. Since 1996, paved roads have been built connecting this township to the coast and Andes. Smaller roads continue to be built linking villages to the main road. Here we discuss remoteness as a categorical variable based on cut-off points of a remoteness score so that communities are either "Close" (near a road), "Medium" (at a medium distance), or "Far" (far from a road).

Social network data & analysis

We collected sociometric and census data during four cross-sectional waves from 15 villages in northern coastal Ecuador, in the province of Esmeraldas, in 2004 and 20 villages (5 added villages) in 2007, 2010, and 2013. We collected data from all consenting community members ≥13 years of age. The study population consisted of primarily Afro-Ecuadorians, Mestizos, and Chachis, an indigenous group of the Cayapas River in the region. Sociometric data was also collected during a single wave in 2010 from all consenting community members ≥13 years in 3 majority Chachi communities. Census data was collected from all communities just prior to each sociometric survey. Compared to village censuses, the average sociometric response rate across communities was approximately 80% each wave. We also collected case-control data from the same communities to estimate risk factors for diarrheal disease in the study region. Census and case-control data was collected yearly from 2003 - 2013, not just during years when the sociometric survey was administered. All study participants provided informed consent and all data collection protocols were approved by institutional review boards at the University of Michigan and the University of San Francisco of Quito.

Participants of the sociometric survey, or *egos*, were asked to identify members of their village outside their household with whom they can discuss important matters. We refer to this as a core discussion network (CDN), as each *edge*, or social tie, indicates a important discussant to the *ego*. The names generated thus created an "ego perceived" network at the individual-level. Previously we've shown that a community's CDN has a consistently protective effect on reducing acute gastrointestinal illness, partially mediated through the adoption of safe WASH practices. Though social influences at the household-level are multidimensional and can be examined through interaction of multiple types of social ties, here we focus on community-level constructs and thus focus on data from the CDN. In supplemental material, we show brief descriptive results from a passing time network though we do not discuss these in the results as there were no notable differences in network statistics in this network (Supplemental Information 1).

Network statistics

We first examined different network measures that would describe gender differences in each community network for each wave of data and within the Chachi communities. To do so, we described the following network measures: 1) gender assortativity, which identifies homophily or the tendency of each gender to prefer ties of the same gender (a positive measure indicates there are more ties between men and between women and a negative measure indicates more ties between genders); 2) cluster modularity by gender, which identifies the number of between gender sub-communities versus within gender sub-communities (a positive measure indicates more sub-communities within gender); 3) edge density by gender, the ratio of the number of edges to the number of possible edges for a sub-graph of all women and sub-graph of all men (a larger value indicates the network is dense with social ties). Edge density is bounded between 0 and 1.

The assortativity coefficient was first defined by Mark Newman [150], and for assessing assortativity for a categorical variable, like gender, is described as

$$r = \frac{(\sum_i e_{i,j}) - (\sum_i a_i b_j)}{(1 - \sum_i a_i b_j)}$$

where $e_{i,j}$ is the fraction of edges connecting vertices, or nodes, of type i and j, $a_i = \sum_j e_{i,j}$, and $b_i = \sum_i e_{i,j}$. When values are assigned to the vertices instead of categories, assortativity is described as for undirected network graphs

$$r = \frac{1}{\sigma_q^2} \sum_{jk} jk(e_{jk} - q_j q_k)$$

where e_{jk} is the joint probability distribution of the remaining degrees of the two vertices, j and k, at either end of a randomly chosen edge. A network with no assortative mixing $e_{jk}= q_j q_k$. Assortativity is bounded by 1 and -1, where a 1 indicates complete assortativity of men only interacted with men and women only interacting with women. Cluster modularity is defined as

$$Q = \frac{1}{2m} \cdot \sum \frac{A_{ij} - k_i k_j}{2m} \Delta_{ij} c_i c_j$$

where m is the number of edges, A_{ij} is the element of the A adjacency matrix in row i and column j, k_i is the degree of i, k_j is the degree of j, c_i is the type of i, c_j is the type of j, the sum goes over all i and j pairs of vertices, and Δ_{ij} is 1 if type x equals type y and 0 otherwise.

We additionally assessed other network measures that are not included in the main analyses as they had minimal informative contribution: 1) overall graph modularity, which identifies sub-communities in the overall network; 2) degree by gender, which indicates the number of social ties, or edges, per individual in a sub-graph of women and sub-graph of men; 3) total triads, which is the total number of three nodes with connected edges between them (i.e. sub-communities between three nodes); 4) global transitivity, which is the ratio of the triangles and the connected triples in the graph or the probability that the adjacent vertices of a vertex are connected; 5) local transitivity, which is the ratio of the triangles connected to the vertex and the triples centered on the vertex. Assessments of these additional measures can be found in supplemental material as we only discuss them briefly in the results (Supplemental Information 2 – Supplemental Information 5).

As there is no agreed upon method for measuring uncertainty of the various network statistics, to estimate confidence intervals of the primary network measures of interest (gender assortativity, cluster modularity by gender, and edge density within gender groups), we compared different techniques including jackknifing, bootstrapping, a method described by Mark Newman to assess error specifically for assortativity [151], and a single null model. The jackknife and the bootstrap are two non-parametric methods which provide estimates bias and variance of an estimator, without any assumption about its statistical distribution [152,153]. The jackknife is based on the observation of the estimator for subsamples, generally of size n-1, obtained from the original sample. The bootstrap is based on the observation of the estimator on size n samples drawn from the original sample. A data set of size *n* has 2^n-1 nonempty subsets and the jackknife method only uses *n* of them, thus some agree the jackknife method, which predates the bootstrap, could improve by using statistics based on more than *n* or even all 2^n-1 subsets, which bootstrap does [154,155].

For assortativity and cluster modularity, we first assessed the jackknifing method, where we removed a percentage *p* of edges from a random sampling of edges and then calculated both network measures. We did this for $p = 0.10, 0.20, 0.40$ to see if there were major differences when the removal threshold is increased. We then did the same thing for a random sampling of nodes, or *egos*, from a random sampling of all nodes. Using the bootstrapping technique, we then calculated the network measures on a random sampling of edges and a random sampling of nodes with replacement. We iterated both the jackknifing and bootstrapping methods 10,000 times to obtain our results and report the 2.5% and 97.5% quantiles of the resultant distributions for our confidence intervals. We also compare our estimates to a null model, which is based on a random redistribution of all edges while maintaining the same degree distribution [156]. Our estimate was obtained after 10,000 niter trials. Lastly, Mark Newman, who originated the method for assessing assortativity [150], suggests using the following to assess assortativity error

$$\sigma_r^2 = \sum_{i=1}^{M} (r_i - r)^2$$

where r is the assortativity estimate estimated from the graph and r_i is the estimated assortativity after a single edge is removed from the graph. According to Newman, jackknifing and removing 1 edge at a time is optimal, though this results in very small error.

For edge density, we assessed the jackknifing method only based on nodes, where we removed 10% of nodes from a random sampling of nodes, as assessing the removal of edges will result in the same edge density statistic. We also used the bootstrapping technique, where we took a random sampling of edges with replacement. We did this for both subgraphs of men and women. We iterated both the jackknifing and bootstrapping methods 10,000 times to obtain our results and report the 2.5% and 97.5% quantiles of the resultant distributions for our confidence intervals.

Finally, to visualize the network assortativity and modularity, we used the multilayer property of time to connect the communities. We use two communities: one close to a road (San Augustin) and one far from a road (San Miguel) to do so. We first estimate modularity of the graph and display the edges and nodes in isometric layers to reposition nodes to reduce the number of ties that cross each other, but also illustrate the distinct sub-communities. We then color the nodes by gender and edges based on whether the tie is between or within gender groups.

Regression statistics

To then quantitatively assess the impact of gender on WASH practices in the household, we estimated the effect of gender assortativity and cluster modularity separately on hygiene, sanitation, and rainwater harvesting using Generalized Estimating Equations (GEE) to account for the multilevel nature of the data. Data on WASH practices was obtained from either the census or case-control surveys, and matched to households in the sociometric dataset by wave of data collection. If a household was missing data for any WASH variable in a wave, we imputed household data from the previous year of data since census or case-control data was collected yearly. We assumed that WASH measures were less likely to change in a year. This was primarily done for the 2004 and 2007 data waves as many of the household WASH observations were collected on the case-control survey, where not every household in a community was surveyed. The 2010 and 2013 waves of data were matched to household data obtained from the census.

For the analysis, we focused on hygiene score, sanitation, and harvesting rain water as these are all primarily individual- or household-level WASH practices that are more likely influenced by women in the household. We measured hygiene as a score based on the proportion of 'Yes' answers to a series of 23 binary response questions related to hygiene conditions inside and outside the household. Sanitation was measured as the proportion of households within a 500 meter radius of each household with improved sanitation. This was done by first determining whether each household in a community had improved sanitation using the 2015 World Health Organization (WHO)/UNICEF Joint Monitoring Program (JMP) definition [131]: a flush or pour flush system to a piped sewage system, septic tank, or pit latrine. Then using this binary indicator per household, we calculated the proportion of households within a 500m radius of each household that had improved sanitation. We refer to this as community sanitation. Harvesting rain water was measured as a binary indicator at the household-level. Another measures accounted for in the models is remoteness, modeled as a categorical variable with the following attributes: "Close" (near a road), "Medium" (at a medium distance), or "Far" (far from a road).

As assortativity and cluster modularity are highly correlated, we estimated their effects on hygiene, community sanitation, and harvesting rainwater at the household-level separately for each wave of data. Given that hygiene and community sanitation are both proportions, here we model them using the following GEE model to estimate the likelihood

$$Y_{ij} = \beta_0 + \beta_1 \text{Remotness cateogry}_\text{j} + \beta_2 \text{Gender network measure}_\text{j}$$

where *i*= household (i = 1,…,N) and *j*= community (j = 1,…,n_j), with a Gaussian distribution and correlation at the household-level accounted for. As the indicator for harvesting rainwater is binary, we use the following GEE model to estimate odds ratios

$$\text{logit}(\text{p}_\text{ij}) = \log\left[\frac{\text{p}_\text{ij}}{1-\text{p}_\text{ij}}\right] = \beta_0 + \beta_1 \text{Remotness cateogry}_\text{j} + \beta_2 \text{Gender network measure}_\text{j}$$

where $Var(Y_{ij}|X_{ij})$ = p_{ijk} $(1 - p_{ijk})$ *(Bernouilli variance)*, *i*= household (i = 1,…,N), *j*= community (j = 1,…,n_j), and correlation is accounted for at the household-level.

Software

Network analyses were conducted in R (v. 3.4.2, R Foundation for Statistical Computing, Vienna, Austria) using the package igraph. Regression analyses were conducted in R (v. 3.4.2) using the package geepack.

5.4 Results

Qualitative

The results of the qualitative analysis fall into two primary themes. First, women experience a distinct set of stressors that inhibit the role they play in social cohesion and water security in rural, coastal Ecuador. Second, men and women identified water insecurity as a primary problem and deterrent to social cohesion. We expand on these themes sequentially.

1) Women experience a distinct set of stressor that inhibit their role in social cohesion and water security

The gender stressors felt by women were socio-politically, spatially, and sexually related (Figure 5.1), all of which affected the way they engaged in improved water practices in the household. We first highlight the socio-political concerns. Across communities, women felt neglected by both the national government and local authorities. Their issues were not being heard or given attention to. They also discussed lacking agency and empowerment or having limited empowerment, and thus being inhibited to take on leadership roles. The role of men in communities to support women was brought up by both men and women, making clear that not having gender equity and gender equitable men was an important stressor. Some men stated that women were key to the success of a community, while others recanted stories of women being used for sex or being raped. The communities with higher female agency and gender-equitable men importantly had higher self-perceived social cohesion.

"Of course women are strong. We would not be able to survive as a community without women. We have had two women presidents." - Afro-Ecuadorian man from a remote community

The spatial stressors included limited access to education because schools did not exist in communities and travel times were too long or only primary education was accessible, resulting in many dropping out of school. For some communities more than others, transportation was a major issue as commutes could take up a whole day sometimes to get to the nearest urban center or healthcare facility. Healthcare access, as a result, was cited as another stressor given distances to the nearest facility and also lack of pharmacies nearby.

The sexual stressors encompassed gender-based violence (GBV) and intimate-partner violence (IPV), and the added stress of birthing and caring for unwanted children. In communities where women experienced severe GBV and IPV, they never cited other issues being problems though water insecurity was a visible issue or discussed issue with men. Instead, much of the time was focused solely on lacking basic rights within the community and wanting to be heard.

The intensity of all stressors experienced by women were modified by ethnicity and remoteness; living environments and access to natural water sources change by both ethnicity and remoteness. In the communities close to a road, the spatial and sexual stressors were not as much of a concern as the socio-political drivers, like feeling neglected by the government. In the more remote communities, the spatial stressors were greater due to a lack of road infrastructure, exorbitant costs for gasoline to reach the nearest township, and limited access to healthcare. The remote communities that are Afro-Ecuadorian, however, had more gender equitable men and female agency compared to all other communities. One community stating having two female presidents of the community in the past and had a women's group, who organized community functions and ran the local bake shop. Among the remote communities, the Chachi women experienced the highest levels of GBV and IPV, and discussed lacking agency and gender equitable men.

"Even if you complain to authorities (about IPV), they say that as a woman you have no rights. The problems the women have are because of very machismo men." - Chachi woman from a remote community

All Chachi woman spoke of violence against woman and children and not having autonomy in the household to make decisions. One woman discussed performing multiple abortions on herself to prevent herself from having more children; she was the mother of six children at the age of thirty. Her husband would not allow her to visit the nearest healthcare facility. Another talked about miscarrying a child after being beaten by her husband.

"I gave myself two abortions because I didn't want anymore children and my husband wouldn't let me go to the clinic in Zapayo Grande." – Chachi woman from a remote community

Though the Afro-Ecuadorian did not discuss GBV of IPV as frequently, they often spoke of their husbands or boyfriends having multiple wives and families. In all the Afro-Ecuadorian communities, regardless of level of social cohesion or remoteness, the women sang *arullos* to share stores of their community. Women in the Chachi communities did not sing but rather discussed how the stressor impacted their ability to engage in improving household hygiene and water security.

2) Men and women identified water insecurity as a primary problem and deterrent to social cohesion

Four major concerns of water insecurity were highlighted by community members: environment, access, reliability, and quality (Figure 5.2). Harvesting rainwater is seasonally dependent resulting in the environment being an issue. Sometimes there was the perception that rainwater is dirty and less optimal than river water for daily use. Also, flooding contributes to less access to both water and agriculture. Some individuals reported sleeping in a canoe in their households during the flooding season. The environment additionally allows for nearby streams or tributaries to fill up.

"When it floods here, we lose everything. We sleep in a canoe in our house because the water levels rise so high." - Chachi woman from a remote community

Concerns of access included the time required to collect water from a water source, like a stream or river, and having any water infrastructure in the household or community. In most of the remote communities, there was no potable water and woman and children had to bring water from the riverbed daily, which included either hiking up a hill or climbing stairs. Reliability concerns not having to worry about the adequacy of the water supply, so that the water supply is constant. Many discussed how quickly the amount of water collected or stored was reduced due to the need of water for various purposes throughout the day like bathing, washing dishes, washing clothes, and drinking. Quality concerns the type of water source used for drinking (protected versus unprotected) and contamination level. In these study communities, river water was often discussed as being contaminated from gold mining pollution upstream, the palm oil industry, and pesticides in roadside communities. Individuals did not want to use dirty water, but felt they have no choice.

> *"There is no water here that is clean. The tube water is the river water, which is contaminated from gold mines." – Afro-Ecuadorian woman from a road community*

The intensity of the issues regarding water insecurity were also modified by remoteness and ethnicity. In the communities close to a road, people spoke more of issues concerning reliability and quality. Mercury contamination of the rivers due to gold mining was a major issue discussed as these communities are located downstream from mines and experience the accumulation of river contamination from upstream. Though women in the more remote communities brought up having infections, both dermal and vaginal, from spending time in the rivers, they were more concerned with access and environmental related issues. Men and women stated not having a centralized water system was a major concern. They only have river or rain water to use, which is seasonally affected. Road communities, on the other hand, stated having malfunctioning tube systems that were built years ago, but that are now contaminated. However, they also said non-governmental agencies have visited to distribute different water treatment methods. Communities at a medium distance from the road felt neglected and also stated access and the environment as major issues. One community said a foreign engineer had visited years back and installed a water tube system connecting spring water to the communities, but it has since broken. The Chachi men discussed quality and access of water security as issues, while the Chachi women focused on GBV and lack of human rights as priority issues before water.

> *"A foreigner came and set up a well and tube-water system. It worked for a few years and then stopped." – Chachi man from a remote community*

Social network

Selecting the right method for measuring uncertainty

For jackknifing, increasing the percentage of nodes removed increased the estimate variance greatly for both assortativity and cluster modularity (Supplementary Figure 5.6 & Supplementary Figure 5.7). Increasing the percentage of edges removed did not vary the uncertainty as much. This was consistent across time and the same trend held within the Chachi communities, where the network sizes were smaller (Supplemental Figure 5.8). Given this and qualitative data in the communities suggesting edges are more likely to break between nodes than nodes being removed, we chose 10% edge removal for the jackknifing method for assortativity and cluster modularity to compare to other uncertainty methods.

Bootstrapping nodes results in confidence intervals that don't necessarily contain the true estimate, presenting more variability than reasonably likely (Supplemental Figure 5.9). Compared to bootstrapping edges, jackknifing edges 10% provided more conservative uncertainty estimates. Using the Newman error calculation method of removing one edge at a time provided very conservative error estimates <0.001. Estimating a null hypothesis to determine uncertainty does not provide error estimates, but rather information that what we observed is different than what is randomly observed. Thus, based on this, we'd determine what we observed was often substantial. The same patterns held true for cluster modularity (Supplemental Figure 5.10). As a result, for both cluster modularity by gender and assortativity we use edge bootstrapping to attain confidence intervals for our estimates.

For edge density, the bootstrapping method provided less conservative confidence intervals than jackknifing, so for this we also report the bootstrapped confidence intervals (Supplemental Figure 5.11).

<u>Network results</u>

Our total study population consisted of 1,616 individuals in 2004, 2,204 individuals in 2007, 2,371 in 2010, 2,326 in 2013, and 274 in the Chachi communities. These individuals made up 15 independent community networks in 2004, 20 from 2007-2013, and 3 in the Chachi communities. The average number of individuals per network was approximately 115 and 91 in the Chachi community. The degree distributions between genders was not markedly different over time (Supplementary Figure 5.2). Across time and remoteness categories, the percentage of women in communities stay relatively consistent (Table 5.1). The ethnic breakdown of communities at close and far distances from a road is more diverse over time than communities at a medium distance from a road, who have a large Mestizo population, but 0% Chachis.

Close communities are more assortative, while communities at a medium and far distance from a road are less assortative (Figure 5.3). In 2013, gender assortativity was on average 0.12 (.03, .32) in close communities compared to -0.05 (-.16, .18) in medium distanced communities and -0.01 (-.26, .24) in far communities, where a positive value means men and women are more assortative and a negative value means men and women interact more. Close communities had more within gender sub-communities than between gender sub-communities as measured by cluster modularity. In 2013, cluster modularity was on average 0.06 (.017, .16) in close communities compared to -0.03 (-.08, .08) in medium distanced communities and -0.005 (-.13, .07) in far communities. For both assortativity and cluster modularity, variability of estimates was observed between medium and far communities, which ranged from negative to positive values.

There was no remarkable difference between edge density measures from subgraphs of men and women, suggesting men and women had an equal number of contacts. For both men and women, edge density was greater in communities at a medium distance from a road. As noted in previous studies, the overall hygiene score and community sanitation improved over time in this study region. Rain water reduced to 4% over time in communities close to a road, while staying relatively the same at ~50% in far and medium distanced communities.

Compared to the Afro-Ecuadorian remote communities sampled at the same time-period, the Chachi communities differed in gender assortativity (Table 5.2). The Chachi communities were on average more assortative (0.06 (-.02, .12)) than the Afro-Ecuadorians (0.006 (-.12, .09)). The Chachis had slightly more within gender sub-communities than between gender sub-communities (0.026 (-.01, .05)) than the Afro-Ecuadorians (0.001 (-.06, .05)). There were no differences in edge density between the ethnic groups. WASH practices, however, were markedly different in Chachi communities compared to the Afro-Ecuadorian communities. The average of proportion of households within a 500 meter radius with improved sanitation was 49% in Chachi communities compared to 79% in Afro-Ecuadorian communities. Eighty-four percent of Chachis harvest rainwater compared to 56.3% of Afro-Ecuadorians.

Using the multilayer property of time between community networks, the isometric layouts and visualizations of San Miguel (Figure 5.4), a far community, and San Augustin (Figure 5.5), a community close to a road, illustrate the difference in assortativity and modularity. San Miguel is less assortative in 2004 (0.087 (-.044, .209)) and becomes less assortative by 2013 (0.025 (-.142, .2)) (Table 5.3). The cluster modularity of San Miguel also decreases over time. The assortativity of San Augustin is larger in 2004 (0.162 (.053, .268)), but decreases over time toward becoming less assortative in 2013 (0.034 (-.068, .128)) (Table 5.4); the number of between gender ties increases over time. Notably, the ethnic breakdown is different in the two communities. The isometric layout and visualization of the Chachi communities also illustrate differences in assortativity and modularity (Figure 5.6).

Based on the GEE models, when remoteness is controlled for, being more assortative by gender in networks resulted in a lower likelihood of improving hygiene, with this effect attenuating over time (Figure 5.7). For every unit increase in community assortativity, a household hygiene score decreases by -0.43 (-.52, -.35) in 2004 (Supplemental Table 5.1). By 2013, for every unit increase in assortativity, a household hygiene score decreases by -0.07 (-.12, -.02). In the Chachi communities, the effect of assortativity on hygiene is not significant. Being more assortative in networks consistently resulted in decreased community sanitation and smaller odds of harvesting rainwater (Figure 5.7). For every unit increase in assortativity, community sanitation around a household decreases by -0.63 (-.71, -.55) in 2004 and decreases by -0.45 (-.49, -.42) in 2013. The odds ratio of harvesting rainwater was 0.18 (.09, .37) times less for households in communities that are more assortative (Supplemental Table 5.2) in 2013. In the Chachi communities, every unit increase in assortativity resulted in a 1.06 (.73, 1.38) increase in community sanitation. The odds ratio of harvesting rainwater was 3.62 (.03, 38.6) times greater for every unit increase in assortativity. The same trends were observed for the effect of cluster modularity on hygiene, community sanitation, and harvesting rainwater in the Afro-Ecuadorian and Chachi communities.

5.5 Discussion

When women lack agency in communities, the provision of public goods and community health is diminished. Empowering women as decision-makers in both the household- and community-levels is critical for social change and improving gender equity in the wider socio-economic environment is crucial for empowerment. Social capital is an umbrella concept that refers to the social resources required to commit collective action for individuals and groups of individuals (e.g. at the household- and community-levels), like cohesion and, on the contrary, lack of conflict, and trust. Gender equity, and more generally power dynamics in social relationships, is another potential social resource, but also a strong component of cohesion as we show here. By integrating, qualitative and quantitative data, this paper elucidates the role of gender equity in reducing community cohesion and thus social capital leading to reduced WASH practices. Indeed, social interaction between men and women, devoid of conflict like gender-based violence, is critical in the context of community progress for WASH measures, a necessary public good for diarrheal disease reduction. Below we summarize the main findings and potential mechanisms using social theory. We then discuss limitations and future directions.

Summary & Mechanisms

In previous studies, we demonstrated the effect of community cohesion, as modeled by social network data from a core discussion network (CDN), on reducing acute gastrointestinal illness (AGI) and showed how this effect is partially mediated through safe WASH practices. Here, we demonstrate the effect of gender equity on community cohesion and WASH through both qualitative and statistical methods, and the effect of water insecurity on community cohesion through qualitative methods (Figure 5.8A). Gender equity and having gender equitable men[59] play a significant role and contribute downstream to female agency and empowerment. Communities with higher agency amongst women experienced high social cohesion, with men playing a critical role in creating gender equity and enhancing agency among women, while communities with less cohesion cited issues of gender-based violence (GBV).

In the remote Afro-Ecuadorian communities we found mostly stronger community cohesion qualitatively, with men discussing the critical role of women in society and how the community would not be able to progress without the success of women. This sentiment illustrates the gender equity required for increased cohesion and interventions and is echoed in the quantitative data through decreased assortativity and cluster modularity. In contrast, in the roadside Afro-Ecuadorian communities social conflict between men and women and within and between communities still exists.

In 1965, Norman Whitten, an anthropologist, visited the same study region and wrote a book titled *Class, Kinship, and Power* primarily about the Afro-Ecuadorian experience. In this book, he described how “social structure change results from socioeconomic mobility” and how women are expected to “trap and hold a man, to make them an stay in one place, and, when that is impossible, to travel with him to set up a new home” [157]. If a woman attempts to leave her husband or have an affair herself, then her husband will beat her severely or may exile her from the household. A woman is only free once a man has relinquished his role in her life. Men, on the other hand, are described as free to do as they please, to leave a woman or find another as they choose, and are valued on being able to find a spouse in any context. Thus, household chores are not valued for men. Even if a man has multiple wives, he represents the role of husband-father in all households and so is responsible for the actions of all women and children. A female is head of a household only when there is no male spouse. Young girls were also cited as being beaten if they were suspected of having sexual intercourse, however, young men were expected to behave irresponsibly, lie, waste money, and get drunk. Being arrested for stealing or rape is no disgrace. Even though Whitten wrote his book more than 50 years ago, some of his keen observations still hold true today.

Just as the more remote communities have found gender equity as a way of improving community cohesion, the roadside communities and those at a medium distance from a road have retained behaviors of conflict described years ago by Norman Whitten. Of note, one roadside community did describe strong female leadership and having a woman as community president, however, this community lacks cohesion as it primarily consists of descendants of two families and when one family is in power the other is in conflict. As such, more extremes, like gender equity and violence, are observed in the Afro-Ecuadorian communities compared to the Chachi communities.

In the study of power dynamics, gender equity has also been described more specifically as having gender equitable men (GEM). Men and women's roles and attitude according to gender are classified as traditional and egalitarian roles. Roles attributed to women in traditional roles consist of non-egalitarian responsibilities such as being responsible for domestic affairs and not being active in a professional life. Roles attributed to men in traditional roles consist of responsibilities such as being the head of the house and being responsible for breadwinning. Egalitarian roles, however, are equal sharing of responsibilities in family, professional, social and educational life.[158–162] Social norms and attitudes which put men in a position of dominance have consequences for women's ability to control their own reproductive and sexual health and also agency and empowerment.

As shown in our data, when given the forum to speak, a majority of women expressed a need for agency and empowerment; giving women an equal voice in the household and community reduces the power dynamics between men and women and leads to female agency and empowerment. The concept of empowerment is explored through three closely interrelated dimensions: agency, resources, and achievements. Agency represents the processes by which choices are made and put into effect. Resources are the medium through which agency is exercised; and achievements refer to the outcomes of agency [163]. Agency in all its forms is critical for women's empowerment. The domains selected to measure women's agency include choice surrounding sexuality, marriage, childbearing, and the exercise of reproductive rights; making decisions in the family; participation in labor, land, and financial markets; and, engagement with collective action and politics. Agency has intrinsic value and invokes an ability to overcome barriers, to question or confront situations of oppression and deprivation, as individuals or with others, and to have influence and be heard in society [164]. Women's agency can thus lead to empowerment to change regressive norms and institutions that perpetuate the subordination of women. When agency doesn't exist, women often are not empowered to act in leading roles in the household and community, as seen in our study population, particularly the Chachi women. Female empowerment and agency alone, however, do not directly lead to gender equity. Men play a critical role in the fight for change in both the household and community.

A key part of achieving gender equality is changing the social norms that both men and women internalize and that influence their practices. Survey research with men and boys in numerous settings has shown how inequitable and rigid gender norms influence men's practices on a wide range of issues, including HIV/STI prevention, contraceptive use, use of physical violence (both against women and between men), domestic chores, care giving, and health seeking behaviors.[165–172] Sample survey research using standardized attitude scales, including the GEM Scale [173], has found that adult and younger men who adhere to more rigid views about masculinity (e.g., believing that men need sex more than women, that men should dominate women, that women are "responsible" for domestic tasks, among others) are more likely to report use of violence against a partner, sexually transmitted infection, previous arrests and drug or alcohol use. As noted in our study population, the lack of gender equity and more misogynistic rhetoric is correlated with behaviors like violence and rape.

Indeed, issues of violence and rape are forms of social conflict, which is importantly not independent of social cohesion and other social constructs like collective efficacy and social capital. Social scientists discuss various conceptualizations of social constructs like collective efficacy, social capital, and social cohesion that gender equity contributes to, and there is much debate about how these constructs relate to each other. General conceptualizations of collective efficacy assume it is a latent construct comprised of a combination of the structural and cognitive components that facilitate a community's shared belief in its ability to come together and execute actions related to a common goal.[49] Social capital, on the other hand, is often conceptualized as features of social structures, such as trust, norms of reciprocity, and mutual aid, that act as resources for individuals to facilitate collective action.[50–52] Social capital is also commonly conceptualized as a component of social cohesion.[53,54] Some conceptualizations of social cohesion conceive it as a bottom-up process with a foundation in local social capital, where the social capital of a community takes on a strong sense of local space, albeit with ambiguous and fluid boundaries.[54] Therefore, social cohesion refers to two broader features of society, 1) the absence of latent social conflict (i.e. presence of social homogeneity); and 2) the presence of strong social bonds.[53]

Thus, it’s important to consider cohesion as not only being an indicator of reduced morbidity and social progress, but also an indicator of social conflict like gender-based violence and intimate partner violence. As seen in the Chachi communities in this study and prior studies, stronger community cohesion results in reduced hygiene and community sanitation, and stronger assortativity of genders results in increased community sanitation, but not hygiene. Social conflict is likely resulting in reduced adoption of WASH practices, as the Chachi women described themselves, they cannot prioritize WASH issues given the severity of the GBV. Furthermore, as Chachi women are unable to engage in household-level WASH practices, assortativity and cluster modularity had a null effect on hygiene, a stronger measure of household activity. Increased separation between men and women, as measured by the network statistics, however, was correlated with increased community sanitation, likely a measure of men, who have more power, building infrastructure.

The network data significantly emphasized the qualitative observations through the measures of assortativity and cluster modularity by gender, good proxies for gender equity. More gender assortativity and cluster modularity in communities (more within gender ties and sub-communities than between gender ties and sub-communities) resulted in reduced hygiene, sanitation, and rainwater harvesting at the household-level. Of note, degree, a common network measure used to describe relational differences, showed limited differences in our egocentric networks, thus limiting how much information we could elucidate from the data. Using the network data allowed us to quantify gender equity through both assortativity and cluster modularity, one term assessing within versus between group ties and the other assessing within versus between group sub-communities. With our qualitative data, we were additionally able to validate the meaning of the network statistics in our dataset. Though network theorists more often use assortativity to examine homophily: the preference of a network node to attach to others that are similar, here we put focus on communities being less assortative by gender to stress the presence gender equity. Lastly, using network data allowed us to note that men don't act much differently than women, with respect to edge density and degree distribution. In fact, it's the interaction between men and women that matters in the context of community progress more than men being different than women; men and women largely have similar behavior socially but a difference exists in how women and men engage together in social ties and sub-communities, which directly translates to gender equity.

Limitations

Our social network data is limited as we do not account for kinship, a key social influence in the household. Additionally, our WASH practice measures of sanitation and harvesting rainwater are indicators of access and not use. Given more precise measures of use, our effect estimates might show stronger associations. Due to limited differences in ethnicity in the majority Afro-Ecuadorian communities, we were unable to look at any meaningful differences between ethnic groups in the same network or community. Though likely a modifier, we do not discuss differences by age groups, though it is probable that issues and experiences differ among youth aged 13 to 18 years in communities.

Additionally, using both qualitative and quantitative data allows us to show rigor of our results, however, qualitative data collected over a 4-month time span like survey data collected at a single time-point does not capture true behavior. Instead, these type of data are hypothesis generating and suggest that a more critical, investigative study be done. Qualitative data also provides context of behavior that is occurring at the individual-, household-, and community-levels, important as our objective focuses on the role of community in household WASH practices.

Future directions

This study could inform both theoretical frameworks for behavior change and context-specific and gender-sensitive interventions, improving both effectiveness and sustainability. If enabled, women could reduce diarrheal disease through education on not only WASH measures, but even treatment management (recognizing symptoms in children, providing zinc). Gender equity is important in community cohesion and WaSH, and gender is a salient example in which individuals may be isolated if they're marginalized within communities.

We also see that gender inequalities are multi-dimensional and cannot be reduced to some single and universally agreed set of priorities. They contain contradictions and imbalances, particularly when there have been changes in the wider socio-economic environment. Unless provision is made to ensure that policy changes are implemented in ways that allow women themselves to participate, to monitor, and to hold policy makers, corporations, and other relevant actors accountable for their actions, this potential is unlikely to be realized. For example, women's access to paid work may give them a greater sense of self-reliance and greater purchasing power, but if it is undertaken in conditions that erode their health and exploit their labor, its costs may outweigh its benefits. The question, therefore, is to what extent the international community is prepared to provide support to women at the grassroots – support which will ensure that they have the collective capabilities necessary to play this role.[163] In the gradient of human connection and action between the human social environment and the natural environment, gender equity exists, and unless practiced can hinder successful actions like WASH practices at the household- and community-levels; respectful engagement of women through leadership in communities is critical and intimate partner violence (IPV) is a critical inhibitor (Figure 5.8B).

A global study in 2013 found over 35% of women worldwide have experienced physical or sexual violence with a partner (IPV) or non-partner sexual violence.[174] IPV, violence at the hands of a husband, boyfriend, or partner, is both the most pervasive form of gender-based violence and one that few governments recognize as a crime[175], though it violates basic human rights and affects the individual, family, and economy. Focusing on men and boys for the nurturing and growth of gender equity in communities is essential.[172] Future studies should additionally investigate other forms of power dynamics and sexual relations including the marginalization of homosexual and transgender populations.

Aside from engaging with community partners on the ground, other future studies could examine diffusion simulations on a multilayer network across time to estimate how much the concept of certain WASH practices spread given particular parameters of the data and estimate how much disease risk results. Novel methods currently exist for Bayesian learning on multilayer networks that could also allow for the estimation of rates of change of assortativity and cluster modularity over time [176]. Lastly, different optimal measures for quantifying the uncertainty in network degree distributions exist based on adapting bootstrapping methods for time series and re-tiling spatial data to random networks, by first sampling a set of multiple ego networks of varying orders that form a patch, or a network block analogue, and then resampling the data within patches. To then select an optimal patch size, researchers developed a new computationally efficient and data-driven cross-validation algorithm. [177] Such types of methods can be explored and further investigated for uncertainty quantification of assortativity and cluster modularity.

Figure 5.1. Gender stressors in the social environment.

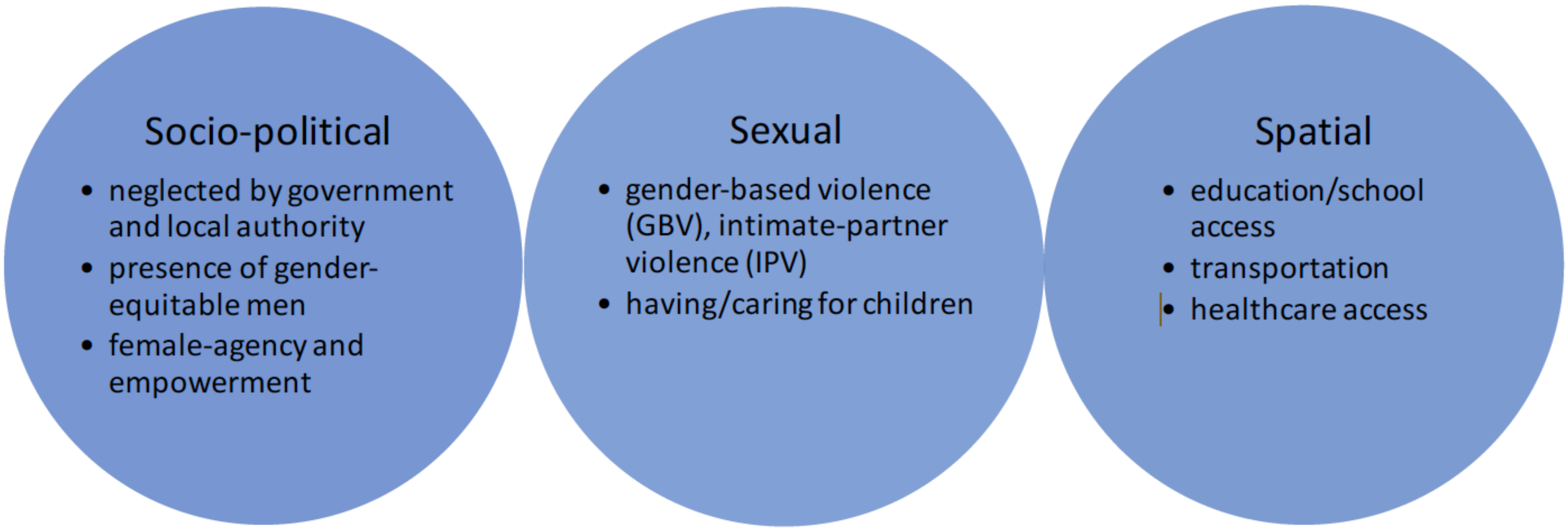


Figure 5.2. Issues of water insecurity.

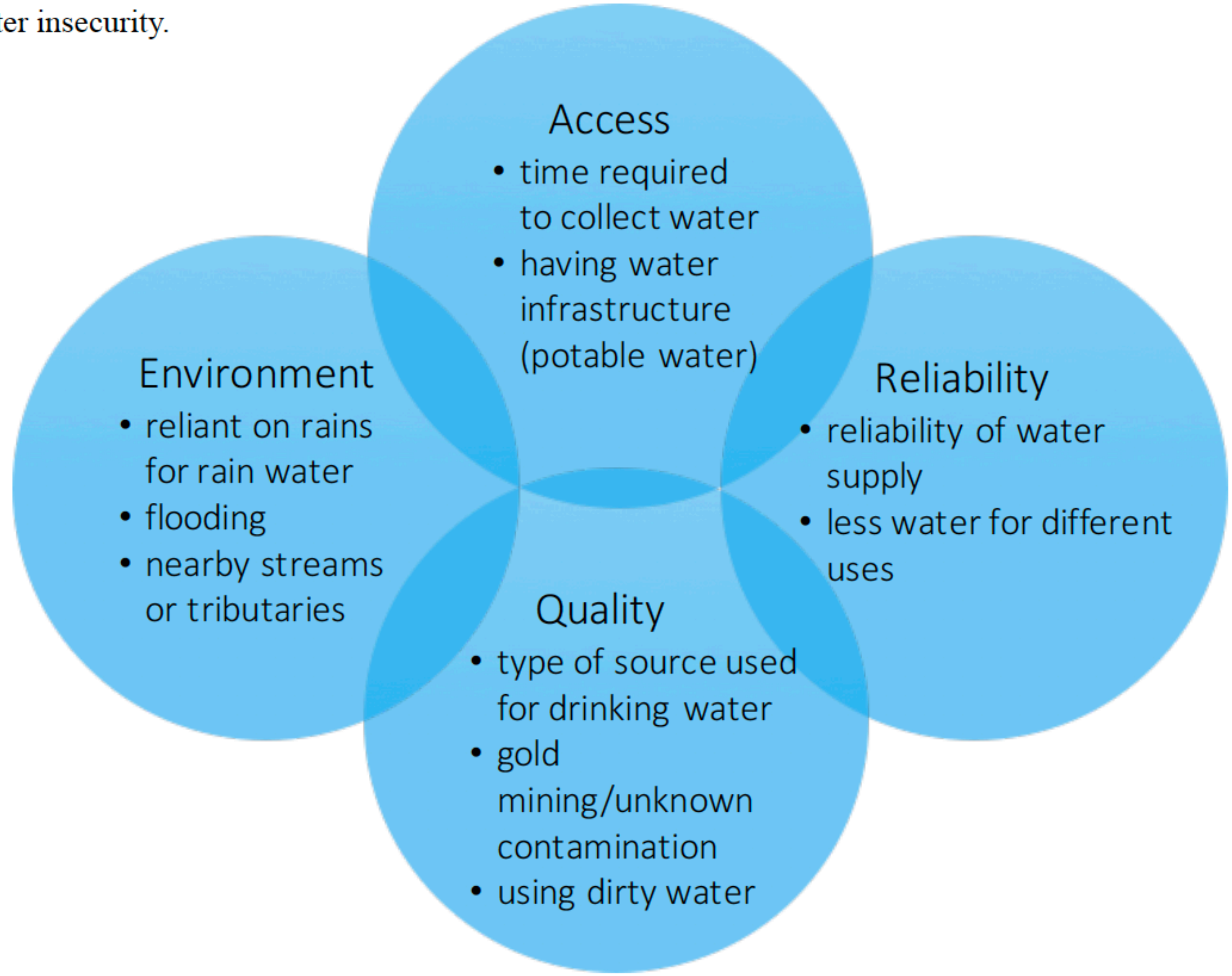
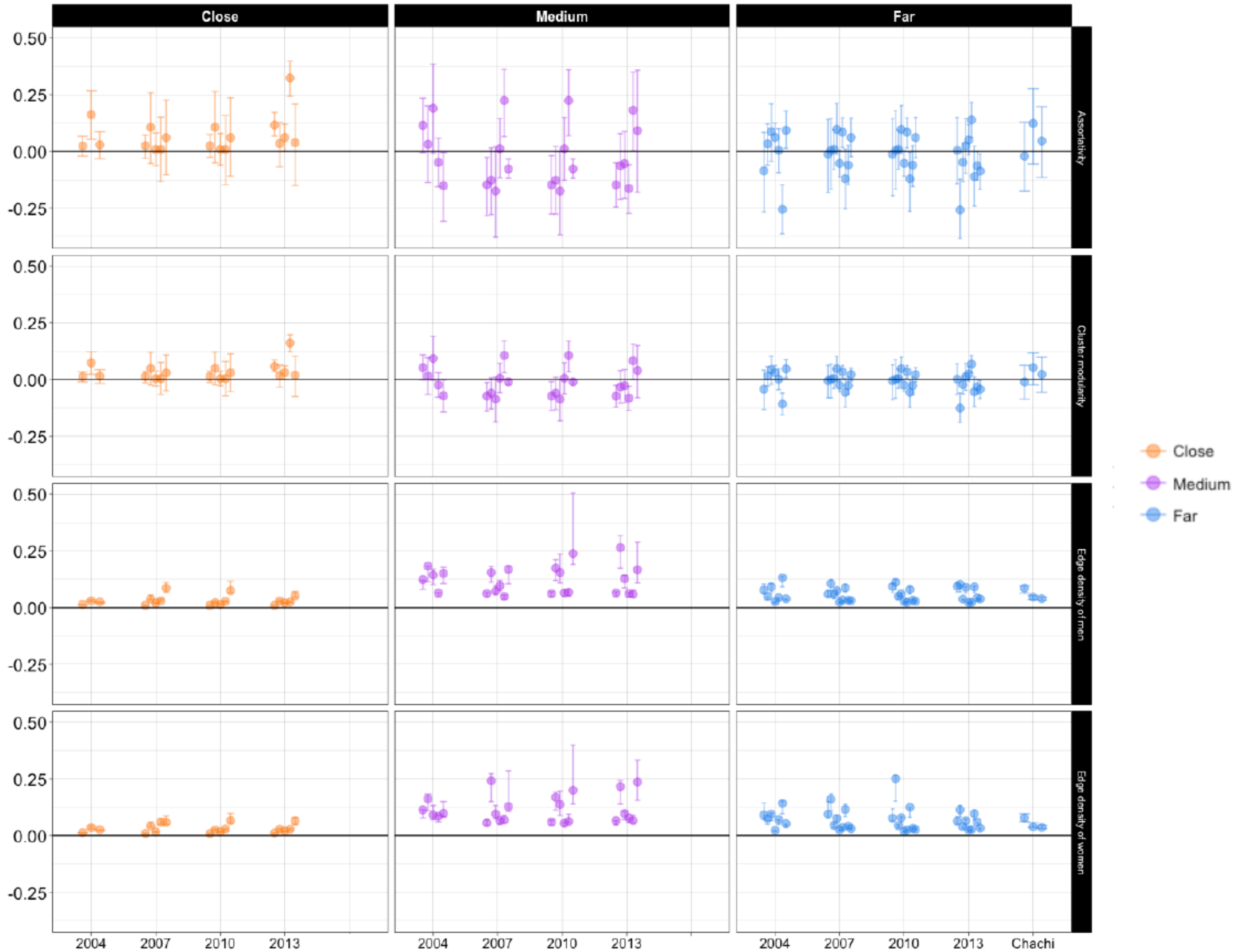


Table 5.1. Descriptive statistics in Afro-Ecuadorian communities

	2004 N=1616 Households=719			2007 N=2204 Households=1005			2010 N=2371 Households=1121			2013 N=2326 Households=1100		
Remoteness-level	Close	Medium	Far	Close	Medium	Far	Close	Medium	Far	Close	Medium	Far
Gender (Female)	47%	46%	45%	51%	49%	47%	54%	51%	51%	55%	51%	51%
Ethnicity												
Afro-Ecuadorian	90%	93%	96%	87.4%	83.8%	94.8%	87.4%	80.9%	94.6%	88.5%	82%	94.7%
Mestizo (mixed)	9%	6%	3%	9.8%	14.6%	3.9%	9.8%	18.2%	4.6%	9.9%	17.4%	4.3%
Chachi	0%	0%	1%	.4%	0%	1.2%	.4%	0%	.5%	.2%	.3%	.8%
Other	1%	1%	0%	2.4%	1.5%	.1%	2.4%	.9%	.3%	1.4%	.3%	.2%
Gender assortativity	0.05 (.02, .16)	0.02 (-.15, .19)	0.03 (-.26, .09)	0.03 (.01, .11)	-0.02 (-.17, .22)	0.007 (-.12, .10)	0.03 (.01, .11)	-0.03 (-.17, .22)	0.006 (-.12, .09)	0.12 (.03, .32)	-.05 (-.16, .18)	-0.01 (-.26, .14)
Cluster modularity by gender	0.02 (.01, .07)	0.01 (-.07, .09)	0.02 (-.11, .05)	0.016 (.004, .05)	-0.01 (-.09, .11)	0.002 (-.06, .05)	0.015 (.004, .05)	-0.01 (-.09, .11)	0.001 (-.06, .05)	0.06 (.017,.16)	-0.03 (-.08, .08)	-0.005 (-.13, ..07)
Number of ties to women	1.15 (0-13)	1.02 (0-11)	1.05 (0-10)	0.77 (0-9)	0.7 (0-7)	1.12 (0-22)	0.91 (0-9)	0.77 (0-9)	0.99 (0-10)	1.1 (0-10)	1.3 (0-11)	1.3 (0-12)
Edge density between women	0.02 (.01, .03)	0.11 (.09, .16)	0.06 (.02, .14)	0.02 (.01, .06)	0.09 (.06, .24)	0.05 (.03, .16)	0.02 (.01, .07)	0.09 (.05, .20)	0.05 (.02, .25)	0.02 (.01, .07)	0.10 (.07, .24)	0.05 (.03, .11)
Edge density between men	0.02 (.01, .03)	0.12 (.07, .18)	0.05 (.03, .13)	0.02 (.01, .09)	0.09 (.05, .17)	0.04 (.02, .11)	0.02 (.01, .07)	0.10 (.06, .24)	0.04 (.02, .11)	0.02 (.009, .05)	0.10 (.06, .26)	0.04 (.02, .10)
WASH measures												
Hygiene score	0.6 (0-1)	0.58 (0-1)	0.6 (0-0.9)	0.6 (0-1)	0.6 (0-0.95)	0.6 (0-1)	0.7 (0.2-1)	0.7 (0.4-0.9)	0.7 (0-1)	0.7 (0-1)	0.7 (0-1)	0.8 (0-1)
Community sanitation	0.41 (0.12-0.52)	0.28 (0-0.54)	0.48 (0.31-1)	0.53 (0-1)	0.49 (0-0.76)	0.70 (0-0.90)	0.75 (0-1)	0.57 (0-1)	0.79 (0-0.94)	0.85 (0-0.97)	0.58 (0-0.75)	0.88 (0.5, 0.96)
Piped water	2.0%	0.009%	21.3%	0.9%	0.9%	2.3%	33.3%	0.0%	22.4%	77.7%	0.0%	16.9%
Rain water	21.2%	14%	28.7%	24.4%	45.7%	53.6%	41.3%	59.8%	56.3%	4.4%	46.8%	53.3%

Figure 5.3. Forrest plot of different network statistics by remoteness



Gender assortativity (bounded by -1 to 1), cluster modularity by gender (bounded by -1 to 1), and edge density (proportion bounded by 0 to 1) statistics of all communities over time, colored by remoteness category. Each point represents a different community and we estimate uncertainty by bootstrapping edges.

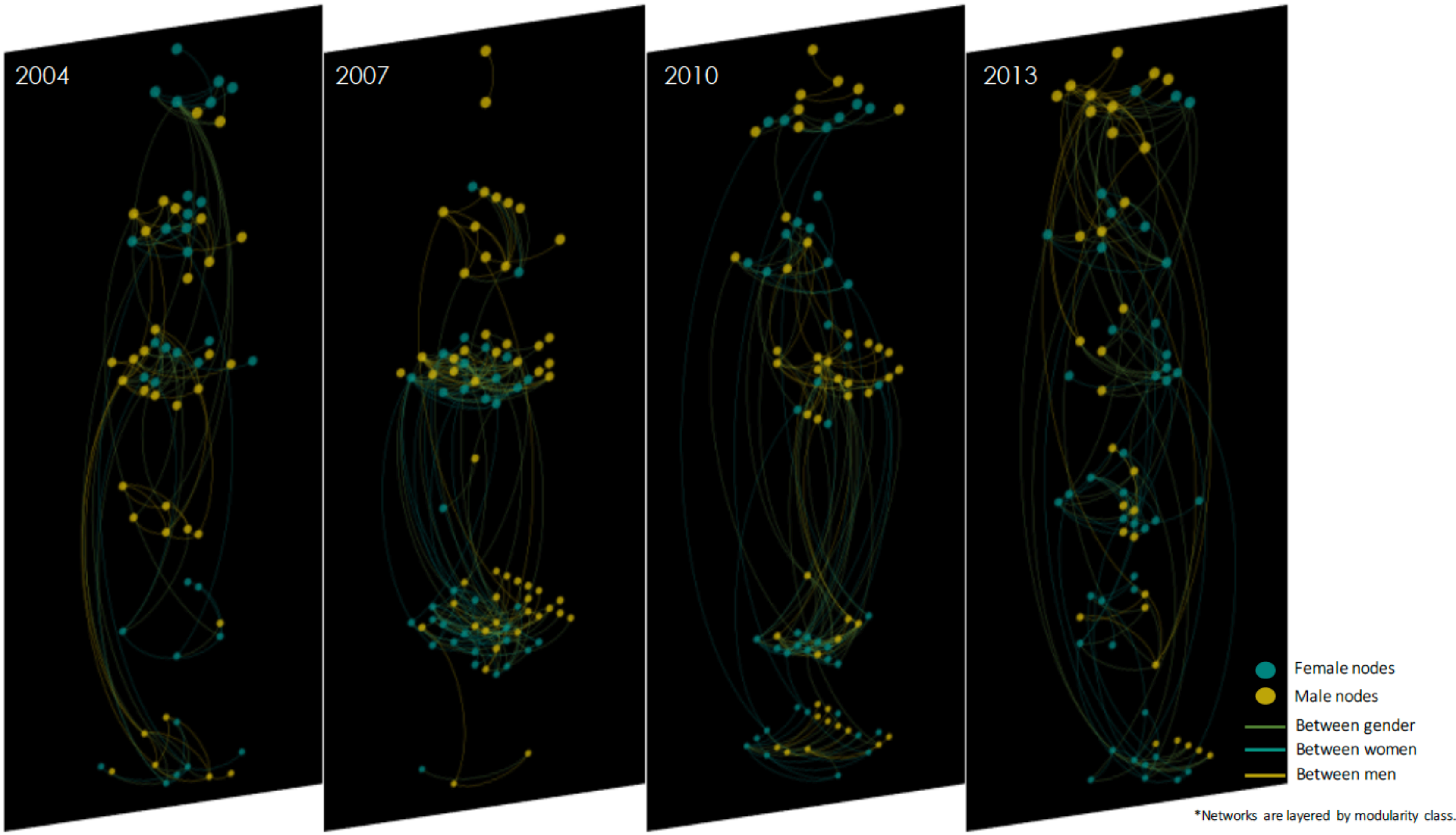
Table 5.2. Descriptive network statistics of Afro-Ecuadorian and Chachi communities

	Afro-Ecuadorian N=978 Households=443	Chachi N=274 Households=93
Gender (Female)	51%	52%
Ethnicity		
Afro-Ecuadorian	94.6%	1.5%
Mestizo (mixed)	4.6%	6.6%
Chachi	.5%	91.9%
Other	.3%	0%
Gender assortativity	0.006 (-.12, .09)	0.059 (-.02, .12)
Cluster modularity by gender	0.001 (-.06, .05)	0.026 (-.01, .05)
Number of ties to women	0.99 (0-10)	0.48 (0-5)
Edge density between women	0.05 (.02, .25)	0.046 (.04, .08)
Edge density between men	0.04 (.02, .11)	0.05 (.039, .085)
WASH measures		
Hygiene score	0.7 (0-1)	0.5 (0.04-0.90)
Community sanitation	0.79 (0-0.94)	0.49 (0-0.77)
Piped water	22.4%	0.0%
Rain water	56.3%	84.2%

*Since data in the Chachi communities was only collected in 2010, we compare data from the remote Afro-Ecuadorian communities in 2010.

Descriptive statistics of Afro-Ecuadorian communities in remote areas compared to Chachi communities in 2010.

Figure 5.4. Network visual of assortativity and modularity for a far community



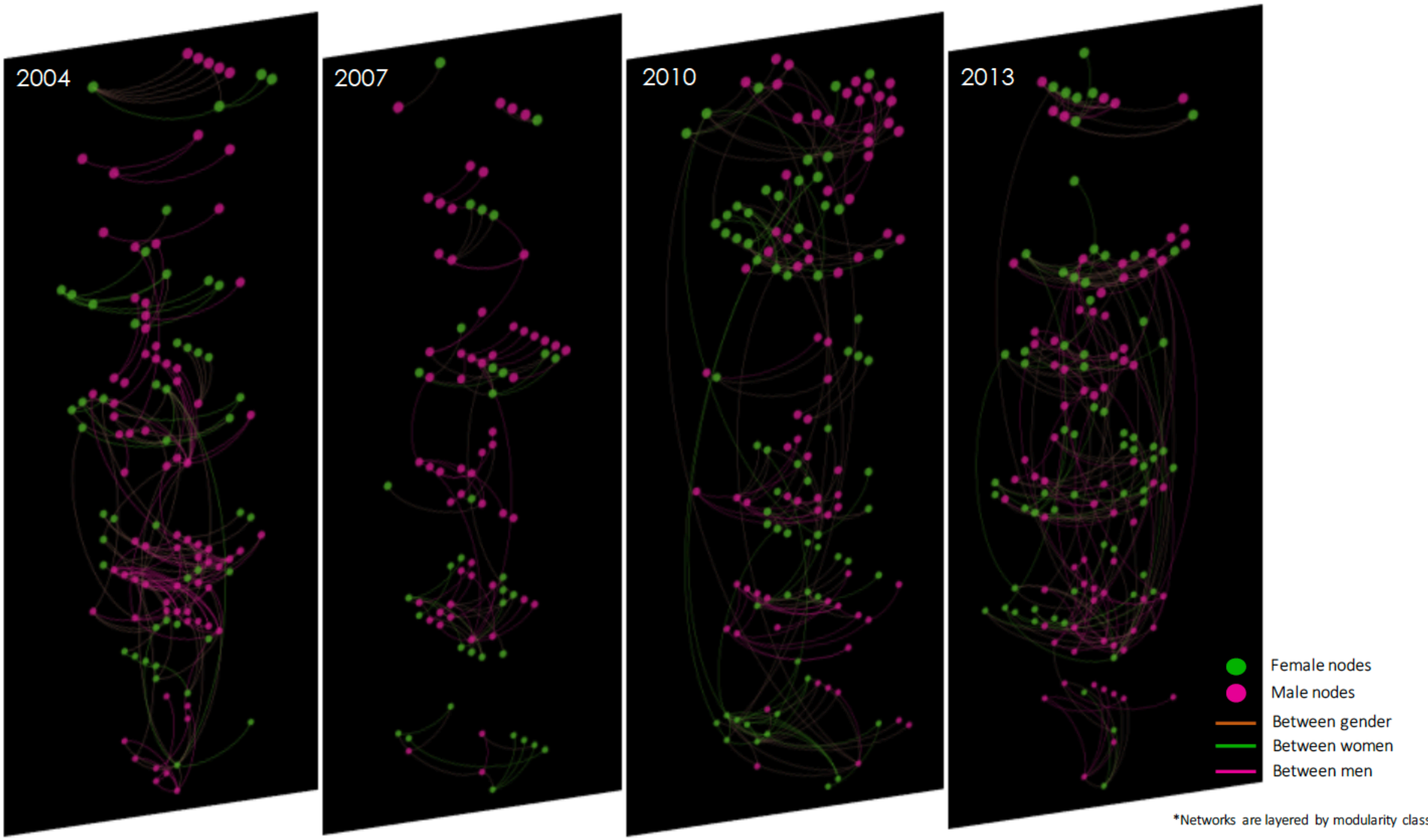
Multilayer networks of San Miguel, a remote community, from 2004 to 2013, shown by modularity layers and gender specific nodes and ties.

Table 5.3. Descriptive network statistics for a far community

	San Miguel			
	2004	2007	2010	2013
Gender (Female)	47%	51%	54%	50%
Ethnicity				
Afro-Ecuadorian	87.2%	89.7%	95.8%	94.8%
Mestizo (mixed)	6.4%	4.4%	1.4%	3.5%
Chachi	6.4%	5.9%	2.8%	1.7%
Other	0%	0%	0%	0%
Gender assortativity	0.087 (-.044, .209)	0.097 (-.013, .211)	0.097 (-.009, .202)	0.025 (-.142, .2)
Cluster modularity by gender	0.043 (-.022, .104)	0.047 (-.006, .101)	0.047 (-.004, .098)	0.012 (-.047, .071)
Number of ties to women	1 (0-10)	1.16 (0-10)	1.21 (0-8)	1.21 (0-5)
Edge density between women	0.095 (.067, .099)	0.076 (.059, .091)	0.080 (.057, .083)	0.066 (.05, .071)
Edge density between men	0.091 (.068, .105)	0.074 (.056, .080)	0.061 (.046, .066)	0.089 (.063, .09)

Descriptive data from San Miguel, a remote community, from 2004 to 2013.

Figure 5.5. Network visual of assortativity and modularity for a roadside community



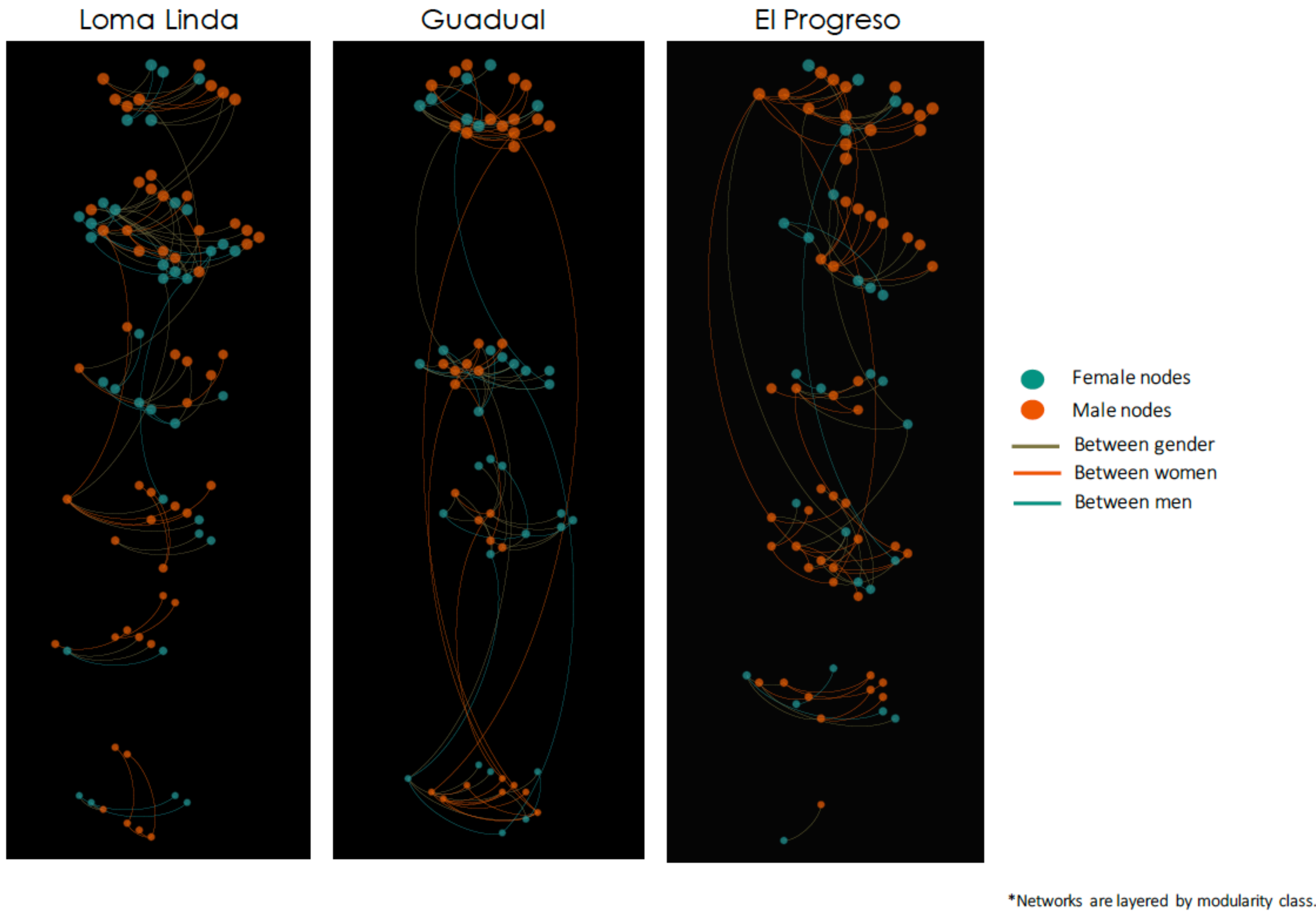
Multilayer networks of San Augustin, a roadside community, from 2004 to 2013, shown by modularity layers and gender specific nodes and ties.

Table 5.4. Descriptive network statistics for a roadside community

	San Augustin			
	2004	2007	2010	2013
Gender (Female)	47%	53%	55%	51%
Ethnicity				
Afro-Ecuadorian	64%	69.7%	59.3%	71%
Mestizo (mixed)	28%	13.9%	24.1%	19.8%
Chachi	8%	16.4%	1.4%	1.5%
Other	0%	0%	15.2%	7.6%
Gender assortativity	0.162 (.053, .268)	0.106 (-.054, .258)	0.106 (-.051, .265)	0.034 (-.068, .128)
Cluster modularity by gender	0.074 (.023, .122)	0.049 (-.024, .119)	0.049 (-.024, .12)	0.017 (-.034, .063)
Number of ties to women	0.66 (0-5)	0.37 (0-5)	0.68 (0-4)	0.85 (0-5)
Edge density between women	0.035 (.031, .044)	0.043 (.038, .061)	0.025 (.022, .03)	0.029 (.024, .031)
Edge density between men	0.030 (.025, .034)	0.037 (.034, .055)	0.022 (.02, .027)	.030 (.025, .034)

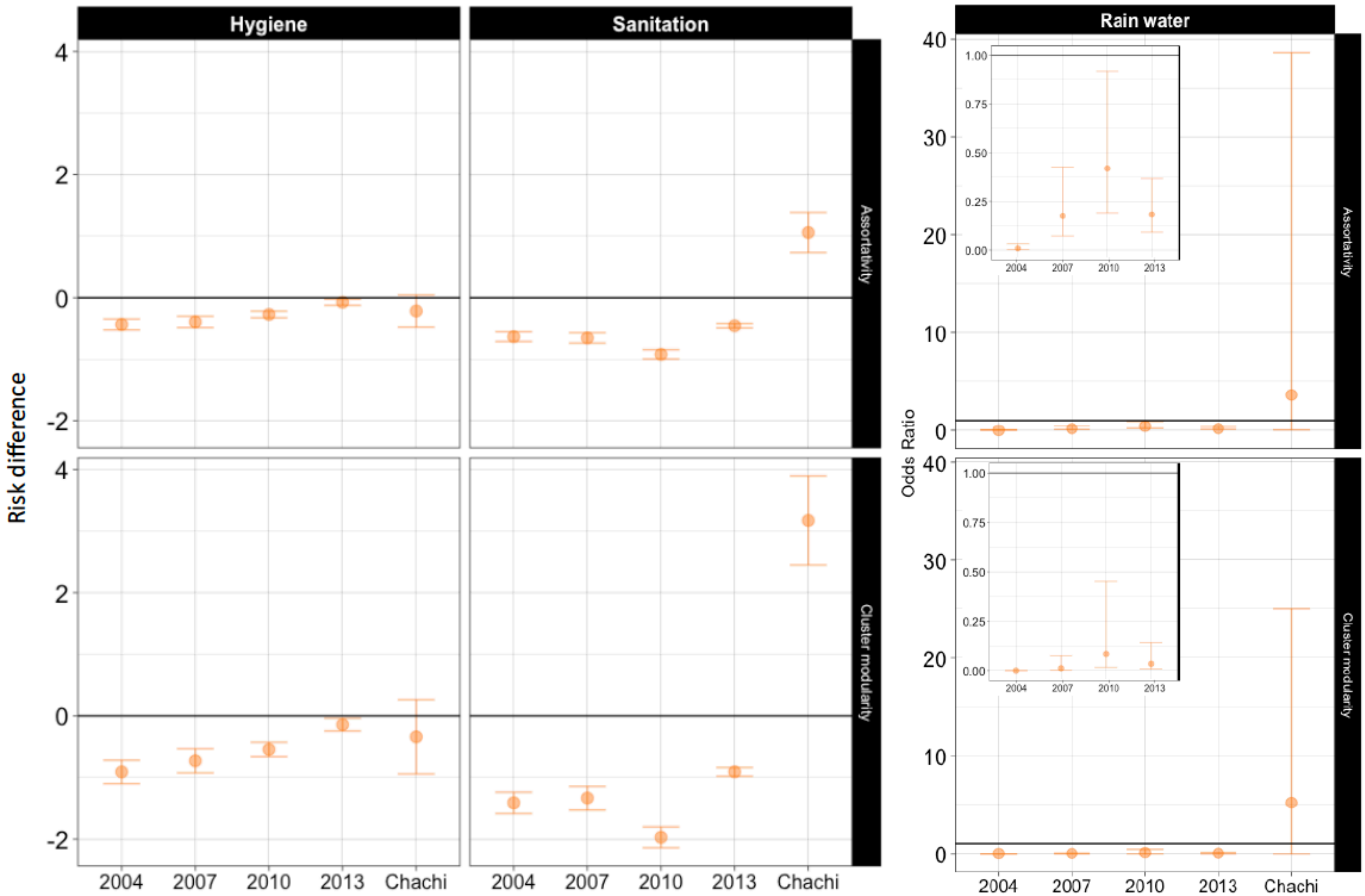
Descriptive data from San Augustin, a roadside community, from 2004 to 2013.

Figure 5.6. Network visual of assortativity and modularity for Chachi communities



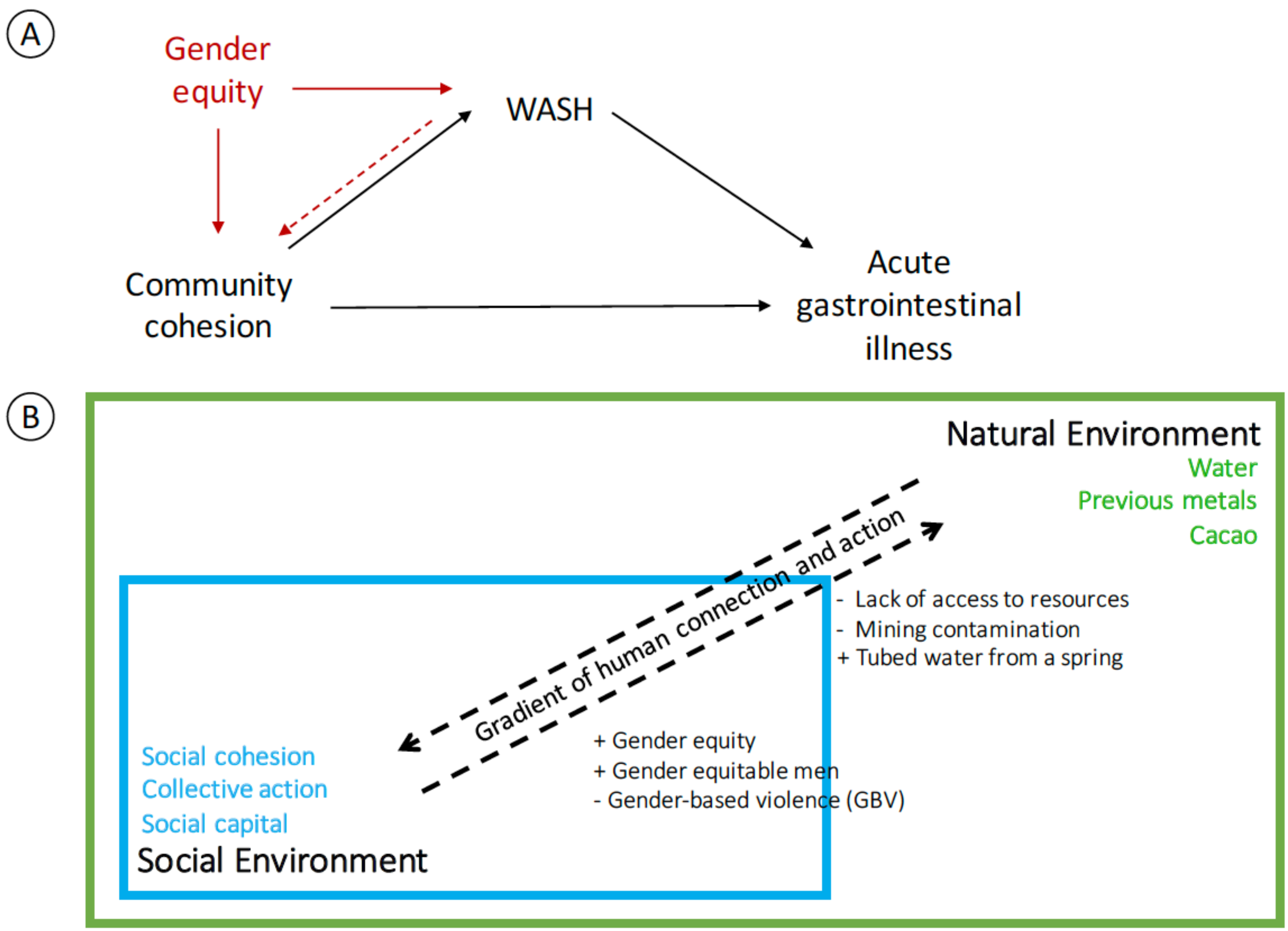
Network plots of the Chachi communities, shown by modularity layers and gender specific nodes and ties.

Figure 5.7. GEE results of effect of network statistics on WASH measures



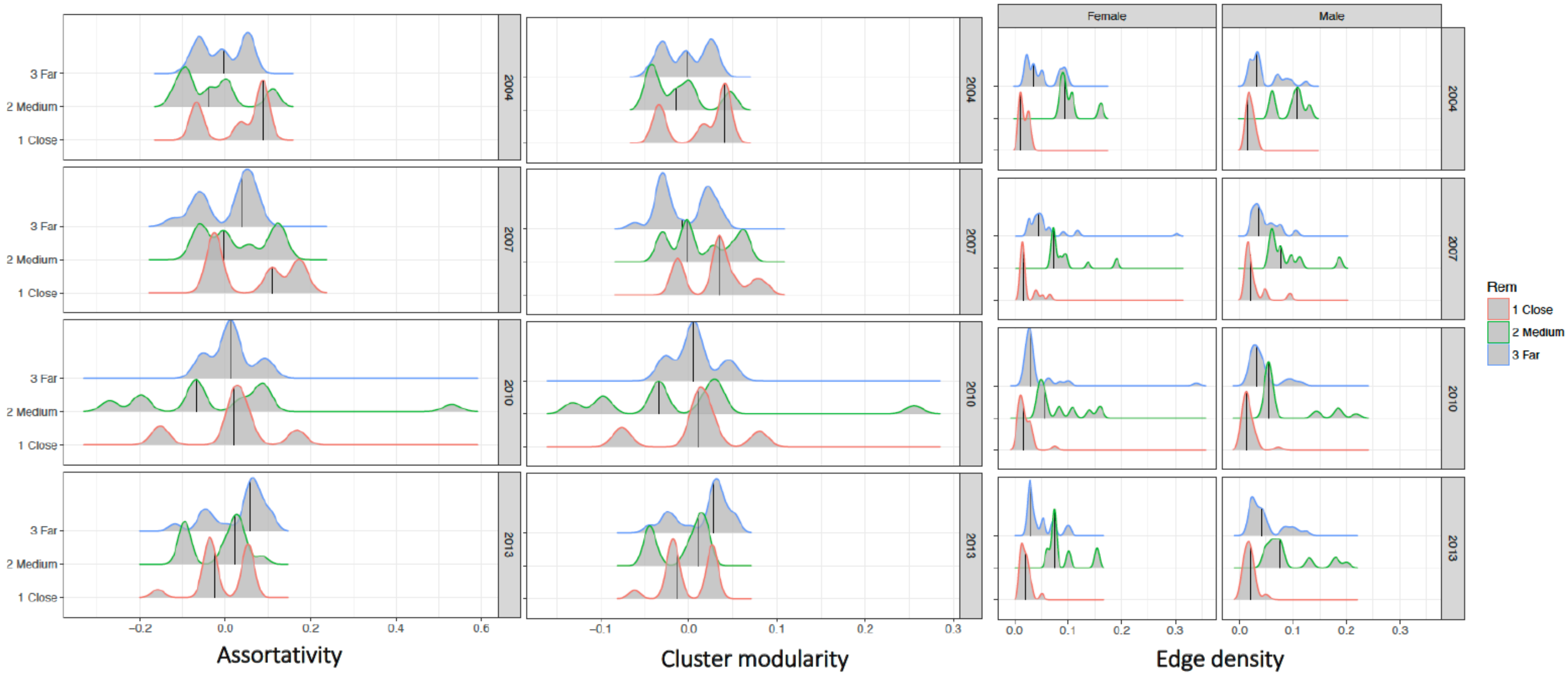
GEE results of the effect of gender assortativity and cluster modularity on hygiene score, community sanitation, and harvesting rain water over time and for the Chachi communities.

Figure 5.8 Conceptual framework of gender equity and WASH



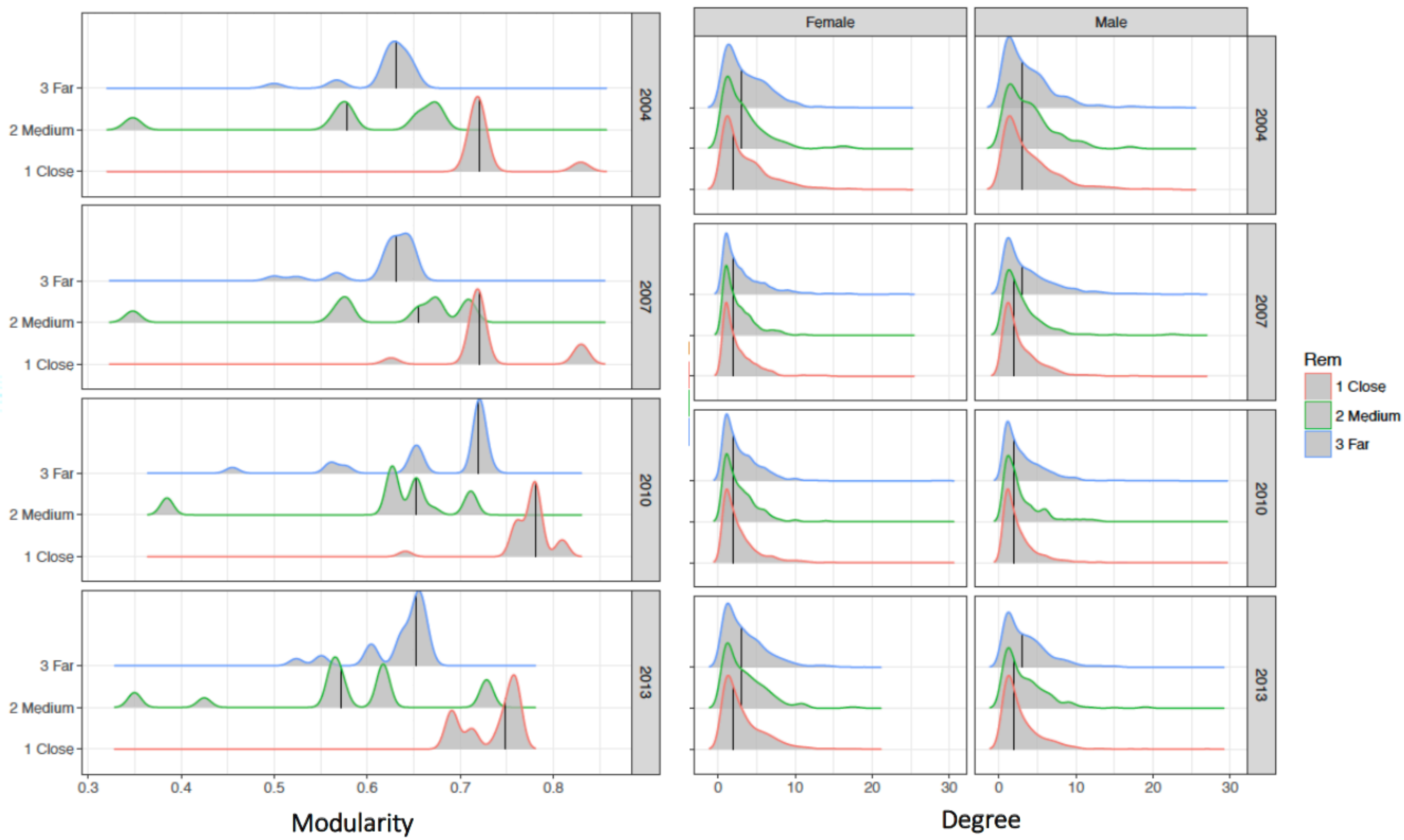
A) A diagram of effects between social cohesion, WASH, illness, and gender equity. B) A diagram of the role of gender equity in the social environment within the context of the natural environment.

Supplementary Figure 5.1. Network statistics for a passing time network



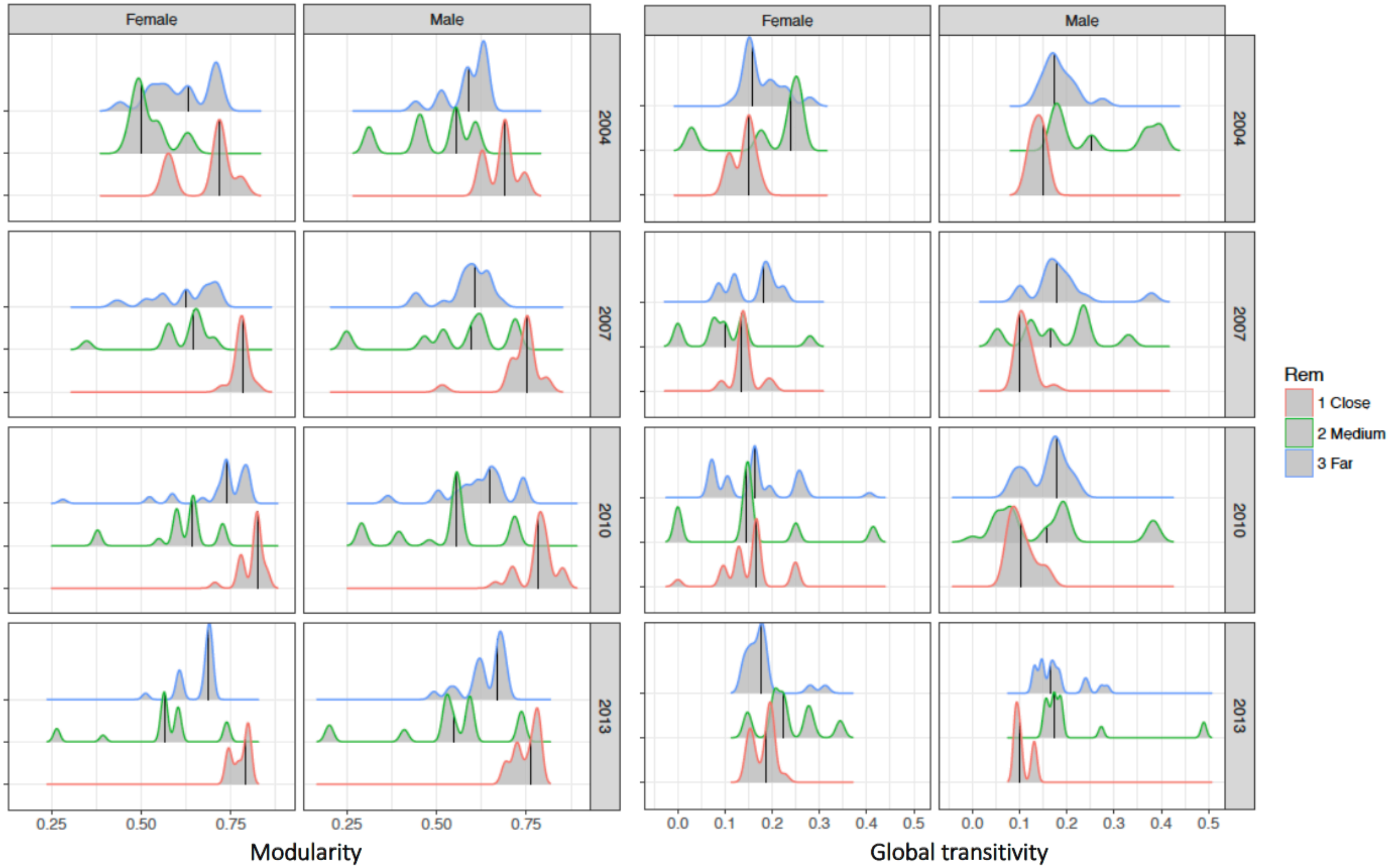
Network statistics over time of assortativity, cluster modularity by gender, and edge density within genders from a network based on ties between individuals that *pass time* together across remoteness categories.

Supplementary Figure 5.2. Other network statistics for a CDN



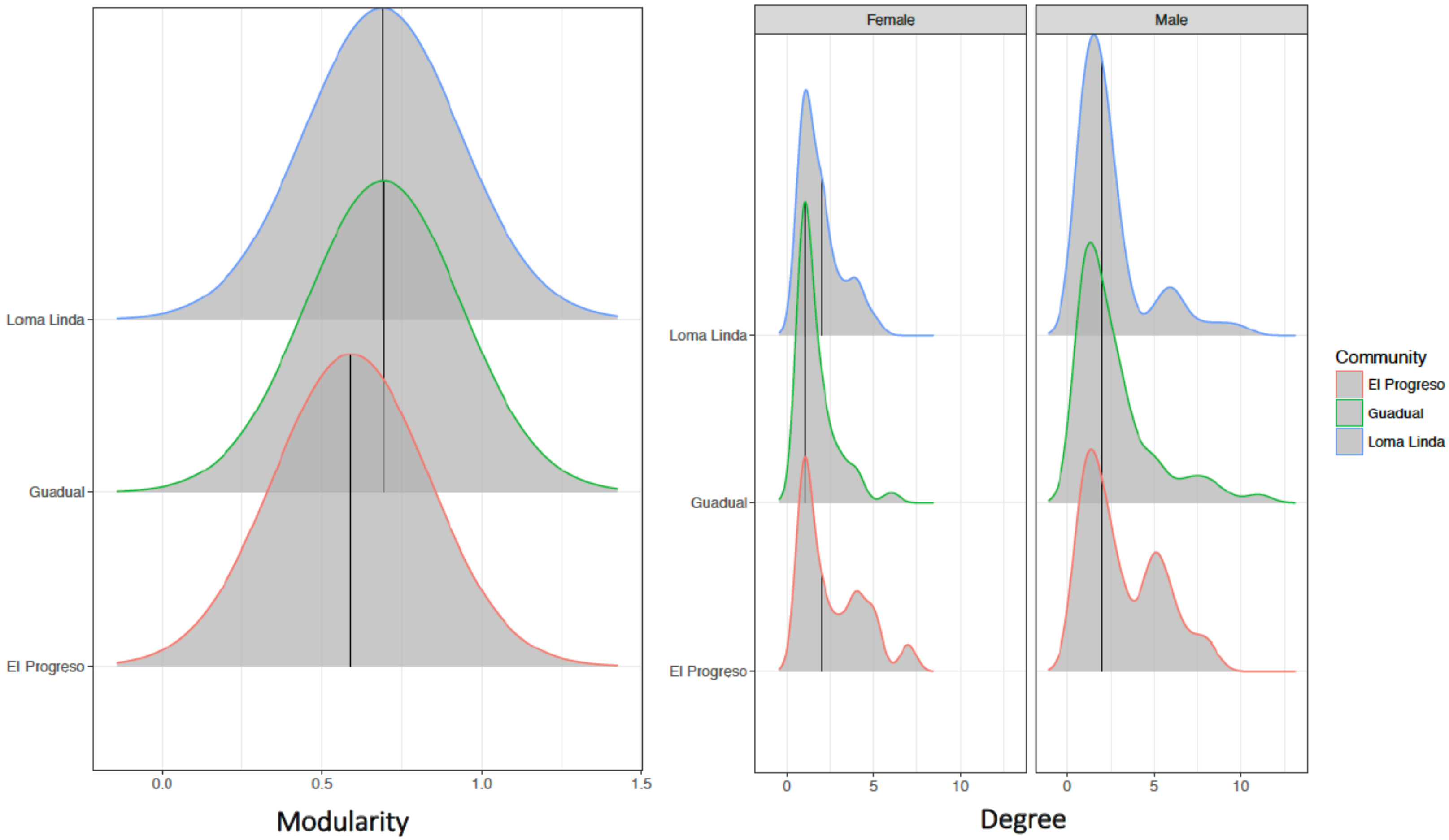
Network statistics over time of overall graph modularity and degree within gender groups from a CDN across remoteness categories.

Supplementary Figure 5.3. Other network statistics between genders for a CDN



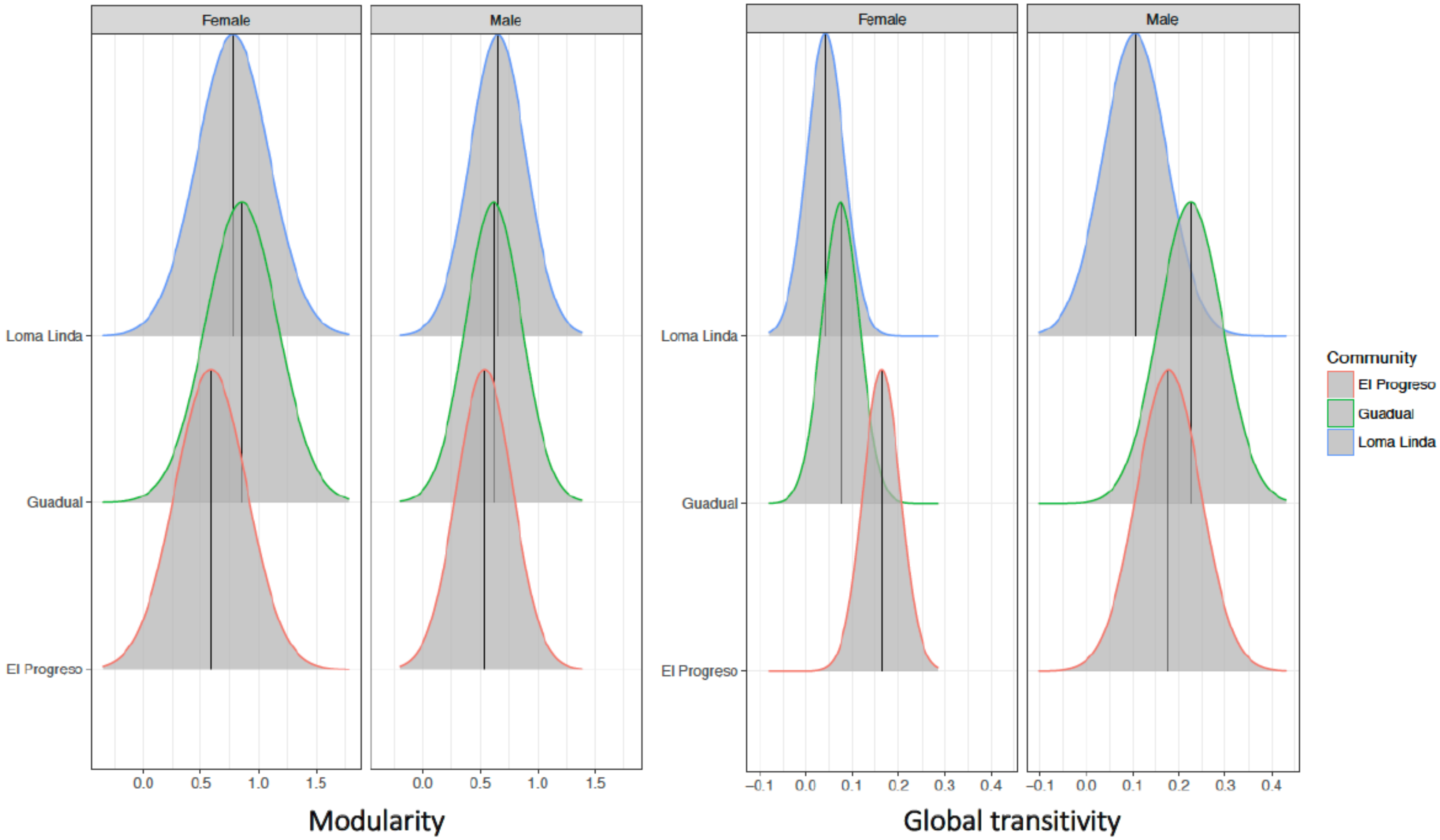
Network statistics over time of modularity within gender groups and global transitivity within gender groups from a CDN across remoteness categories.

Supplementary Figure 5.4. Other network statistics for a CDN in Chachi communities



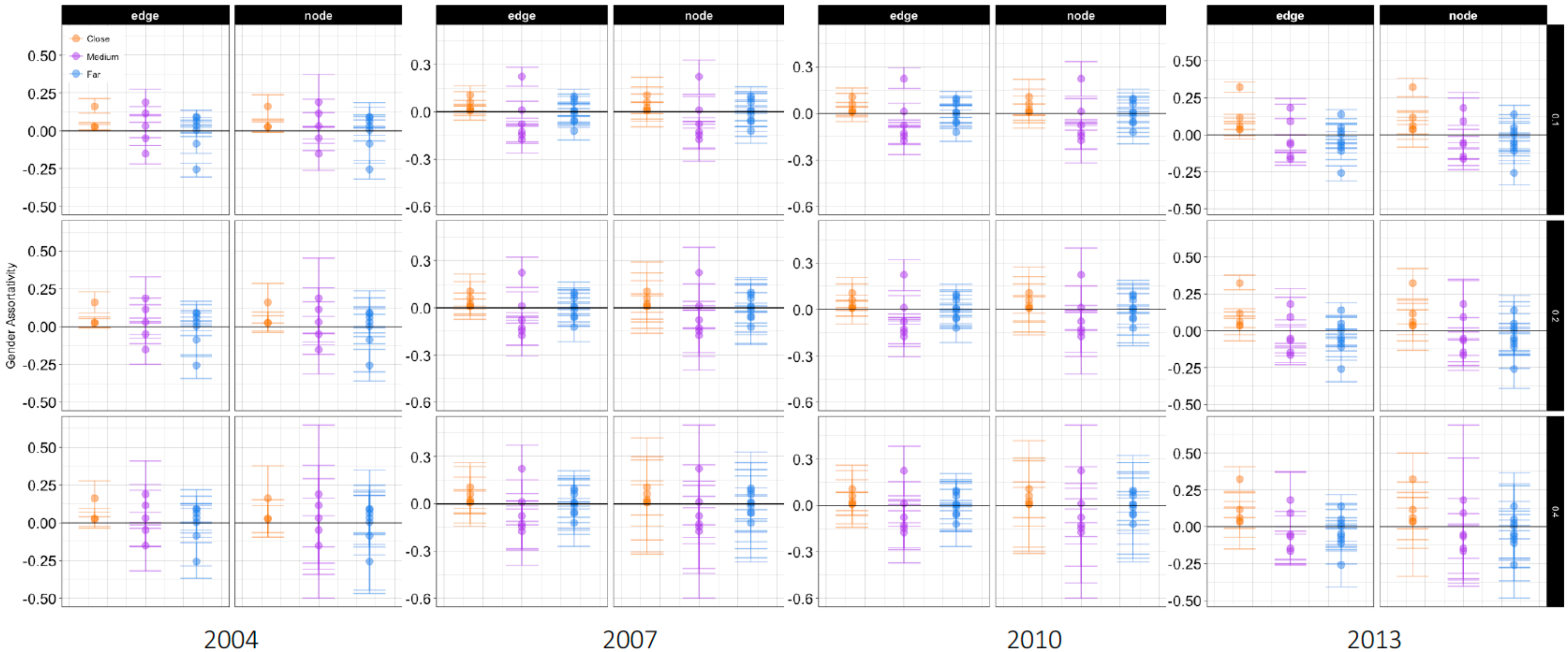
Network statistics over time of overall graph modularity and degree within gender groups from a CDN across remoteness categories for the Chachi communities, the indigenous communities of the Cayapas river basin.

Supplementary Figure 5.5. Other gender network statistics for a CDN for Chachis



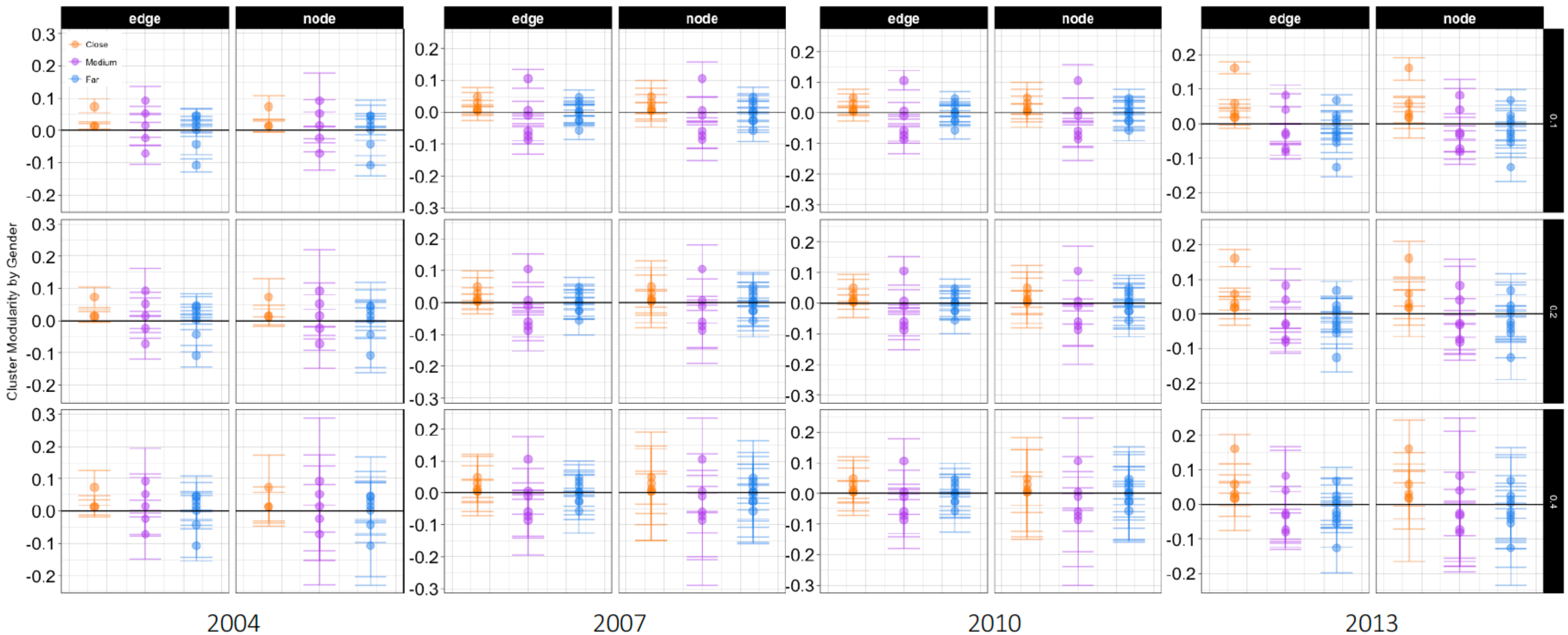
Network statistics over time of modularity within gender groups and global transitivity within gender groups from a CDN across remoteness categories for the Chachi communities, the indigenous communities of the Cayapas river basin.

Supplementary Figure 5.6. Jackknifing results for assortativity uncertainty



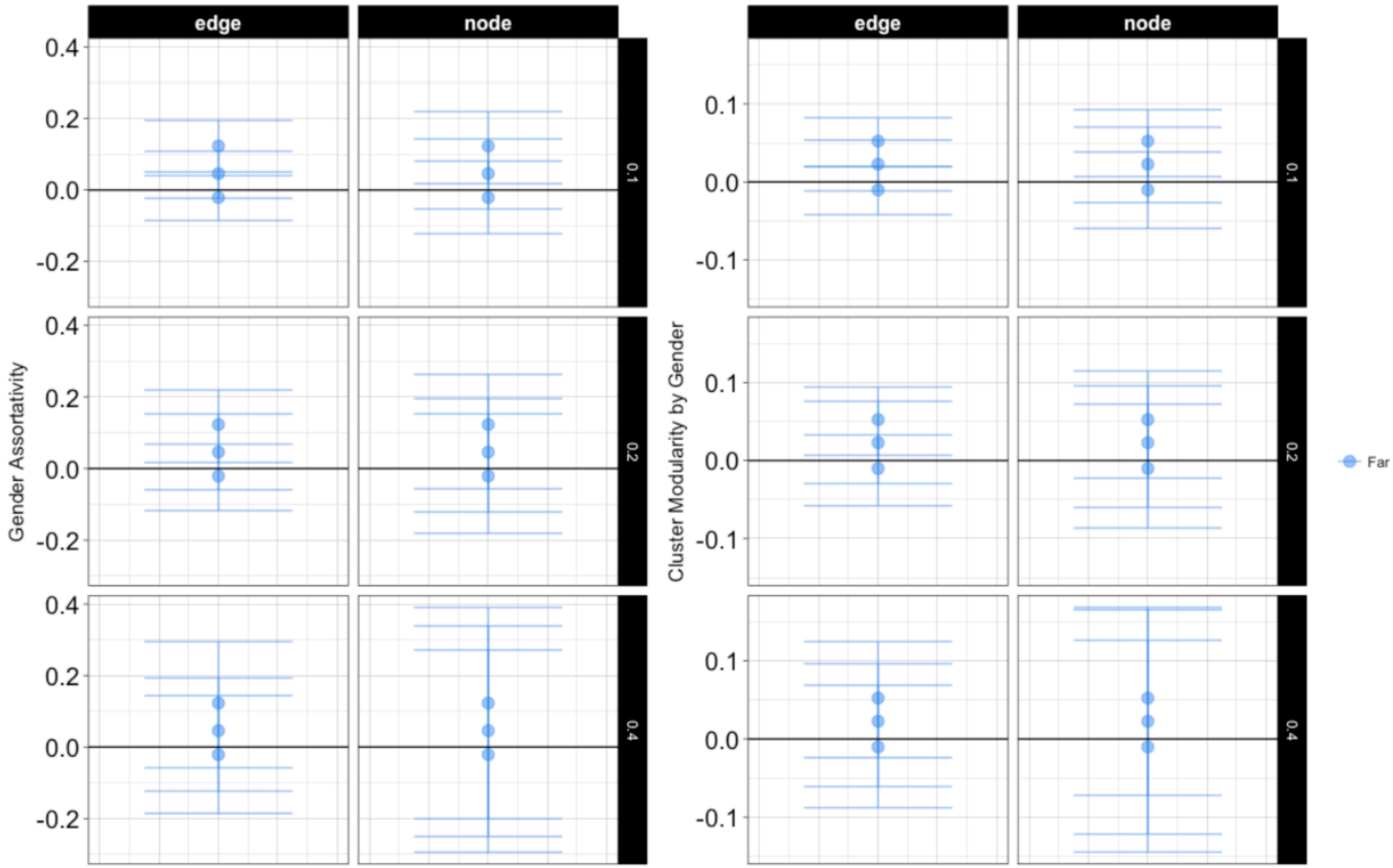
Examining the difference between the removal of different percentages of nodes and edges using the jackknifing method for assessing uncertainty of gender assortativity by remoteness category.

Supplementary Figure 5.7. Jackknifing results for cluster modularity uncertainty



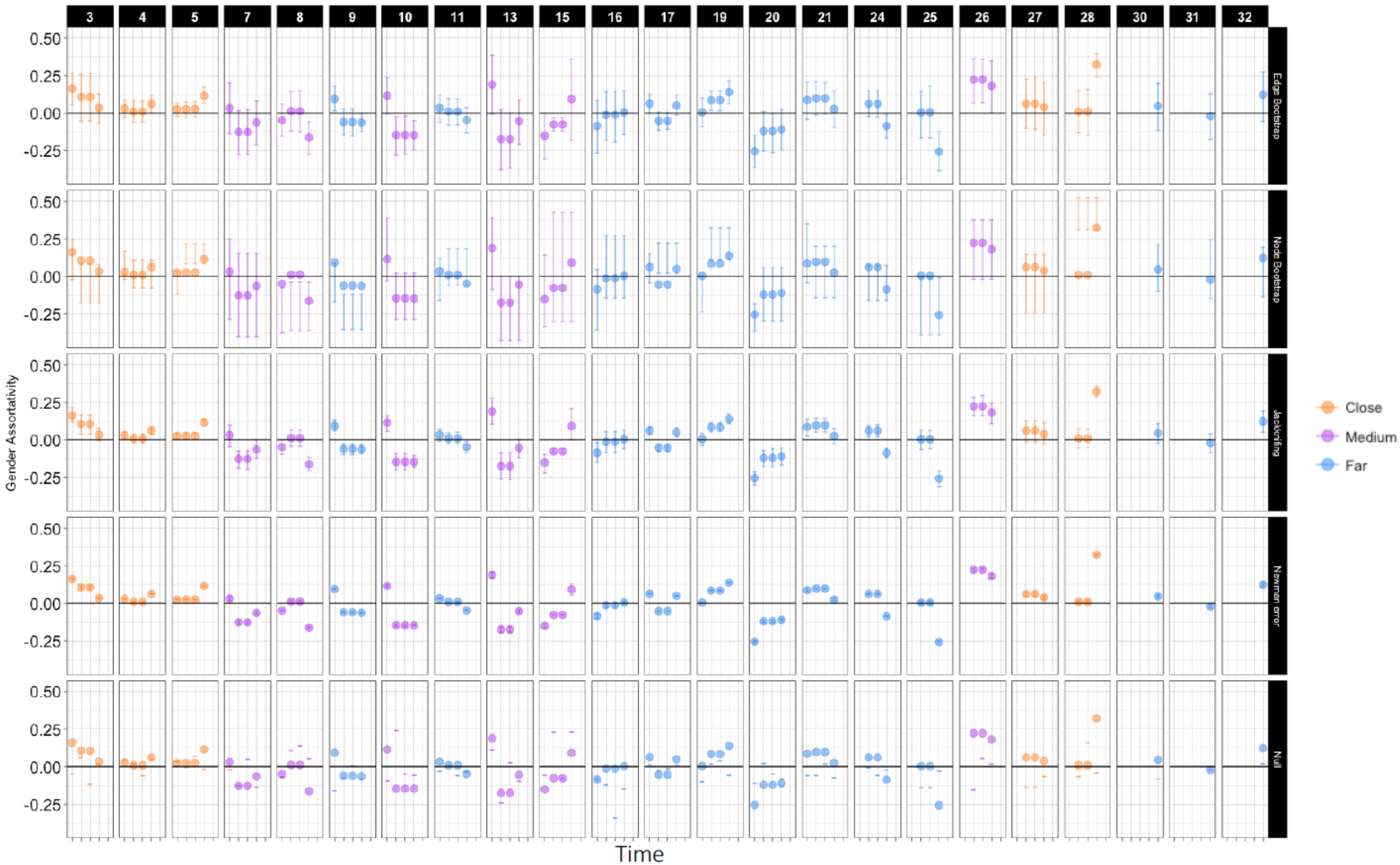
Examining the difference between the removal of different percentages of nodes and edges using the jackknifing method for assessing uncertainty of cluster modularity by gender by remoteness category.

Supplementary Figure 5.8. Jackknifing results for network uncertainty for Chachis



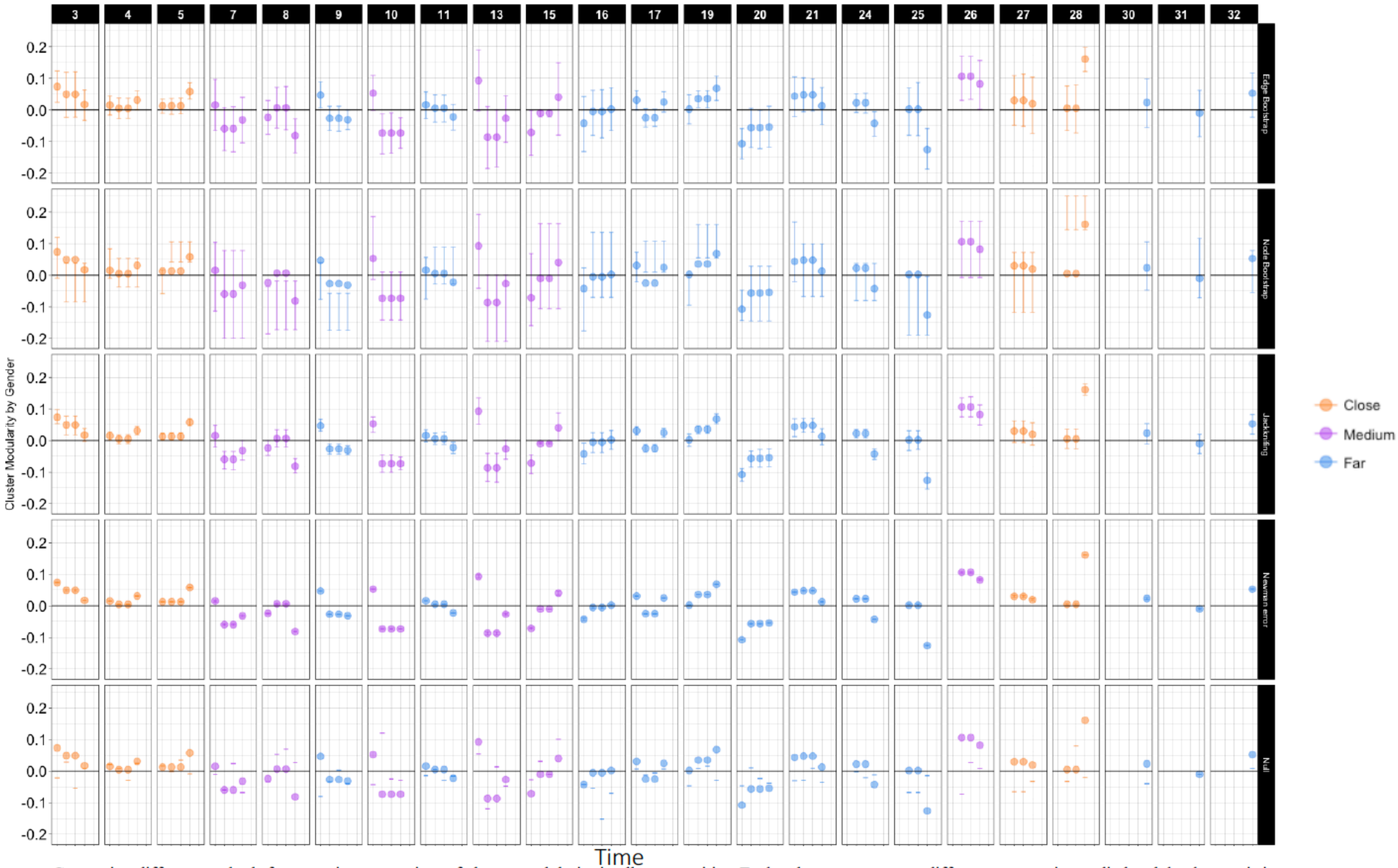
Examining the difference between the removal of different percentages of nodes and edges using the jackknifing method for assessing uncertainty of gender assortativity and cluster modularity in the Chachi communities, where network size is relatively smaller.

Supplementary Figure 5.9. Results of uncertainty assessment for network measures



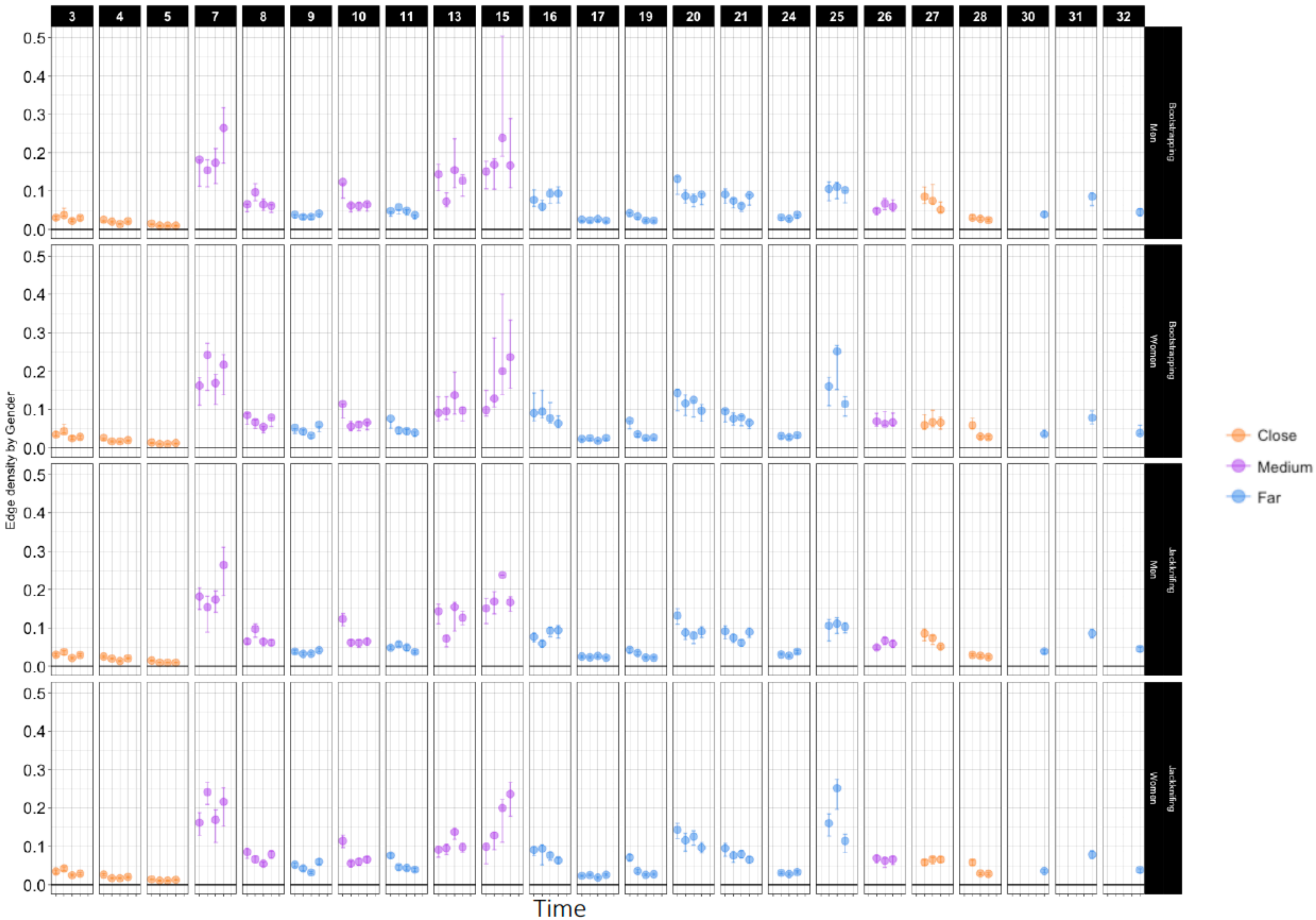
Comparing different methods for assessing uncertainty of gender assortativity in all communities. Each column represents a different community studied and the the x-axis is time where each tick mark within each column represents a year. Thus, here we can compare the network statistics and uncertainty measures within each community across the four time-points. Communities 30-32 are the Chachis.

Supplementary Figure 5.10. Results of uncertainty assessment for network measures



Comparing different methods for assessing uncertainty of cluster modularity in all communities. Each column represents a different community studied and the the x-axis is time where each tick mark within each column represents a year. Thus, here we can compare the network statistics and uncertainty measures within each community across the four time-points. Communities 30-32 are the Chachis.

Supplementary Figure 5.11. Results of uncertainty assessment for network measures



Comparing different methods for assessing uncertainty of edge density for sub-graphs on men and women in all communities. Each column represents a different community studied and the the x-axis is time where each tick mark within each column represents a year. Thus, here we can compare the network statistics and uncertainty measures within each community across the four time-points. Communities 30-32 are the Chachis.

Supplemental Table 5.1. GEE results of effect of network statistics on WASH

		Assortativity			Cluster modularity		
		RR	LL	UL	RR	LL	UL
2004	**Hygiene**						
	Remoteness Close	Ref	Ref	Ref	Ref	Ref	Ref
	Medium	0.02	0.00	0.03	0.02	0.00	0.03
	Far	-0.02	-0.05	0.00	-0.02	-0.05	0.00
	Network Statistic	-0.43	-0.52	-0.35	-0.91	-1.10	-0.72
	Sanitation						
	Remoteness Close	Ref	Ref	Ref	Ref	Ref	Ref
	Medium	0.06	0.05	0.07	0.06	0.05	0.07
	Far	-0.18	-0.19	-0.16	-0.18	-0.20	-0.16
	Network Statistic	-0.63	-0.71	-0.55	-1.41	-1.58	-1.24
2007	**Hygiene**						
	Remoteness Close	Ref	Ref	Ref	Ref	Ref	Ref
	Medium	-0.04	-0.06	-0.02	-0.04	-0.06	-0.02
	Far	0.00	-0.01	0.02	0.00	-0.01	0.02
	Network Statistic	-0.39	-0.48	-0.30	-0.73	-0.93	-0.53
	Sanitation						
	Remoteness Close	Ref	Ref	Ref	Ref	Ref	Ref
	Medium	-0.09	-0.11	-0.07	-0.08	-0.10	-0.06
	Far	0.14	0.14	0.15	0.14	0.14	0.15
	Network Statistic	-0.65	-0.74	-0.57	-1.33	-1.53	-1.14
2010	**Hygiene**						
	Remoteness Close	Ref	Ref	Ref	Ref	Ref	Ref
	Medium	0.01	0.00	0.02	0.01	0.00	0.02
	Far	0.02	0.01	0.03	0.02	0.01	0.03
	Network Statistic	-0.27	-0.33	-0.22	-0.54	-0.66	-0.43
	Sanitation						
	Remoteness Close	Ref	Ref	Ref	Ref	Ref	Ref
	Medium	-0.23	-0.25	-0.21	-0.23	-0.25	-0.21
	Far	0.02	0.02	0.03	0.02	0.01	0.03
	Network Statistic	-0.92	-1.00	-0.84	-1.97	-2.14	-1.80
2013	**Hygiene**						
	Remoteness Close	Ref	Ref	Ref	Ref	Ref	Ref
	Medium	0.00	-0.02	0.02	0.00	-0.02	0.02
	Far	0.02	0.01	0.03	0.02	0.01	0.03
	Network Statistic	-0.07	-0.12	-0.02	-0.14	-0.24	-0.04
	Sanitation						
	Remoteness Close	Ref	Ref	Ref	Ref	Ref	Ref
	Medium	-0.33	-0.35	-0.32	-0.33	-0.35	-0.32
	Far	-0.03	-0.04	-0.03	-0.03	-0.04	-0.03
	Network Statistic	-0.45	-0.49	-0.42	-0.91	-0.98	-0.84
Chachi	**Hygiene**						
	Network Statistic	-0.21	-0.48	0.05	-0.34	-0.94	0.26
	Sanitation						
	Network Statistic	1.06	0.73	1.38	3.17	2.45	3.89

GEE results of the effect of gender assortativity and cluster modularity on hygiene score and community sanitation over time and for the Chachi communities. Risk ratios are reported.

Supplemental Table 5.2. GEE results of effect of network statistics for Chachis

			Assortativity			Cluster modularity		
			OR	LL	UL	OR	LL	UL
2004	**Rain**							
	Remoteness	Close	Ref	Ref	Ref	Ref	Ref	Ref
		Medium	1.36	1.11	1.67	1.37	1.12	1.68
		Far	0.46	0.33	0.64	0.45	0.32	0.63
	Network Statistic		0.01	0.00	0.03	0.00	0.00	0.00
2007	**Rain**							
	Remoteness	Close	Ref	Ref	Ref	Ref	Ref	Ref
		Medium	1.82	1.48	2.23	1.79	1.45	2.19
		Far	2.53	2.16	2.96	2.49	2.12	2.91
	Network Statistic		0.18	0.07	0.43	0.01	0.00	0.08
2010	**Rain**							
	Remoteness	Close	Ref	Ref	Ref	Ref	Ref	Ref
		Medium	3.34	2.81	3.98	3.28	2.76	3.91
		Far	2.78	2.47	3.14	2.75	2.44	3.11
	Network Statistic		0.42	0.19	0.92	0.09	0.02	0.45
2013	**Rain**							
	Remoteness	Close	Ref	Ref	Ref	Ref	Ref	Ref
		Medium	16.73	12.97	21.59	16.67	12.89	21.55
		Far	18.51	14.89	23.03	18.53	14.89	23.05
	Network Statistic		0.18	0.09	0.37	0.03	0.01	0.14
Chachi	**Rain**							
	Network Statistic		3.62	0.03	38.60	5.20	0.00	25.00

GEE results of the effect of gender assortativity and cluster modularity on harvesting rainwater over time and for the Chachi communities. Odds ratios are reported.

Chapter VI

Conclusion

6.1 Review of major findings

This dissertation first demonstrates the utility of a two-stage Bayesian modeling approach to account for both separated and highly correlated data using the study data as a motivating example. The two-stage approach involves fitting a Bayesian hierarchical model to account for correlation using priors derived from parameter estimates from a Firth-corrected logistic regression model to account for separation. When we compared these estimates from the two-stage approach to standard regression methods that only account for either separation or correlation, we found that correctly accounting for separation and correlation when both are present can provide better inference.

Using the two-stage Baysiean hierarchical model, this dissertation secondly demonstrates that having a larger community network of people to discuss important matters with, having trust in one's community, and participating in institutional organization becomes more protective against acute gastrointestinal illness (AGI) over time in rural Ecuador. Effect modification of networks occurs within households as having more individuals to pass time with relative to other community members becomes protective as the household network of individuals to discuss important matters with grows. By 2013, the household networks become a greater risk for AGI and we observe synergistic effects as the people an individual passes time with becomes the people they go to for important matters. Different network types contribute to the multidimensionality of social processes that occur at the household-level and that in turn influences individual health. Having a strong community network of ties to individuals to discuss important matters with is importantly protective against AGI. To validate our findings, we used qualitatively data and show 1) the introduction of a wage economy and changing infrastructure likely contributed to the effect of cohesion changing over time and 2) communities that lacked external influence from government and NGOs felt neglected and therefore felt they lacked social cohesion.

In investigating the mechanism behind the protective effect of the community-level social network variable (i.e community cohesion) on AGI, we thirdly show partial mediation of the protective effect of community cohesion on AGI by community sanitation and improved water use over time. This suggested the importance of social constructs at the community-level for intervention implementation and in turn the reduction of diarrheal disease. Our results also underscored the important role of infrastructure development in changing access to WASH facilities over time; communities that are closest to a road have gained access to WASH infrastructural development while communities that are farther away and have not provided the government or outside agencies with economic incentives (like cacao or gold) have been consistently ignored. In contrast, in the indigenous communities (the Chachis), having greater community cohesion resulted in smaller likelihood of community sanitation and rain water use, though both community sanitation and rain water were protective against AGI. This indicated that the concept of community cohesion is represented differently in these communities and therefore has the opposite effect. Qualitative data suggested it could be due to gender inequities.

Using qualitative data analysis as the primary analytic approach, we lastly showed communities with higher agency amongst women experienced high social cohesion, with men playing a critical role in creating gender equity and enhancing agency among women. Both men and women, however, identified water insecurity as a primary stress and deterrent to social cohesion. The Chachis experienced more severe gender inequity than other ethnicities, including severe gender-based violence and intimate partner violence. Cohesion in the Chachi communities is an indicator of social conflict, resulting in the reduced adoption of WASH practices in the household but continued infrastructure development by men power roles. Additionally, we showed that the network statistic of assortativity is a good proxy for measuring gender equity or gender equitable men on networks. Communities that were more assortative between genders were less likely to engage in WASH practices at the household-level, with this effect decreasing over time.

6.2 Discussion of findings

Prior to this dissertation, mechanisms by which social relationships affect diarrheal disease and WASH measures remained relatively unexplored. As diarrheal disease is a persistent global threat and clean, safe water increasingly hard to access, this dissertation has the potential to inform context-specific community and gender-sensitive interventions in the WASH sector and also underscores the importance of social influences in public health. Furthermore, this dissertation emphasizes the importance of using both quantitative and qualitative methods to answer research questions in public health.

Networks can capture social connectedness and complex individual interactions, and social network data collection is often cumbersome, however, as we've shown, social networks provide strong measures of social constructs; social network measures provide stronger measures of social constructs than both self-reported trust and participation in institutional organization. Measures derived from social network data encompass a plurality of social influences, including core discussants who are most influential and gender equity. Though not shown in this body of work, this dissertation has led to additional analyses investigating the different social properties ascertained from different types of network data using Structural Equation Modeling (SEM). These results not only demonstrated the stability of the important matters or CDN network over time compared to a passing time network, but that ties in a social network represent the social constructs of "influence", "attachment", and "resilience" through different structural network measures. "Resilience", in particular, is why we find the CDN so protective.

In the first few chapters of this dissertation, we use variations of the ego-centric network measure degree and not structural network measures like assortativity or centrality. Given our own prior analyses on the data and our interest to describe social influences in a multi-level framework of consisting of the individual-, household-, and community-levels, we opted to use degree, a standard indication of social ties of an individual, and aggregated this measure at the household- and community- levels. The measure of household degree deviance specifically was derived from prior work in sociology and the idea that the effect of social ties on the individual is dependent on the surrounding environment, in other words, social influence is only relative to the community. In the ancillary SEM study described above, we expand on the difference between the ego-centric and structural network measures and their effect on human health through the example of diarrheal disease, though more research is warranted to investigate the relative contributions of different social network measures to answer different public health questions.

Nonetheless, through the network statistics used, we were able to identify the strong protective effects of a CDN over time and articulate that community leaders play a critical, influential role in these study communities for both mitigating risk of diarrheal disease and increasing safe WASH practices like hygiene and rainwater harvesting, and increasing sanitation infrastructure. In fact, community leaders are likely essential for intervention implementation, through buy-in and dissemination across the community, in order to change behavioral norms and reach sustainability for not only WASH but other public health prevention measures as well. As we learned through the qualitative data and network statistics, gender equity and having gender equitable men is equally important for the success of interventions and community engagement; gender-based violence is a serious detriment for women and WASH. Having women in key roles in the community could greatly influence positive action within households and collective action in communities.

This key piece of qualitative data led to the realization that collecting and using qualitative data allowed for a deeper understanding of the quantitative measures assessed through statistical methods. Indeed, it is doubtful as much information and insight would have been garnered from the statistics computed alone. The mixed methods approach is an optimal tool for better understanding social influences and the culture in communities that exists behind social networks. However, it should be noted that qualitative data, like quantitative data obtained from survey tools, cannot fully capture the human experience, especially when the data collection period is cross-sectional and for short duration. Longer observation periods and longitudinal data collection can provide better information.

Engaging with communities not only raised significant modifying factors that would have otherwise gone unnoticed, like gender equity and water insecurity due to gold mining, but also emphasized the importance of Community-Based Participatory Research (CBPR). All interviewed community members were ready and willing to engage when they realized they were being asked about the problems in their community. They were asked to list issues from their perspective instead of being told by an external investigator what their problems were. As such, we should be working with communities to find solutions worth engaging in and not just share our own interests alone. This will lead to greater relationships with community partners and likely sustainable solutions.

A major advantage in answering such research questions is having the data that allows you to do so. Given the longitudinal structure of the social network data across a multitude of communities, we had the advantage of being able to answer the questions posed and given the remoteness of the communities, we were able to assess social processes in relatively closed populations. We could observe infrastructure development through the building of roads and the growing economy as well as the resulting disparity over time. As such, these results are not only generalizable across this study population and for the study of WASH in other rural, low-resource settings, but also across other emerging middle-income countries in Latin America where there is growing disparity between rural and urban centers and marginalized ethnic groups. Indeed, our finding that the individual-level protection from AGI in the initial years of the study was attenuated over time, highlights the increasingly harmful role of infrastructure development as a result of socio-political and economic influences that lead to increased socio-economic disparity and marginalization in a changing global economy.

Given the interdisciplinary nature of this dissertation work, it additionally has the potential to inform novel research that reaches across different domains of applied science. This dissertation research borrows from social network literature from both physics and sociology, social theory, anthropology, epidemiology, and biostatistics, and suggests the importance of interdisciplinary work for answering complex research questions in public health.

6.3 Future directions

The larger implications of our findings have roots in marketing and economics whereby understanding individual influence and decision-making is critical for addressing behavior change. In our growing world, it's become more important to understand the important role of social relationships and environment to better predict disease risk and prevention.
We are not isolated beings, but rather influenced daily by those around us and the environment around us – others influence how we think, what we say, and ultimately how we act. Additionally, for interventions to be sustainable it's critical to not only engage with community members to have a better understanding of the cultural context, but to conduct CBPR to have a better understanding of what work is possible, what solutions exist for them, and then go to the drawing board to think of innovative solutions. Just as the field of marketing adapts a product for

each global context based on market research, so should we for WASH interventions. CBPR should further remind us to consider that data points are individuals and that our analyses and findings should fit in the context of the people they are about.

Because of our qualitative data, we were able to identify gold mining as another major deterrent to water security in the study region. This is an issue that community members asked us to research. The Afro-Ecuadorians were brought as slaves by the Spanish from West Africa approximately 400 years ago to pan for gold. Since, they have continued the mining as the government provides conditional incentives and outside jobs are limited. As a result, this population has suffered the consequences of heavy metal contamination. Gold mining is done by alluvial mining and mercury amalgamation, and deposits waste in the nearby environment including in the atmosphere and water. However, the primary benefiting parties, the government and external agencies, are neither providing solutions to clean the environment nor providing access to clean water. As such, we have conducted a comprehensive search of the literature in order to write a review and policy brief assessing the methods of mining used in the study area, contamination that exists, and health impacts. We conducted our literature search using the databases: Pubmed, ProQuest (Technology Collection), Embase, One Mine, and Google Scholar for health related articles and a second search on ProQuest and One Mine to identify articles on mining methods and technology. We have also done an exposure assessment assessing the impact of methylated mercury in consumed fish due to mining.

As gold mining pollution contributes to the global water crisis, studies beyond our review should include finding solutions to curb contamination or stop mercury amalgamation altogether. The global water crisis threatens humanity and is a complex problem that needs interdisciplinary attention to meet the challenges of rapid population growth and urbanization, limited water resources, changing environmental capacity, and climate change. Though considerable research has been done on interventions, few have been done on the sustainability of interventions to address these challenges or the nuanced context in which interventions are effective. For example, like researching the importance of community cohesion for intervention adoption. The UN Sustainable Development Goals (SDGs) call for access to clean water and sanitation for all by 2030. Achieving this goal, however, is estimated to cost trillions of dollars. Current levels of funding from developing countries and donors are not sufficient to fill this gap, and at the current rate of growth will not meet the SDGs.

Though microfinancing water and sanitation in communities has been done and demonstrated economic and social value to microfinance institutions with minimal risk profiles, no study currently identifies individual- or community-level social constructs that enable successful microfinancing or the involvement of women's groups as both lenders and borrowers. Additionally, little data exists on how to leverage microfinancing to make WASH interventions effective, defined by compliance, sustainability, health outcomes, and female agency, in addition to cost-efficiency. However, based on our results and the results of other studies on the effect of social influences in WASH, we should change the framing of success for microfinance and cost-effectiveness evaluations for WASH infrastructure to reduction in health outcomes and female agency. This seems imperative for understanding the social market for long term adoption of public health interventions.

Of note, microfinance doesn't necessarily result in female empowerment and agency. Research on this scheme in southern India points out that women's loans often end up invested in assets that are primarily controlled by spouses or are used for household production or consumption, neither of which improve women's capacity to expand their own businesses, or to comply with stringent repayment schedules. Therefore, more research on the social constructs that lead to empowerment is necessary[178], and on flexible micro-loans that target a particular market, like water and sanitation or women's collectives. Many households appear to find it difficult to save, and many households face significant credit constraints. The existence of such short-term credit constraints, therefore, points to the value of more research on the value of flexible micro-payments that can be adapted to a household's individual circumstances, such as daily fluctuations in income[179], which could be administered at the community-level.

Lastly, a key future direction is returning results to community members and connecting with local non-governmental organizations who have been on the ground addressing gender-based violence issues in the Chachi communities and water contamination issues along the river basins. I would like to team up with local academic institutions as well to share the study findings and engage local Ecuadorian students in both the social and quantitative sciences to pursue these futures studies.

6.3 Public health significance

By understanding how social cohesion affects intervention practices and diarrheal disease risk, we can leverage social networks to influence positive behavior change and WASH infrastructure. Target 6 of the United Nations Sustainable Development Goals (SDGs) is dedicated to ensuring availability and sustainable management of water and sanitation for all. Target 6b focuses on leveraging the support and participation of local communities in improving water and sanitation. As such, this dissertation may influence different intervention trials in pursuit of achieving SDG 6, and add to the discussion of sustainability versus compliance at the community-level. Though current intervention campaigns exist that leverage community attributes through community-based participatory research [180], few exist in the WASH sector.

Through the qualitative work, this dissertation also has the potential to impact the local Ecuadorian population by initiating more studies on water quality (chemical versus pathogen contamination) and gender-related work in WASH as SDG 5 aims to achieve gender equality and empower all women and girls. With its more nuanced analysis of the relative contributions of social organization across communities on the risk and prevention of diarrheal disease through the use of social network data, this study could provide important insights for not only the study population but other low-resource settings.

Appendices

Appendix A

Other important figures and tables

Figure A.1. Map of the study region in rural Ecuador

Table A.1. Additional network statistics over time between genders

	2007			2010			2013		
	close	**medium**	**far**	**close**	**medium**	**far**	**close**	**medium**	**far**
Gender (Female)	48%	44%	39%	48%	51%	49%	51%	53%	53%
Mean degree	3.24 (1-19)	3.37(1-23)	4.36(1-25)	3.18 (1-30)	3.34 (1-16)	3.70 (1-29)	3.78 (1-31)	4.53 (1-23)	4.52 (1-28)
Degree gender ratio (women/men)	0.99	1.08	1.03	1.08	1.00	0.96	0.99	1.01	0.98
Assortativity by degree (median)	0.071 (-.06,.09)	0.065 (-.23,.30)	-0.002 (-.13,.14)	0.033 (-.09,.08)	0.016 (-.04,.31)	0.007 (-.21,.09)	0.068 (.07,.20)	-0.021 (-.33,.015)	0.036 (-.12,.17)
Assortativity by sex (median)	0.003 (-.14,.07)	0.02 (-.09,.14)	0.02 (-.21,.19)	0.033 (-.04,.11)	-0.001 (-.11,.15)	0.019 (-.04,.13)	0.098 (-.02,.36)	-0.042 (-.14,.11)	-0.001 (-.17,.08)
Betweenness gender ratio (women/men)	0.92	1.33	1.08	1.04	0.97	0.93	1.12	0.85	0.89
Closeness gender ratio (women/men)	0.83	0.68	1.65	0.94	1.08	0.78	1.09	1.04	1.10
Average path length gender ratio (women/men)	0.98	1.04	0.99	1.00	1.00	0.98	0.98	0.97	0.99
Triads gender ratio (women/men)	1.00	1.08	1.07	1.14	1.54	0.81	0.92	0.94	0.90
Eigenvector centrality gender ratio (women/men)	0.86	1.13	1.17	0.95	0.76	0.94	1.04	1.12	0.96
Burt's holes gender ratio (women/men)	0.97	0.93	1.00	0.94	0.98	1.00	1.02	0.93	0.98
Individual deviance from community mean (std women)	-0.19 (-.89,4.14)	-0.26 (-1.0,4.54)	-0.19 (-1.1,4.5)	-0.14 (-.91,5.1)	-0.27 (-1.0,3.62)	-0.18 (-1.1,4.13)	-0.27 (-.96,6.7)	-0.26 (-1.2,4.3)	-0.18 (-1.2,4.6)
Individual deviance from community mean (std men)	-0.19 (-.89,5.14)	-0.30 (-1.0,4.12)	-0.33 (-1.1,4.6)	-0.33 (-.91,9.5)	-0.25 (-1.0,4.16)	-0.22 (-1.1,7.29)	-0.27 (-.96,4.9)	-0.22 (-1.2,4.0)	-0.24 (-1.2,6.6)

Additional network statistics on the social network data from 2007-2013, highlighting differences between men and women.

Table A.2. Additional network statistics between Afro-Ecuadorians and Chachis

	Far communities 2010	
	Afro-Ecuadorian	**Chachi**
Gender (Female)	49%	40%
Mean degree	3.70 (1-29)	2.82 (1-11)
Degree gender ratio (women/men)	0.96	0.99
Assortativity by degree	0.007 (-.21,.09)	0.061 (.033,.112)
Assortativity by sex	0.019 (-.04,.13)	-0.051 (-.23,.07)
Betweenness gender ratio (women/men)	0.93	1.20
Closeness gender ratio (women/men)	0.78	1.02
Average path length gender ratio (women/men)	0.98	1.02
Triads gender ratio (women/men)	0.81	0.90
Eigenvector centrality gender ratio (women/men)	0.94	0.89
Burt's holes gender ratio (women/men)	1.00	0.96
Individual deviance from community mean (women)	-0.18 (-1.1,4.13)	-0.31 (-1.13,3.52)
Individual deviance from community mean (men)	-0.22 (-1.1,7.29)	-0.33 (-1.13,4.00)

Additional network statistics comparing the Afro-Ecuadorian and Chachi communities in 2010.

Appendix B

Qualitative data collection material

B.1. Data collection discussion guide for focus groups and in-depth interviews.

Discussion Guide

INTRO

1. Do social connections define a community or are there government borders that define your village?

   Probe: is this village divided into two or three communities? why are these divided?

2. What does a person need to do in order to be a member of this community? Is there legal identification to buy or sell land or participate in community decisions?

3. How important is having social connections and interaction to your community?

   Probe: does it help change happen?

4. How would you describe the social cooperation and willingness to help each other in this community?

   Probe: why?

ORGANIZATIONS

5. What organizations or groups exist in this community?

   Probe: communal of agriculture, committees, mingas, about water, led by women, religious, which development agencies exist, sports, crafts

6. What are the different types of projects the different organizations engage in?

   Probe: development projects, agriculture, projects that deal with community problems

7. How would you rank the importance of these communities starting with 1 as the most important organization?

8. What kind of funding do these organizations provide or receive?

   Probe: development agency funding, government, is there a registry

9. What are the reasons why older organizations have lasted so long in this community and others don't last?

Probe: ask for examples, what are the oldest organizations in this community? what are organizations that have ended and why did they end?

10. Why do people form organizations?

    Probe: mingas, for the betterment of the community, because have similar interests or traits

11. How do organizations form?
12. What are the most important problems that your community faces?

    Probe: why is it a problem?

13. What are the actions of groups, organizations, or people that have been formed to deal with these problems?

    Probe: events, organizations, committees

14. What is needed in this community to deal with these problems?
15. What are reasons why projects have been successful in dealing with problems in this community?

    Probe: ask for examples and give examples of what was said, what are these projects, why are some successful and others not

WATER

16. How is water a part of daily activities or festivals in this community?

    Probe: potable water, washing you face, bathing, games, going to the bathroom

17. Have there been issues with having access to water in this community?

    Probe: clean water issues, river water, rain water, water pump, wells

18. Whose job is it to ensure there is always access to water?

    Probe: government, the community, the people

19. How is water collected in households?

    Probe: women, girls, potable water, rain water

20. Do people share agriculture, water or sanitation in this community? They share in the same house and with other houses?

    Probe: Is there a household or community based system for water, taking care of sewage, land rights?

21. Is there an existing water system in this community to deal with water distribution, sewage system, or storm water from the rains? Describe it.

    Probe: why or why not, who put this system there (government, community, development group)

22. How have your water and sewage systems changed in the past 10 years?

    Probe: do you like what you have now

23. If someone in the community developed an idea to build a community wide well or rain water collection system, what would happen next? How would this idea spread?

    Probe: how do you strategize, ask a development agency or government?, how does one person influence others in a community?

24. Are there any strategies now to promote good water practices like hand washing, rain water collection, water treatment or pit latrines in this community?

    Probe: how did these start? how long did it last? Or why not?

25. Are people influenced by watching the behavior of their neighbors or influenced by watching the behavior of their friends?

    Probe: if your neighbor or friend has rain water collection, would you use rain water collection

26. How does people's relationship with water change across the seasons?

    Probe: During the rains, flooding, dry season

27. How do organizations change across seasons?

KINSHIP & MOVEMENT

28. What is the difference in the relationship between persons who are socially connected in the community, who live in the same house, and who are neighbors?

Probe: family or friends, what do you with people you spend time with or go to for important matters, what do you do with people that live in your house, what do you do with your neighbors

29. Are there familial relations in this community? How are people in this community related?

30. What is your relationship with the neighboring communities?

    Probe: School, work, travel, share organizations or committees

31. Are there close or influential relationships with people in other communities?

32. Are there major community events that most of the community participates in?

    Probe: Festivals, seasonal events, gather for food, what are these events, who is coming and who is going

33. How often do people leave this community permanently or come to join this community?

34. Do some people have multiple households in this community or also have households in other communities?

    Probe: why, do people have multiple wives or multiple husbands?

GENDER

35. Do women have a strong presence in community organizations?

    Probe: participation, are they leaders, do they form organizations

36. How do women influence community organization and change in this community?

    Probe: Do they make sure projects or organizations are sustained?, what are women good at doing

37. Are there differences in how women deal with issues in the community versus men?

    Probe: Personality, because women have certain tasks in the household

38. What jobs do the women have in the community and in the household?

CLOSING

39. Do you have trust in your community?

Probe: to make change and deal with problems, to help each other

40. What makes your community different than other communities?

Probe: pride in your community, trust, social support, organization, social interaction, ability to face challenges

41. Where does the empowerment come from for change to happen in this community?

Probe: strength, social support

42. Where does social support come from for people of this community?

Probe: emotional, economic (supporting local economy), organizations, friends, neighbors

43. How successful is your community in dealing with the problems it faces?

44. How did this community form?

45. Lastly, do you have any songs that you sing as a community? Can you sing them?

B.2. Key informant interview guide

Nombre de informante clave ______________________
Comunidad ___________________ Fecha de nacimiento ______________________
Fecha ___________________ Edad ______________________
Investigador ___________________

1. ¿Cómo se formaron esta comunidad?

2. ¿Existe alguna comisión u otra forma de organización en la comunidad?
   Si o No

3. Cuales son los organizaciones o grupos que esta comunidad tiene ?
   ____
   ____
   ____
   ____

4. ¿Actualmente existe alguna comisión u otra forma de organización de la comunidad para resolver los problemas de los residentes?
   Si o No

5. ¿En el último año, la comunidad jamás se organizado para resolver un problema que afectaba a todos?
   Si o No

6. En el último año, ¿con qué frecuencia se reunen los habitantes del pueblo con el fin de discutir los problemas comunes?
   - Cada cuantos dias
   - Una vez a la semana
   - Cada dos semanas
   - Una vez al mes
   - Una vez cada dos meses
   - Unas pocas veces al año

7. En comparación con otros pueblos de esta zona, cual es el éxito que tienen los vecinos de este pueblo en la búsqueda de soluciones a los problemas que enfrenta la comunidad?
   - Menos éxito
   - Sobre lo mismo
   - Más exitoso

   5a) Por qué piensas eso…

8. En su opinión, ¿cuáles son los tres problemas principales que los habitantes de esta comunidad le gustaria resolver?

   Problema 1: ______________________________

   Problema 2: ______________________________

   Problema 3: ______________________________

9. ¿Cuáles son las acciones principales que los habitantes de la comunidad han logrado en los últimos 3 años?

   Acción 1: ______________________________

   Acción 2: ______________________________

10. En los últimos 2 años la comunidad ha hecho algun plan para resolver problemas?

11. ¿Ha habido planes en esta comunidad para mejorar la salud de las personas y la interacción con el agua?

    (Como la construcción de un centro de tratamiento de agua, promover el lavado de manos, promover el uso de pozos o la recolección de agua de lluvia, mejora el saneamiento, soluciones para los problemas de inundación)

    Por favor describa cada plan y cuando fue.

12. Si la comunidad no ha hecho ningún plan, por qué no?

13. ¿Cuál es la capacidad de organización de los habitantes de esta comunidad?

14. ¿Cuál es su percepción de las intervenciones y el movimiento para el cambio entre las personas de la comunidad?

15. ¿Cuál es su confianza en la comunidad?

16. ¿Cuál es la relación de esta comunidad con las comunidades vecinas? ¿La gente de esta comunidad se comunican con otros en otras comunidades o frecuencia de viajes o el comercio?

17. ¿Existe un registro de la organización en esta comunidad o con las comunidades vecinas? Si o No

18. ¿Existen agencias de desarrollo que operan en la zona? ¿De ser asi, cuales? Si o No
____
____
____
____

19. ¿Quienes son los lideres claves de la comunidad? (Nombres y lugares)

20. ¿Si hay lideres mujeres? (Nombres y lugares)

    Si o No

21. ¿El pueblo esta dividido en dos o tres comunidades?

22. ¿Dónde está el límite geográfico de este pueblo y los comunidades diferentes? ¿Cómo se define el límite? Es por lazos sociales o líneas de gobierno?

    Puedes dibujar para nosotros en un mapa? (Participa en el mapeo social, comienza dibujando edificios o casas bien conocidos en la comunidad para definir las fronteras del pueblo. Conserve este documento.)

    ¿Nos puede mostrar los límites del pueblo? (Usa el GPS. Usted no tiene que tener un punto de GPS a cada paso, pero debería ser suficiente para mapear la comunidad. Recuerde que no es necesario tener un número excesivo de puntos de GPS. El objetivo es entender la forma de la pueblo, incluyendo los bordes dentados, como se define por los líderes del pueblo. Esto puede requerir a veces muchos puntos y otras veces menos. Si una parte de los límites del pueblo es una línea recta, sólo es necesario tener un punto.)

La mapa:

B.3. Qualitative data collection protocol.

Qualitative Data Collection

The survey tool covers 4 main themes: The discussion guide currently includes approximately 40 questions in total across the four different themes, but as is done in qualitative data collection, the specific questions under each theme were subject to change. Qualitative data collection and discussion guides go through an iterative process in order to fully understand a topic and reach saturation in data collection.

Recruitment was conducted within each of the 18 communities in EcoDess; we spent approximately 3 consecutive days in each community to conduct both recruitment and data collection. There were 3 different types of study participants: a key informant, focus group participants, and in-depth interview participants. Four focus groups were conducted in each community with community leaders, women only, men only, and youth aged 13-18 years. Each focus group consisted of 6-8 persons. The in-depth interviews were conducted with adult men and adult women, approximately 5 interviews per gender.

The key informant was either the community health promoter, who previously worked with EcoDess, or a community leader suggested by the health promoter. Such a recruitment strategy can be referred to as word-of-mouth. The key informant responde in real-time, in-person whether he or she wanted to participate in the interview in the subsequent 3 days the study team was in the community. To identify community leaders, we asked the key informant to name 6-8 persons who were leaders of organizations, influential, etc. We then visited these persons and recruited them to participate in the focus group with community leaders. To then recruit men and women for both focus groups and in-depth interviews, we either used word of mouth or went around to households and asked. The field assistants began on opposite ends of each community and visited households to recruit participants. This followed until a representative community sample of men and women was reached.

Qualitative data collection commonly uses purposive sampling. As such, youth participants were recruited intentionally from the same households as adults in order to see if youth opinions on discussion topics differed from adults. Six to eight youth, aged 13-18 years old, were recruited to participate. Recruitment was done in real-time, in-person during the 3 days stay in each community. At the time of recruitment, screening was done to ensure the persons being recruited were knowledgeable about social organization in the community.

The consent process took place after recruitment. Field assistants reviewed the consent document with each participant and obtained a written signature. The consent form included a description of the purpose of the research, why the individual was being invited to participate, what was expected of the participant (including that their voice will be recorded during interviews), risks and benefits, privacy, anonymity, and confidentiality, future use of the information, right not to participate and withdraw, principle of compensation, and contact information of the investigators should the participant have questions later. Consent for youth was sought from parents during the time of recruitment. If the person being asked to consent could not write, we sought oral consent and asked the field assistant plus a witness to sign the consent form.

The interviews and focus groups were conducted after recruitment and the consent process. The key informant interview was conducted first in each community in either the key informant's household or place of work. The survey tool for the key informant is slightly different, shorter, and referred to as the Key Informant Discussion Guide [Appendix 2]. It includes a shorter set of questions on the same 4 themes: social organization, water, kinship, and gender. Towards the end of the key informant interview, the key informant was asked to describe the social boundaries of his/her community, at which time the field assistant drew a social map of the community with the assistance of the key informant. The key informant then was asked to accompany the field assistant for a walk around the community to collect GPS points of the village boundary. The entire key informant interview was recorded on a voice recorder.

Both the in-depth interviews and focus groups used a lengthier discussion guide. After the consent was administered, in-depth interviews were conducted in a quiet location, often within the household of the participant and not in the presence of other family members. The field assistant first shared an opening statement regarding the protocol/method of an in-depth interview and specifically what it would be about. The field assistant then began and used the discussion guide to help facilitate the conversation but did not necessarily stick strictly to the questions on the guide or the order to get at the information ultimately needed to answer our research question. Similarly, each focus group was conducted after the consent of each individual was received. We started by reading an opening statement to welcome participants and reviewed the protocol/method of a focus group and what it would be about. The field assistant then began the focus group and used the same discussion guide as the in-depth interview to help facilitate the discussion. Both the in-depth interviews and focus groups were voice recorded from the time the first question was asked to the time the last response was given. After the voice recorder was stopped for both the in-depth interviews and focus groups, a closing statement reiterating the confidentiality of the interview and sharing our gratitude was read to the participants. Female field staff were hired, who speak Chapalachi, the language of the Chachi people, were hired in each of the three Chachi communities.

Bibliography

1. Aiello AE. Invited commentary: Evolution of social networks, health, and the role of epidemiology. *Am J Epidemiol*. 2017;185(11). doi:10.1093/aje/kwx076.
2. Holt-Lunstad J, Robles TF, Sbarra DA. Advancing social connection as a public health priority in the United States. *Am Psychol*. 2017;72(6):517-530. doi:10.1037/amp0000103.
3. Valente TW, Pitts SR. An Appraisal of Social Network Theory and Analysis as Applied to Public Health: Challenges and Opportunities. *Annu Rev Public Health*. 2017;38(1):103-118. doi:10.1146/annurev-publhealth-031816-044528.
4. BERKMAN LF, SYME SL. SOCIAL NETWORKS, HOST RESISTANCE, AND MORTALITY: A NINE-YEAR FOLLOW-UP STUDY OF ALAMEDA COUNTY RESIDENTS. *Am J Epidemiol*. 1979;109(2):186-204. doi:10.1093/oxfordjournals.aje.a112674.
5. Walker CL, Rudan I, Liu L, et al. Global burden of childhood pneumonia and diarrhoea. *Lancet*. 2013;381(9875):1405-1416.
6. Liu L, Johnson HL, Cousens S, et al. Global, regional, and national causes of child mortality: an updated systematic analysis for 2010 with time trends since 2000. *Lancet*. 2012;379(9832):2151-2161. doi:10.1016/S0140-6736(12)60560-1.
7. Bay G, Miller T, Faijer DJ. *Level & Trends in Childhood Mortality: Estimates Developed by the UN Inter-Agency Group for Child Mortality Estimation*.; 2012.
8. Eisenberg JNS, Trostle J, Sorensen RJD, Shields KF. Toward a systems approach to enteric pathogen transmission: from individual independence to community interdependence. *Annu Rev Public Health*. 2012;33:239-257. doi:10.1146/annurev-publhealth-031811-124530.
9. Szreter S, Woolcock M. Health by association? Social capital, social theory, and the political economy of public health. *Int J Epidemiol*. 2004;33(4):650-667. doi:10.1093/ije/dyh013.
10. Dreibelbis R, Winch PJ, Leontsini E, et al. The Integrated Behavioural Model for Water, Sanitation, and Hygiene: a systematic review of behavioural models and a framework for designing and evaluating behaviour change interventions in infrastructure-restricted settings. *BMC Public Health*. 2013;13(1):1015. doi:10.1186/1471-2458-13-1015.
11. Cameron L, Olivia S, Shah M. Initial conditions matter: social capital and participatory development. 2015;(9563):34 pp. doi:10.2139/ssrn.2704614.
12. Eisenberg JNS, Scott JC, Porco T. Integrating Disease Control Strategies: Balancing Water Sanitation and Hygiene Interventions to Reduce Diarrheal Disease Burden. *Am J Public Heal*. 2007;97(5):846-852.

doi:10.2105/ajph.2006.086207.
13. Zelner 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. *Am J Public Health*. 2012;102(12):2233-2239. doi:10.2105/AJPH.2012.300795.
14. Bahl MR. Impact of piped water supply on the incidence of typhoid fever and diarrhoeal diseases in Lusaka. *Med J Zambia*. 10(4):98-99. http://www.ncbi.nlm.nih.gov/pubmed/1023699. Accessed February 18, 2017.
15. Ryder R, Reeves W, Singh N. The childhood health effects of an improved water supply system on a remote Panamanian island. *Am J Trop Med Hyg*. 1985;(34):921–24.
16. Zeng-sui W, Shepard DS, Yun-cheng Z, et al. Reduction of enteric infections disease in rural China by providing deep-well tap water. *Bull World Heal Organ*. 1989;67:171-180.
17. Daniels DL, Cousens SN, Makoae LN, Feachem RG. A case-control study of the impact of improved sanitation on diarrhoea morbidity in Lesotho. *Bull World Health Organ*. 1990;68(4):455-463. http://www.ncbi.nlm.nih.gov/pubmed/2208559. Accessed February 18, 2017.
18. United Nations. *Proclamation of the International Drinking Water Supply and Sanitation Decade*.; 1980. http://www.un.org/documents/ga/res/35/a35r18e.pdf.
19. Stanton BF, Clemens JD. An educational intervention for altering water-sanitation behaviors to reduce childhood diarrhea in urban Bangladesh. II. A randomized trial to assess the impact of the intervention on hygienic behaviors and rates of diarrhea. *Am J Epidemiol*. 1987;125(2):292-301. http://www.ncbi.nlm.nih.gov/pubmed/3812435. Accessed February 18, 2017.
20. Shahid NS, Greenough WB, Samadi AR, Huq MI, Rahman N. Hand washing with soap reduces diarrhoea and spread of bacterial pathogens in a Bangladesh village. *J Diarrhoeal Dis Res*. 1996;14(2):85-89. http://www.ncbi.nlm.nih.gov/pubmed/8870400. Accessed February 18, 2017.
21. Wilson JM, Chandler GN, Muslihatun, Jamiluddin. Hand-washing reduces diarrhoea episodes: a study in Lombok, Indonesia. *Trans R Soc Trop Med Hyg*. 85(6):819-821. http://www.ncbi.nlm.nih.gov/pubmed/1801364. Accessed February 18, 2017.
22. Aziz KM, Hoque BA, Hasan KZ, et al. Reduction in diarrhoeal diseases in children in rural Bangladesh by environmental and behavioural modifications. *Trans R Soc Trop Med Hyg*. 84(3):433-438. http://www.ncbi.nlm.nih.gov/pubmed/2260182. Accessed February 18, 2017.
23. Sobel J, Mahon B, Mendoza CE, et al. Reduction of fecal contamination of street-vended beverages in Guatemala by a simple system for water

purification and storage, handwashing, and beverage storage. *Am J Trop Med Hyg*. 1998;59(3):380-387.
24. Sobsey MD. *Managing Water in the Home: Accelerated Health Gains from Improved Water Supply*.; 2002.
25. Handzel T. The Effect of Improved Drinking Water Quality on the Risk of Diarrhoeal Disease in an Urban Slum of Dhaka, Bangladesh: A Home Chlorination Intervention Trial. *Chapel Hill Univ NC*. 1998.
26. Luby SP, Agboatwalla M, Painter J, et al. Combining drinking water treatment and hand washing for diarrhoea prevention, a cluster randomised controlled trial. *Trop Med Int Heal*. 2006;11(4):479-489. doi:10.1111/j.1365-3156.2006.01592.x.
27. Opryszko MC, Majeed SW, Hansen PM, et al. Water and hygiene interventions to reduce diarrhoea in rural Afghanistan: a randomized controlled study. *J Water Health*. 2010;8(4).
28. Arnold B, Arana B, Mausezahl D, Hubbard A, Colford JM. Evaluation of a pre-existing, 3-year household water treatment and handwashing intervention in rural Guatemala. *Int J Epidemiol*. 2009;38(6):1651-1661. doi:10.1093/ije/dyp241.
29. Hoque BA, Juncker T, Sack RB, Ali M, Aziz KM. Sustainability of a water, sanitation and hygiene education project in rural Bangladesh: a 5-year follow-up. *Bull World Health Organ*. 1996;74(4):431-437. http://www.ncbi.nlm.nih.gov/pubmed/8823966. Accessed February 18, 2017.
30. Luby SP, Mendoza C, Keswick BH, Chiller TM, Hoekstra RM. Difficulties in bringing point-of-use water treatment to scale in rural Guatemala. *Am J Trop Med Hyg*. 2008;78(3):382-387. http://www.ncbi.nlm.nih.gov/pubmed/18337330. Accessed February 18, 2017.
31. Fewtrell L, Kaufmann RB, Kay D, Enanoria W, Haller L, Colford JM. Water, sanitation, and hygiene interventions to reduce diarrhoea in less developed countries: a systematic review and meta-analysis. *Lancet Infect Dis*. 2005;5(1):42-52. doi:10.1016/S1473-3099(04)01253-8.
32. Arnold BF, Colford JM. Treating water with chlorine at point-of-use to improve water quality and reduce child diarrhea in developing countries: a systematic review and meta-analysis. *Am J Trop Med Hyg*. 2007;76(2):354-364. http://www.ncbi.nlm.nih.gov/pubmed/17297049. Accessed February 18, 2017.
33. Luby SP, Agboatwalla M, Bowen A, Kenah E, Sharker Y, Hoekstra RM. Difficulties in maintaining improved handwashing behavior, Karachi, Pakistan. *Am J Trop Med Hyg*. 2009;81(1):140-145. http://www.ncbi.nlm.nih.gov/pubmed/19556579. Accessed February 18,

2017.
34. Waddington H, Snilstveit B. Effectiveness and sustainability of water, sanitation, and hygiene interventions in combating diarrhoea. *J Dev Eff*. 2009;1(3):295-335. doi:10.1080/19439340903141175.
35. Clasen T, Boisson S, Routray P, et al. Effectiveness of a rural sanitation programme on diarrhoea, soil-transmitted helminth infection, and child malnutrition in Odisha, India: a cluster-randomised trial. *Lancet Glob Heal*. 2014;2(11):e645-e653. doi:10.1016/S2214-109X(14)70307-9.
36. Luby S. Is targeting access to sanitation enough? *Lancet Glob Heal*. 2014;2(11):e619-e620. doi:10.1016/S2214-109X(14)70326-2.
37. Clasen T, Schmidt W-P, Rabie T, Roberts I, Cairncross S. Interventions to improve water quality for preventing diarrhoea: systematic review and meta-analysis. *BMJ*. 2007;334(7597):782. doi:10.1136/bmj.39118.489931.BE.
38. Enger KS, Nelson KL, Rose JB, Eisenberg JNS. The joint effects of efficacy and compliance: A study of household water treatment effectiveness against childhood diarrhea. 2013. doi:10.1016/j.watres.2012.11.034.
39. Wolf J, Prüss-Ustün A, Cumming O, et al. Systematic review: Assessing the impact of drinking water and sanitation on diarrhoeal disease in low- and middle-income settings: systematic review and meta-regression. *Trop Med Int Heal*. 2014;19(8):928-942. doi:10.1111/tmi.12331.
40. Figueroa ME, Kincaid DL. *Social, Cultural and Behavioral Correlates of Household Water Treatment and Storage*.; 2010.
41. Sweat M, Kerrigan D, Moreno L, et al. Cost-effectiveness of environmental-structural communication interventions for HIV prevention in the female sex industry in the Dominican Republic. *J Health Commun*. 2006;11(Suppl 2):123-142. doi:10.1080/10810730600974829.
42. Clasen T, Schmidt W, Rabie T, Roberts I, Cairncross S. Interventions to improve water quality for preventing diarrhoea: systematic review and meta-analysis. *Br Med J*. 2007;(334):782. doi:10.1136/bmj.39118.489931.BE.
43. Huda TM, Unicomb L, Johnston RB, Halder AK, Yushuf Sharker MA, Luby SP. Interim evaluation of a large scale sanitation, hygiene and water improvement programme on childhood diarrhea and respiratory disease in rural Bangladesh. *Soc Sci Med*. 2012;75(4):604-611. doi:10.1016/j.socscimed.2011.10.042 [doi].
44. Wen M, Browning CR, Cagney KA. Poverty, affluence, and income inequality: Neighborhood economic structure and its implications for health. *Soc Sci Med*. 2003;57(5):843-860. doi:10.1016/S0277-9536(02)00457-4.
45. Browning CR. Moving beyond Poverty : Neighborhood Structure , Social Processes , and Health Author (s): Christopher R . Browning and Kathleen A . Cagney Source : Journal of Health and Social Behavior , Vol . 44 , No . 4

(Dec ., 2003), pp . 552-571 Published by : A. 2016;44(4):552-571.
46. Diez-Roux A. Multilevel analysis in public health research. *Annu Rev Public Health*. 2000;21:171-192.
47. Marmot M, Friel S, Bell R, Houweling TA, Taylor S. Closing the gap in a generation: health equity through action on the social determinants of health. *Lancet*. 2008;372(9650):1661-1669. doi:10.1016/S0140-6736(08)61690-6.
48. McLeroy KR, Bibeau D, Steckler A, Glanz K. Ecological Perspective on Promotion Programs. *Health Educ Q*. 1988;15(4):351-377. doi:10.1177/109019818801500401.
49. Bandura A. Collective efficacy. In: *Self-Efficacy: The Exercise of Control*. Vol 50. ; 1997:477-525. doi:10.1177/0957154X9400501708.
50. Berkman LF, Glass T. Social integration, social networks, social support, and health. *Soc Epidemiol*.:137-173. http://www.scopus.com/inward/record.url?eid=2-s2.0-0002294480&partnerID=tZOtx3y1.
51. Coleman JS. *Foundation of Social Theory*.; 1990. doi:10.2307/2579680.
52. Putnam RD. The Prosperous Community: Social Capital and Public Life. *Am Prospect*. 1993;13(13):35-42. http://www.citeulike.org/group/1702/article/1032759.
53. Kawachi I, Berkman LF. Social cohesion, social capital, and health. In: *Social Epidemiology*. ; 2000:174-190. http://books.google.com.au/books?id=wows50o7cw4C&dq=isbn:019987967&source=gbs_navlinks_s%5Cnhttps://books.google.com.au/books?hl=en&lr=&id=i0qWxy5-X1gC&oi=fnd&pg=PA174&dq=Social+cohesion,+social+capital,+and+health&ots=NIS35gus8d&sig=B-YkR-V_Aupm0gWEO_BKi.
54. Forrest R, Kearns A. Social Cohesion, Social Capital and the Neighbourhood. *Urban Stud*. 2001;38(12):2125-2143.
55. Rogers EM. *Diffusion of Innovations*.; 1983. doi:citeulike-article-id:126680.
56. Sahoo KC, Hulland KRS, Caruso BA, et al. Sanitation-related psychosocial stress: A grounded theory study of women across the life-course in Odisha, India. *Soc Sci Med*. 2015;139:80-89. doi:10.1016/j.socscimed.2015.06.031.
57. Hulland KRS, Chase RP, Caruso BA, et al. Sanitation, stress, and life stage: A systematic data collection study among women in Odisha, India. *PLoS One*. 2015;10(11). doi:10.1371/journal.pone.0141883.
58. Foster JA, Rinaman L, Cryan JF. Stress & the gut-brain axis: Regulation by the microbiome. *Neurobiol Stress*. March 2017. doi:10.1016/j.ynstr.2017.03.001.
59. Bharwani A, Mian MF, Foster JA, Surette MG, Bienenstock J, Forsythe P. Structural & functional consequences of chronic psychosocial stress on the

microbiome & host. *Psychoneuroendocrinology*. 2016;63:217-227. doi:10.1016/j.psyneuen.2015.10.001.
60. Riis JL, Granger DA, Minkovitz CS, Bandeen-Roche K, DiPietro JA, Johnson SB. Maternal distress and child neuroendocrine and immune regulation. *Soc Sci Med*. 2016;151:206-214. doi:10.1016/j.socscimed.2015.12.043.
61. United Nations Development Program (UNDP). *Gender Mainstreaming in Practice*.; 2007. https://www.k4health.org/sites/default/files/6D8DEA3D-F203-1EE9-B2E46DEFBD98F3F1.pdf. Accessed August 30, 2017.
62. Klovdahl A., Graviss E., Yaganehdoost A, et al. Networks and tuberculosis: an undetected community outbreak involving public places. *Soc Sci Med*. 2001;52(5):681-694. doi:10.1016/S0277-9536(00)00170-2.
63. Jolly AM, Muth SQ, Wylie JL, Potterat JJ. Sexual networks and sexually transmitted infections: a tale of two cities. *J Urban Health*. 2001;78(3):433-445. doi:10.1093/jurban/78.3.433.
64. Newman M. Spread of epidemic disease on networks. *Phys Rev E*. 2002;66(1):016128. doi:10.1103/PhysRevE.66.016128.
65. Pahl-Wostl C, Craps M, Dewulf A, Mostert E, Tabara D, Taillieu T. Social learning and water resources management. *Ecol Soc*. 2007;12(2). http://www.scopus.com/inward/record.url?eid=2-s2.0-34447323232&partnerID=tZOtx3y1.
66. Hunt RC, Hunt E. Canal irrigation and local social organization. *Curr Anthropol*. 17(3):389-411. http://www.scopus.com/inward/record.url?eid=2-s2.0-0344225012&partnerID=tZOtx3y1.
67. Wallace R. A synergism of plagues: "Planned shrinkage," contagious housing destruction, and AIDS in the bronx. *Environ Res*. 1988;47(1):1-33. doi:10.1016/S0013-9351(88)80018-5.
68. Cohen JM, Wilson ML, Aiello AE. Analysis of social epidemiology research on infectious diseases: historical patterns and future opportunities. *J Epidemiol Community Health*. 2007;61(12):1021-1027. doi:10.1136/jech.2006.057216.
69. Trostle JA, Hubbard A, Scott J, Cevallos W, Bates SJ, Eisenberg JNS. Raising the level of analysis of food-borne outbreaks: food-sharing networks in rural coastal Ecuador. *Epidemiology*. 2008;19(3):384-390. doi:10.1097/EDE.0b013e31816a9db0.
70. Bates SJ, Trostle J, Cevallos WT, Hubbard A, Eisenberg JNS. Relating diarrheal disease to social networks and the geographic configuration of communities in rural Ecuador. *Am J Epidemiol*. 2007;166(9):1088-1095. doi:10.1093/aje/kwm184.
71. GBD Diarrhoeal Diseases Collaborators, Troeger C, Forouzanfar M, Rao PC,

et al. Estimates of global, regional, and national morbidity, mortality, and aetiologies of diarrhoeal diseases: a systematic analysis for the Global Burden of Disease Study 2015. *Lancet Infect Dis*. 2017;17(9):909-948. doi:10.1016/S1473-3099(17)30276-1.

72. Albert A, Anderson JA. On the existence of maximum likelihood estimates in logistic regression models. *Biometrika*. 1984;71(1):1-10. doi:10.1093/biomet/71.1.1.
73. Santner TJ, Duffy DE. A note on A. Albert and J. A. Anderson's conditions for the existence of maximum likelihood estimates in logistic regression models. *Biometrika*. 1986;73(3):755-758. doi:10.1093/biomet/73.3.755.
74. Heinze G, Schemper M. A solution to the problem of separation in logistic regression. *Stat Med*. 2002;21(16):2409-2419. doi:10.1002/sim.1047.
75. Day NE, Kerridge DF. A general maximum likelihood discriminant. *Biometrics*. 1967;23(2):313-323. doi:10.2307/2528164.
76. Webb MC, Wilson JR, Chong J. An Analysis of Quasi-complete Binary Data with Logistic Models: Applications to Alcohol Abuse Data. *J Data Sci*. 2004;2:273-285. http://www.jds-online.com/files/JDS-155.pdf.
77. Firth D. Bias reduction of maximum likelihood estimates. *Biometrika*. 1993;80(1):27-38. doi:10.1093/biomet/80.1.27.
78. Abrahantes JC, Aerts M. A solution to separation for clustered binary data. *Stat Model An Int J*. 2012;12(1):3-27. doi:10.1177/1471082X1001200102.
79. Hubbard AE, Ahern J, Fleischer NL, et al. To GEE or Not to GEE. *Epidemiology*. 2010;21(4):467-474. doi:10.1097/EDE.0b013e3181caeb90.
80. Neuhaus JM, Kalbfleisch JD, Hauck WW. A comparison of cluster-specific and population-averaged approaches for analyzing correlated binary data. *Int Stat Rev*. 1991;59(1):25-35. doi:10.2307/1403572.
81. Gelman A. Prior Distribution for Variance Parameters in Hierarchical Models. *Bayesian Anal*. 2006;1(3):515-533. doi:10.1214/06-BA117A.
82. Gelman A, Jakulin A, Pittau MG, Su YS. A weakly informative default prior distribution for logistic and other regression models. *Ann Appl Stat*. 2008;2(4):1360-1383. doi:10.1214/08-AOAS191.
83. Allison P, Altman M, Gill J, McDonald MP. Convergence Problems in Logistic Regression. In: *Numerical Issues in Statistical Computing for the Social Scientist*. ; 2003:238-252. doi:10.1002/0471475769.ch10.
84. Garthright WE, Archer DL, Kvenberg JE. Estimates of incidence and costs of intestinal infectious diseases in the United States. *Public Health Rep*. 1988;103(2):107-115. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1477958&tool=pmcentrez&rendertype=abstract.
85. Roderick P, Wheeler J, Cowden J, et al. A pilot study of infectious intestinal

disease in England. *Epidemiol Infect*. 1995;114(2):277-288. http://www.ncbi.nlm.nih.gov/pubmed/7705491. Accessed April 24, 2018.
86. Payment P, Richardson L, Siemiatycki J, Dewar R, Edwardes M, Franco E. A randomized trial to evaluate the risk of gastrointestinal disease due to consumption of drinking water meeting current microbiological standards. *Am J Public Health*. 1991;81(6):703-708. doi:10.2105/AJPH.81.6.703.
87. Roy SL, Beach MJ, Scallan E. The rate of acute gastrointestinal illness in developed countries. *J Water Health*. 2006;04(Suppl 2):31. doi:10.2166/wh.2006.017.
88. Majowicz SE, Hall G, Scallan E, et al. A common, symptom-based case definition for gastroenteritis. *Epidemiol Infect*. 2008;136(7):886-894. doi:10.1017/S0950268807009375.
89. Friedkin NE. Social Cohesion. *Annu Rev Sociol*. 2004;30:409-425. doi:10.1146/annurev.soc.30.012703.110625.
90. Marsden P V. Core Discussion Networks of Americans. *Am Sociol Rev*. 1987;52(1):122. doi:10.2307/2095397.
91. Eisenberg JNS, Desai MA, Levy K, et al. Environmental determinants of infectious disease: A framework for tracking causal links and guiding public health research. *Environ Health Perspect*. 2007;115(8):1216-1223. doi:10.1289/ehp.9806.
92. Hoffman MD, Gelman A. The No-U-Turn Sampler: Adaptively Setting Path Lengths in Hamiltonian Monte Carlo. November 2011. http://arxiv.org/abs/1111.4246. Accessed May 23, 2018.
93. Geweke J. Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. *Bayesian Stat 4*. 1992:169–193. doi:1176289.
94. Greenland S. Putting background information about relative risks into conjugate prior distributions. *Biometrics*. 2001;57(3):663-670. doi:10.1111/j.0006-341X.2001.00663.x.
95. Small ML. Weak ties and the core discussion network: Why people regularly discuss important matters with unimportant alters. *Soc Networks*. 2013;35(3):470-483. doi:10.1016/j.socnet.2013.05.004.
96. Perry BL, Pescosolido BA. Functional specificity in discussion networks: The influence of general and problem-specific networks on health outcomes. *Soc Networks*. 2010;32(4):345-357. doi:10.1016/j.socnet.2010.06.005.
97. Meyers LA, Pourbohloul B, Newman MEJ, Skowronski DM, Brunham RC. Network theory and SARS: Predicting outbreak diversity. *J Theor Biol*. 2005;232(1):71-81. doi:10.1016/j.jtbi.2004.07.026.
98. Forrest R, Kearns A. Social Cohesion, Social Capital and the Neighbourhood. *Urban Stud*. 2001;38(12):2125-2143.

doi:10.1080/00420980120087081.
99. Shakya HB, Christakis NA, Fowler JH. Association Between Social Network Communities and Health Behavior: An Observational Sociocentric Network Study of Latrine Ownership in Rural India. *Am J Public Health*. 2014;104(5):930-937. doi:10.2105/AJPH.2013.301811.
100. Merluzzi J, Burt RS. How many names are enough? Identifying network effects with the least set of listed contacts. *Soc Networks*. 2013;35(3):331-337. doi:10.1016/j.socnet.2013.03.004.
101. Eisenberg JNS, Cevallos W, Ponce K, et al. Environmental change and infectious disease: how new roads affect the transmission of diarrheal pathogens in rural Ecuador. *Proc Natl Acad Sci U S A*. 2006;103(51):19460-19465. doi:10.1073/pnas.0609431104.
102. Griffith G, Rose C. Technical notes: Calculating the human development indices. *Hum Dev Rep 2016*. 2016;37(1):14. doi:10.1115/1.1731317.
103. Silvapulle MJ. On the Existence of Maximum Likelihood Estimators for the Binomial Response Models. *J R Stat Soc Ser B*. 1981;43(3):310-313. http://www.jstor.org/stable/2984941.
104. Kamada T, Kawai S. An algorithm for drawing general undirected graphs. *Inf Process Lett*. 1989;31(1):7-15. doi:10.1016/0020-0190(89)90102-6.
105. Fruchterman TMJ, Reingold EM. Graph drawing by force-directed placement. *Softw Pract Exp*. 1991;21(11):1129-1164. doi:10.1002/spe.4380211102.
106. Newman MEJ. Modularity and community structure in networks. *Proc Natl Acad Sci*. 2006;103(23):8577-8582. doi:10.1073/pnas.0601602103.
107. Goldstick JE, Trostle J, Eisenberg JNS. Ask when - Not just whether - It's a risk: How regional context influences local causes of diarrheal disease. *Am J Epidemiol*. 2014;179(10). doi:10.1093/aje/kwu034.
108. Kraay ANM, Trostle J, Brouwer AF, Cevallos W, Eisenberg JNS. Determinants of Short-term Movement in a Developing Region and Implications for Disease Transmission. *Epidemiology*. 2018;29(1):117-125. doi:10.1097/EDE.0000000000000751.
109. Pustejovsky JE, Spillane JP. Question-order effects in social network name generators. *Soc Networks*. 2009;31(4):221-229. doi:10.1016/j.socnet.2009.06.001.
110. Shakya HB, Christakis NA, Fowler JH. An exploratory comparison of name generator content: Data from rural India. *Soc Networks*. 2017;48:157-168. doi:10.1016/j.socnet.2016.08.008.
111. Small ML, Sukhu C. Because they were there: Access, deliberation, and the mobilization of networks for support. *Soc Networks*. 2016;47:73-84. doi:10.1016/j.socnet.2016.05.002.

112. Small ML, Deeds Pamphile V, McMahan P. How stable is the core discussion network? *Soc Networks*. 2015;40:90-102. doi:10.1016/j.socnet.2014.09.001.
113. Rimé B. Comment: Social Integration and Health: Contributions of the Social Sharing of Emotion at the Individual, the Interpersonal, and the Collective Level. *Emot Rev*. 2018;10(1):67-70. doi:10.1177/1754073917719330.
114. Harber KD, Podolski P, Dyer L. Hearing Stories That Violate Expectations Leads to Emotional Broadcasting. *J Lang Soc Psychol*. 2014;33(1):5-28. doi:10.1177/0261927X13502793.
115. Rimé B. Emotion elicits the social sharing of emotion: Theory and empirical review. *Emot Rev*. 2009;1(1):60-85. doi:10.1177/1754073908097189.
116. Parkinson B, Fischer AH, Manstead ASR. *Emotion in Social Relations: Cultural, Group, and Interpersonal Processes*.; 2004. doi:10.4324/9780203644966.
117. Neville F, Reicher S. The experience of collective participation: Shared identity, relatedness and emotionality. *Contemp Soc Sci*. 2011;6(3):377-396. doi:10.1080/21582041.2012.627277.
118. Gasparre A, Bosco S, Bellelli G. Cognitive and social consequences of participation in social rites: Collective coping, social support, and post-traumatic growth in the victims of Guatemala genocide. *Rev Psicol Soc*. 2010;25(1):35-46. doi:10.1174/021347410790193513.
119. Swann WB, Jetten J, Gómez Ã, Whitehouse H, Bastian B. When group membership gets personal: A theory of identity fusion. *Psychol Rev*. 2012;119(3):441-456. doi:10.1037/a0028589.
120. Kanyangara P, Rimé B, Philippot P, Yzerbyt V. Collective rituals, emotional climate and intergroup perception: Participation in “gacaca”tribunals and assimilation of the Rwandan genocide. *J Soc Issues*. 2007;63(2):387-403. doi:10.1111/j.1540-4560.2007.00515.x.
121. Rimé B, Kanyangara P, Yzerbyt V, Paez D. The impact of Gacaca tribunals in Rwanda: Psychosocial effects of participation in a truth and reconciliation process after a genocide. *Eur J Soc Psychol*. 2011;41(6):695-706. doi:10.1002/ejsp.822.
122. Páez D, Basabe N, Ubillos S, González-castro JL. Social sharing, participation in demonstrations, emotional climate, and coping with collective violence after the March 11th Madrid bombings. *J Soc Issues*. 2007;63(2):323-337. doi:10.1111/j.1540-4560.2007.00511.x.
123. Hawdon J, Ryan J. Well-Being After the Virginia Tech Mass Murder. *Traumatology (Tallahass Fla)*. 2012;18(4):3-12. doi:10.1177/1534765612441096.
124. Ramugondo EL, Kronenberg F. Explaining collective occupations from a

human relations perspective: Bridging the individual-collective dichotomy. *J Occup Sci*. 2015;22(1):3-16. doi:10.1080/14427591.2013.781920.
125. EASTERLY W, RITZEN J, WOOLCOCK M. SOCIAL COHESION, INSTITUTIONS, AND GROWTH. *Econ Polit*. 2006;18(2):103-120. doi:10.1111/j.1468-0343.2006.00165.x.
126. WHO and UNICEF. *Progress on Drinking Water, Sanitation and Hygiene WHO*.; 2017. doi:10.1111 / tmi.12329.
127. Clasen TF, Alexander KT, Sinclair D, et al. Interventions to improve water quality for preventing diarrhoea. *Cochrane database Syst Rev*. 2015;(10). doi:10.1002/14651858.CD004794.pub3.
128. Pickering AJ, Djebbari H, Lopez C, Coulibaly M, Alzua ML. Effect of a community-led sanitation intervention on child diarrhoea and child growth in rural Mali: A cluster-randomised controlled trial. *Lancet Glob Heal*. 2015;3(11). doi:10.1016/S2214-109X(15)00144-8.
129. Parvez SM, Azad R, Rahman M, et al. Achieving optimal technology and behavioral uptake of single and combined interventions of water, sanitation hygiene and nutrition, in an efficacy trial (WASH benefits) in rural Bangladesh. *Trials*. 2018;19(1). doi:10.1186/s13063-018-2710-8.
130. Kempe D, Kleinberg J, Tardos É. Maximizing the spread of influence through a social network. In: *Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD '03*. ; 2003:137. doi:10.1145/956755.956769.
131. WHO/UNICEF. *Progress on Sanitation and Drinking Water: 2015 Update and MDG Assessment*.; 2015. doi:10.1007/s13398-014-0173-7.2.
132. VanderWeele TJ. Mediation Analysis: A Practitioner's Guide. *Annu Rev Public Health*. 2016;37(1):17-32. doi:10.1146/annurev-publhealth-032315-021402.
133. Crawley MJ. *The R Book*.; 2007. doi:10.1002/9780470515075.
134. Vanderweele TJ. A unification of mediation and interaction: A 4-way decomposition. *Epidemiology*. 2014;25(5):749-761. doi:10.1097/EDE.0000000000000121.
135. VanderWeele TJ, Vansteelandt S. Odds ratios for mediation analysis for a dichotomous outcome. *Am J Epidemiol*. 2010;172(12):1339-1348. doi:10.1093/aje/kwq332.
136. McGranahan G. Realizing the Right to Sanitation in Deprived Urban Communities: Meeting the Challenges of Collective Action, Coproduction, Affordability, and Housing Tenure. *World Dev*. 2015;68:242-253. doi:10.1016/J.WORLDDEV.2014.12.008.
137. Simiyu S, Swilling M, Cairncross S. Decision-making on shared sanitation in the informal settlements of Kisumu, Kenya. *Int J Environ Health Res*.

2017;27(5):377-393. doi:10.1080/09603123.2017.1350261.
138. Graham JP, Polizzotto ML. Pit latrines and their impacts on groundwater quality: A systematic review. *Environ Health Perspect*. 2013. doi:10.1289/ehp.1206028.
139. Routray P, Schmidt W-P, Boisson S, Clasen T, Jenkins MW. Socio-cultural and behavioural factors constraining latrine adoption in rural coastal Odisha: an exploratory qualitative study. *BMC Public Health*. 2015;15(1). doi:10.1186/s12889-015-2206-3.
140. Dzwairo B, Hoko Z, Love D, Guzha E. Assessment of the impacts of pit latrines on groundwater quality in rural areas: A case study from Marondera district, Zimbabwe. *Phys Chem Earth*. 2006. doi:10.1016/j.pce.2006.08.031.
141. Knappett PSK, Escamilla V, Layton A, et al. Impact of population and latrines on fecal contamination of ponds in rural Bangladesh. *Sci Total Environ*. 2011;409:3174-3182. doi:10.1016/j.scitotenv.2011.04.043.
142. Muruka C, Fagbamigbe FA, Muruka A, et al. The Relationship between Bacteriological Quality of Dug-Wells & Pit Latrine Siting in an Unplanned Peri-Urban Settlement: A Case Study of Langas – Eldoret Municipality, Western Kenya. *Public Heal Res*. 2012;2012(2):32-36. doi:10.5923/j.phr.20120202.06.
143. VanderWeele TJ. Direct and indirect effects for neighborhood-based clustered and longitudinal data. *Sociol Methods Res*. 2009. doi:10.1177/0049124110366236.
144. MacKinnon DP, Valente MJ. Mediation from multilevel to structural equation modeling. In: *Annals of Nutrition and Metabolism*. ; 2014. doi:10.1159/000362505.
145. Clasen T, Fabini D, Boisson S, et al. Making sanitation count: Developing and testing a device for assessing latrine use in low-income settings. *Environ Sci Technol*. 2012. doi:10.1021/es2036702.
146. Prüss-Ustün A, Bartram J, Clasen T, et al. Burden of disease from inadequate water, sanitation and hygiene in low- and middle-income settings: A retrospective analysis of data from 145 countries. *Trop Med Int Heal*. 2014;19(8). doi:10.1111/tmi.12329.
147. Ercumen A, Arnold BF, Kumpel E, et al. Upgrading a Piped Water Supply from Intermittent to Continuous Delivery and Association with Waterborne Illness: A Matched Cohort Study in Urban India. *PLoS Med*. 2015;12(10). doi:10.1371/journal.pmed.1001892.
148. Sweat M, Kerrigan D, Moreno L, et al. Cost-Effectiveness of Environmental–Structural Communication Interventions for HIV Prevention in the Female Sex Industry in the Dominican Republic. *J Health Commun*. 2006;11(sup2):123-142. doi:10.1080/10810730600974829.

149. Hennink M, Hutter I, Bailey A. *Qualitative Research Methods*. SAGE Publications Limited; 2011.
150. Newman MEJ. Assortative Mixing in Networks. *Phys Rev Lett*. 2002. doi:10.1103/PhysRevLett.89.208701.
151. Newman MEJ. *Mixing Patterns in Networks*.; 2003. https://arxiv.org/pdf/cond-mat/0209450.pdf. Accessed October 5, 2018.
152. Efron B. Bootstrap Methods: Another Look at the Jackknife. *Ann Stat*. 1979. doi:10.1214/aos/1176344552.
153. Chavance M. Jackknife and bootstrap. *Rev Epidemiol Sante Publique*. 1992.
154. Dixon PM. The Bootstrap and the Jackknife. *Des Anal Ecol Exp*. 2001. doi:10.1007/0-387-30623-4_3.
155. Wu CFJ. Jackknife, Bootstrap and Other Resampling Methods in Regression Analysis. *Ann Stat*. 1986. doi:10.1214/aos/1176350161.
156. Farine DR. A guide to null models for animal social network analysis. Freckleton R, Rands S, eds. *Methods Ecol Evol*. 2017;8(10):1309-1320. doi:10.1111/2041-210X.12772.
157. Whitten NE. *Class, Kinship, and Power in an Ecuadorian Town*. First. Stanford, California: Stanford University Press; 1965.
158. Akin A, Demirel S. Social gender concept and its effects on health. *Cumhur Üniversitesi Tip Fakültesi Derg*. 2003;(25):73–82. https://scholar-google-com.proxy.lib.umich.edu/scholar?cluster=2509001963342288248&hl=en&as_sdt=80000005&sciodt=0,23. Accessed November 9, 2018.
159. Basow SA. *Gender: Stereotypes and Roles*. 3rd ed. Belmont, CA: Thomson Brooks Cole Publishing ; 1992. http://psycnet.apa.org.proxy.lib.umich.edu/record/1992-97754-000. Accessed November 9, 2018.
160. Dokmen Y. Gender social psychological explanations [Toplumsal cinsiyet: Sosyal psikolojik açıklamalar.]. *Istanbul Sist Yayıncılık*. 2004. https://scholar-google-com.proxy.lib.umich.edu/scholar?hl=en&as_sdt=0%2C23&q=Dokmen+2004+gender&btnG=. Accessed November 9, 2018.
161. Mahaffy KA, Ward SK. The gendering of adolescents' childbearing and educational plans: Reciprocal effects and the influence of social context. *Sex Roles*. 2002. doi:10.1023/A:1020413630553.
162. Lindsey LL. *Gender Roles: A Sociological Perspective*. 6th ed. London and New York: Routledge Taylor & Francis Group; 2015. https://books.google.com/books?hl=en&lr=&id=qjjbCgAAQBAJ&oi=fnd&pg=PP1&dq=Lindsey+1990+Gender+roles:+a+sociological+perspective.&ots=U4hB85H8Uw&sig=8tW8-Hum93MPBIsLHAtVymSwRfc#v=onepage&q=Lindsey 1990 Gender

roles%3A a sociological perspective. Accessed November 9, 2018.
163. Kabeer N. Gender equality and women's empowerment: A critical analysis of the third millennium development goal 1. *Gend Dev*. 2005;13(1):13-24. doi:10.1080/13552070512331332273.
164. Hanmer L, Klugman J. Exploring Women's Agency and Empowerment in Developing Countries: Where do we stand? *Fem Econ*. 2016;22(1):237-263. doi:10.1080/13545701.2015.1091087.
165. Marsiglio W. Adolescent Male Sexuality and Heterosexual Masculinity: A Conceptual Model and Review. *J Adolesc Res*. 1988;3(3-4):285-303. doi:10.1177/074355488833005.
166. Kaufman PT. Library Leadership: Does Gender Make a Difference? *J Libr Adm*. 1993;18(3-4):109-128. doi:10.1300/J111v18n03_08.
167. Rivers K, Aggleton P, Elizondo J, et al. Gender relations, sexual communication and the female condom. *Crit Public Health*. 1998;8(4):273-290. doi:10.1080/09581599808402916.
168. Rivers K, Aggleton P. *Gender and the HIV Epidemic Adolescent Sexuality, Gender and the HIV Epidemic*. http://www.pwhce.ca/hivaids/scenario2/Gender_and_the_HIV_Epidemic.pdf. Accessed November 9, 2018.
169. A G, M K, J L. Men masculinities and development: broadening our work towards gender equality. 2000. https://www.popline.org/node/172497. Accessed November 9, 2018.
170. Barker G. Gender equitable boys in a gender inequitable world: Reflections from qualitative research and programme development in Rio de Janeiro. *Sex Relatsh Ther*. 2000;15(3):263-282. doi:10.1080/14681990050109854.
171. Barker G, Ricardo C, Nascimento M. *Engaging Men and Boys in Changing Gender-Based Inequity in Health: Evidence from Programme Interventions*. http://apps.who.int/iris/bitstream/handle/10665/43679/9789241595490_eng.pdf?sequence=1&isAllowed=y. Accessed November 9, 2018.
172. Barker G, Ricardo C. *Young Men and the Construction of Masculinity in Sub-Saharan Africa: Implications for HIV/AIDS, Conflict, and Violence*. http://documents.worldbank.org/curated/en/481401468101357773/pdf/327120rev0PAPER0AFR0young0men0WP26.pdf. Accessed November 9, 2018.
173. Pulerwitz J, Barker G. The Gender-Equitable Men Scale (GEM Scale). *Men Masc*. 2007. doi:10.1037/t03149-000.
174. World Health Organization. Global and regional estimates of violence against women: prevalence and health effects of intimate partner violence and non-partner sexual violence. *2013*. 2013:57. doi:10.1007/s13398-014-0173-7.2.
175. Klugman J, Hanmer L, Twigg S, Hasan T, McCleary-Sills J, Santamaria J.

Voice and Agency: Empowering Women and Girls for Shared Prosperity. The World Bank; 2014. doi:10.1596/978-1-4648-0359-8.

176. Durante D, Steorts RC. *Bayesian Learning of Dynamic Multilayer Networks*. Vol 18.; 2017. http://jmlr.org/papers/v18/16-391.html. Accessed October 5, 2018.
177. Gel YR, Lyubchich V, Ramirez Ramirez LL. Bootstrap quantification of estimation uncertainties in network degree distributions. *Sci Rep*. 2017;7(1):5807. doi:10.1038/s41598-017-05885-x.
178. Chant S. Women, girls and world poverty: empowerment, equality or essentialism? *Int Dev Plan Rev*. 2016;38(1):1-24. doi:10.3828/idpr.2016.1.
179. Collins D, Morduch J, Rutherford S, Ruthven O. Portfolios of the poor: How the world's poor live on $2 a day. *Princet Univ Press*. 2009;23(4):431-432. doi:10.1111/j.1747-7093.2009.00237_2.x.
180. Wallerstein N, Duran B. Community-based participatory research contributions to intervention research: The intersection of science and practice to improve health equity. *Am J Public Health*. 2010;100(SUPPL. 1):40-46. doi:10.2105/AJPH.2009.184036.