Distribution Agreement

In presenting this thesis or dissertation as a partial fulfillment of the requirements for an advanced degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis or dissertation in whole or in part in all forms of media, now or hereafter known, including display on the world wide web. I understand that I may select some access restrictions as part of the online submission of this thesis or dissertation. I retain all ownership rights to the copyright of the thesis or dissertation. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation.

Signature:

________________________________ _______________

Forest Altherr Date

Integrating a Geographic Information System to Explore the Effect of Water, Sanitation, and Hygiene on Trachoma at Aggregate Spatial Scales

By

Forest Altherr
Master of Public Health

Epidemiology

____
Karen Levy, PhD, MPH
Committee Chair

____
Julie Clennon, PhD, MSc
Committee Member

____
Scott Nash, PhD
Committee Member

Integrating a Geographic Information System to Explore the Effect of Water, Sanitation, and Hygiene on Trachoma at Aggregate Spatial Scales

By

Forest Altherr

Bachelor of Science
University of California, San Diego
2011

Thesis Committee Chair: Karen Levy, PhD, MPH

An abstract of
A thesis submitted to the Faculty of the
Rollins School of Public Health of Emory University
in partial fulfillment of the requirements for the degree of
Master of Public Health
in Epidemiology
2017

Integrating a Geographic Information System to Explore the Effect of Water, Sanitation, and Hygiene on Trachoma at Aggregate Spatial Scales

Forest Altherr[1,2], Eshetu Sata[3], Aisha E. P. Stewart[1], Tigist Astale[3], Mulat Zerihun[3], Demelash Gessesse[3], Gedefaw Ayenew[3], Melsew Chanyalew[4], Berhanu Melaku[3], Zerihun Tadesse[3], Julie A. Clennon[2], Karen Levy[2], E. Kelly Callahan[1], Scott D. Nash[1]

[1] The Carter Center, Atlanta, Georgia, United States of America; [2]Rollins School of Public Health, Emory University, Atlanta, GA, United States of America, [3] The Carter Center, Addis Ababa, Ethiopia, [4] Amhara Regional Health Bureau, Bahirdar, Ethiopia

*Email: scott.nash@cartercenter.org

Short Title: Spatial Effects of WASH on Trachoma

Abstract

Integrating a Geographic Information System to Explore the Effect of Water, Sanitation, and Hygiene on Trachoma at Aggregate Spatial Scales
By Forest Altherr

Trachoma is an infectious disease responsible for a large proportion of the global burden of preventable blindness. A multipronged intervention strategy challenged the distribution of the disease by focusing on surgery, antibiotics, facial cleanliness, and environmental improvements (SAFE). The F and E arms have proven efficacy in combating the disease, but few studies have attempted to explain the effect of geographic oscillations in these programs. This study addresses whether the regional differences in water, sanitation, and hygiene (WASH) interventions are associated with the spatial distribution of trachomatous inflammation-follicular (TF) in the Amhara Regional State of Ethiopia.

The spatial pattern of TF was found to exhibit global dependence with neighboring evaluation units. Using spatial methods, the bias of neighboring observational units in close geographic proximity was corrected to yield a reliable spatially lagged regression model. Additionally, clusters of high TF were identified at the gott and the woreda scale of analysis using aggregated weighted estimates of the prevalence of the disease with a geographic information system and the Getis-Ord G_i*(d) statistic for local clustering. Socio-demographic, community, and geoclimatic factors thought to promote the clustering of the disease were modeled with logistic regression.

The prevalence of TF among children aged 1 to 9 years in Amhara ranged from a district prevalence of 1.2% to 73.9%. Clustering was identified at two spatial scales, district and village. Percent of children without nasal or ocular discharge as well as percent of households with access to a water source in less than 30 minutes were statistically significantly associated with a reduced prevalence of TF. The aforementioned variables and household access to a latrine were significant predictors in a model where spatial lag was applied to the outcome.

This study demonstrated that water and hygiene are important factors in the clustering of trachoma within a hyperendemic area. Intensified promotion of structural and behavioral interventions to increase WASH coverage may be necessary to eliminate trachoma as a public health problem in Amhara.

Integrating a Geographic Information System to Explore the Effect of Water, Sanitation, and Hygiene on Trachoma at Aggregate Spatial Scales

By

Forest Altherr

Bachelor of Science
University of California, San Diego
2011

Thesis Committee Chair: Karen Levy, PhD, MPH

A thesis submitted to the Faculty of the
Rollins School of Public Health of Emory University
in partial fulfillment of the requirements for the degree of
Master of Public Health
in Epidemiology
2017

Acknowledgements

I would like to dedicate this work to the sincere support of friends and family who helped me throughout the process. Without these pillars to lean on the lessons learned along the way would not have been apparent.

To Dr. Scott Nash at the Carter Center as who took me under his wing and provided limitless support for my success on this project. Furthermore, the Carter Center's Trachoma Control Program whose team guided me through rough patches and inspired me to pursue a creative approach to answering my question. To my colleague Andrew Nute who helped me in the process as much as anyone, and would intently listen and provide an insightful perspective on my ideas.

To Dr. Karen Levy of the Rollins School of Public Health who helped me to find my first global field experience and who was understanding when I wanted to pursue my academic interests in spatial epidemiology with the Carter Center.

To Dr. Julie Clennon and Dr. Lance Waller of the Rollins School of Public Health who have forever changed the way I will look at solving problems. Their mantras, "Tell a story with your data," and "You can't map what you don't know" will guide the research I continue to pursue in my career.

Table of Contents

List of Tables

List of Figures

Literature Review

Introduction

Trachoma, caused by the bacterium *Chlamydia trachomatis*, is the leading form of preventable blindness worldwide(1). Globally, trachoma is endemic in 42 countries focal to Sub-Saharan Africa and the disease is suspected to be responsible for visual impairment in 1.9 million people(1). In 1998 an agenda to eliminate trachoma by the year 2020 was established by the World Health Organization (WHO) and involved the implementation of the Global Alliance to Eliminate Blinding Trachoma by 2020 (GET 2020). Multiple countries have been able to achieve national elimination targets since this meeting through the use of a multipronged strategy of surgery, antibiotics, facial cleanliness, and environmental improvements (SAFE) (2). Population based prevalence surveys around the world have found that the prevalence of trachoma exhibits regional heterogeneity (3). The Carter Center supports the efforts of partner countries, including the Amhara National Regional State in Ethiopia to eliminate trachoma as a public health problem. Mapping the observed prevalence of trachoma at the woreda (district) level in Amhara is used to inform the implementation of interventions, and it has been suggested that spatial dependency of sociodemographic and environmental factors impact the efforts to reduce the prevalence of the disease (4). Advancing the understanding of the spatial epidemiology of trachoma will yield programmatic benefits by accounting for the within-region risk factor differences. A plausible pathway exists between the environment and water, sanitation, and hygiene (WASH) that amplifies areas of localized transmission (5). Although many studies have assessed the impact of various interventions on trachoma, relatively few have considered the spatial dependency of the disease to investigate the effectiveness regionally specific WASH interventions.

Biology and Transmission of Trachoma

Trachoma is caused by infection with the bacterium *Chlamydia trachomatis*. While a complete understanding of transmission mechanisms remains undetermined, it is believed that transmission is through direct contact with infected eyes and noses, indirectly with fomites (towels, bedsheets, etc.), and by the *Musca sorbens* fly serving as a mechanical vector (6, 7). Additionally, trachoma is most prevalent in crowded areas with poor access to clean water and sanitation and disproportionately affects women and children (8, 9). Children likely exhibit a high prevalence of trachoma due to passing the infection between siblings and playmates through direct contact (10). Women are more likely to be exposed to trachoma in part due to the traditional gender role as a caretaker interacting with children who exhibit higher prevalence of inflammatory trachoma. Furthermore, the prevalence of trachoma is found to be greater in groups with lower literacy rates and areas that are more economically disadvantaged (11, 12). Mild irritation and signs of inflammation typically present in the early stages of the disease but repeated and prolonged episodes of infection lead to blinding later in life. Progressive scaring eventually leads to the in turning of the eyelid causing the eyelashes to scratch the surface of the cornea which when combined with dryness and secondary infections leads to corneal opacification. Active infection with trachoma is most prevalent in young children and declines with age (13). Inversely, the scarring features associated with advanced stages of trachoma are most prevalent in adults illustrating the cumulative progression of the disease (14).

Evaluation Methods

Trachoma is commonly assessed with the WHO's simplified grading system which was developed in 1987 as a standardized way to grade the clinical presentation of trachoma symptoms during surveys (15). The clinical diagnosis involves inverting the eyelid and examining the conjunctiva for quantifiable physical symptoms using a 2.5X ocular loop and flashlight. Field survey teams undergo rigorous training in identification of the clinical signs of trachoma and are only considered qualified to grade trachoma if they are able to achieve an 80% agreement rate with experienced graders during evaluation (13). Trachomatous inflammation-follicular (TF) is described as the mildest form of the disease and is identified by the presence of 5 or more follicles in the upper tarsal conjunctiva. The next stage of the disease is trachomatous inflammation intense (TI) which is identified with a pronounced inflammation and thickening of the upper tarsal conjunctiva that obscures more than half of the deep tarsal vessels. Trachomatous scarring (TS) occurs after repeated episodes and is identified by the presence of scarring in the tarsal conjunctiva and the tarsal plate begins to thicken. Trachomatous trichiasis (TT) occurs when the scarring has worsened to the point where the eyelid turns inward in the process of entropion resulting in at least one eyelash touching the eyeball. The blinding stage of the disease is corneal opacity (CO) which is identifiable by visible opacity over the pupil (16). The benefit of diagnosing trachoma by clinical features is that the cost is lower than PCR and the diagnosis is rapid. On the other hand, PCR has a better discriminatory capability and has recently been used in studies to confirm clinical diagnosis, but cannot determine prior infection (17, 18).

Population-based prevalence surveys are frequently employed to determine woredas where trachoma is a public health problem. Cluster random sampling is considered the most robust method for estimation and these surveys are based on methods designed to weight the selection of villages according to the population size to generate a representative sample for a woreda (19). Each woreda is surveyed by the same sampling method. Villages within the woreda are selected based on a probability proportional to estimated size and the sample size of surveyed households is estimated to reflect the prevalence of trachoma in the entire population. These cross sectional survey designs are used across many programs to monitor progress and plan interventions.

Interventions

The current trachoma intervention strategy is the WHO recommended SAFE strategy. Surgery by incising the tarsal plate is recommended by the WHO in trachoma endemic communities and is frequently offered to anyone presenting with TT (20). While this intervention has not been causally associated with prevention of progression to CO it mechanistically could prevent the eyelashes from abrading the cornea and thus prevent visual loss. Surgical interventions have been challenged by cost, fear, lack of awareness of services, and transportation barriers (21). In areas where the prevalence of TF among children ages 1 to 9 is > 10% the WHO recommends annual community-wide mass drug administration (MDA) for three years with a target coverage of ≥80% of the population in order to reduce the infection rates with *C. trachomatis*. Factors such as the baseline prevalence or the percent of the population reached with antibiotics influences the effectiveness of these treatment campaigns (22). The Carter Center has assisted the Amhara Regional Health Bureau to collect data used to determine the effectiveness of MDA programs in Amhara since 2010 with trachoma impact surveys (TIS). The method of survey and use of antibiotics is dependent on the prevalence of trachoma using the WHO guidelines (23).

- Woreda-level TF ≥ 30%: Continued MDA and a TIS performed at least five years after implementation of the full SAFE strategy
- Woreda-level TF >10%: MDA to continue for 3 years followed by a second TIS
- Woreda-level TF 5 to 9.9%: MDA and continued support for facial hygiene and environmental improvements
- Woreda-level TF <5%: MDA stopped but other SAFE interventions continued along with surveillance

The promotion of facial cleanliness is aimed at reducing the transmission of trachoma by ocular and nasal secretions which are an important reservoir of infection (24). A previous meta-analysis found that children presenting with clean faces, as defined as the absence of ocular or nasal discharge, are significantly associated with reduced odds of trachoma graded clinically and by presence of infection (25). The promotion of face washing may be achieved with health education activities that incorporate the communities perceptions the importance of face washing (26). Additionally increasing water accessibility to provide households with a supply to perform washing activities may help to reduce the prevalence of the disease (27). Environmental improvements promote better health while improving living conditions to attenuate the role that the environment plays in the transmission of trachoma. Increasing access to latrines, development of sustainable water systems, vector control, and hygiene promotion all fit within this arm of the intervention strategy. The individual components of the SAFE strategy have shown success in reduction of trachoma prevalence, however an integrated strategy has been shown to be the most effective (28-30). Despite the availability of interventions to control trachoma, sustainability reducing trachoma prevalence is difficult to achieve in hyperendemic settings with antibiotics alone (31). Trachoma is capable of rebounding to pre-intervention levels and the success of programs frequently exhibit heterogeneous effects that differ across communities (22). Differential success of interventions may be the result of spatial variance in large scale social or environmental factors. For example, one study suggested that accounting for the properties of the natural environment might help future project managers find locations where latrines would be sustainable by incorporating soil type, slope, and rainfall along with other environmental factors into a model to predict sanitation coverage (32). The differences in uptake of interventions in conjunction with the spatial distribution of communities could be influencing the geographic spread of trachoma.

The Role of Flies in the Transmission of Trachoma

The entomology of the vector for trachoma transmission, *M. sorbens*, requires further study. A small number of studies have demonstrated that the fly could be verified as a vector through the established scaffolding, known as Barnett's criteria, required in other vector based studies (33):

1. The suspected vector must be demonstrated to repeatedly contact the host
2. The vector must be found to be geographically associated with the infection
3. The pathogen must be demonstrated to be contained within or upon the vector under natural conditions

4. Under experimental conditions the pathogen must be demonstrated to be transmitted

M. sorbens is attracted to the watery surface of the eye mucosa as its food source and ocular secretions are known to contain *C. trachomatis* (34). In the laboratory environment *M. sorbens* was demonstrated to transmit *C. trachomatis* by having the fly feed on an infected egg yolk (35). Subsequently, *C. trachomatis* was detected on the proboscis and in the stomach of the flies, and using a guinea pig model, researchers were able to produce transmission from one animal to another (35). In order to demonstrate that the *M. sorbens* fly captured under natural conditions could harbor the pathogen, one study used sticky paper to trap flies from the faces of children (36). The authors found 15 of 103 captured flies tested positive for *C. trachomatis* by PCR, whereas none of the controls were positive, indicating that *M. sorbens* harbors the pathogen in natural conditions. A study by Emerson *et al.* provided evidence of frequent *M. sorbens* contacts with the eyes of young children (37). The results from this study showed that *M. sorbens* was responsible for over 90% of the fly-eye contact events with a median frequency of 3 fly-eye contacts every fifteen minutes. Two flies trapped during this experiment tested positive by PCR for *C. trachomatis* DNA. Additionally, this study found that children assessed to have an ocular or nasal secretion attracted more flies than their counterparts who did not present with these symptoms. The vector host relationship was further substantiated by a study that looked at the prevalence of trachoma in both the wet and dry seasons and three months following an insecticide spraying for fly control (38). This study found that fly control with the provision of latrines and insecticide spray reduced the number of contacts between flies and children's eyes as well as reducing trachoma prevalence. Insecticide spraying is not a viable solution due to the necessity for prolonged intervention, but latrine provision may have a more sustainable effect.

Climate and the Fly

The ability for M. sorbens to serve as a mechanical vector for trachoma depends on the environmental factors that influence the fly's ability to reproduce, the amount of flies that carry the infection, the number of fly contacts with the human face, and the ability to transmit infection. Flies maintain temperature homeostasis through the input of the external environment and are sensitive to environmental fluctuations. Temperature and moisture influence reproductive success and lifespan in many flies including *M. sorbens* (39-41). Some regions have climate conditions that are more hospitable for the fly and in these locations, an intervention to inhibit breeding capability or transmission potential of the fly may affect trachoma prevalence to a greater extent than in areas without the same climate suitability.

The Effect of WASH on Trachoma Prevalence

Over a number of studies, the lack of access to water, and lack of latrine facilities have been associated with trachoma (25). *M. sorbens* is found to have a high preference for breeding in human feces and it is suspected that removing human feces would help to disrupt the breeding environment for the vector(42). This study found that the *M. sorbens* emerging from human feces was larger than when emerging from other breeding mediums (i.e. cow, dog, and goat) and the authors postulated that the increased size led to a more robust life cycle. Flies of similar sizes to those that emerged from human feces were trapped around the eyes of children. While this study suggests that human feces on the soil surface does influence the population of *M. sorbens* the transmission of trachoma by fly-to-eye contact would still persist due to the ability of the fly to breed in other varieties of feces. A method described by Fuller *et al.* has been used to exam the community protection offered from sanitation interventions (43). These studies found that while sanitation access at the household level was a significant predictor of child growth the household's sanitation also contributed to herd protection in the whole community through an indirect effect. An analogous study could be evaluated to assess the community benefits of individual household water access and hygiene behaviors. Using a similar methodology, households with latrines will provide protection to neighboring households with the assumption that a good sanitary environment will protect the community from trachoma by reducing the favored breeding medium for the fly and thus reducing the fly population. Mechanistically more latrines lead to fewer flies which in turn reduces the number of flies that contact eyes infected with *C. trachomatis* and subsequently fewer infections result. Important behavioral and environmental risk factors for inflammatory trachoma were found to be poor facial cleanliness and household fly density in a cluster random survey in Tanzania (44).

Ocular and nasal discharge on children's faces has tested positive for bacterial DNA and it has been observed that these secretions draw more flies than children's faces without these secretions (37, 45, 46). These findings suggest that behavioral interventions targeting facial hygiene, in order to remove an environmental reservoir for disease, may exert an important influence on reducing transmission. To achieve facial hygiene a household requires access to water to perform simple face washing procedures that can remove these secretions. A study that evaluated the effect of water use found that households with trachoma were likely to allocate less water for face washing than households that were trachoma free (47). A household's distance to water has been evaluated to explain trachoma in a variety of studies which have frequently found that increasing distance to water is positively associated with trachoma (48-50).

Many studies have sought to investigate the associations between trachoma and environmental improvements in the water, sanitation, and hygiene sector in support of the F and E components of the SAFE strategy (51). The studies have found varying results of the effect of WASH on the odds of the prevalence of TF. A meta-analysis by Stocks *et al.* found that across all studies sanitation access, distance to water source, and observations of clean faces resulted in a decreased odds of trachoma (25). Some of the results are borderline significant indicating that a greater understanding of the effects of WASH interventions would benefit programs that are seeking to reduce the burden of disease. WASH is a critical component of the 'F &E' part of the SAFE strategy because trachoma is the result of unsanitary conditions especially where access to water is challenged. Increased education, sanitation, and hygiene have suggested efficacy against the disease in Nigeria where implementation of this strategy is cited as a major contributing factor to the reduction of the disease (52). Additionally, investments in WASH provide co-benefits for the protection from other diseases leading to improved overall health (53).

Spatial Analysis in Trachoma

The properties of spatial location drive human health mechanistically through exposures to the built and natural environments. The properties of space, composed of political, social, cultural, and historic influences, affect the distribution and determinants of disease (53). Spatial patterns in health outcomes are observed because similar people tend to aggregate and their shared group characteristics exert effects over and beyond individual characteristics. Tobler's Law is a fundamental principle of spatial epidemiology and is stated as, "Everything is related to everything else, but near things are more related than distant things" (54). The influence of spatial relationships is an important mechanism to consider in health studies where neighborhood properties can influence health outcomes. Most spatial analyses assume complete spatial randomness and then proceed with statistical tests to validate whether the observed patterns are more structured than would be expected by chance alone. This is an important consideration because if there is a spatial association between the observations, independence cannot be assumed thus violating one of the key tenants of generalized linear models. Trachoma has been observed to lack spatial randomness as studies have found clustering of the disease at multiple levels of analysis (45, 55) . Furthermore, intervention with antibiotics is expected to cause clustering in the disease and in many locations antibiotics have been applied for many years (56, 57). As such, controlling for spatial dependency is an important consideration when undertaking epidemiological studies in this domain.

Mapping health data has evolved into a vital tool to support the distribution of health goods and services by visualizing areas at greatest risk for a disease. Publications applying geographic information systems (GIS) to population health studies have surged since the early 2000's as computer processing power and GIS technologies advanced the ability of researchers to undertake complex analysis (58). There are numerous GIS based studies that have described the geographical properties of neglected tropical diseases (NTDs) although most extant literature is focused on schistosomiasis and malaria. The two primary categories of GIS and Health studies are disease mapping and disease modeling (58). Disease mapping allows for the determination of the distribution of disease and is most useful for hypothesis generation and evaluation of programmatic interventions. On the other hand, disease modeling is used to predict disease distribution, analyze patterns in geographic data, and determine the most effective locations for interventions. This review focuses on spatial measurements and statistics used by previous studies to describe and understand how geography affects the prevalence of trachoma. Maps are instrumental in the design and decision making process for targeting interventions against disease, and mapping NTDs provides estimates of the number of people infected which can be used to predict the prevalence of the disease in areas that are difficult to sample. The Global Atlas of Trachoma has coalesced the research efforts across multiple surveys and geographic regions to describe the distribution of trachoma at a fine scale and identify areas where further data is required (59). The efforts to describe location and extent of trachoma have been successful and are critical for visualization of geographic heterogeneities. Few studies to date have applied spatial analysis to predict the epidemiology of trachoma in terms of causality, transmission dynamics, or analysis of disease patterns.

Literature review methodology

A literature search was carried out with no date restrictions to assess the extent of previous studies which have applied spatial analysis techniques towards the understanding of the geographic properties of trachoma. Google Scholar, PubMed, and EBSCOhost databases were used to search for publications up to November 2016. Searches were performed using combinations of terms including trachoma, trachomatis, pattern, proximity, geographic information, geostatistics, spatial analysis, GIS, mapping, and Bayesian statistics. While multiple studies of descriptive disease mapping exist this review was focused on locating studies that involved the use of statistical inference to test the validity of observed patterns in the distribution of trachoma. Additional inclusion criteria for this review were if the outcome under measurement was trachoma and if the study used or reviewed geographic positioning systems to locate points of interest. No exclusion criteria were set for geographic location, age of participants, or journal in which the study was published.

Of the 14 selected publications 11 were cross sectional studies, two studies were literature reviews, and one was a case control study. All selected publications referenced the simplified WHO trachoma grading system to assess the clinical stage of trachoma. The studies were initially grouped into three primary topic areas. First, multiple studies show the location and extent of trachoma by descriptively mapping prevalence. Studies of this nature were not included in this review. The second category of studies are those that were focused on describing the distribution of trachoma with statistical evidence of a pattern or shape most often focusing on clustering. Finally, studies addressed spatial association to determine if the qualities of a geographic property could be statistically associated with the prevalence of trachoma. Within the studies of spatial association two key focal areas were identified, studies addressing environmental variables that drive trachoma prevalence and studies that describe an association by distance to a location. Environmental variables were considered as any of the climatic or physical features of a place that could be related to trachoma prevalence. A study was defined as distance to location study if it measured time or geographical distance between cases/controls and the physical location of a place. Relevant epidemiologic outcome information was extracted from each of these domains and tabulated to summarize the trends in the measures of association for different variables obtained in these studies and to provide an overall view of the previous work that has been completed in this domain.

Separately, a search of spatial methods that have been applied to other neglected tropical diseases was performed to enhance the understanding of techniques that have been applied to other diseases. The results found in these studies were used to generate hypotheses from analogous studies that utilized data similar to that held by The Amhara Regional Health Bureau and The Carter Center. Furthermore, the results from these studies were used to design the analytic framework for assessing a spatial hypothesis involving trachoma. These articles were left uncategorized. Malaria and Schistosomiasis were the most commonly researched diseases using GIS to investigate and describe spatial patterns. It was observed that studies frequently attempted to correlate environmental factors to help describe the distribution of disease.

Results

Eight of the studies reviewed incorporated at least one climate variable in their analysis although frequently some climate variables were dropped prior to finalizing a model. These studies are listed in a table in *appendix F*. Precipitation was the most commonly studied climate variable with five studies evaluating this covariate while temperature and altitude were examined in four studies. Precipitation was quantified differently in the five studies with average annual precipitation being found to be the most common. In all but one study, higher precipitation was found to be associated with a reduced odds of trachoma although only two studies incorporated rainfall as a significant predictor in their final multivariate models. Temperature was measured on a variety of temporal scales from average daily, monthly, to annual and was either measured as land surface temperature (LST) or air temperature. Two studies showed that increasing temperature resulted in a decreased odds ratio for trachoma (45, 60), one study showed that higher temperature increased the odds of trachoma (50), and one study showed increased odds when using residual LST and decreased odds when using mean annual temperature (4).

Altitude was included in one study's final model showing a significant trend for increased odds of trachoma with an increase in altitude (61). Of the three remaining studies, increasing altitude, in simple logistic models, was found to be associated with a reduced odds of trachoma in one model (50) while the other two models showed that higher elevation was associated with an increased odds of trachoma (4, 60). Variation in land cover using satellite imagery was evaluated in two studies which both suggested that savanna type land cover increased the odds of trachoma when compared to other land cover types (4, 62). Similarly, the Smith *et al.* 2015 study also investigated the vegetation cover of the study region and showed in a univariate analysis that decreasing vegetation cover, using the enhanced vegetation index (EVI), increased the odds of trachoma.

Finally, four studies evaluated additional covariates that could not be grouped to compare across studies. First, lower latitude was significantly associated with a decreased risk of trachoma in a multivariate model and that the effect of relative humidity was varied in a simple analysis (50). Additionally, Hagi *et al.* found the average monthly sunshine fraction decreased the odds of trachoma with increased sunshine duration in a Bayesian hierarchal model. In the analysis presented by Koukounari, a multivariate model found that air pressure was significantly associated with a decreased odds of trachoma. Finally, one study assessed urbanization as a covariate with a random effects model and found reduced odds when compared to rural classification (4).

Five studies, listed in the appendix, applied statistical methods to identify and compare the geographic proximity of clusters. Multiple statistical methods were used to analyze clustering patterns including the κ-function analysis, random effects logistic regression, Monte Carlo simulation, Kulldorf spatial scan statistic, and Bayesian hierarchal logistic models. All five studies assessed clustering at the household level and two studies investigated clustering at the bedroom level. All studies found significant clustering of cases at the household level and the studies that evaluated clustering at the bedroom level were also significant. One study looked at the difference between clustering at the level of child, caretaker, household, and village levels and found that significant clustering existed at all levels (45). Two studies evaluated the change in clustering 2 to 42 months following administration of MDA programs (56, 57). Both studies observed increased clustering following MDA and the distance between clusters varied with time. The reviewed studies did not use a homogenous disease indicator as two studies used PCR to diagnose infection after finding no significant clustering using clinical diagnosis (56, 57), two studies used clinical diagnosis by the WHO grading system (45, 55), and one study evaluated clustering using PCR diagnostic alone (63). Two studies found no evidence of spatial clustering (45, 55) while the remaining four found evidence of significant spatial clustering at a variety of distances up to a 283-meter radius encompassing 24 households.

Seven studies used distance to location as covariates when modeling the predictors of trachoma. Six of these studies used logistic regression while only the *Bailey* *et al. 1989* study used a Monte Carlo simulation to evaluate a spatial relationship. Two studies used time as a proxy for distance in survey questions to evaluate the relationship between distance to water source (63, 64). A statistically significant association between time to water source and trachoma was observed in the Polack *et al. 2005* study which found that the longest reported time to collect water, >30 minutes, was associated with an increase odds of trachoma (OR 4.9, $CI_{95\%}$ 1.1-21.7) although this result is borderline significant. Studies that measured geodesic distance from household to water source all found a significant increase in the odds of trachoma with increased distance to water source (4, 50, 61). Two studies reported null associations between trachoma and distance to latrine but neither study reported a measure of association or illustrated the results of their statistical tests (55, 63). The study by Montgomery *et al.* examined the role of household social isolation on trachoma using distance to bars/cafes, religious establishments, and government centers to evaluate the outcome. The houses that were more than 1400 meters from bars, cafes, and religious establishments were found to be at an increased odds of trachoma.

Discussion

The goal of this literature review was to increase the understanding of statistical methods that have been applied to assess the spatial distribution of trachoma. The current body of literature on spatial analyses in trachoma research is mostly focused on point patterns of infection, predictive modeling, and mapping of disease in relation to a target location. Differential results were observed across a variety of scales for both the geographic regions studied as well as the effects of the variables of interest. One example is the studies that found clustering at the household level but, when analyzed at large scale, was applied failed to produce meaningful associations of clustering between villages (45). Most studies included age and gender into their models as women and children are disproportionately affected by trachoma. These variables should be considered as important covariates in future studies. Multiple studies have attempted to understand the large-scale pattern of trachoma by considering one or more climate variables. The results of these studies are frequently mixed when assessing whether the environmental variables are important predictors of trachoma prevalence.

The results from this review indicate that a climate with a sustained higher temperature might exert a protective effect by drying feces faster restricting the breeding habitat for *Mucosa sorbens*. Overall, increased precipitation was shown to be negatively associated with the odds of trachoma. Few studies have explored the utility of GIS for quantifying land cover and no studies have looked for the associations between trachoma prevalence and environmental changes over time. Population centers in Africa and Asia are expected to see the biggest growth in urbanization over the next 30 years (65). This may provide a protective effect against trachoma if the water and sanitation infrastructure is developed effectively, but few studies have attempted to quantify this relationship. This review found one study reported a protective effect of urbanization, odds ratio 0.27 ($CI_{95\%}$ 0.13-0.52), when comparing urban to rural areas(4).

When comparing the physical location of trachoma cases in a region there is a frequently observed propensity for cases to cluster. Places in close proximity having greater similarity than areas farther apart and spatial autocorrelation should be accounted for in epidemiologic analysis. Two studies compared clustering when using PCR to diagnose infection and when using the WHO clinical grading system. Both studies found that when using PCR diagnostics, significant clustering was found whereas when evaluating clustering with the clinical grading scale, clustering was not significant (56, 57). A future inquiry comparing the effectiveness of diagnostics at different spatial scales when making comparisons of clustering could be addressed by future studies.

The most common distance-to-point of interest study was an examination of the distance of cases and non-cases to a water source. Access to water at the household level has frequently been studied because it is likely that a more accessible water supply improves hygiene and reduces the transmission of infection (66). Distance to water supply is a feasible proxy for assessing access and the results from the four studies that produced significant results showed that a greater distance to water was associated with an increased odds of trachoma. Social isolation could also be a risk factor for trachoma as two of the three household to point of interest models showed a significant association with the greatest distance to a point of interest and the odds of trachoma. It is possible that households which are more isolated from the rest of the community are less likely to be a part of community interventions or may have more difficulty in accessing community services. Only two studies considered the distance to latrine as a risk factor for trachoma and neither reported a significant association. Improving latrine access aides transmission interruption of trachoma (2) and as such there may be a significant geographical pattern between latrine access and trachoma prevalence which could be analyzed further.

Conclusion

Multiple opportunities exist to apply methods from spatial epidemiology to statistically assess the variation in disease seen throughout trachoma endemic regions. In particular, the assessment of the geographical variation in exposure and the corresponding relationships with disease can be modeled and visualized in a geographic information system to improve survey design and define the determinants of trachoma so that the most efficient method of applying interventions can be implemented. Three further studies were identified during the course of this review that were proposed for a thesis aimed at determining the underlying risk factors for the spatial heterogeneity of trachoma in Amhara, Ethiopia. First, there are a limited number of large scale spatial analysis studies on trachoma. As a result, a study analogous to *Clements et al. 2010* to develop a risk map for the Amhara Regional State, would allow the comparison of predicted geographic risk while incorporating a variety of climatic and social factors. Second, a large-scale cluster analysis of the geographic relationships between communities would help explain the factors most driving the observed regional clustering of trachoma. The aim of these studies is to provide evidence of a geographic pattern from observational data collected by The Amhara Regional Health Bureau and The Carter Center that will enhance the programmatic interventions to fight the disease.

Introduction

Caused by infection with the *Chlamydia trachomatis* bacterium, trachoma leads to blindness or visual impairment and affects millions of people around the world (3). Poor communities in low-income countries are at greatest risk for trachoma (67). The Amhara National Regional State of Ethiopia, a region of nearly 18 million individuals, has many districts hyperendemic for trachoma (68). Scale up of the SAFE strategy for trachoma began in Amhara in 2007, and eventually reached all districts within the region. SAFE stands for surgery for trichiasis, antibiotics to clear infection, facial cleanliness, and environmental improvements. Since the trachoma program began, the prevalence of TF in children aged 1 to 9 (TF_{1to9}) has decreased in Amhara regionally, and a number of woredas (districts) have achieved the elimination target of TF less than 5% in children in children aged 1 to 9 years. Despite these successes, trachoma persists with the prevalence of $TF_{1to9} \geq 10\%$ in many woredas despite eight to ten years of the SAFE strategy. This disparate slow return on resource intensive interventions warrants an enhanced understanding of the spatial distribution of the disease.

Prior studies presented evidence that trachoma spatially clusters within households and between communities (45, 63, 69). While rarely statistically confirmed, the spatial patterns of trachoma prevalence frequently appear regionally heterogeneous. The results from various population-based prevalence surveys (PBPS) used for the Global Trachoma Mapping Project (GTMP) have found TF_{1to9} prevalence estimates to vary within county political divisions in sub-Saharan Africa. Examples include the range of TF_{1to9} prevalence in Kaduna State, Nigeria 0.03% to 8% (70), Oromia State, Ethiopia 1.1% to 48.7% (71), Darfur and Khartoum, Sudan 0.0% to 19.8% (72), Niger State, Nigeria 0.0% to 11.7% (70), Tigray Region, Ethiopia 9.3% to 41.4% (73), Upper Nile and Eastern

Equatorial States, South Sudan 35.45% to 77.05% (74), Ghana 0.14 to 2.81% (75). These broad evaluation units present the picture of trachoma at a coarse scale and hide the conglomerate detail available at a finer scale of analysis. Some studies have used health clusters or villages to explain the prevalence of TF_{1to9}, but the use of these aggregated estimates for health promotion decisions is limited (76-79).

Children are the putative reservoir for trachoma and the nasal and ocular discharge observed on children's faces is a source of infection (44). Mechanistic transmission of *C. trachomatous* infection occurs by contact with infectious ocular fluid either directly through human-to-human contact (E.g. children playing together), fomites (E.g. A shawl used to wipe the eyes of an infected person and another), or indirectly by the vector *M. sorbens.* The F&E arms of the SAFE strategy focus on modification of the environmental and behavioral risk factors for trachoma. In particular, the use of latrines in a community helps to decrease the abundance of *M. sorbens*, improving water supply allows households to perform bathing more frequently, and routine face washing behavior removes a potential source of infection from the environment (30). Water, sanitation, and hygiene (WASH) interventions hold substantial potential to disrupt the chain of transmission in the Amhara Region of Ethiopia. Local access to WASH is dependent on geography and the intrinsic contextual influences of geoclimatic, cultural, and political factors inherently related with location. Regional variations in access to WASH are heterogeneous and the influence of large-scale access to these services remains unexplained.

The distribution of behaviors and environmental conditions that may influence the pattern of TF_{1to9} are complicated by phenomenon inherent with working with geographically dependent data. Spatial autocorrelation occurs when locations that are closer in space are more related than locations that are more distant. Frequently, epidemiologic modelling does not explicitly account for the dependence of neighboring observations, which leads to biased estimates. Locating clusters of high TF_{1to9} in a hyperendemic region is not feasible without the use of local spatial statistics. Spatial analysis provides confirmation of the hypothesis that the observed patterns are meaningfully different from what would have occurred under complete spatial randomness. Identification of geographic areas, where the pattern of the disease is found to significantly cluster, can be a tool for programs to identify the areas in greatest need of intervention. Furthermore, regression techniques used to model the relative importance of multiple predictors can be used to explain the mechanisms by which these clusters are formed.

This study aims to better understand the geographic distribution of trachoma in Amhara, to explore the most significant WASH factors that explain the clustering of TF_{1to9} prevalence, and to develop a mathematical model that accounts for the influence of village and district level neighborhoods. In order to extricate this relationship, a variety of social, behavioral, and climate factors were identified as important confounders. Insight into the drivers of the spatial pattern of the disease in Amhara will help to target interventions to have the greatest impact.

Methods

Dataset

The dataset used in this study was derived from a series of cluster randomized surveys, which were performed to assess the impact of SAFE interventions on the prevalence of trachoma throughout the Amhara Region. Trachoma impact surveys (TIS) have been implemented as a tool for monitoring and evaluation in order to meet the WHO guidelines of regular program assessment following implementation of SAFE in trachoma endemic regions. The data collected from The Carter Center supported TIS evaluated the impact of 3 to 5 years of SAFE implementation and identified factors that contribute to the community prevalence of the disease. Baseline surveys were initiated in all ten zones with the ultimate goal of assessing all woredas in Amhara (80). The number of annual rounds of MDA required before the next TIS was determined by disease prevalence estimates from baseline and monitoring surveys. The geographic distribution of surveys described by *Figure 1* highlights the date of the most recent TIS survey for each woreda within the Amhara.

Survey Timing

Current WHO guidelines for trachoma monitoring provide a series of recommendations for conducting surveys dependent on disease prevalence estimates from repeat cluster random surveys (23). In areas where trachoma is suspected, but no previous interventions were initiated, a baseline evaluation at the woreda level or higher is applied using a cluster random sample survey. A TIS through cluster random sampling at the woreda level is required after at least five years of implementation of the SAFE strategy where the prevalence of TF_{1to9} is ≥ 30%. In woredas where the TF prevalence among children aged 1 to 9 is between 10 and 29.9%, the TIS is repeated after three years of antibiotic distribution and continued support for facial hygiene and environmental improvements. When the prevalence is found to be between 5% and 9.9% an additional year of woreda wide antibiotic distribution followed by a repeated impact survey is recommended. Prevalence of TF in children 1 to 9 years old <5% suggests suspension of antibiotic distribution although facial hygiene and environmental improvements are recommended to persist along with pre-elimination surveillance (23). As a result of the response of TF prevalence to the SAFE strategy, availability of the drug, and other social/political factors, some areas in Amhara have received more rounds of MDA or more frequent surveys than others. Cluster-random TIS have been conducted in Amhara to monitor impact since the beginning of the program in 2001. Surveys are generally conducted twice per year and have generally employed the same sampling and data collection methodologies. Survey data from 2011 through 2016 used for this analysis incorporated only the most recent survey data from each woreda in order to observe the spatial patterns of disease in Amhara, an area served by a mature trachoma control program. .

Sample Size

Cluster randomized surveys were powered to calculate woreda level prevalence estimates for TF_{1to9}. The sampling methods used in the TIS presented here were detailed in a publication by *King et al.* and a publication currently in press (68, 81). From 2011 to 2014, woreda level sample size was proportional to relative size of the woreda based on the most recent population estimates. The sample size was calculated to estimate a 20% decrease in TF prevalence. The number of clusters ranged from 99 (Survey 8) to 360

(Survey 1) in woredas surveyed during these years. Since 2014, two approaches to sample size calculations were used in Amhara based on TF prevalence estimated from the earlier survey. For woredas where the prevalence of TF was ≥ 10%, it was determined that in order to estimate a prevalence of 20% ± 10% with a design effect of 3.04 (calculated from the results of earlier surveys) and a non-response rate of 15%, a sample size of 215 children ages one to nine years was required. Children in this age range make up 27% of the population of Amhara. There are 4.1 persons per household; therefore, a total number of 194 households were required to be surveyed. In order to achieve a representative sample for the woreda, 30 households from 8 clusters were randomly selected. For those woredas with a <10% prevalence of TF_{1to9}, and assuming a 3% prevalence ± 2%, a design effect of 3.04 and a non-response rate of 15%, the sample size required was 977 children aged 1 to 9 years. To reach the sample size 16 clusters of 60 households were sampled. Woreda level data from all years are population-based estimates, with the newest surveys having increased precision in general. Trachoma outcomes were weighted at the cluster level by the inverse of the probability of selection.

Sample selection

In order to provide a representative sample for each woreda a two stage cluster random sample technique was applied. The gott (village) was the primary sampling unit (cluster) because it was the smallest administrative unit for which there was population data available. In the second stage of sampling, households were selected using a modified segmentation approach (Compact segment sampling) described by *Turner et al.* that has become a standard practice in cluster random probability sampling for trachoma (19, 82).

Selection of clusters

All gotts within a woreda, their geographical position, and the populations were listed in geographical order. A sampling interval was calculated by dividing the total population of the woreda by the number of gotts to be selected based on the survey design described above. The first cluster was randomly selected using row randomization in Microsoft Excel and each additional cluster was selected by adding the sampling interval to the previous cluster.

Selection of households

Within each gott a method of selecting households was implemented that subdivided the cluster into segments called development teams. Development teams were segmented into units approximately the size of 30 households. Subsequently, one or two segments were randomly selected by the village leader based on the desired sample size. All individuals within all households within each segment were then surveyed. A verbal consent was performed with the household head and all members of the household who were eligible and provided consent were surveyed for clinical signs of trachoma.

Trachoma grading

The WHO simplified trachoma grading system is a long established examination technique to quantify the scale of the disease and support interventions in endemic regions (83). This grading system allows for non-specialists to be trained to provide programmatic assistance in population-based prevalence surveys that generate assessments of the disease. The assessment is focused on identifying distinct clinical features associated with different stages of the disease. The upper tarsal conjunctiva (eyelid) is flipped open and the graders examine the eyelid for signs of TF, trachomatous inflammation - intense (TI), tarsal plate scaring (TS), trichiasis (TT), or corneal opacity (CO). Active trachoma is synonymous with TF and is most prevalent in children although all members of the household were examined in the TIS surveys utilized in this study.

Training of Examiners

Experienced examiners were selected for assessing study subjects and grading trachoma signs. The comparability of clinical examination results over time was stabilized by conducting training prior to each survey round. Before entering the field, examiners were required to undergo training over the course of a week to review the survey methodology, to practice grading in the field, and to confirm competence for trachoma diagnosis. Following the training, the trainees were required to grade a series of 50 images from a standardized image set that comprises all grades of trachoma. Only those with comparable grades to the 'gold standard' grade for the slide set moved on to the field reliability exam. Subsequently, the trainees evaluated 50 eyes from children in a non-study village, and the consensus grade of three examiners was used as the gold standard. In order to be considered competent for field diagnosis the trainees had to achieve an inter-observer agreement of at least 84% agreement on the clinical sign of TF and a kappa statistic of 0.7 and above.

Training of Recorders

Recorders are selected from a pool of qualified university students or recent graduates who assist the field team by obtaining verbal consent, performing interviews, and collecting electronic survey data. Recorders undergo a week-long training on survey methodology, consent procedures, data collection, standardization of interviews, and translation. Prior to launching the survey, the recorders undergo a field test to demonstrate that the sampling methods and tools are standardized.

Household Questionnaire

This survey tool was used to quantify SAFE uptake and implementation at the household and individual levels. The survey questions were refined over the course of multiple trachoma impact surveys, but most questions only varied slightly and could be related across surveys. The most recent survey questionnaire is included in the appendix. The head women of the household was interviewed in Amharic and responses to questions were recorded in the questionnaire using an electronic tablet with custom built software (81). The household survey assessed many water, sanitation, and hygiene indicators. Water availability was assessed with a question about how long it took the household to fetch water. Questionnaire choices included less than 30 minutes, 30 to 60 minutes, and greater than 60 minutes. This variable was dichotomized to less than 30 minutes and ≥30 minutes. Additionally, a question was asked about whether the household was able to access an improved water source. Improved water source was defined according to the WHO's Joint Monitoring Program definition as a water source that is either piped water into dwelling, piped water to yard/plot, public tap or standpipe, tube well or borehole, protected dug well, protected spring, rainwater (84). Household access to a latrine was directly observed by the field team and indicated on the survey as presence or absence of a latrine facility. Data on the use of latrines was not included in all surveys and was excluded from the present analysis.

Sociodemographic variables

The average number of household wealth indicators reported owned by the respondents during the household survey were calculated for each woreda and village as per the method described for Amhara in *Oswald et al* (32). Households reported a maximum of five wealth indicators from a list including a radio, television, electricity, mobile phone, or an iron roof on the dwelling and a proportion of these five indicators was calculated for each household. Household education has also been studied as a risk factor for trachoma (85, 86). To derive a dichotomous variable that represents the highest education any adult in the household has attended the seven educational groups from the survey were collapsed into any education or no education for 151 of the 152 woredas. The most recent TIS from the woreda, Andabet, was an old survey format that asked for the respondent's education rather than the highest educational level completed by any household member. For this woreda only, the respondent's educational level was used to represent the highest household education. Crowding had previously been defined as greater than 5 persons per household regardless of household size and this definition was retained for the present analysis (71). The number of rounds of azithromycin distributed to each woreda as recorded in the International Trachoma Initiative's Trachoma Expert Committee's meeting records was applied to each woreda and each gott within the woreda (87). Survey variables about the presence of a health care facility within a gott and presence of a paved road within the gott were also considered as variables to control for in the analysis and were subset into the final dataset.

Household Census

The objective of the household census was to estimate the prevalence of all clinical signs of trachoma through conjunctival examination. All consenting participants residing in study households had both eyes examined for the presence or absence for clinical signs of trachoma using the simplified grading system. After everting each eyelid, examination was performed using sterile technique and visual inspection was enhanced with a 2.5x binocular loop and light. Examiners sterilized hands with alcohol between each eye. A recorder then entered the diagnosis into the survey form. During the examination, the recorder obtained demographic information (age and sex). Facial cleanliness was directly observed by field teams. A clean face was defined as absence of nasal or ocular discharge, whereas dirty faces were defined as having any of these secretions. At the conclusion of the visit, individuals found to have active trachoma were offered antibiotic treatment (either 1% tetracycline eye ointment or a single dose of oral azithromycin; 20mg/kg) while those presenting with trichiasis were referred for surgery at the nearest health facility.

Descriptive Mapping

The Carter Center has used maps to visualize the impact of interventions on the prevalence of trachoma for many years. Additionally, the Global Trachoma Mapping Project developed an atlas of the global distribution of the disease (88). In order to understand the extent of the disease and the distribution of SAFE measures throughout the region, maps of TF_{1to9} and the hypothesized predictors were generated using weighted estimates from the most recent TIS. A total of 152 woredas have been surveyed by TIS and make up the entire region of Amhara. Household level covariates were weighted to represent the aggregate proportion of the factors at the woreda level and gott levels using STATA version 14.0 (STATA Corporation, College Station TX, USA).

Spatial autocorrelation

To assess the distance of spatial influence an incremental spatial autocorrelation assessment was performed at the woreda and gott levels (89). For each woreda the geographic centroid of the area was used whereas the gotts were geographically positioned by the average latitude and longitude from the surveyed households in the evaluated health cluster. A global Moran's I was run at increasing distances in order to understand the scale of the intensity of clustering in ArcMap version 10.4.1 (ESRI, Redlands, CA, USA). To determine the starting distance, the smallest distance that provided each feature with at least one neighbor was identified. The z-scores from the Moran's I were plotted against distance to illustrate any peaks in the intensity of clustering. When multiple peaks were identified the distance that reflected a distinct neighborhood was preferred although cluster analysis was examined at multiple distance scales. For each distance a spatial weights matrix was calculated for use with hotspot analysis and geographically weighted analysis discussed below.

For gotts, the neighborhood relationship was conceptualized with a fixed distance band so that neighborhoods would have a consistent scale of analysis throughout the region that accounted for a peak in spatial autocorrelation. To define the neighborhood for the woredas, connectivity maps were evaluated for a fixed distance band of 110 to 200 km and 8-12 nearest neighbor relationships in GeoDa version 1.8 (Luc Anselin, Chicago, IL, USA). The eight nearest neighbor relationship was identified as the best conceptualization of the woreda neighborhood because the large woredas in the west of Amhara had enough neighbors for the statistical tests yet did not have too large a distance as to introduce bias by including too many neighbors. A spatial weights matrix was exported for standardized usage among all other spatial statistics.

Mapping Climate and Physical Environment Variables

Raster surfaces for average annual temperature, average annual precipitation, and altitude were obtained from BioClim Global Climate Datasets (Available at: http://worldclim.org/bioclim), in a method that has become common practice for estimating local climate conditions (72). To summarize these variables for each woreda, the raster was converted to point data containing 358,160 points within the study area for each variable using ArcMap. The points were then projected to the coordinate system of the data frame and the identity tool assigned woreda ID's to each point from the raster. The points were clipped to the study area. Subsequently, summary statistics were calculated for each woreda including values for minimum, maximum, range, mean, standard deviation, and the number of points within each area. The calculated summaries were then table joined back to each woreda and an Excel file was exported for analysis in SAS version 9.4 (SAS Institute Inc., Cary, NC, USA). The geographic point location for each gott served as the location for which temperature, precipitation, and altitude values were extracted.

Linear Regression Model Selection

A forward regression model selection procedure was used to find a multiple linear regression model that fit the hypothesized predictors with the outcome TF_{1to9} prevalence in SAS. Initially, household access to sanitation, children presenting for examination with a clean face, household access to a water source within 30 minutes, and household access to an improved water source, climate, and economic status variables were tested for inclusion in the model. The primary selection criteria included the explanatory variable with the highest Pearson correlation coefficient with the outcome. Subsequently, additional models were run testing to see which covariate had the most significant partial t-test to include in the next stage of the modeling process. When no more variables were found to significantly improve the model a final model was indicated. Following the selection process that considered all possible variables a parallel model was developed that forced in water sanitation and hygiene indicators and then considered the stepwise inclusion of economic and climate variables using the most significant p-value for inclusion (90).

Spatial Regression

Using the best fit linear model created through the process above, geographically weighted regression, spatial error, and spatial lag models were compared to select a best linear model for predicting the association between WASH factors and trachoma at the woreda scale that controlled for autocorrelation. Initially, the dataset was loaded into ArcMap to perform geographically weighted regression (GWR) to control for spatial autocorrelation of the residuals (91). The residuals from the aspatial ordinary least squares (OLS) model were mapped to the corresponding woredas and then a Global Moran's I test was performed to assess the presence of spatial autocorrelation (92). In order to control for autocorrelation a variety of spatial regression models were applied to account for the lack of independence in the model and to generate local regression coefficients from the model.

Originally, a model was run that included the WASH factors and the mean household wealth score, but the results from the GWR indicated that additional factors needed to be accounted for to increase the predictive ability of the model. The model selection process was repeated in SAS. With the inclusion of climate variables for average annual precipitation, average annual temperature, altitude, community-reported presence of paved roads, and community-reported presence of a health facility. The model selection procedure indicated that multicollinearity was present between altitude and temperature, which was expected as temperature should be lower at higher altitudes. Temperature was selected to alleviate the model design issues.

The selected model was evaluated by numerous criteria. First, model residuals were required to be randomly distributed geographically. The residuals from the GWR were plotted and a Global Moran's I was once again performed to illustrate the reduction of spatial dependency of the residuals. The geographically weighted regression model allowed for the visualization of the regional variation in the model predictors of TF_{1to9} prevalence. The coefficients from the model were assessed for statistical significance so that each selected variable was determined to be an important predictor of TF_{1to9}. The relationships between the variables were assessed to ensure that the relationship between the predictor and the outcome could be explained in a defensible way. Multicollinearity was assessed from the GWR output in ArcMap to ensure that each variable was independent of all other predictors. Additionally, the model required normally distributed residuals so that high and low values performed equally well. Testing for residual normality was performed with the Jarque-Bera test (93). The coefficients from the model, the model r^2, and the condition index were mapped to the corresponding woredas to visualize the models for each woreda. Finally, the condition index values were observed for each woreda.

Additional spatial models were designed to control for the effects of autocorrelation using a method adapted from spatial econometrics (94). Models designed to control for spatial error and spatial lag were constructed and run in SpaceStat Version 4.0.21 (BioMedware, Anne Arbor, MI, USA) using the same predictors defined in the best fit OLS regression model. These models used the 8 nearest neighbor spatial weight matrix to define each neighborhood. Model fit was assessed using r^2 values, the AIC, and the regression coefficients for each predictor. The Breusch-Pagan test was used to assess whether any model was able to control for spatial autocorrelation (95). The output from the SpaceStat models was exported and then table joined to the original shapefiles in ArcMap.

Hot Spot Identification

To determine the risk factors associated with hot spots in Amhara, local spatial statistics were used to identify high prevalence clusters, or "hot spots" and regression analysis was applied to reveal the significant predictors among the 152 woredas and 1,558 gotts included in the analysis in ArcMap. Using the weighted prevalence estimates the Getis-Ord Gi* statistic was calculated to evaluate local clustering of high and low values (89). The spatial weights matrix for a 25km fixed distance band was applied for the gotts and the weights matrix using 8 nearest neighbors was applied for the woredas. The false discovery rate (FDR) accounted for multiple testing and reduced the number of false positives returned by the Getis-Ord G_i* (d) statistic (89). Hotspots identified with 95% confidence for gotts and 90% confidence for woredas were coded with a new dichotomous variable. Finally, the attribute table was exported to Microsoft Excel and then uploaded for logistic regression analysis in SAS.

Hot Spot Regression Analysis

Two regression analysis were performed to evaluate the drivers of clustering at both the woreda and the gott scales of analysis. Identical model selection guidelines were used. The values of the covariates and the outcome corresponded to either the gott or the woreda aggregate unit of analysis. The outcome for logistic regression analysis was residence in a hotspot woreda and residence in a hotspot gott. In order to assess the factors that are important predictors of residing in a hotspot a stepwise regression modeling strategy was applied. Univariate and multivariate regression analysis were performed to assess the potential factors contributing to residence in a hotspot at both levels. Variables were considered significant if the p-value was less than a 5% level of significance. All primary exposure variables were applied to the model as predictors prior to the assessment of confounding. Collinearity was assessed prior to model selection using a condition index greater than 30 and a variable decomposition factor greater than 0.5 as the thresholds for collinearity (96). A backwards stepwise selection process was then preformed to assess how different combinations of confounding variables affected the estimates of WASH variables and the model fit. Model fit was assessed with the Akaike information criterion (AIC) where the best fit model was selected as having the lowest AIC. Once all combinations of confounders were assessed the model fit was again examined with different combinations of the exposure variables in order to find a best overall model. The final models were assessed for discriminatory performance using an ROC curve, and the model's fit to the data was confirmed using the Hosmer-Lemeshow test. The regression coefficients from the best fit model were used to generate a predictive map using the point estimates that had been converted to rasters to determine the probability of residence in a high TF_{1to9} cluster in a method similar to Vazquez Prokopec *et al* (97).

Results

Study area and population

Ethiopia is located on the horn of Africa and Amhara is a National Regional State that borders Sudan. Amhara is subdivided into 10 zones, 152 administrative units, and is home to approximately 18 million persons. A map of Amhara and the administrative boundaries is included in the appendix. The woredas that border Sudan are the warmest regions in Amhara while most of the woredas to the northeast of Lake Tana are some of the driest, *Figure 2*. When observing the gott estimates for average annual precipitation the mountains create a rain shadow on the northern and the eastern border. The temperature estimates follow trends in altitude, and this relationship is most noticeable in the highlands with low average annual temperatures especially in East Gojam and the South Wollo woredas. Surveys conducted from 2011 through 2016 reached 1,558 gotts in all 152 administrative units of Amhara enumerating 282,400 individuals of whom 202,312 (71.6%) were examined for clinical signs of trachoma. Among all individuals enumerated, 75,144 were children between the ages of 1 and 9 years and 69,236 (92.14%) of these children were examined for clinical signs of trachoma.

Overall prevalence and geographic distribution

The woreda prevalence estimates of TF_{1to9} ranged from 1.2% in Dangela Town ($CI_{95\%}$ 0.49, 2.29) to 73.9% in Abergele ($CI_{95\%}$ 49.65, 89.06). The distribution of TF_{1to9} exhibits slight right skewedness but was not a grossly non-normally distributed variable at the woreda level of analysis, whereas the distribution of TF_{1to9} was significantly right skewed when examining the variable at the gott level, *Figure* 3. The woredas with large landmass in the northwestern corner of Amhara and the smallest woredas, that are often unified townships, frequently exhibited the lowest prevalence of TF_{1to9}, *Figure 4,* Figure 5. The woredas with prevalence in the highest two quintiles tended to aggregate in a north to south orientation to the east of Lake Tana, *Figure 5*. This area featured the most rugose landscapes with a distribution of geoclimatic variables that are influenced by the mountainous characteristics of the region, *Figure 5*.

Risk factors and their geographic distribution

The weighted percent of children presenting to the examiners without nasal and ocular discharge (clean face) was 86.4% in the region as a whole, and ranged from 36.1% ($CI_{95\%}$ 32.0, 40.545) in Sekota Town to 99.1% ($CI_{95\%}$ 95.8. 99.8) in Dangela Town. The prevalence of clean face was highest to the West and South of Lake Tana. Household access to a latrine ranged from 12.1% ($CI_{95\%}$ 3.4-35.0) in Wogera to 94.0% ($CI_{95\%}$ 89.0, 96.8) in Banja. Additionally, some of the woredas surveyed had no households with access to an improved water source (Abergele, Menz Keya, and Sehale Seyemt) whereas household access to an improved water source in Dessie Town was 100%. Sehale Seyemt was found to have the lowest percentage of households with access to water in less than 30 minutes (1.8% $CI_{95\%}$ 0.2, 13.8) and water access was within 30 minutes for all residents within Kobo Town and West Armachiho. A hotspot analysis confirmed that the central corridor woredas had higher proportions of unclean faces in relation to the global average, *Figure 7*. The woredas within the central corridor of the region had received greater numbers of rounds of antibiotics, *Figure 9*. These areas are where the program started distributing antibiotics and also the woreda prevalence of TF_{1to9} had been persistently high enough to warrant sustained treatment. The woredas to the northeast of Lake Tana had the most statistically significant clusters of low percentage access to water, latrines, low wealth, and education, *figures 10-13.* Descriptive mapping of the hypothesized predictors at the gott level shows more granularity and outlying observations when comparing neighboring gotts using the Getis-Ord G_i*(d) statistical test for clustering, *figures* Figure 14 - Figure 19. The minimum distance to give each woreda at least one neighbor was 78km while the minimum one neighbor distance for each gott was 25km.

Linear Regression Analysis

Using simple linear regression to explain the TF_{1to9} prevalence in all 152 woredas, clean face was the most explanatory variable with an r^2 value of 0.50 and a -1.03 percent change in woreda-level prevalence of TF_{1to9} for a one percent increase in the woreda prevalence of clean faces in children, *Table 1*. Household wealth (r^2=0.33) and number of rounds of MDA distributed to the woreda (r^2=0.20) were also strong linear predictors of TF_{1to9} prevalence at the woreda level. The remaining WASH factors; household access to a latrine (r^2=0.12), household access to any water source (r^2=0.13), and household access to an improved water source (r^2=0.14) also showed univariate linear relationships with the outcome. Gotts with a paved road (r^2=0.13) and average annual precipitation (r^2=0.15) were independent predictors of TF_{1to9}. Average annual temperature, average altitude, highest household education, household crowding, and presence of a health facility in the gott did not exhibit linear associations with the outcome.

The best ordinary least squares (OLS) regression model identified with a manual forward stepwise linear regression model selection procedure identified six explanatory variables. These variables accounted for 64% of the variation in TF_{1to9} with clean face, number of previous rounds of MDA, and paved roads as the most significant model predictors in the multivariate model *Table 2*. The residuals from the model were found to be normally distributed and free from heteroscedasticity, but the OLS model did not adequately control for the spatial autocorrelation of the residuals. The Global Moran's I 3.56 (p-value <0.01) indicated spatial dependency in the residuals and these over and under predictions from the model were most significant at a distance of approximately 150km, *Figure* 20. The peaks in the figure are the distances where spatial autocorrelation is most influential. The presence of significant autocorrelation drove the continued spatial modeling procedures by exploration in geographically weighted regression followed by spatial error and spatial lag modeling procedures to account for the lack of independence in the model.

The geographically weighted regression procedure allowed for each woreda to use its nearest spatial neighbors and calculate a multivariate linear regression for each feature. The optimized GWR for model 1 improved the adjusted Pearson correlation coefficient from the aspatial model ($r^2 = 0.70$) and improved the model fit (AIC =104.07). Household access to water within 30 minutes and access to an improved water source were not statistically significant in the global OLS model or the GWR model but this may have been due to the regional differences in positive and negative associations between predictors and the outcome. The maps of the woreda level coefficients for the statistically significant model predictors, clean face, rounds of MDA distributed, and paved roads in the woreda show strong variation. The GWR model showed statistically significant negative linear associations with the predictors clean face (Coefficient range -0.58 to -0.87) and gotts with paved roads (Coefficient range -0.66 to -3.06) while the number of rounds of MDA showed significant positive associations with TF (Coefficient range 1.56-2.71). The regional variation in clean face exhibited the strongest negative linear associations with TF_{1to9} trending northeast across the region, *Figure 21*. The geographic relationship identified with the GWR model indicated that clean face has a stronger linear relationship in the woredas identified as having significantly high or low prevalence of TF_{1to9}. This model is effective at breaking the spatial dependency of the residuals (Moran's I 0.053 p-value 0.13) but the tradeoff comes at the expense of local multicollinearity in the estimates especially in the western terminus of the region. 92.1% of woredas exhibited a condition number greater than 30 indicating that this model suffered from severe multicollinearity design issues. This diagnostic served as the motivation to test other spatial modelling techniques by showing that there was spatial heterogeneity in the explanatory variables.

The analysis workflow evolved after finding that the GWR model was insufficient, and prompted the consideration of both spatial error and spatial lag models to control for spatial dependency of the model residuals. Spatial dependence in variables and error terms was assessed using two hypothesized neighborhood relationships that exhibited varied efficacy at improving the model fit, *Table 9*. Spatial dependence was controlled in all four iterations of the model. The model designed to control for spatial error accounted for spatial autocorrelation but reduced the amount of variation explained by the model to less than the OLS model. The spatial error model with a fixed neighborhood distance of 110km reduced the Pearson's correlation coefficient (r^2) to 0.55 while the neighborhood defined with 8 nearest neighbors decreased the model r^2 to 0.49. Both spatial lag models improved model fit by reducing the Akaike Information Criterion (AIC) from 1133.39 in the OLS model to 1119.86 in the 110km fixed distance neighborhood and 1111.78 in the 8-nearest neighbor model. Additionally, the models designed to control for spatial lag improved the r^2 over the OLS model from 0.64 to 0.68

using a fixed distance and 0.69 using the 8 nearest neighbor's relationship. The overall model fit was best when using the 8-nearest neighbor spatial relationship and was considered to be the best localized conceptualization of woreda level spatial influence. As such, the spatial lag model was identified as the best model to describe the linear relationship between TF_{1to9} and WASH related factors while controlling for other influential variables. *Figure 22*, shows how spatial autocorrelation in the final model was influenced by increasing distance. The first peak is not statistically significant whereas, the second peak is statistically significant but occurs at a distance band of 261km. If the model were to account for autocorrelation at this distance it would require a neighborhood size that would span the entire regional state and is thus not a meaningful way to conceptualize the spatial relationship. As such the decision was validated that the model with spatial lag of the dependent variable would control for autocorrelation and provide a valid conceptualization of the relationship between WASH and TF_{1to9}.

Gott Hotspot Analysis

The prevalence of TF_{1to9} was assessed in a hotspot analysis to determine which of the 1,558 gotts had significantly higher prevalence in relation to neighboring gotts. The cluster analysis performed using the Getis-Ord G_i*(d) statistical test and a fixed neighborhood distance band of 25km identified 325 gotts (20.9%) that had a statistically significantly higher local prevalence of TF_{1to9} than was expected by chance ($P < 0.05$). This analysis also identified 371 (23.8%) cold spots or locations where there was a significantly lower local prevalence than would have been expected to occur by chance. The geographic distribution of hotspots occurs exclusively to the east of Lake Tana. The gott level hot spot analysis provides additional granularity by looking at the finer detail of the distribution of highs and lows and identifies gotts in the southern boundary of Amhara along the border with the regional state of Oromiya, *Figure 26*. This detail was likely obscured when the gott estimates were aggregated to the woreda level and is a good example of the modifiable areal unit problem (MAUP) (98). Gotts with low percentage of clean face were again identified to the east of Lake Tana, however some of the gotts in the southern region of Amhara and in the Oromiya zone, were identified as significant cold spots of clean face. The hotspot gotts for high percent of household latrine access appeared to overlap with the cold TF_{1to9} gotts to the south of Lake Tana, *Figure 16*. On the other hand, the gotts in the North Wollo zone had high household latrine access that overlapped with high prevalence of TF_{1to9}. The gotts surrounding Lake Tana formed some of the most noticeable clusters of high average number of household wealth indicators in Amhara, *Figure 19*, while the regions to the east of Lake Tana along the border with Tigray were found to have lower numbers of household wealth indicators than would have occurred if the spatial distribution was completely random. Descriptive mapping provides some insight into the where the hypothesized features tend to aggregate with the outcome but in order to validate these associations logistic regression was employed.

Gott Hotspot Logistic Regression

Gotts identified as hotspots (n=325) served as the outcome for a series of logistic regression equations. Combinations of significant explanatory variables at this scale of analysis were explored in various models. Univariate odds ratios between the hypothesized predictors and the outcome are outlined in *table 1*.

The significant univariate predictors of being a hotspot gotts were clean face (OR 0.97; $CI_{95\%}$ 0.97, 0.977), household access to a latrine (OR 0.990; $CI_{95\%}$ 0.986, 0.994), household access to a water source within 30 minutes (OR 0.995; $CI_{95\%}$ 0.992, 0.998), number of household wealth indicators owned (OR 0.442; $CI_{95\%}$ 0.375, 0.521), presence of a health facility in the gott (OR 0.639; $CI_{95\%}$ 0.446, 0.876), presence of a paved road within the gott (OR 0.601; $CI_{95\%}$ 0.460, 0.785), the number of previous rounds of MDA distributed to the woreda (OR 1.780; $CI_{95\%}$ 1.615, 1.962), and average annual precipitation (OR 0.997 $CI_{95\%}$ 0.996, 0.997). Multivariate logistic regression analysis determined how the different combinations of variables would affect the estimates and fit of the model. The explanatory variable, clean face, was an extremely robust predictor across all the models assessed never exhibiting a p-value > 0.01. The distribution of percent of clean face is highly left skewed at the gott, and the median for the variable clean face is lower for the hotspot gotts than for all other gotts, *Figure 28*. Additionally, whether the household reported any education, previous rounds of antibiotics distributed to the woreda, average annual precipitation, and average annual temperature were identified as significant explanatory variables across multiple iterations of the model selection process. Thus, these variables were important confounders to control for when addressing the relationship between exposure to WASH related factors and the relationship with membership in a hotspot gott. The AIC was used to assess the quality of the models in relation to the other models tested. The selected model had the best quality (AIC=1185.921), exhibited good discriminatory performance from the receiver operator curve (ROC) (AUC=0.83; *Figure 27*), and showed no evidence of a lack of fit to the data (Hosmer-Lemeshow Test $\chi^2 = 10.49$; p-value 0.23). The odds ratio of the association between the only remaining significant WASH exposure, clean face, and the odds of residence in a hotspot gott exhibited a negative relationship. For a one percent increase in clean face at the gott level the odds of a gott being a hotspot decreased by 1.9% controlling for household wealth, household education, antibiotic distribution, average annual precipitation, and average annual temperature.

Woreda Hotspot Analysis

Following the assessment of significant predictors of gott hotspots a similar analysis was adjusted to explore the factors that determine hotspot affiliation at the woreda level. The Getis-Ord G*i**(d) statistical test allowed for the analysis of the spatial clustering of woredas with high and low values of TF_{1to9} and other predictors in relation to each feature's nearest neighboring eight woredas. In order to be identified as a hotspot the local sum of each woredas and its neighbors was compared to the expected sum given the global sum of all features. Twelve woredas were identified as statistically significant hotspots of TF_{1to9} prevalence with ≥90% confidence while 11 woredas were identified as statistically significant cold spots with the same level of confidence. The woredas with significantly low prevalence in neighborhoods of low prevalence were predominately located along the western border with Sudan and directly southeast of Lake Tana, *Figure 23*. Furthermore, the highest concentration of statistically significant hotspots of high TF_{1to9} prevalence were located to the east and north of Lake Tana, along the border with Tigray. One hot spot woreda, Mida Woremo, was located in the south of the region in the North Shewa zone. Overall, 7.9% of the woredas in Amhara were identified as hotspots and these woredas served as the outcome for the logistic regression analysis.

Woreda Hotspot Logistic Regression

Univariate logistic regression summarized in *table 4* showed that statistically significant univariate relationships are expressed between TF_{1to9} and the individual predictor clean face as well as the household predictors access to water in less than 30 minutes, access to an improved water source and mean number of household items owned. The percent presence of paved roads in gotts and the average annual precipitation in the woreda both had significant negative linear association with TF_{1to9}. All of the hypothesized predictors were input into a multivariate regression model and a multiple backwards model selection criterion identified three candidate models. Clean face was identified to be a statistically significant model predictor in every model that was evaluated and *Figure 25* highlights that the percent of children presenting with clean face is lower among hotspots than among all other woredas. The data driven model was best fit with the predictors clean face, water within 30 minutes of the residence, household access to an improved water source and presence of a health facility in the gott (AIC =48.867), *Table 5*. This model had good fit (Hosmer-Lemeshow Test = 2.28; p-value = 0.97) and exhibited good discriminatory power from the ROC curve generated for the model (AUC=0.94). The model selected based on forced fit of the same factors from the gott hotspot analysis showed a lower relative model quality (AIC=53.307), good discriminatory power (AUC=0.95), and had no evidence of a lack of fit (Hosmer-Lemeshow Test = 1.85; p-value=0.99). This model contained clean face, the percent of households reporting any formal education, number of household items owned, and the number of rounds of MDA distributed to the woreda. Finally, a model that was driven by the hypothesized influence of WASH related factors being the most important model factors required confounders to be considered for removal from the selection process prior to consideration of the WASH variables. This model depended on the predictors of clean face, household access to water within 30 minutes, mean number of household items owned, percent of households reporting any formal education, and the percent of gotts with paved roads in the woreda, *Table 6*. This model had the best relative fit (AIC=41.84), exhibited no evidence of a lack of fit (Hosmer-Lemeshow Test = 0.33; p-value 1.00).

Hotspot Model Comparison

The factors that explain affiliation in a hotspot woreda or a hotspot gott in logistic regression are varied at the two analysis scales. The hotspot gotts identified in this analysis were found to be significantly determined by the prevalence of clean face and the effect of this variable may be influenced by interaction. *Figure 29,* illustrated that the effect of increased clean face prevalence varies by high and low temperature. By dichotomizing annual temperature in the state of Amhara at the median there is a noticeable change in the effect of clean faces on the odds ratio of residence in a hotspot woreda and adding an interaction term to the final model was statistically significant ($p<0.01$). As the percent of children with clean faces increases at the gott level, the effect is stronger above the median temperature when compared to the odds ratio at the same percent of clean faces at an average annual temperature below the median.

The three strategies for model selection were compared and used to select the most appropriate hotspot woreda model. The changes in the odds ratio, for a one percent increase in the woreda percent of clean face, and found the model determined by a purely statistically driven selection process and hypothesis driven models performed similarly. The data driven model was slightly more conservative exhibiting the shallowest curve in response to changing the percent of clean face in the woreda, *Figure 24* . The model that was forced to fit with the same predictors as those selected from the gott level analysis showed the most dramatic decline in the odds of trachoma among those exposed to higher levels of clean face indicating that a 10% increase in clean face could reduce the odds ratio close to zero. The most conservative assessment is a better approximation for reality as TF_{1to9} is likely to be more responsive to improvements in the percent of children with clean faces in some geographic locals than in others.

Discussion

Trachoma impact surveys covered all 152 administrative woredas throughout the Amhara National Regional State. Of the woredas, 117 (77%) had TF_{1to9} prevalence ≥ 10% which requires immediate implementation of the full SAFE strategy and confirms the previous findings that trachoma remains endemic within Amhara (99). The distribution of trachoma was spatially clustered in both the woreda and the gott analysis. Furthermore, identification of hotspot gotts allows the program to identify the areas in greatest need at a spatial scale that is frequently masked when using the larger, woreda unit, for intervention decisions. Clean faces and access to water within 30 minutes were significant predictors of clustering and trachoma prevalence. These findings suggest that hygiene promotion and water availability are important contributes to the spatial pattern of trachoma throughout Amhara. As such, sustainable trachoma control requires increased F&E investments that are designed to accommodate context dependent infrastructure, behaviors, and community perceptions.

Clean Face as a Predictor

Percentage of children with a clean face was a robust predictor of both TF prevalence and likelihood of residence in a trachoma hotspot. These findings support prior research that suggests facial cleanliness and hygiene behavior are significantly negatively associated with the odds of trachoma in children (25). Mechanistically, nasal and ocular discharge on children's faces serves as a reservoir of infection that can be transmitted by direct contact, by wiping of faces with the shawl of a caretaker, or through the transmission from person-to-person by the vector *M. sorbens* (46). The model controlling for spatial lag suggests that the impact of improving the percentage of children with a clean face may result in a reduction of the trachoma prevalence at the woreda administrative unit. At a large spatial scale of analysis this is a novel finding, and the results can be utilized by administrators considering distributing funding for F&E. Additionally, trachoma hotspot woredas are associated with percent of clean face. Utilizing the hotspot analysis allows programs to adapt intervention strategies to target the eleven hotspot woredas for health promotion activities with behavioral change models that can be sustainably adopted (100).

Extensions of the construct of herd immunity from vaccine research to the WASH field are gaining traction across recent publications (43, 101, 102). The Trachoma Amelioration in Northern Amhara (TANA) study found herd protection through treatment with antibiotics of only the children in a community could be a sufficient antibiotic target (103). Furthermore, Oswald *et al*. examined the effect of community sanitation and found a significant decrease in the prevalence odds of trachoma in communities with ≥60% community latrine coverage compared to <20% coverage (104). These studies exemplify the possibility of community benefits by controlling the environmental suitability of the infection. Although additional research is needed, given the community scale of the current analysis and the strength of the aggregate measure of clean face across multiple models it is reasonable to consider herd protection from trachoma may exist in gotts and woredas that exhibit high facial cleanliness. Promoting knowledge about hygiene behavior is not sufficient to sustainably control trachoma, rather distinct regional interventions must be designed to adapt to the community in which they are launched in order to create a lasting behavioral change (26). In particular, due to funding limitations, programs should incorporate prior research and utilize community engaged participatory research as the cornerstone to influence individual behaviors that help the community as a whole. Geographically targeting hotspot gotts allows for these resource intensive projects, targeting improvement to facial hygiene, to efficiently use funding and locate the areas with greatest need.

Evaluation of hotspots at a fine spatial scale aids in the analysis by reducing the modifiable areal unit problem that results in the systematic variation when a large spatial scale is used for analysis (105). These findings support the evidence base that community acceptance of positive hygiene behavior collectively reduces the burden of TF_{1to9} (26). When comparing the change in the odds ratio for a one percent change in clean face on TF_{1to9} across the grouped climate variables the measure of effect was significantly different for high and low temperature groups ($P<0.01$). The effect of clean face is greater in gotts with high average annual temperature compared to low average annual temperatures. For a 10% increase in clean faces the odds of residence in a hotspot gott is decreased by 14% in gotts that with an average annual temperature below the median and 20% in gotts that have an average annual temperature above the median. The observed interaction could be attributable to a scenario where the areas with higher average annual temperature have less barriers to performing face washing. Regional differences in psychosocial, contextual, and technological barriers may inhibit habituation of hygiene behaviors and the ability to create a sustained impact on trachoma (26). Gotts that have average annual temperatures below the median may have a community risk perception that bathing activities leave children cold and more likely to develop an illness (106). In colder regions caretakers may more rigorously dry their children's faces with contaminated towels than in areas where the temperature is warmer and caretakers are less likely to be concerned with drying faces after washing. On the other hand, the availability of towels as a resource to dry children following washing may be perceived as a luxury item. Community attitudes of ownership of these items may alter the behavior associated with face washing (107). Under these circumstances the same fly density and play activities at a warmer region may be more conducive to transmission. Furthermore, families spending more time indoors could exacerbate the effects of crowding by increasing the frequency of contacts between infected individuals (108). In gotts that share similar low average annual temperature, children who receive a higher exposure to indoor cooking or heating fires may have increased ocular irritation that renders them susceptible to ocular infection (109). In communities that are colder on average it is reasonable to suspect that families spend more time indoors in close proximity to smoke and other infected individuals. Communities where facial cleanliness is challenged by behavioral practices may have more infectious children and a higher burden of disease.

Household Water Access

A consensus has not yet been reached about the benefit of improving household water access and the effect on the prevalence of active trachoma. A recent literature review demonstrated that distance to water across 38 studies was not found to be statistically significantly associated with trachoma (25). This observation could be due to the bias of self-reported distance to water source although it is a frequently used measure in trachoma studies. However, few studies have considered the community effect of household access to water on the odds of trachoma.

The current study found that increased distance to water was significantly associated with residence in a trachoma hotspot gott or a hotspot woreda when controlling for other variables in the model. Furthermore, woredas with a higher percentage of households with access to a water source within 30 minutes were linearly associated with a reduced prevalence of trachoma, indicating that distance to a water source is an important predictor of trachoma at the gott and the woreda level of analysis supporting other studies (71, 110).

Water access in close proximity to households can promote face washing behavior by allowing households to more easily collect and store water for washing. Behaviors such as washing at the water source can be enabled to be more flexibly integrated into one's daily schedule if water sources are more easily accessible. Furthermore, increased water access yields the co-benefit of freeing up time for the care taker to perform washing activities (107). The survey question did not allow the respondent to relate whether the water supply functioned at a capacity that provided consistent access. This could affect the water quantity households collect or the time of collection which in turn could influence other hygiene behaviors. While this is a frequently used measure throughout the literature, directly observed household water quantity may be a better health indicator in future studies (111). Access to an improved water source was not a significant community predictor of TF_{1to9} prevalence or hotspot residence possibly indicating that the type of water source used for hygiene behaviors is not as important as having access to a water source in general. This study does not advocate against the investment in improved water sources because improved water promotes child health multiple ways (112). Rather, the regions where the only access to a water source is an unimproved source could develop hygiene promotion activities that focus on using any source of water for face washing.

Household Latrine Access

The community proportion of households with access to a latrine was found to be significantly linearly associated with TF_{1to9} prevalence in the model that controlled for spatial lag, but not in hotspot analysis. There is likely a community protective effect of sanitation access that is masked when using household presence of a latrine rather than considering the influence that neighboring households have on creating a hospitable environment for the breeding of *M. sorbens* (104). Various studies have found that household latrine access is protective against active trachoma, but far fewer publications have accounted for latrine use (25). Provision of latrines rarely results in behavioral modification so observation of the presence or absence of a latrine without accounting for usage may obscure the exposure/disease relationship that is present (113). In the current study, household latrine use was omitted due to survey modification over time that affected comparability. Future surveys will be able to evaluate the effect of household use of a latrine at similar spatial scales. Latrine usage may be further investigated to incorporate a granular analysis that attempts to structure interventions based on community participatory programs. Sanitation coverage within a community may be a more important predictor of hotspots than individual household access to a latrine and future studies could evaluate the role of the presence of used latrine facilities in relation to surrounding households. While the effect of sanitation is still unclear, the literature suggests that community led total sanitation programs should be investigated (114). This approach not only increases coverage of latrines but can facilitate sustainable behavioral change that engages the community to take collective action (26).

Spatial Analysis

Spatial data analysis requires information about a geographic position in space and links the attributes of a geographic feature to a dataset. The polygon data provided by the woredas and the point data derived from individual gotts help to explain the distribution of trachoma in Amhara. The aggregate measures at the gott and woreda spatial scales enhanced the analysis because interventions applied by the Amhara Regional Health Bureau are based off of the estimates at these scales. This analysis aggregates weighted estimates from different levels but the implications from this study only pertain to the level of analysis at which the model was designed to estimate. Due to the aggregate nature of the data the woreda level relationships cannot be expected to explain the gott relationships between explanatory predictors and the disease nor vice versa. The distribution of TF_{1to9} is representative of both programmatic success and the tenacity of the disease. The quantitative analysis of patterns is a valuable tool to comprehend the complex interrelationships inherent within human populations. These methods enhance surveillance and intervention strategies by identifying the local determinants of disease and accounting for the spatial relationships in modeling procedures.

Determining a biologically plausible and meaningful conceptualization of a neighborhood is one of the main challenges with performing a spatial analysis. Exploratory analysis experimented with different weights, connectivity, and conceptualizations of spatial relationships to define the neighborhood surrounding each woreda and determine the best spatial weights matrix for all analysis. Spatial analysis frequently makes inferences using polygons and a relationship that is based on contiguous edges defined by the way a rook or a queen would move on a chess board

(115). This conceptualization was attempted but due to underlying problems with the digitization of the shapefile for Amhara this conceptualization was unavailable. As a result, alternative methods of defining the relationship between woredas were evaluated. Multiple threshold distance bands were assessed by using the geographic centroids for each polygon feature. Distance bands starting with the minimum one neighbor distance of 80 km were incrementally assessed up to 250km. Subsequently, the number of neighbors and connectivity between each feature were assessed with a histogram followed by visualization of the feature neighborhood with a connectivity map in GeoDa (GeoDa version 1.8.16.4). The smallest distance bands did not allow the larger features in northwest Amhara to have as many neighbors as smaller features south of Lake Tana. Large distance bands created neighborhoods that provided more neighbors for large woredas but the bias from the extent beyond second order neighbor relationships could not be justified. Another attempt to assess compatible neighborhood conceptualizations of k-nearest neighbors was applied so that each feature would use the centroids from the k-nearest neighbors that are closest to the reference feature's centroid. As the number of neighbors was increased features along the edges of the map would have neighborhoods that would extend beyond second order relationships. The best conceptualization of spatial relationship between woredas was identified as the 8-nearest neighbor conceptualization. This relationship gave an average distance between woredas of 54.4 km and required a maximum distance of 132.6 km to give the feature with the largest neighborhood 8 neighbors. Conceptualizing the relationship in a meaningful way allows for inferences drawn from equations that rely on the spatial weights matrix to be influenced greater by neighbors that are closer than those that are further away. A justifiable neighborhood allowed for the calculation of the Getis-Ord

$G_i^*(d)$ measure for clustering. The neighborhoods of woredas throughout Amhara are likely to influence trachoma outcomes through political, social, climatic, and historical features that have caused people of similar characteristics to aggregate into these regions. The regional similarities between woredas are more likely to be similar between nearby woreda than those separated further across the regional state of Amhara.

GWR Model

In Wag Hemra, a zone where all the woredas are statistically significant hotspots, a one percent increase in woreda level clean face decreases the prevalence of TF_{1to9} by 0.87% holding all other variables constant. Comparing the change in prevalence in Wag Hemra to the western half of North Gondar zone where a one percent increase in children presenting with clean faces results in a 0.58% change in TF_{1to9} prevalence. This suggests that a face washing intervention may provide greater benefit in the Wag Hemra zone. This effect is more noticeable when the percent of children with clean face is increased by 20% in the two regions and a change in TF_{1to9} prevalence can be expected to be reduced by >5% in the Wag Hemra zone.

The difference in the effect of facial cleanliness in the two regions may be the result of the role that flies play in the transmission process (6). This could be the case if Wag Hemra is a more suitable geographic region for fly growth than western North Gondar, but there are no entomological studies to support this hypothesis. Further studies could pursue this question by investigating the role of *M. sorbens* in the Wag Hemra woredas. Additionally, the GWR model indicates that infrastructure investment in roads may provide the best benefit in the southern most regions of the East Gojam zone where a one percent increase in paved road access leads to a three percent decrease in trachoma prevalence controlling for other variables in the model. This could be because some gotts are harder to access due to the road infrastructure and that access to preventive services such as educational programs or antibiotic distribution are more limited.

Spatial Lag Model

The model that controls for spatial lag was selected as the best fit model as a result of disentangling the effects of spatial lag on the outcome and the model residuals. The spatial lag model assumes that there is a diffusion process occurring by which TF_{1to9} prevalence in one woreda is explained in part by the values of TF_{1to9} in neighboring woredas. The row standardized spatial weights matrix defines the relationship between neighboring values of TF_{1to9} prevalence and this relationship is modified by a lag parameter included in the model. The conceptualization modified from Anselin *et al.* and shown in the appendix, indicates how the spatial lag parameter helps to regulate the influence of neighboring woredas (94). Trachoma does not remain constrained by political boundaries. Infection spillover is assumed to occur within geographic locations that are closer in space unless physical boundaries such as extreme differences in topography, transportation networks, or other contextual processes influence the comingling between gotts and woredas. The spatial autoregressive model was able to break the spatial dependence to allow the model to meet the assumption of independent observations and uncorrelated error terms. This model provides a better global fit for predicting the prevalence of TF_{1to9} than the GWR model explained above.

Multiple WASH variables retained a significant role in the model with negative linear associations with TF_{1to9}. Controlling for other variables in the model, a one percent increase in woreda level household water access within 30 minutes corresponds to a 0.09% decrease in TF_{1to9} (p-value 0.03). Similarly, controlling for the other variables in the model a one-unit increase in woreda level household access to a latrine predicted a 0.08% decrease in TF_{1to9} prevalence. Household access to an improved water source was not a significant predictor in this model. The strongest association between a WASH variable and TF_{1to9} was a 0.64% decrease in TF_{1to9} prevalence for a 1% increase in woreda level clean face. Additionally, increasing access to paved roads by one percent at the woreda level while controlling for other variables in the model resulted in a 1.7% decrease in TF_{1to9} prevalence and increasing MDA by one round in the woreda was likely to be linearly associated with a 1.8% increase in TF_{1to9}. The model coefficients suggest that some WASH interventions might be more successful than others across Amhara. The positive association with the number of rounds of MDA and TF_{1to9} is explained by woredas where TF has been less responsive to antibiotics. The woredas where TF_{1to9} prevalence is higher receive more annual rounds of Azithromycin and therefore are positively associated. These observations further strengthen the argument to promote the F&E arms of the SAFE strategy.

Woreda Hotspot Logistic Regression Model

The woreda hotspot analysis identified the Wag Hemra zone as having one of the highest neighborhoods of TF_{1to9} prevalence throughout the region. The neighborhood conceptualization limited the number of features considered hotspots. The false discovery rate (FDR) was used to identify as many significant comparisons as possible while maintaining a low false positive rate due to multiple testing. It was determined that the conservative approach of adjusting the critical p-values and missing some statistically significant results was preferable to performing the tests in isolation and ignoring the inherent bias of multiple testing. This provided 90% confidence that the hotspots identified were the most significant in the Amhara Regional State. The logistic model used to predict residence in a hotspot woreda was best fit when including three of the four WASH predictors although only children presenting with clean faces and household access to water within 30 minutes were statistically significant.

Collinearity affected many of the predictors that could be included in the final model as seven of the hypothesized confounders were dropped. The observed collinearity was likely due to the similarity in estimates when weighted and aggregated to the woreda level. Many of the confounders were either related to economic status or geoclimatic variables which could be directly related with water availability. The model indicates that with a 10% increase in the percent of children with clean faces the odds of residing in a hotspot woreda were reduced by 64% holding constant household access to water within 30 minutes, household access to an improved water source, and controlling for percent of gotts in the woreda with a healthcare facility. Alternatively, a 10% increase in water access less than 30 minutes from the household leads to a 42% decrease in the odds of residing in a hotspot woreda when holding all other variables in the model constant. These findings suggest that woreda level investments in both improving access to water as well as increasing behavioral practices associated with face washing likely yield the most substantial benefits to reducing the frequency of hotspots.

Gott Hotspot Logistic Regression Model

The analysis of the factors that explain whether a gott is high prevalence and surrounded by high prevalence when compared to the global average identified one significant WASH predictor and five social economic or geoclimatic variables to control. Gott level prevalence estimates of TF_{1to9} were highly variable ranging from gotts where TF in children was completely absent to gotts where all children were estimated to have the disease. Hotspot gotts were best explained by the percent of children with clean faces when controlling for economic, social, and geoclimatic variables. Within gotts a 10% increase in children with clean faces reduces the odds of residence in a hotspot by 17% when controlling for all other variables in the model. Average annual temperature and average annual precipitation were dichotomized at their median values because no previous studies supported a consistent value for these variables and no noticeable variation in prevalence was observed when contrasting TF_{1to9} with these categorized variables.

Strengths and Limitations

Strengths:

This study takes place in a hyperendemic setting for trachoma and the Amhara National Regional State has the highest burden of active trachoma in Ethiopia (116). Few studies have sought to disaggregate the influence of spatial relationships when investigating the epidemiology of trachoma. This is the first study that evaluated spatial relationships at large scale, between woredas, and compared the results to a more fine-scale analysis. The size of the dataset, 69,236 children aged 1 to 9 representing all districts in the Amhara National Regional State, benefited the aggregate estimates by helping to improve the precision in both the gott and the woreda analysis. The geographic position of all study households was recorded whenever possible and gott averaged latitude and longitude protected the anonymity of survey respondents. Selection bias was minimized as all study participants were selected to be representative of the entire region, and a probability sample was used allowing for all individuals to have a known probability of selection. Outcome misclassification was likely to be negligible because the high quality training which was standardized across all survey rounds and performed prior to each survey to minimize potential drift that could occur in grading clinical signs. The survey was modified over several rounds of implementation to more effectively gauge the important behavioral determinants and standardized across other trachoma surveys (71, 117). The use of direct observations of clean faces and latrines helped to reduce bias is associated with self-reported indicators (118). Effective electronic survey data collection methods were employed and the success in being able to adequately report geographic position was verified.

The dataset allowed for a global view of the entire regional state which few other mapping surveys have been able to achieve with such a large geographic area. Furthermore, aggregate spatial analysis are uncommon when looking at trachoma distribution spatially as only one study identified clustering between villages (45) and few studies considered clustering outside of the household scale (55-57, 63). The present analysis far outweighed the village clustering study by Hagi *et al.* where only 203 villages were utilized whereas this study uses aggregate data from 1,558 gott evaluation units. The extant literature supports the use of spatial analysis for identification of locations in greatest need of intervention and continued surveillance (62). Other studies have found the effect of temperature to be a significant explanatory variable in multivariate regression models, but none of these studies were conducted in a region with average altitude and average annual temperature similar to Amhara (4, 45, 50, 60). Utilizing two spatial scales of clustering allowed for a comparison of predictors at the level where health decisions are frequently made. The spatial analysis of clustering of both the outcome and the hypothesized risk factors confirmed the hypothesis that geographic variability exists between communities that could be utilized to pinpoint intervention strategies. Finally, adaptation of spatial econometrics theory to develop a model to control for regional influences was a novel technique in trachoma research (119). Use of this model may help administrators predict TF_{1to9} prevalence based on woreda reports of F&E infrastructure to design a framework for an intervention strategy based on woreda-based need.

Limitations:

There were several limitations to this study. Inherently, cross sectional studies based on survey methods do not allow for any determination about causality because both the exposure and the outcome data are collected at the same time. Despite the finding that clean faces in the community are significantly associated with TF_{1to9}, when controlling for other variables, the conclusion that community clean face is the cause of trachoma clinical signs cannot be drawn. Most studies in trachoma utilize cross sectional data and are limited by the same conditions (25) . The second limitation with the survey data is that exposure information for water access is self-reported, although this is frequently measured across the literature in this manner (49, 120). There are many validity issues when considering self-reported exposures. Interviewers could have prodded the subjects into responding one way or another when they were considering the question, but training protocols would likely limit these effects. Second, it is a possibility that social desirability bias was present when interviewees responded to questions. The interviewees may have answered in a way they deemed to be correct or their response may have been influenced by a desire to present a situation that would please the interviewer. This misclassification of exposure is likely to be non-differential and bias would be towards the null. Furthermore, when considering the results from all models, generalizing the results to individuals within these geographic domains would inflict the ecologic fallacy (121).

The use of long term climate averages from the years 1950 to 2000 does not account for annual climate variability. Studies reliant on remote sensing in areas where historical weather data is incomplete or unreliable have used these interpolated averages (62). The Amhara region is known to experience distinct seasonality with heavy rains and seasonal drought in a biannual pattern (122). The natural fluctuations may change the seasonal abundance of flies and water availability. The TIS was designed to capture the greatest number of participants by surveying based on times when the people would not be sowing or harvesting crops. The time of year that these groups are surveyed could result in differential misclassification of exposures and disease. For instance, the self-reported distance to a water source could be greater in the dry season as local water supplies dry up and the households in a gott are forced to travel further to collect water. Additionally, the seasonal abundance of *M. sorbens* may fluctuate and cause seasonal spikes in active disease, but entomological studies supporting this assumption are non-existent. Furthermore, some influential geoclimatic variables were not included in this analysis and warrant further investigation. Greenness may be an important explanatory variable as a proxy for fly abundance and some measure to account for this influence should be included in future research (123). Humidity is thought to be an important modulator of fly populations in a close relative of *M. sorbens* but weather station data in Amhara was too incomplete to draw conclusions about this climate factor (124). A calculation of terrain ruggedness would be of value to include in future models as the landscapes that have more dramatic elevation changes in short distances might be more isolated from receiving interventions or create pockets where the disease can thrive.

Dichotomizing variables such as highest household education, household access to a latrine, and crowding may obscure the detail inherent in some of the associations with TF_{1to9}. The decision to dichotomize variables was made after evaluating the distribution of the variables independently and the relationships between the levels of each explanatory factor and the outcome. No glaring differences were evident and these categorizations were supported elsewhere in the trachoma literature supporting the assumption that any bias inherent in the categorization of these variables would be minimal to the overall association (25, 125).

Latrine access at the household was not an important predictor of trachoma hotspots; however, further research is needed to address whether household use of a latrine or household use of an improved latrine would benefit a community better (104). Additionally, other studies have suggested that there is a benefit of community sanitation as households may have the greatest health benefits if their neighbors have access and use latrines (43, 126). Therefore, the way latrines were utilized in this analysis should not occlude the possibility that safe containment of human feces by community latrine coverage can also be a significant predictor of trachoma hotspots.

Despite regimented training, trachoma grading is subject to a variability. The WHO simplified grading scheme increases reliability and helps to build the number of qualified examiners but the evaluation is still subject to reliability error. Laboratory diagnostic tests are increasingly being used to validate diagnosis in trachoma regions and recently the implementation of photographic validation is suspected to improve the reliability of diagnostics (127). As reported in a recent study, the relationship between lack of nasal and ocular discharge and TF clinical signs is complicated by the reciprocal relationship between the outcome and the exposure (25). An acyclic mechanism is proposed where ocular and nasal discharge can be the effect of the inflammation caused by trachoma, and this contagious substance can in turn cause infection challenging future studies investigating causality. Additionally, clean face defined as the lack of nasal or ocular discharge exhibited agreement among observers in a clinical trial, but the ability to predict the frequency of face washing was challenged (46). Despite this limitation, a standardized observation of clean faces is a more robust indicator than using self-reporting of hygiene behaviors because self-reporting can be inherently subject to bias (128).

Public Health Significance and Recommendations

Moving from implementation to intervention requires a thoughtful process that considers not only the areas in greatest need for an intervention, but also the consideration of where provision of goods and services will yield the greatest benefit. Meeting the goals of any program requires attempts at maximizing control of the disease with the least cost and the most efficient use of resources. Spatial analysis is an effective technique for visualizing the overall picture of disease and going further by allowing inferences about targeted interventions. This study supports prior research in the associations between water access, hygiene interventions, and trachoma but, in particular, isolates a formula for determining where these interventions can yield a substantial impact.

In order to ensure continued progress towards elimination, targeted efforts must be sustained over time. Regional hotspots require special emphasis in disease control. Finding a successful intervention template in these regions will promote the future progress towards control and elimination in areas where the disease is more resilient to antibiotics. Contextual factors are geographically grouped to similar locations as a result of historical and current social pressures that mold the landscape of trachoma. The groups of individuals residing in neighboring gotts or woredas are more likely to interact and share similar behavioral traits that influence their susceptibility to the disease. Further research into how hygiene norms differ by religion and ethnicity may be informative in Amhara. Interventions against trachoma must incorporate all the facets of SAFE in order to achieve sustained success, but some regions warrant a greater focus on the F&E components. In particular, the woredas with the highest prevalence of TF_{1to9} clustered around Wag Hemra would benefit the most by promoting facial hygiene and developing water resources that are more accessible to households. Furthermore, the gotts in high prevalence clusters are likely to have neighbors that were not surveyed which may exhibit high prevalence of the disease. Utilization of the gott level hotspot analysis can help teams predict where other clusters are located through the development of risk maps (62).

Infrastructure improvements (provision of water and latrines) should be coupled with behavioral modifications adapted to local behaviors and perceptions for the greatest returns on investment (26). Projects that aim to provide communities with safe water resources should earmark funding so that projects are coupled with the promotion of knowledge that only a limited quantity of water is required to perform washing activities (107). Previous studies have shown that some communities perceive face washing as a luxury because time to collect extra water for this practice is limited (107, 129). Currently, The Carter Center is participating in the sanitation, water, and instruction in face-washing for trachoma (SWIFT) cluster randomized clinical trial to learn more about the effects of WASH on communities (130). This research was motivated by findings that antibiotics alone are not capable of preventing the return of the disease in hyperendemic areas and that sustainable trachoma control depends on infrastructure and behavioral changes. The findings from the present study suggest the following interventions based on the two significant WASH related risk factors.

Communities with high water access and large proportions of children with clean faces are best positioned to protect children from TF. Communities that have low percentage of household with access to water less than 30 minutes away but high clean face are possibly promoting the conservative use of water for face washing and the importance of facial hygiene. Areas with high water access but low clean face would benefit the most from a face washing behavioral campaign that works with community groups to design a successful intervention. Communities with low water access and low proportions of clean face would likely need dual intervention strategies to improve water access while simultaneously teaching about the importance of hygiene behavior. Future studies of the risk factors for clusters of low proportions of clean faces could help explain the factors that contribute to communities and neighboring villages that share similarities which are driving a lack of face washing. Currently some programs do not use clean face as an indicator although the findings from the present study illustrate its efficacy. Integrating trachoma control with other WASH operations could help communities through the co-benefit of improving point of use drinking water quality as well as motivating a behavior that is likely to be preventative for trachoma.

At least two studies have supported the finding that ambient temperature influences face washing behaviors (106, 107). Assuming the efficacy of facial hygiene is temperature dependent a strategy to achieve facial cleanliness in the colder regions would be beneficial. People might be more willing to wash their children's faces if they experience less discomfort or social stigmatization. One study found that the members of a community did not have towels available to dry children and focus groups showed that mothers were concerned that children would become ill in wet and cold conditions (107). Micro investments in solar projects might benefit the gotts where behavioral adoption is most challenged. A series of cost effective kits with solar panels and a water kettle could be piloted in the regions where resistance to face washing is due to discomfort when using cold water. Prior research has shown that people learn how to adapt hygiene behaviors by listening to the strategies outlined by others in community discussions and as such the community's voice should be considered when implementing WASH strategies (106).

It is critical to note that while there are many woredas above the WHO recommended threshold of TF_{1to9} prevalence in the Amhara National Regional State. However, Amhara has made tremendous progress by steeply reducing the prevalence of the disease in many regions. During the baseline survey in 2007, 100% of the woredas were suspected of having TF prevalence ≥10% (80) and required interventions. In the current study, two woredas had a TF prevalence <5% and thus no longer required MDA. The current study shows that the distribution of TF in children aged 1 to 9 years is not spatially random at the woreda and the gott levels of analysis, and the underlying spatial structure of the region can be taken into account to target interventions to specific areas. The diffusion process of the disease, behaviors, and possibly interventions increase the likelihood of occurrence of similar events at neighboring locations. Due to the endemicity of trachoma, the woreda is frequently the administrative level that supports interventions, while gotts are used as the distribution point for intervention campaigns.

Future surveys should consider the investment in geolocating water resources and categorizing them according to the JMP guidelines. This strategy would be more cost intensive as personnel would be required to work with the communities to locate water sources when they are not located within the household. However, this would alleviate the bias associated with self-reporting of water access and the type of water source households have access to. Additionally, the data from such a project would help partners working on development projects in the region so that a geospatial strategy to improve access to water can be efficiently maximized. Furthermore, Ethiopia is likely to experience climate change that may shift the distribution of water resources in the region (131). Being able to dynamically adjust interventions will help Amhara adapt to water scarcity and water redistribution issues that will likely affect the region in the coming decades. Fortunately, changes to face washing behavior may be sustained in the region once more people are educated on the limited quantity of water required to effectively perform washing (107).

Risk mapping is becoming increasingly important for use with diseases as they approach elimination (132). Additionally, accessing all of the gotts in Amhara can be extremely challenging given the rugged terrain of the region. As the program continues to close in on elimination targets ensuring that remote populations are reached with MDA and other interventions will require increased efficiency in targeting resources. Developing risk maps will increase the precision of locating the final garrisons of the disease. Descriptive and inferential maps should be routinely updated to track progress as interventions are applied and the disease responds. Automated mapping programs will provide policy makers with a tool that allows for near real-time and less responsive decision making.

The fight to control trachoma in Amhara is far from over, but with geospatial surveillance and targeted implementation of the SAFE strategy progress towards disease elimination may be in sight. The communities in which individuals live have a dramatic impact on the overall health of their neighbors. No single intervention is going to be independently effective across all communities. Community based participatory strategies will help to elucidate the local drivers of trachoma. Trachoma is not restricted to geographic boundaries. Likewise, health behaviors can be equally diffuse. With the results of this investigation in mind, targeting of water and hygiene interventions should help to achieve trachoma elimination, and with time healthy practices will disperse to become the norm.

References

1. WHO. Trachoma. 2016. (http://www.who.int/mediacentre/factsheets/fs382/en/). (Accessed 20 Oct 2016).
2. Kuper H, Solomon AW, Buchan J, et al. A critical review of the SAFE strategy for the prevention of blinding trachoma. *The Lancet infectious diseases* 2003;3(6):372-81.
3. Mariotti SP, Pascolini D, Rose-Nussbaumer J. Trachoma: global magnitude of a preventable cause of blindness. *British Journal of Ophthalmology* 2009;93(5):563-8.
4. Smith JL, Sivasubramaniam S, Rabiu MM, et al. Multilevel Analysis of Trachomatous Trichiasis and Corneal Opacity in Nigeria: The Role of Environmental and Climatic Risk Factors on the Distribution of Disease. *PLoS Negl Trop Dis* 2015;9(7):e0003826.
5. Ramesh A, Bristow J, Kovats S, et al. The impact of climate on the abundance of Musca sorbens, the vector of trachoma. *Parasites & vectors* 2016;9(1):48.
6. Emerson PM, Lindsay SW, Alexander N, et al. Role of flies and provision of latrines in trachoma control: cluster-randomised controlled trial. *The Lancet* 2004;363(9415):1093-8.
7. Gambhir M, Basáñez M-G, Turner F, et al. Trachoma: transmission, infection, and control. *The Lancet infectious diseases* 2007;7(6):420-7.
8. Abdou A, Nassirou B, Kadri B, et al. Prevalence and risk factors for trachoma and ocular Chlamydia trachomatis infection in Niger. *British Journal of ophthalmology* 2007;91(1):13-7.
9. West SK, Munoz B, TURNER VM, et al. The epidemiology of trachoma in central Tanzania. *International Journal of Epidemiology* 1991;20(4):1088-92.
10. West SK, Congdon N, Katala S, et al. Facial cleanliness and risk of trachoma in families. *Archives of Ophthalmology* 1991;109(6):855-7.
11. Mathur G, Sharma R. Influence of some socio-economic factors on the prevalence of trachoma. *Indian journal of medical sciences* 1970;24(6):325-33.
12. Habtamu E, Wondie T, Aweke S, et al. Trachoma and relative poverty: a case-control study. *PLoS Negl Trop Dis* 2015;9(11):e0004228.
13. King JD, Ngondi J, Gatpan G, et al. The burden of trachoma in ayod county of southern Sudan. *PLoS Negl Trop Dis* 2008;2(9):e299.
14. Courtright P, Sheppard J, Schachter J, et al. Trachoma and blindness in the Nile Delta: current patterns and projections for the future in the rural Egyptian population. *British Journal of Ophthalmology* 1989;73(7):536-40.
15. Thylefors B, Dawson CR, Jones BR, et al. A simple system for the assessment of trachoma and its complications. *Bulletin of the World Health Organization* 1987;65(4):477.

16. MARIOTTI S. Geographical information system and (Expected) results in the fight against trachoma. *Revue internationale du trachome et de pathologie oculaire tropicale et subtropicale et de santé publique* 2004;81:21-38.
17. See CW, Alemayehu W, Melese M, et al. How Reliable Are Tests for Trachoma?—A Latent Class Approach. *Investigative Ophthalmology & Visual Science* 2011;52(9):6133-7.
18. Bobo L, Viscidi R, Quinn T, et al. Diagnosis of Chlamydia trachomatis eye infection in Tanzania by polymerase chain reaction/enzyme immunoassay. *The Lancet* 1991;338(8771):847-50.
19. Ngondi J, Reacher M, Matthews F, et al. Trachoma survey methods: a literature review. *Bulletin of the World Health Organization* 2009;87(2):143-51.
20. Rajak SN, Collin JRO, Burton MJ. Trachomatous Trichiasis and its Management in Endemic Countries. *Survey of Ophthalmology* 2012;57-341(2):105-35.
21. Rajak SN, Habtamu E, Weiss HA, et al. Why Do People Not Attend for Treatment for Trachomatous Trichiasis in Ethiopia? A Study of Barriers to Surgery. *PLoS Negl Trop Dis* 2012;6(8):e1766.
22. West SK, Bailey R, Munoz B, et al. A randomized trial of two coverage targets for mass treatment with azithromycin for trachoma. *PLoS Negl Trop Dis* 2013;7(8):e2415.
23. Organization WH. Technical consultation on trachoma surveillance: meeting report. September 11−12, 2014, Task Force for Global Health, Decatur, USA. 2015(2015):23.
24. Mabey DC, Solomon AW, Foster A. Trachoma. *The Lancet* 2003;362(9379):223-9.
25. Stocks ME, Ogden S, Haddad D, et al. Effect of water, sanitation, and hygiene on the prevention of trachoma: a systematic review and meta-analysis. *PLoS Med* 2014;11(2):e1001605.
26. DELEA MG, SOLOMON H, FREEMAN MC. INTERVENTIONS TO IMPROVE FACIAL CLEANLINESS & ENVIRONMENTAL IMPROVEMENT FOR TRACHOMA PREVENTION AND CONTROL.
27. West S, Lynch M, Turner V, et al. Water availability and trachoma. *Bulletin of the World Health Organization* 1989;67(1):71.
28. Ngondi J, Gebre T, Shargie EB, et al. Evaluation of three years of the SAFE strategy (Surgery, Antibiotics, Facial cleanliness and Environmental improvement) for trachoma control in five districts of Ethiopia hyperendemic for trachoma. *Transactions of the Royal Society of Tropical Medicine and Hygiene* 2009;103(10):1001-10.
29. Astle WF, Wiafe B, Ingram AD, et al. Trachoma control in Southern Zambia—an international team project employing the SAFE strategy. *Ophthalmic epidemiology* 2006;13(4):227-36.

30. Emerson PM, Cairncross S, Bailey RL, et al. Review of the evidence base for the ‘F’and ‘E’components of the SAFE strategy for trachoma control. *Tropical medicine & international health* 2000;5(8):515-27.
31. Porco TC, Gebre T, Ayele B, et al. Effect of mass distribution of azithromycin for trachoma control on overall mortality in Ethiopian children: a randomized trial. *Jama* 2009;302(9):962-8.
32. Oswald WE, Stewart AE, Flanders WD, et al. Prediction of Low Community Sanitation Coverage Using Environmental and Sociodemographic Factors in Amhara Region, Ethiopia. *The American journal of tropical medicine and hygiene* 2016;95(3):709-19.
33. Barnett H. The incrimination of arthropods as vectors of disease. Presented at Proceedings of the 11th International Congress of Entomology1960.
34. Darougar S, Forsey T, Jones B, et al. Isolation of Chlamydia trachomatis from eye secretion (tears). *British Journal of Ophthalmology* 1979;63(4):256-8.
35. Forsey T, Darougar S. Transmission of chlamydiae by the housefly. *British Journal of Ophthalmology* 1981;65(2):147-50.
36. Miller K, Pakpour N, Yi E, et al. Pesky trachoma suspect finally caught. *British journal of ophthalmology* 2004;88(6):750-1.
37. Emerson PM, Bailey RL, Mahdi OS, et al. Transmission ecology of the fly Musca sorbens, a putative vector of trachoma. *Transactions of the Royal Society of Tropical Medicine and Hygiene* 2000;94(1):28-32.
38. Emerson PM, Lindsay SW, Walraven GE, et al. Effect of fly control on trachoma and diar rhoea. *The Lancet* 1999;353(9162):1401-3.
39. Farmer KJ, Sohal RS. Effects of ambient temperature on free radical generation, antioxidant defenses and life span in the adult housefly, Musca domestica. *Experimental gerontology* 1987;22(1):59-65.
40. Gage KL, Burkot TR, Eisen RJ, et al. Climate and vectorborne diseases. *American journal of preventive medicine* 2008;35(5):436-50.
41. HAFEZ M, Attia M. Studies on the ecology of Musca sorbens Wied. in Egypt (Diptera: Muscidae). *Bulletin de la Societe Entomologique d'Egypte* 1958;42.
42. Emerson P, Bailey R, Walraven G, et al. Human and other faeces as breeding media of the trachoma vector Musca sorbens. *Medical and veterinary entomology* 2001;15(3):314-20.
43. Fuller JA, Eisenberg JN. Herd Protection from Drinking Water, Sanitation, and Hygiene Interventions. *The American journal of tropical medicine and hygiene* 2016:15-0677.
44. Taylor HR, West SK, Mmbaga B, et al. Hygiene factors and increased risk of trachoma in central Tanzania. *Archives of Ophthalmology* 1989;107(12):1821-5.

45. Hägi M, Schémann J-F, Mauny F, et al. Active Trachoma among Children in Mali: Clustering and Environmental Risk Factors. *PLoS Negl Trop Dis* 2010;4(1):e583.
46. King JD, Ngondi J, Kasten J, et al. Randomised trial of face-washing to develop a standard definition of a clean face for monitoring trachoma control programmes. *Transactions of the Royal Society of Tropical Medicine and Hygiene* 2011;105(1):7-16.
47. Bailey R, Downes B, Downes R, et al. Trachoma and water use; a case control study in a Gambian village. *Transactions of the Royal Society of Tropical Medicine and Hygiene* 1991;85(6):824-8.
48. Hoechsmann A, Metcalfe N, Kanjaloti S, et al. Reduction of trachoma in the absence of antibiotic treatment: evidence from a population-based survey in Malawi. *Ophthalmic epidemiology* 2001;8(2-3):145-53.
49. Prost A, Negrel A. Water, trachoma and conjunctivitis. *Bulletin of the World Health Organization* 1989;67(1):9.
50. Schémann J-F, Laffly D, Sacko D, et al. Trichiasis and geoclimatic factors in Mali. *Transactions of the Royal Society of Tropical Medicine and Hygiene* 2007;101(10):996-1003.
51. Prüss A, Mariotti SP. Preventing trachoma through environmental sanitation: a review of the evidence base. *Bulletin of the World Health Organization* 2000;78(2):267-73.
52. Burki T. The broad benefits of trachoma elimination. *The Lancet Infectious Diseases* 2016;16(5):530.
53. Bartram J, Lewis K, Lenton R, et al. Focusing on improved water and sanitation for health. *The Lancet* 2005;365(9461):810.
54. Miller HJ. Tobler's first law and spatial analysis. *Annals of the Association of American Geographers* 2004;94(2):284-9.
55. Bailey R, Osmond C, Mabey D, et al. Analysis of the household distribution of trachoma in a Gambian village using a Monte Carlo simulation procedure. *International Journal of Epidemiology* 1989;18(4):944-51.
56. Broman AT, Shum K, Munoz B, et al. Spatial clustering of ocular chlamydial infection over time following treatment, among households in a village in Tanzania. *Investigative ophthalmology & visual science* 2006;47(1):99-104.
57. Yohannan J, He B, Wang J, et al. Geospatial Distribution and Clustering of Chlamydia trachomatis in Communities Undergoing Mass Azithromycin TreatmentGeospatial Clustering of C. trachomatis. *Investigative ophthalmology & visual science* 2014;55(7):4144-50.
58. Lyseen AK, Nøhr C, Sørensen EM, et al. A review and framework for categorizing current research and development in health related geographical information systems (GIS) studies. *Yearb Med Inform* 2014;9(1):110-24.

59. Hürlimann E, Schur N, Boutsika K, et al. Toward an open-access global database for mapping, control, and surveillance of neglected tropical diseases. *PLoS Negl Trop Dis* 2011;5(12):e1404.
60. Koukounari A, Touré S, Donnelly CA, et al. Integrated monitoring and evaluation and environmental risk factors for urogenital schistosomiasis and active trachoma in Burkina Faso before preventative chemotherapy using sentinel sites. *BMC infectious diseases* 2011;11(1):1.
61. Baggaley R, Solomon A, Kuper H, et al. Distance to water source and altitude in relation to active trachoma in Rombo district, Tanzania. *Tropical Medicine & International Health* 2006;11(2):220-7.
62. Clements ACA, Kur LW, Gatpan G, et al. Targeting Trachoma Control through Risk Mapping: The Example of Southern Sudan. *PLoS Negl Trop Dis* 2010;4(8):e799.
63. Polack S, Solomon A, Alexander N, et al. The household distribution of trachoma in a Tanzanian village: an application of GIS to the study of trachoma. *Transactions of the Royal Society of Tropical Medicine and Hygiene* 2005;99(3):218-25.
64. Edwards T, Smith J, Sturrock HJ, et al. Prevalence of trachoma in Unity State, South Sudan: results from a large-scale population-based survey and potential implications for further surveys. *PLoS Negl Trop Dis* 2012;6(4):e1585.
65. Neiderud C-J. How urbanization affects the epidemiology of emerging infectious diseases. *Infection ecology & epidemiology* 2015;5.
66. Hu VH, Harding-Esch EM, Burton MJ, et al. Epidemiology and control of trachoma: systematic review. *Tropical Medicine & International Health* 2010;15(6):673-91.
67. Kasi PM, Gilani AI, Ahmad K, et al. Blinding trachoma: a disease of poverty. *PLoS Med* 2004;1(2):e44.
68. Aisha E. Stewart MZ, Demelash Gessesse, Berhanu Melak, Eshetu Sata, Tigist Astale, Tekola Endeshaw, Zerihun Tadesse, E. Kelly Callahan, Melsew Chanyalew, Birhan Gaudie, Paul M. Emerson, Jonathan D. King, Scott D. Nash. Trachoma impact surveys after at least 5 years of SAFE strategy in hyper-endemic Amhara National Regional State, Ethiopia Presented at American Society of Tropical Medicine & Hygiene Annual Meeting, Atlanta, GA, November 14, 2016 2016.
69. Katz J, ZEGER SL, Tielsch JM. Village and household clustering of xerophthalmia and trachoma. *International journal of epidemiology* 1988;17(4):865-9.
70. Muhammad N, Mpyet C, Adamu MD, et al. Mapping Trachoma in Kaduna State, Nigeria: Results of 23 Local Government Area-Level, Population-Based Prevalence Surveys. *Ophthalmic epidemiology* 2016:1-9.
71. Bero B, Macleod C, Alemayehu W, et al. Prevalence of and risk factors for trachoma in Oromia Regional State of Ethiopia: results of 79 population-

based prevalence surveys conducted with the global trachoma mapping project. *Ophthalmic epidemiology* 2016;23(6):392-405.
72. Elshafie BE, Osman KH, Macleod C, et al. The epidemiology of trachoma in Darfur States and Khartoum State, Sudan: results of 32 population-based prevalence surveys. *Ophthalmic Epidemiology* 2016;23(6):381-91.
73. Sherief ST, Macleod C, Gigar G, et al. The Prevalence of Trachoma in Tigray Region, Northern Ethiopia: Results of 11 Population-Based Prevalence Surveys Completed as Part of the Global Trachoma Mapping Project. *Ophthalmic Epidemiology* 2016:1-6.
74. Ngondi J, Onsarigo A, Adamu L, et al. The epidemiology of trachoma in Eastern Equatoria and Upper Nile States, southern Sudan. *Bulletin of the World Health Organization* 2005;83(12):904-12.
75. Yayemain D, King JD, Debrah O, et al. Achieving trachoma control in Ghana after implementing the SAFE strategy. *Transactions of the Royal Society of Tropical Medicine and Hygiene* 2009;103(10):993-1000.
76. Melese M, Alemayehu W, Lakew T, et al. Comparison of annual and biannual mass antibiotic administration for elimination of infectious trachoma. *Jama* 2008;299(7):778-84.
77. Lakew T, House J, Hong KC, et al. Reduction and return of infectious trachoma in severely affected communities in Ethiopia. *PLoS Negl Trop Dis* 2009;3(2):e376.
78. West SK, Munoz B, Mkocha H, et al. Infection with Chlamydia trachomatis after mass treatment of a trachoma hyperendemic community in Tanzania: a longitudinal study. *The Lancet* 2005;366(9493):1296-300.
79. Gebre T, Ayele B, Zerihun M, et al. Comparison of annual versus twice-yearly mass azithromycin treatment for hyperendemic trachoma in Ethiopia: a cluster-randomised trial. *The Lancet* 2012;379(9811):143-51.
80. Emerson PM, Ngondi J, Biru E, et al. Integrating an NTD with one of "the big three": combined malaria and trachoma survey in Amhara Region of Ethiopia. *PLoS Negl Trop Dis* 2008;2(3):e197.
81. King JD, Teferi T, Cromwell EA, et al. Prevalence of trachoma at sub-district level in Ethiopia: determining when to stop mass azithromycin distribution. *PLOS Negl Trop Dis* 2014;8(3):e2732.
82. Turner AG, Magnani RJ, Shuaib M. A not quite as quick but much cleaner alternative to the Expanded Programme on Immunization (EPI) Cluster Survey design. *International journal of epidemiology* 1996;25(1):198-203.
83. Taylor HR, West SK, Katala S, et al. Trachoma: evaluation of a new grading scheme in the United Republic of Tanzania. *Bulletin of the world health organization* 1987;65(4):485.
84. WHO U. Improved and unimproved water sources and sanitation facilities. *Geneva: WHO/UNIFEC Joint Monitoring Programme (JMP) for Water Supply and Sanitation (h ttp://www wssinfo org/definitions-methods/watsan-categories/, accessed 30 November 2015)* 2016.

85. Schemann J, Sacko D, Malvy D, et al. Risk factors for trachoma in Mali. *International journal of epidemiology* 2002;31(1):194-201.
86. Luna EJA, Medina NH, Oliveira MB, et al. Epidemiology of Trachoma in Bebedouro State of São Paulo, Brazil: Prevalence and Risk Factors. *International Journal of Epidemiology* 1992;21(1):169-77.
87. Jimenez V, Gelderblom HC, Flueckiger RM, et al. Mass drug administration for trachoma: how long is not long enough? *PLoS Negl Trop Dis* 2015;9(3):e0003610.
88. Trachoma Atlas. Decatur, GA: International Trachom Initiative; 2017. (http://www.trachomaatlas.org/). (Accessed April 2 2017).
89. Getis A, Ord JK. The analysis of spatial association by use of distance statistics. *Geographical analysis* 1992;24(3):189-206.
90. Kleinbaum D, Kupper L, Nizam A, et al. *Applied regression analysis and other multivariable methods*. Nelson Education; 2013.
91. Brunsdon C, Fotheringham AS, Charlton ME. Geographically weighted regression: a method for exploring spatial nonstationarity. *Geographical analysis* 1996;28(4):281-98.
92. Moran PA. Notes on continuous stochastic phenomena. *Biometrika* 1950;37(1/2):17-23.
93. Jarque CM, Bera AK. Efficient tests for normality, homoscedasticity and serial independence of regression residuals. *Economics letters* 1980;6(3):255-9.
94. Anselin L, Baltagi B. A companion to theoretical econometrics. *Companion to Theoretical Econometrics* 2001:310-30.
95. Breusch TS, Pagan AR. A simple test for heteroscedasticity and random coefficient variation. *Econometrica: Journal of the Econometric Society* 1979:1287-94.
96. David K, Mitchel K. Logistic regression: A self learning text. New York: Springer–Verlag Inc, 1994.
97. Vazquez-Prokopec GM, Spillmann C, Zaidenberg M, et al. Spatial heterogeneity and risk maps of community infestation by Triatoma infestans in rural northwestern Argentina. *PLoS Negl Trop Dis* 2012;6(8):e1788.
98. Waller LA, Gotway CA. *Applied spatial statistics for public health data*. John Wiley & Sons; 2004.
99. Mengitsu B, Shafi O, Kebede B, et al. Ethiopia and its steps to mobilize resources to achieve 2020 elimination and control goals for neglected tropical diseases: Spider webs joined can tie a lion. *International health* 2016;8(suppl 1):i34-i52.
100. Hubley J, Gilbert C. Eye health promotion and the prevention of blindness in developing countries: critical issues. *British journal of ophthalmology* 2006;90(3):279-84.

101. Carter RC. Can and should sanitation and hygiene programmes be expected to achieve health impacts? *Waterlines* 2017;36(1):92-103.
102. Andres L, Briceño B, Chase C, et al. Sanitation and externalities: evidence from early childhood health in rural India. *Journal of Water Sanitation and Hygiene for Development* 2017:washdev2017143.
103. Ying R, Williams PD, Dorratoltaj N. Prospects for trachoma elimination through mass treatment targeted at children. Presented at Proceedings of the Agent-Directed Simulation Symposium2016.
104. Oswald WE, Stewart AE, Kramer MR, et al. Active trachoma and community use of sanitation, Ethiopia. *Bull World Health Organ* 2017.
105. Openshaw S, Taylor PJ. A million or so correlation coefficients: three experiments on the modifiable areal unit problem. *Statistical applications in the spatial sciences* 1979;21:127-44.
106. Lynch M, West SK, Muñoz B, et al. Testing a participatory strategy to change hygiene behaviour: face washing in central Tanzania. *Transactions of the Royal Society of Tropical Medicine and Hygiene* 1994;88(5):513-7.
107. McCauley AP, Lynch M, Pounds MB, et al. Changing water-use patterns in a water-poor area: lessons for a trachoma intervention project. *Social science & medicine* 1990;31(11):1233-8.
108. Ejere HO, Alhassan MB, Rabiu M. Face washing promotion for preventing active trachoma. *The Cochrane Library* 2014.
109. Zambrano AI, Muñoz BE, Mkocha H, et al. Exposure to an indoor cooking fire and risk of trachoma in children of Kongwa, Tanzania. *PLoS Negl Trop Dis* 2015;9(6):e0003774.
110. Muhammad N, Mohammed A, Isiyaku S, et al. Mapping trachoma in 25 local government areas of Sokoto and Kebbi states, northwestern Nigeria. *British Journal of Ophthalmology* 2014;98(4):432-7.
111. Stelmach RD, Clasen T. Household water quantity and health: a systematic review. *International journal of environmental research and public health* 2015;12(6):5954-74.
112. Montgomery MA, Elimelech M. Water and sanitation in developing countries: including health in the equation. *Environ Sci Technol* 2007;41(1):17-24.
113. Bracken P, Kvarnström E, Ysunza A, et al. Making sustainable choices–the development and use of sustainability oriented criteria in sanitary decision making. Presented at Third International Conference on Ecological Sanitation, Durban, South Africa2005.
114. Galvin M. Talking shit: is Community-Led Total Sanitation a radical and revolutionary approach to sanitation? *Wiley Interdisciplinary Reviews: Water* 2015;2(1):9-20.
115. Anselin L, Rey S. Properties of tests for spatial dependence in linear regression models. *Geographical analysis* 1991;23(2):112-31.

116. Berhane Y, Worku A, Bejiga A, et al. Prevalence and causes of blindness and Low vision in Ethiopia. *The Ethiopian Journal of Health Development (EJHD)* 2016;21(3).
117. Adera TH, Macleod C, Endriyas M, et al. Prevalence of and risk factors for trachoma in Southern Nations, Nationalities, and Peoples' Region, Ethiopia: results of 40 population-based prevalence surveys carried out with the Global Trachoma Mapping Project. *Ophthalmic Epidemiology* 2016:1-10.
118. Gorber SC, Tremblay MS. Self-Report and Direct Measures of Health: Bias and Implications. *The Objective Monitoring of Physical Activity: Contributions of Accelerometry to Epidemiology, Exercise Science and Rehabilitation*: Springer, 2016:369-76.
119. Anselin L. Spatial econometrics. *A companion to theoretical econometrics* 2001;310330.
120. Golovaty I, Jones L, Gelaye B, et al. Access to water source, latrine facilities and other risk factors of active trachoma in Ankober, Ethiopia. *PLoS One* 2009;4(8):e6702.
121. Diez-Roux AV. Bringing context back into epidemiology: variables and fallacies in multilevel analysis. *American journal of public health* 1998;88(2):216-22.
122. Bewket W, Conway D. A note on the temporal and spatial variability of rainfall in the drought-prone Amhara region of Ethiopia. *International Journal of Climatology* 2007;27(11):1467-77.
123. Ramesh A, Kovats S, Haslam D, et al. The impact of climatic risk factors on the prevalence, distribution, and severity of acute and chronic trachoma. *PLoS Negl Trop Dis* 2013;7(11):e2513.
124. Dakshinamurty S. The common House-fly, Musca domestica, L., and its behaviour to temperature and humidity. *Bulletin of entomological research* 1948;39(03):339-57.
125. Harding-Esch EM, Edwards T, Mkocha H, et al. Trachoma prevalence and associated risk factors in the gambia and Tanzania: baseline results of a cluster randomised controlled trial. *PLoS Negl Trop Dis* 2010;4(11):e861.
126. King D, Emerson PM, Callahan EK, et al. Active trachoma and community use of sanitation, Ethiopia.
127. West SK, Taylor HR. Reliability of photographs for grading trachoma in field studies. *British journal of ophthalmology* 1990;74(1):12-3.
128. Freeman MC, Stocks ME, Cumming O, et al. Systematic review: hygiene and health: systematic review of handwashing practices worldwide and update of health effects. *Tropical Medicine & International Health* 2014;19(8):906-16.
129. West S, Muñoz B, Lynch M, et al. Impact of face-washing on trachoma in Kongwa, Tanzania. *The Lancet* 1995;345(8943):155-8.

130. Fry D. SWIFT: Sanitation, Water, and Instruction in Face-Washing for Trachoma. UCSF F.I. Proctor Foundation, 2017.
131. Hailemariam K. Impact of climate change on the water resources of Awash River Basin, Ethiopia. *Climate Research* 1999;12(2-3):91-6.
132. Baker M, Mathieu E, Fleming F, et al. Mapping, monitoring, and surveillance of neglected tropical diseases: towards a policy framework. *The Lancet* 2010;375(9710):231-8.

Tables

Table 1. Simple linear regression analysis of factors associated with trachomatous inflammation follicular in children aged 1 to 9 years, Amhara, Ethiopia. Univariate linear associations between model explanatory factors and the outcome expressed as unadjusted Pearson's Correlation Coefficients (N=152).

Variable	Intercept Estimate	Slope Estimate	P Value for Test of Slope	Adjusted R2
Household latrine access	40.19	-0.29	<0.0001	0.12
Household access to an improved water source	39.1	-0.25	<0.0001	0.13
Household access to a water source in less than 30 minutes	43.31	-0.3	<0.0001	0.14
Absence of nasal or ocular discharge	108.07	-1.03	<0.0001	0.50
Household crowding	16.04	0.24	0.0864	0.01
Households reporting no education	22.24	0.07	0.3720	0.00
Mean wealth indicator score	43.18	-12.24	<0.0001	0.33
Villages with access to any health facility	27.94	-0.15	0.0622	0.02
Villages with a paved road*	48.12	-6.91	<0.0001	0.13
Average annual precipitation	51.19	-0.02	<0.0001	0.15
Average annual temperature	24.42	0.00	0.9333	0.00
Average altitude	23..66	0.00	0.8025	0.00
Number of previous rounds of MDA distributed to the woreda	-16.58	4.89	<0.0001	0.2

* Variable was log transformed to create a normal distribution

Table 2. Model estimates from an ordinary least squares model using TF prevalence among children age 1 to 9 as the outcome

Model Properties		Model Parameters							
R-squarred	AIC	Intercept (p-value)	Water within 30 minutes (p-value)	Clean Face (p-value)	latrine (p-value)	Improved water (p-value)	Previous MDA (p-value)	Log of paved roads (p-value)	Morans I (8 NN)
0.64	1133.39	90.28 (<0.01)	-0.12 (0.01)	-0.77 (<0.01)	-0.09 (0.04)	-0.02 (0.63)	2.20 (<0.01)	-2.48 (<0.01)	3.56 (<0.01)

Table 3. Models run in SpaceStat to assess model forms and neighborhood relationships that account for spatial dependency at different scales compared with the ordinary least squares model

Model Properties						Model Parameters							
Model Type	Neighborhood	R-squared	AIC	Rho (p-value)	Lambda (p-value)	Intercept (p-value)	Water within 30 minutes (p-value)	Clean Face (p-value)	latrine (p-value)	Improved water (p-value)	Previous MDA (p-value)	Log of paved roads (p-value)	Spatial Breusch-Pagan Test* (p-value)
Ordinary Least Squares	--	0.64	1133.39	--	--	90.28 (<0.01)	-0.12 (0.01)	-0.77 (<0.01)	-0.09 (0.04)	-0.02 (0.63)	2.20 (<0.01)	-2.48 (<0.01)	--
Spatial Lag	8 Nearest Neighbors	0.69	1111.78	0.40 (<0.01)	--	69.11 (<0.01)	-0.09 (0.03)	-0.65 (<0.01)	-0.08 (0.04)	-0.03 (0.36)	1.81 (<0.01)	-1.71 (0.02)	10.73 (0.097131)
Spatial Error	8 Nearest Neighbors	0.49	1119.34	--	0.55 (<0.01)	83.46 (<0.01)	-0.08 (0.06)	-0.66 (<0.01)	-0.14 (<0.01)	-0.04 (0.31)	1.75 (<0.01)	-1.52 (0.06)	11.70 (0.069124)
Spatial Lag	110km	0.68	1119.86	0.43 (<0.01)	--	73.13 (<0.01)	-0.07 (0.08)	-0.71 (<0.01)	-0.10 (0.02)	-0.03 (0.42)	1.90 (<0.01)	-1.93 (0.01)	9.56 (0.143211)
Spatial Error	110km	0.55	1121.94	--	0.68 (<0.01)	88.30 (<0.01)	-0.04 (0.31)	-0.70 (<0.01)	-0.15 (<0.01)	-0.04 (0.31)	1.67 (<0.01)	-2.07 (0.01)	10.23 (0.115357)

* A regression diagnostic for heteroskedasticity

** Spatial lag dependence on weights

Table 4. Analysis variables weighted to the woreda and stratified by the outcome of residing in a hotspot versus residing in a woreda that was not a statistically significant hotspot identified using the Getis-Ord G_i(d) statistic (n=152)*

Variable	All Woredas (n=152) Mean (sd)	Non-hotspot Woredas (n=140) Mean (sd)	Hot Spot Woredas (n=12) Mean (sd)	Univariate Odds Ratio OR (95%)
Individual				
Trachomatous inflammation - follicular (%)	25.08 (16.19)	23.08 (14.53)	48.54 (16.79)	1.132 (1.065, 1.204)
Children with clean faces (%)	80.58 (11.13)	81.63 (10.31)	68.30 (13.39)	0.910 (0.863, 0.961)
Household				
Access to a latrine	52.22 (19.95)	53.04 (20.08)	42.60 (16.24)	0.974 (0.945, 1.004)
Access to a water source within 30 minutes (%)	61.58 (20.73)	63.48 (19.52)	39.34 (22.43)	0.946 (0.916, 0.976)
Access to an improved water source (%)	55.07 (23.35)	56.70 (22.55)	36.10 (25.18)	0.962 (0.936, 0.988)
Crowding (>6 persons living in the home) (%)	38.22 (9.54)	37.88 (9.49)	42.19 (9.70)	1.048 (0.985, 1.115)
Mean number of household items owned	1.48 (0.76)	1.56 (0.74)	0.52 (0.32)	0.014 (0.002, 0.110)
Any reported formal education (%)	41.03 (16.99)	40.73 (16.49)	44.51 (22.69)	1.013 (0.978, 1.049)
Gott				
Health facility (%)	22.53 (17.19)	24.32 (16.24)	15.70 (23.37)	0.965 (0.920, 1.012)
Paved road (%)	39.64 (24.19)	40.78 (24.18)	20.26 (14.99)	0.940 (0.893, 0.991)
Woreda				
Previous Rounds of MDA	8.52 (1.50)	8.47 (1.51)	9.08 (1.16)	1.354 (0.876, 2.092)
Climate variables				
Average annual precipitaion (mm)	1135.00 (278.37)	1162.90 (271.19)	809.50 (94.27)	0.981 (0.971, 0.991)
Average annual temperature (°C)	18.41 (3.05)	18.29 (3.09)	19.78 (2.25)	1.157 (0.968, 1.384)
Altitude (m)	2057.91 (477.36)	2067.04 (488.55)	1951.42 (313.98)	1.000 (0.998, 1.001)

Table 5. Summary of the model selection procedure for finding the best fit logistic model that predicts membership in a hotspot woreda using a data driven procedure where the least significant variable was dropped from the model regardless of the hypothesized relationship.

	Variables analyzed						Model fit	
Model	Clean	Latrine	H2O < 30	iH2O	Education	HC Facility	AIC	LRT
1	-X**	+X NS	-X*	-X NS	+X NS	-X NS	51.390	29.942
2	-X**	+X NS	-X **	-X NS	+X NS	—	65.590	30.372
3	-X**	+X NS	-X*	-X NS	—	-X NS	50.656	28.676
4	-X**	+X NS	-X**	-X NS	—	—	63.906	30.056
5	-X**	—	-X *	-X NS	+X NS	-X NS	49.857	29.475
6	-X**	—	-X **	-X NS	—	-X NS	48.867	28.465
7	-X**	—	-X **	-X NS	—	—	63.691	28.271
8	-X**	+X NS	-X**	—	+ NS	-X NS	52.454	26.879
9	-X*	+X NS	-X**	—	—	-X NS	52.843	24.489
10	-X**	—	-X**	—	—	-X NS	50.843	24.489
11	-X**	—	-X**	—	+X NS	-X NS	50.731	26.601
12	-X**	—	-X**	—	—	—	63.620	26.342

Symbols: X (Variable tested in model); — (Variable not included in model) ; - (Negative association); + (Positive association) ; ** (P≤0.01); * (0.05≤P<0.01)

Variables: Clean (Children who presented to examiners without nasal or ocular discharge), Latrine (Household has a latrine of any type), H2O<30 (The household has access to a water source within 30 minutes of the home), iH2O (The household has access to an improved water source where improved is defined using the sustainable development goals criteria for an improved source), Crowd (The household has more than six residents), Wealth (The mean number of household wealth indicators), Education (The percentage of households that reported having attended any type of formal education), HC Facility (The presence of any type of health facility within the surveyed gott), Paved Road (The presence of a paved road within the gott), Prev MDA (The number of previous rounds of antibiotic distributed to the woreda during MDA campaigns), Precip (The average annual precipitation in millimeters for the gott), Temp (The average annual temperature in degrees celcius for the gott), altitude (The altitude for the gott center)

Model fit: AIC (Akaike information criterion), LRT (Likelihood ratio test)

Table 6. Summary of the multivariate model selection procedure for selecting the best explanatory factors based on the hypothesized relationship that WASH variables would best account for the relationship with hotspot residence. The variables household latrine access, water within 30 minutes of the household, household access to an improved water source, and clean face were not considered to be removed from the model until all other covariate relationships had been assessed.

	Variables analyzed									Model fit	
Model	Clean	Latrine	H2O < 30	iH2O	SES	Education	Health Facility	Road	MDA	AIC	LRT
1	-X *	+X NS	-X NS	+X NS	-X NS	+X NS	+X NS	-X NS	+X NS	48.447	33.116
2	-X *	+X NS	-X NS	-X NS	-X NS	+X NS	—	-X NS	+X NS	47.601	32.192
3	-X *	+X NS	-X NS	-X NS	-X NS	+X NS	—	-X NS	—	45.649	32.144
4	-X *	+X NS	-X NS	-X NS	-X **	+X NS	—	—	—	54.441	43.521
5	-X *	+X NS	-X NS	-X NS	-X **	—	—	—	—	52.501	43.461
6	-X *	+X NS	-X NS	—	-X NS	+X NS	+X NS	-X NS	+X NS	46.447	33.115
7	-X *	+X NS	-X NS	—	-X NS	+X NS	—	-X NS	+X NS	45.664	32.129
8	-X *	+X NS	-X NS	—	-X NS	+X NS	—	-X NS	—	43.705	32.088
9	-X *	—	-X NS	—	-X NS	+X NS	—	-X NS	—	41.842	31.952
10	-X *	—	-X NS	—	-X**	+X NS	—	—	—	52.518	41.444
11	-X *	—	-X NS	—	-X**	—	—	—	—	50.555	41.407

Symbols: X (Variable tested in model); — (Variable not included in model) ; - (Negative association); + (Positive association) ; ** (P ≤0.01); * (0.05≤P<0.01)

Variables: Clean (Children who presented to examiners without nasal or ocular discharge), Latrine (Household has a latrine of any type), H2O<30 (The household has access to a water source within 30 minutes of the home), iH2O (The household has access to an improved water source where improved is defined using the sustainable development goals criteria for an improved source), Crowd (The household has more than six residents), Wealth (The mean number of household wealth indicators), Education (The percentage of households that reported having attended any type of formal education), HC Facility (The presence of any type of health facility within the surveyed gott), Paved Road (The presence of a paved road within the gott), Prev MDA (The number of previous rounds of antibiotic distributed to the woreda during MDA campaigns), Precip (The average annual precipitation in millimeters for the gott), Temp (The average annual temperature in degrees celcius for the gott), altitude (The altitude for the gott center)

Model fit: AIC (Akaike information criterion), LRT (Likelihood ratio test)

Table 7. Univariable analysis of factors that are hypothesized to determine whether or not a gott is a statistically significant hotspot of TF prevalence in children aged 1 to 9 in Amhara, Ethiopia.

Variable	All Villages Mean (sd)	Non-hotspot Villages Mean (sd)	Hot Spot Villages Mean (sd)	Univariate Odds Ratio OR (95%)
Individual				
Trachomatous inflammation - follicular (%)	24.79 (22.72)	19.35 (19.87)	45.42 (20.96)	1.053 (1.046, 1.060)
Children with clean faces (%)	81.16 (19.24)	83.70 (17.06)	71.52 (23.55)	0.971 (0.965, 0.977)
Household				
Access to a latrine	52.19 (30.18)	54.06 (29.80)	45.09 (30.60)	0.990 (0.986, 0.994)
Access to a water source within 30 minutes (%)	62.39 (38.71)	64.12 (37.77)	55.83 (41.49)	0.995 (0.992, 0.998)
Access to an improved water source (%)	54.49 (44.21)	55.79 (43.38)	49.52 (47.00)	0.997 (0.994, 1.000)
Crowding (>6 persons living in the home) (%)	62.96 (15.82)	63.34 (16.06)	61.51 (14.82)	0.993 (0.985, 1.000)
Mean number of household items owned	1.48 (1.07)	1.63 (1.08)	0.94 (0.87)	0.442 (0.375, 0.521)
Any reported formal education (%)	54.17 (27.30)	54.89 (26.65)	51.44 (29.52)	0.995 (0.991, 1.000)
Gott				
Health facility (%)	23.04 (42.12)	24.57 (43.07)	17.23 (2.34)	0.639 (0.466, 0.876)
Paved road	37.29 (48.37)	39.65 (48.94)	28.31 (45.12)	0.601 (0.460, 0.785)
Woreda				
Previous Rounds of MDA	8.65 (1.67)	8.39 (1.67)	9.65 (1.22)	1.780 (1.615, 1.962)
Climate variables				
Average annual precipitaion (mm)	1166.10 (272.68)	1204.76 (282.35)	1019.47 (164.10)	0.997 (0.996, 0.997)
Average annual temperature (°C)	17.61 (3.44)	17.64 (3.68)	17.49 (2.33)	0.988 (0.953, 1.024)
Altitude (m)	2195.88 (544.04)	2182.30 (580.62)	2247.40 (370.23)	1.000 (1.000, 1.000)

Table 8. Factors derived from multivariate logistic regression models and the association with residence in a cluster of high trachoma prevalence in the Amhara Region of Ethiopia.

	Variables analyzed													Model fit	
Model	Clean	Latrine	H2O < 30	iH2O	Crowd	Wealth	Education	HC Facility	Paved Road	Prev MDA	Precip	Temp	Altitude	AIC	LRT
1	-X**	-X NS	+X NS	+X NS	-X NS	-X *	-X **	-X NS	+X NS	+X **	-X **	-X NS	-X NS	1190.914	423.780
2	-X**	-X NS	+X NS	+X NS	-X NS	-X *	-X **	-X NS	+X NS	+X **	-X **	-X **	—	1189.504	432.190
3	-X**	-X NS	-X NS	+X NS	—	-X *	-X **	-X NS	+X NS	+X **	-X **	-X NS	-X NS	1191.291	430.403
4	-X**	-X NS	+X NS	+X NS	-X NS	-X **	-X **	—	+X NS	+X **	-X **	-X NS	-X NS	1190.089	431.605
5	-X**	+X NS	+X NS	+X NS	-X NS	-X *	-X **	-X NS	—	+X **	-X **	-X NS	-X NS	1190.938	430.756
6	-X**	-X NS	+X NS	+X NS	—	-X *	-X **	-X NS	+X NS	+X **	-X **	-X **	—	1189.400	429.854
7	-X**	-X NS	+X NS	+X NS	-X NS	-X *	-X **	—	+X NS	+X **	-X **	-X **	—	1188.626	431.067
8	-X**	+X NS	+X NS	+X NS	-X NS	-X **	-X **	-X NS	—	+X **	-X **	-X **	—	1189.451	430.243
9	-X**	-X NS	+X NS	+X NS	—	-X *	-X **	—	+X NS	+X **	-X **	-X **	—	1189.074	428.620
10	-X**	-X NS	+X NS	+X NS	—	-X *	-X **	-X NS	—	+X **	-X **	-X **	—	1189.610	428.084
11	-X**	-X NS	+X NS	+X NS	-X NS	-X *	-X **	—	—	+X **	-X **	-X **	—	1188.458	429.236
12	-X**	-X NS	+X NS	+X NS	—	-X *	-X **	—	—	+X **	-X **	-X **	—	1188.726	426.968
13	-X**	—	+X NS	+X NS	—	-X *	-X **	—	—	+X **	-X **	-X **	—	1186.754	426.940
14	-X**	-X NS	—	+X NS	—	-X *	-X **	—	—	+X **	-X **	-X **	—	1187.290	426.404
15	-X**	-X NS	+X NS	—	—	-X NS	-X **	—	—	+X **	-X **	-X **	—	1189.070	424.624
16	-X**	—	—	+X NS	—	-X *	-X **	—	—	+X **	-X **	-X **	—	1185.299	426.395
17	-X**	—	+X NS	—	—	-X NS	-X **	—	—	+X **	-X **	-X **	—	1187.070	424.624
18	-X**	+X NS	—	—	—	-X NS	-X **	—	—	+X **	-X **	-X **	—	1187.914	423.780
19	-X**	—	—	—	—	-X NS	-X **	—	—	+X **	-X **	-X **	—	1185.921	423.773
20	-X**	—	—	—	—	—	-X **	—	—	+X **	-X **	-X **	—	1186.945	420.748

Symbols: X (Variable tested in model); — (Variable not included in model) ; - (Negative association); + (Positive association) ; ** ($P \leq 0.01$); * ($0.05 \leq P < 0.01$)

Variables: Clean (Children who presented to examiners without nasal or ocular discharge), Latrine (Household has a latrine of any type), H2O<30 (The household has access to a water source within 30 minutes of the home), iH2O (The household has access to an improved water source where improved is defined using the sustainable development goals criteria for an improved source), Crowd (The household has more than six residents), Wealth (The mean number of household wealth indicators), Education (The percentage of households that reported having attended any type of formal education), HC Facility (The presence of any type of health facility within the surveyed gott), Paved Road (The presence of a paved road within the gott), Prev MDA (The number of previous rounds of antibiotic distributed to the woreda during MDA campaigns), Precip (The average annual precipitation in millimeters for the gott), Temp (The average annual temperature in degrees celcius for the gott), altitude (The altitude for the gott center)

Model fit: AIC (Akaike information criterion), LRT (Likelihood ratio test)

Table 9. Models run in SpaceStat to assess model forms and neighborhood relationships that account for spatial dependency at different scales compared with the ordinary least squares model

Model Properties						Model Parameters							
Model Type	Neighborhood	R-squared	AIC	Rho (p-value)	Lambda (p-value)	Intercept (p-value)	Water within 30 minutes (p-value)	Clean Face (p-value)	latrine (p-value)	Improved water (p-value)	Previous MDA (p-value)	Log of paved roads (p-value)	Spatial Breusch-Pagan Test* (p-value)
Ordinary Least Squares	--	0.64	1133.39	--	--	90.28 (<0.01)	-0.12 (0.01)	-0.77 (<0.01)	-0.09 (0.04)	-0.02 (0.63)	2.20 (<0.01)	-2.48 (<0.01)	--
Spatial Lag	8 Nearest Neighbors	0.69	1111.78	0.40 (<0.01)	--	69.11 (<0.01)	-0.09 (0.03)	-0.65 (<0.01)	-0.08 (0.04)	-0.03 (0.36)	1.81 (<0.01)	-1.71 (0.02)	10.73 (0.097131)
Spatial Error	8 Nearest Neighbors	0.49	1119.34	--	0.55 (<0.01)	83.46 (<0.01)	-0.08 (0.06)	-0.66 (<0.01)	-0.14 (<0.01)	-0.04 (0.31)	1.75 (<0.01)	-1.52 (0.06)	11.70 (0.069124)
Spatial Lag	110km	0.68	1119.86	0.43 (<0.01)	--	73.13 (<0.01)	-0.07 (0.08)	-0.71 (<0.01)	-0.10 (0.02)	-0.03 (0.42)	1.90 (<0.01)	-1.93 (0.01)	9.56 (0.143211)
Spatial Error	110km	0.55	1121.94	--	0.68 (<0.01)	88.30 (<0.01)	-0.04 (0.31)	-0.70 (<0.01)	-0.15 (<0.01)	-0.04 (0.31)	1.67 (<0.01)	-2.07 (0.01)	10.23 (0.115357)

* A regression diagnostic for heteroskedasticity

** Spatial lag dependence on weights

Figures and Figure Legends

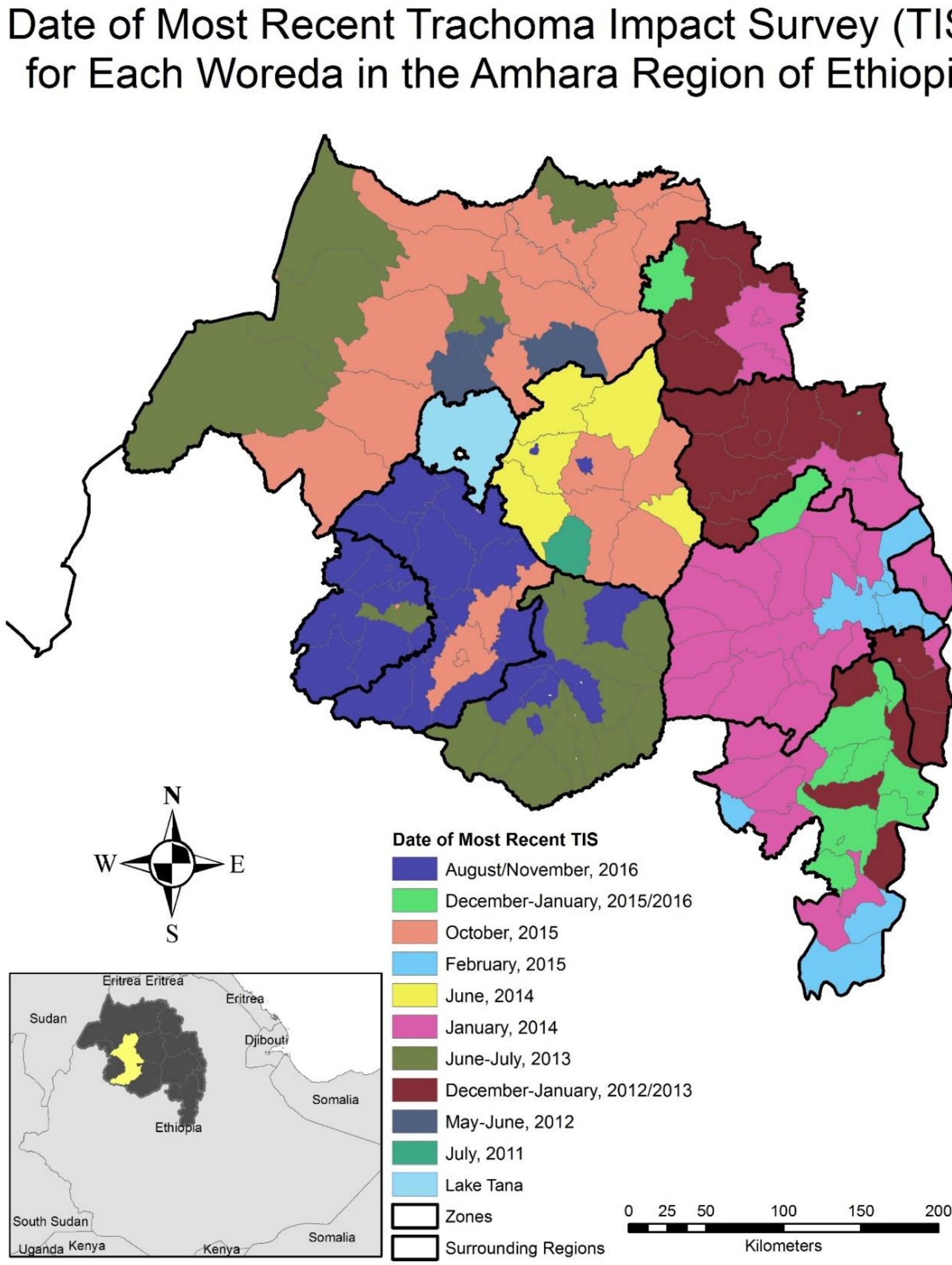


Figure 1. Woredas represented by survey date throughout the Amhara Region of Ethiopia

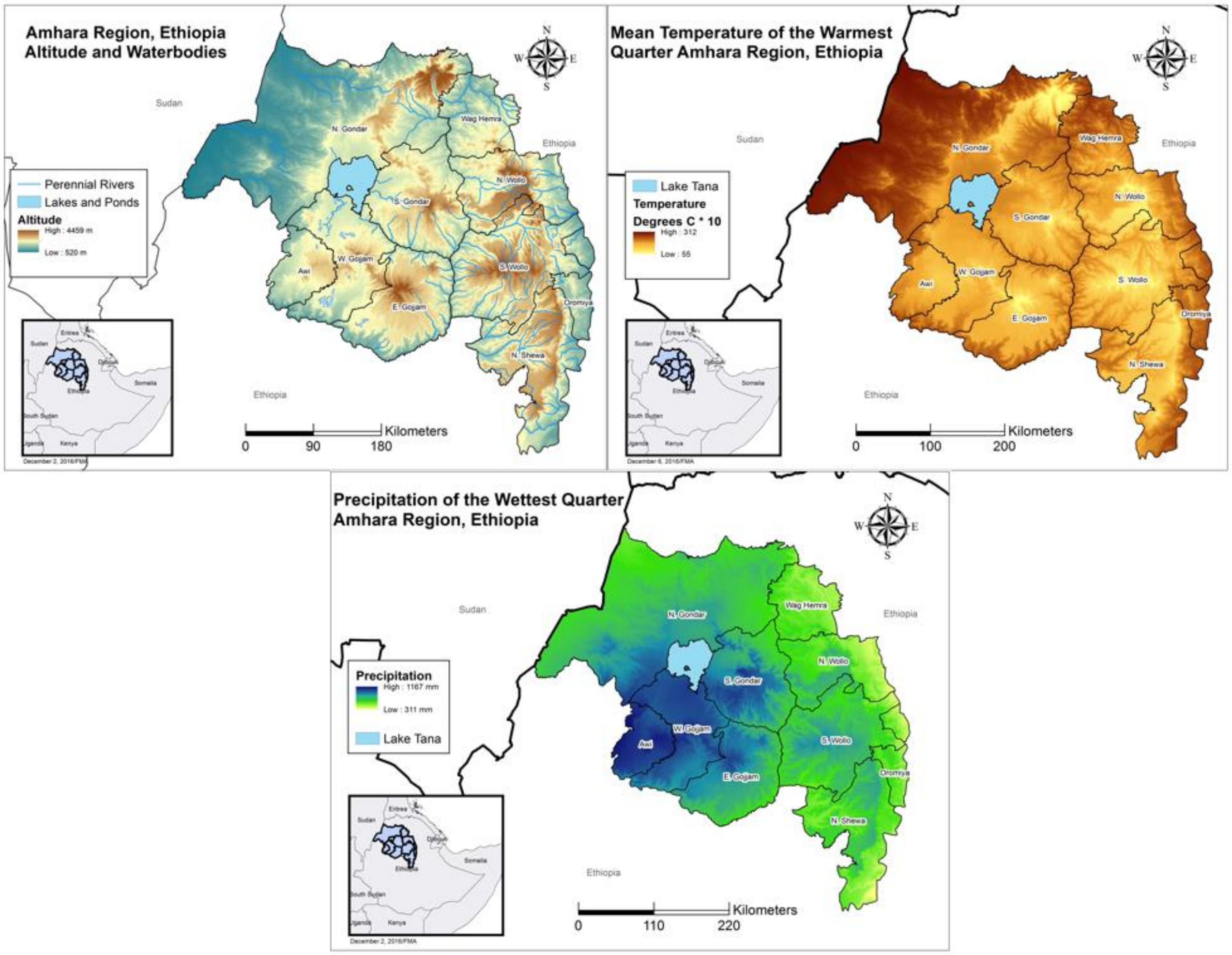


Figure 2. Geographic distribution of climate variables in the Amhara National Regional State derived from interpolated surfaces obtained from the BioClim open source data

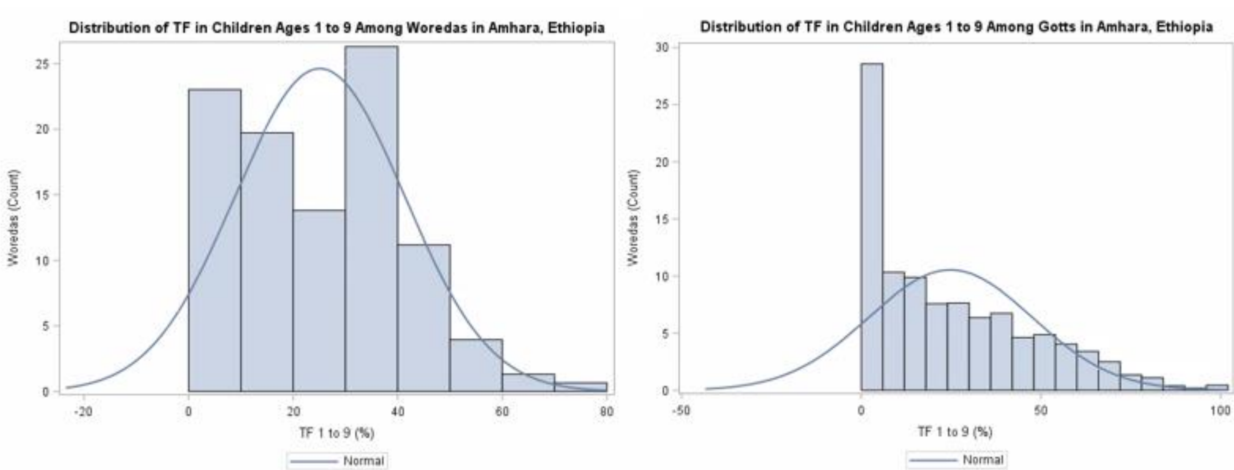


Figure 3. Distribution of TF among children in comparing the distribution between woreda and gotts

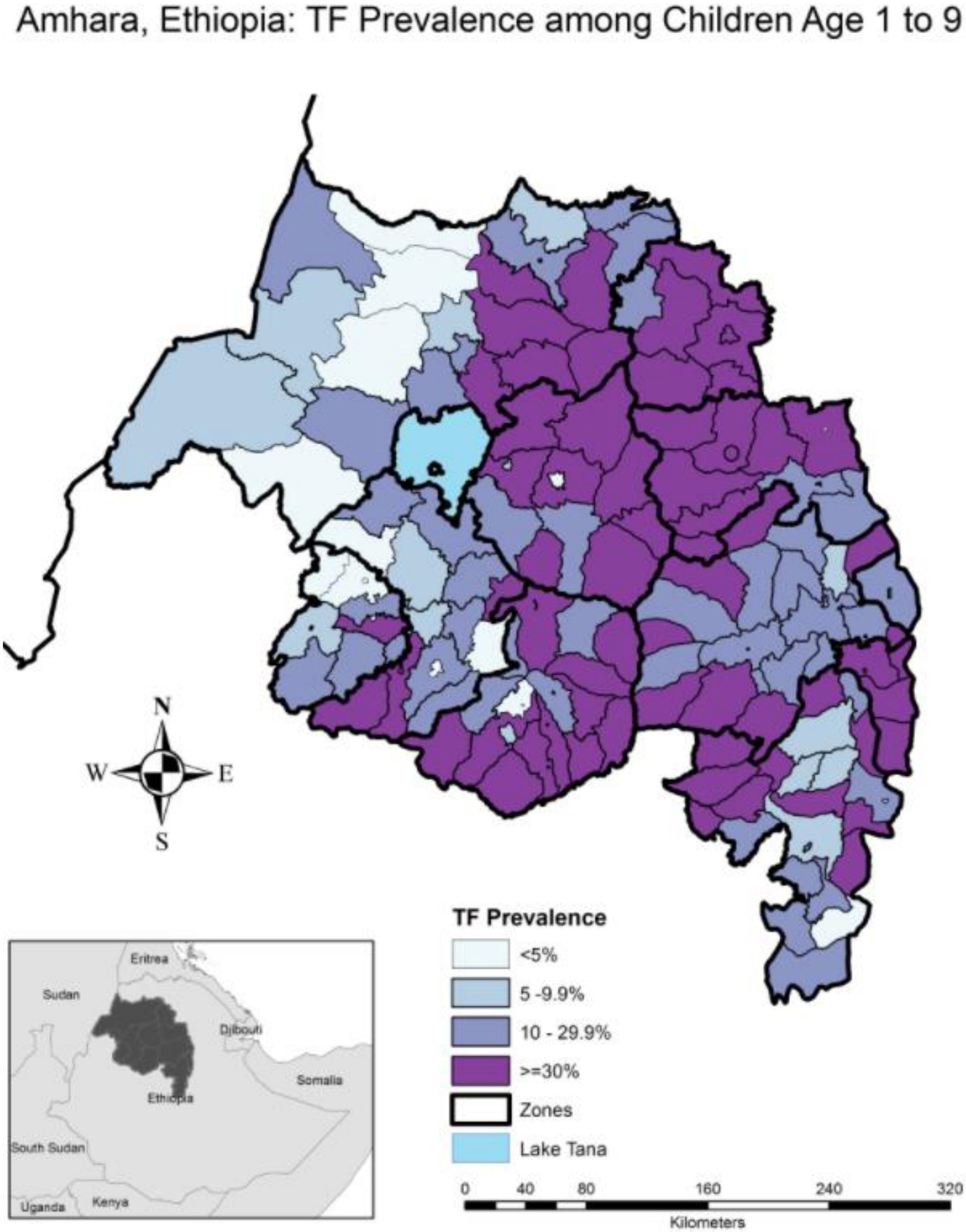


Figure 4. Estimates of trachomatous inflammation follicular among children aged 1 to 9 weighted by the population of the woreda across the Amhara National Regional State of Ethiopia.

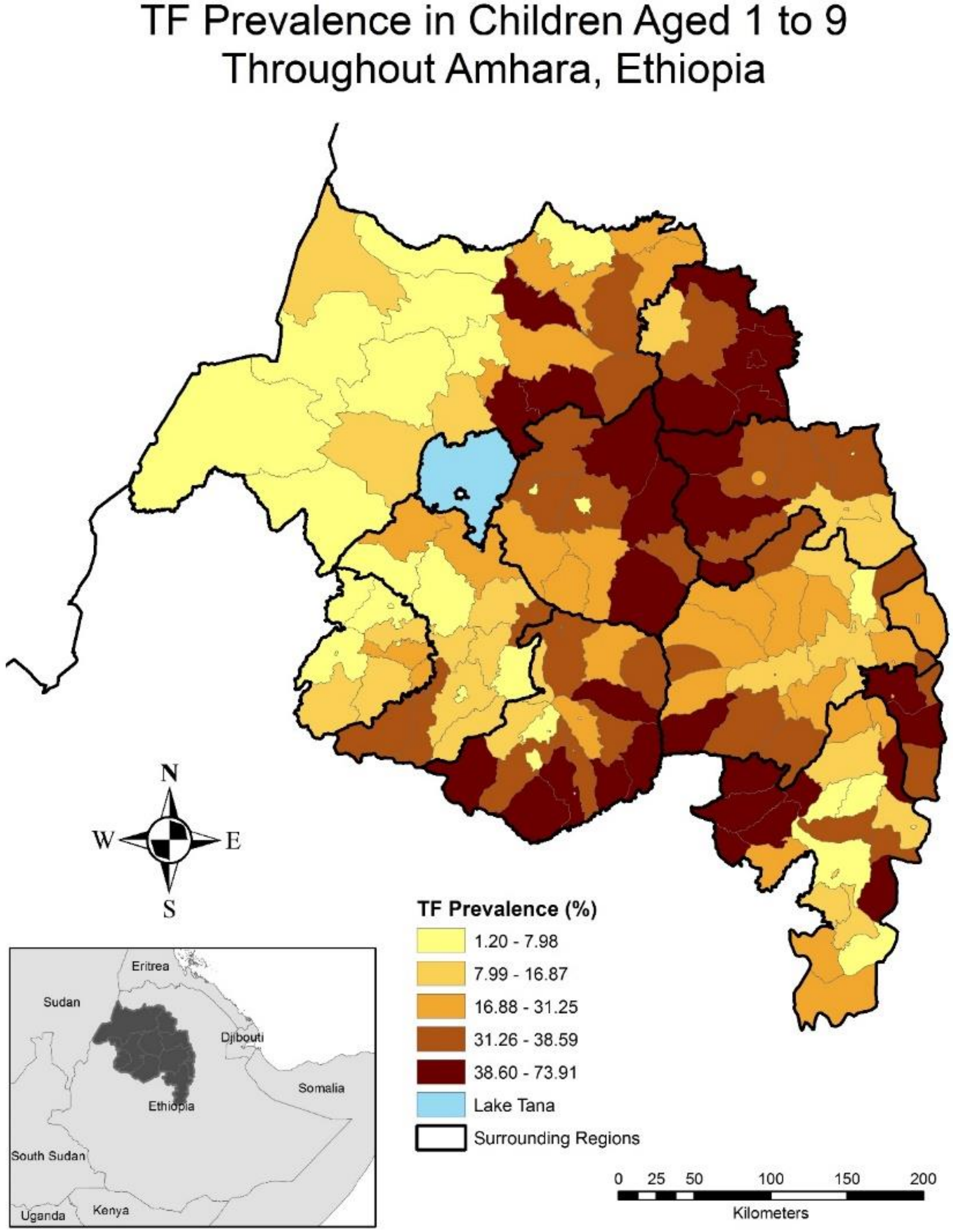


Figure 5. Estimates of trachomatous inflammation follicular among children aged 1 to 9 weighted by the population of the woreda, by quintile across the Amhara National Regional State of Ethiopia.

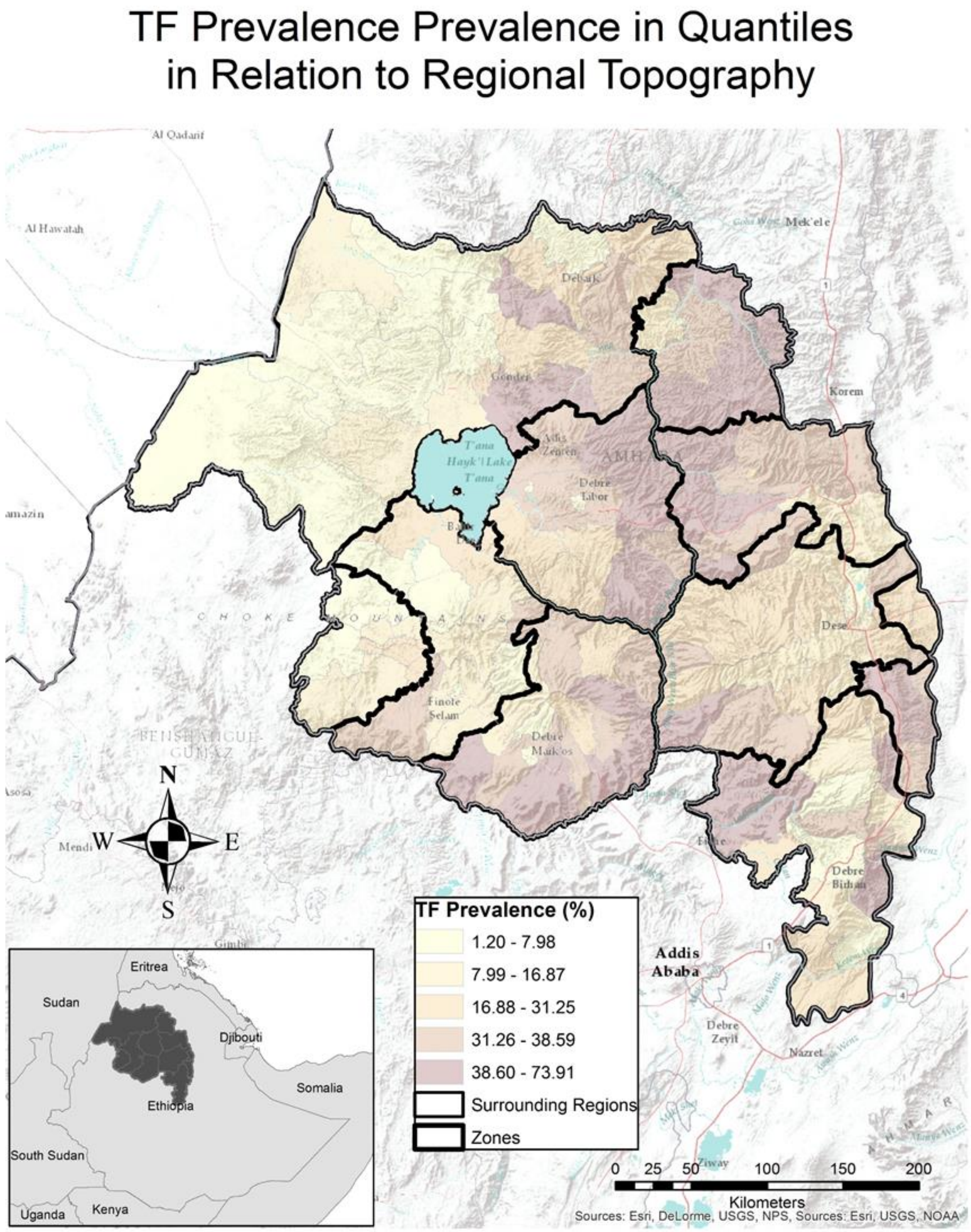


Figure 6. Topographic map of the Amhara Region of Ethiopia with weighted TF prevalence estimates by quintile

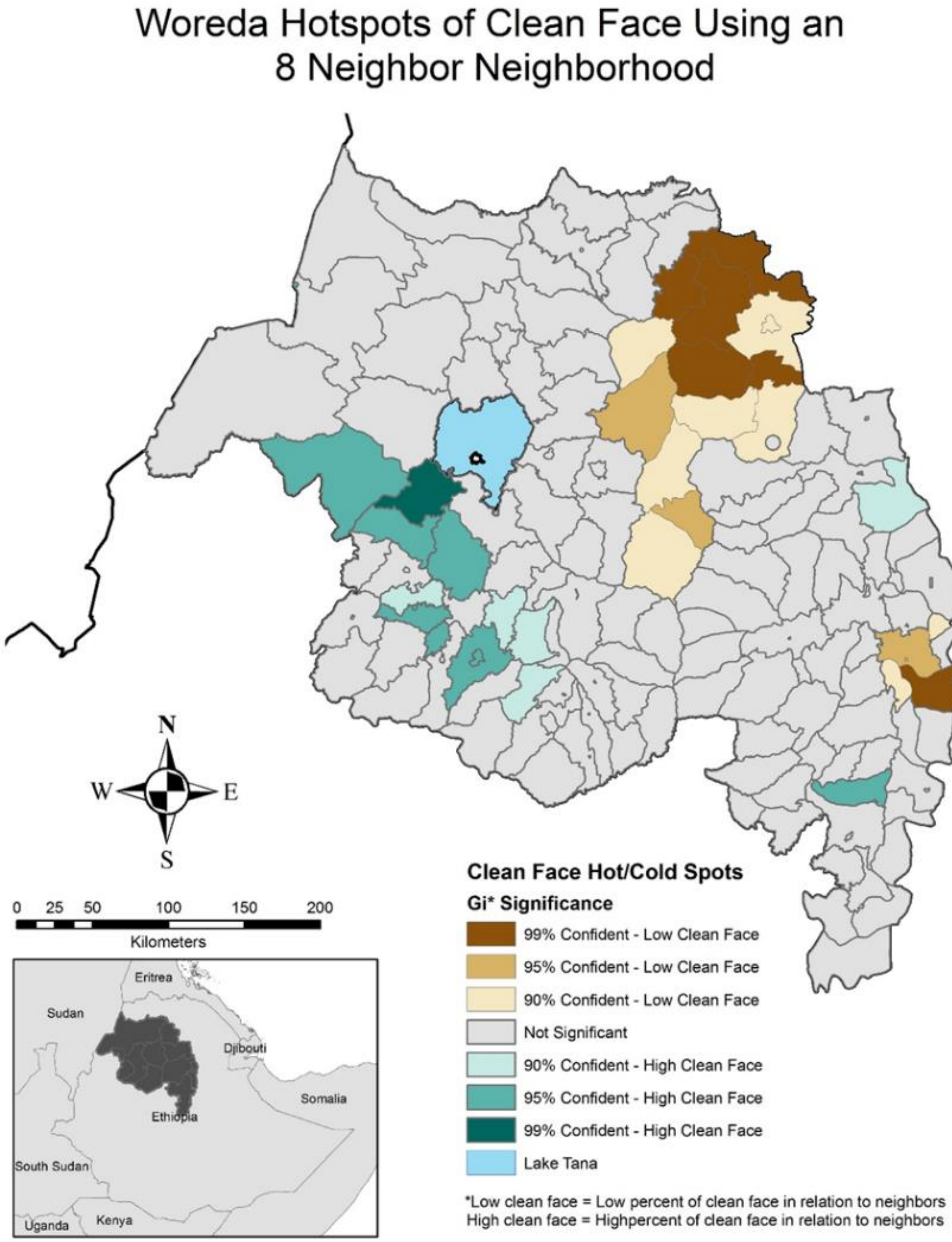


Figure 7. Woredas with high or low weighted percentages of children presenting with clean faces upon examination in comparison to neighboring woredas.

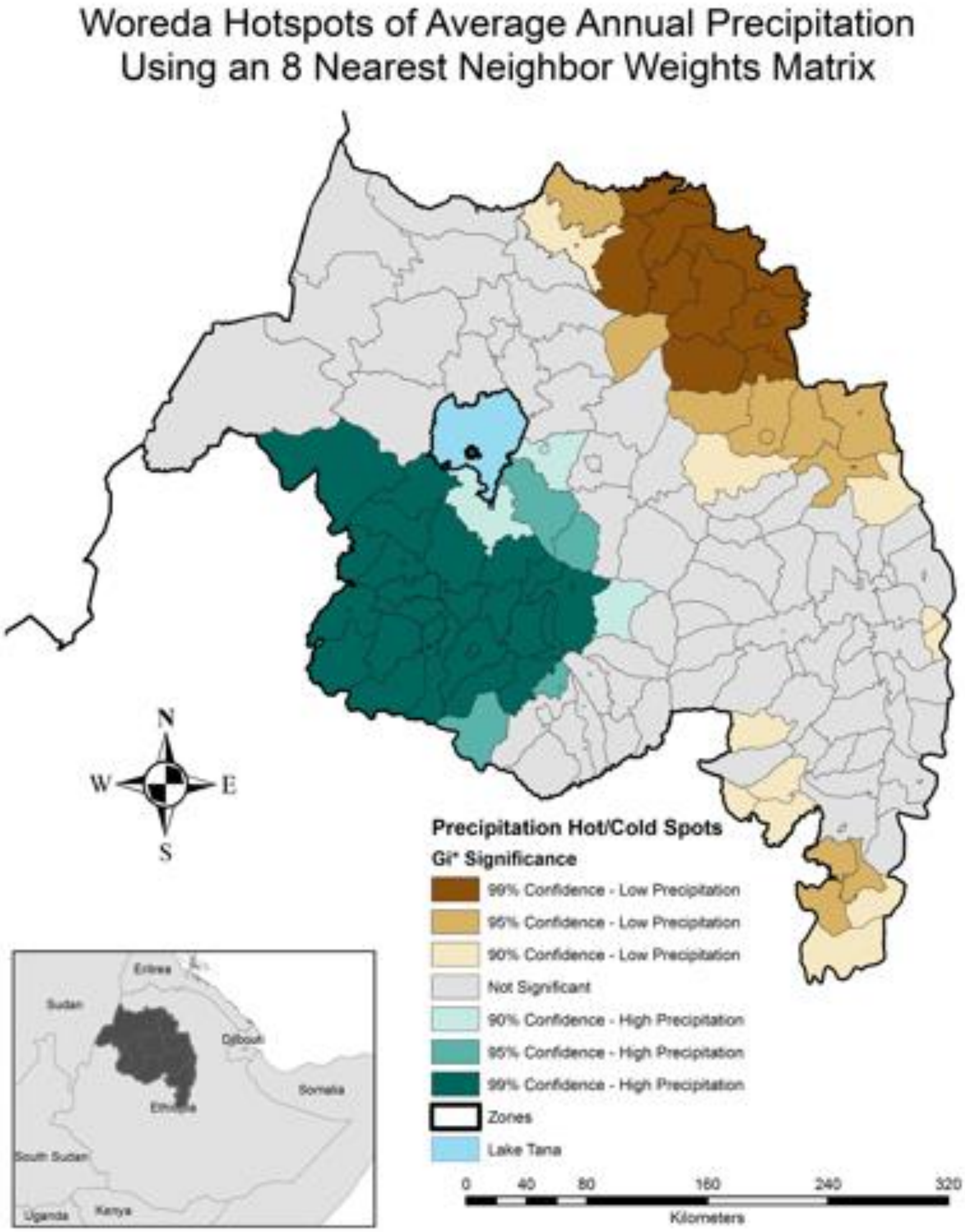


Figure 8. High and low clusters of average annual precipitation in the Amhara Region using the Getis-Ord G_i(d) statistic to compare each woreda to its neighbors*

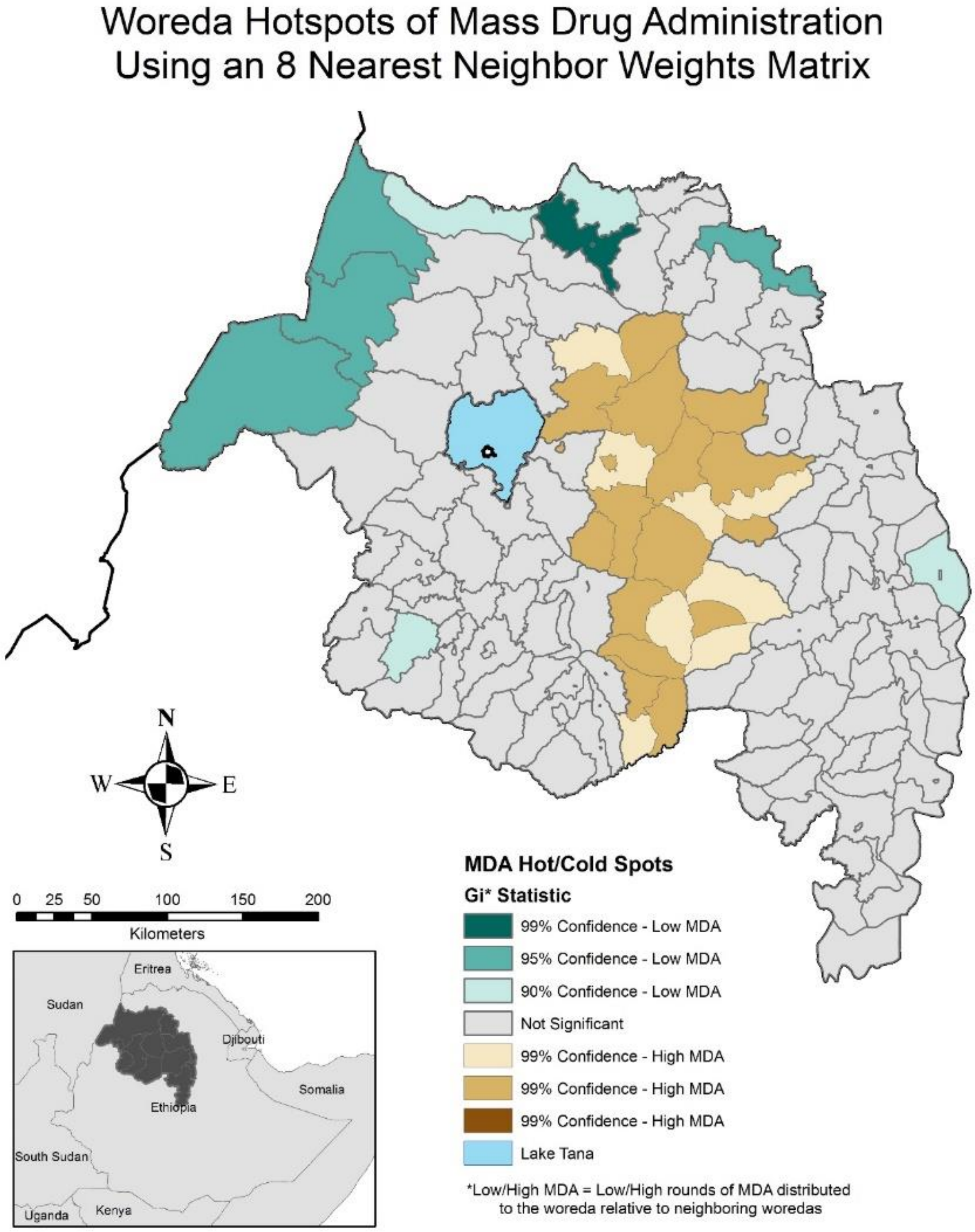


Figure 9. Clusters of high and low counts for the numbers of rounds of antibiotics distributed in each woreda.

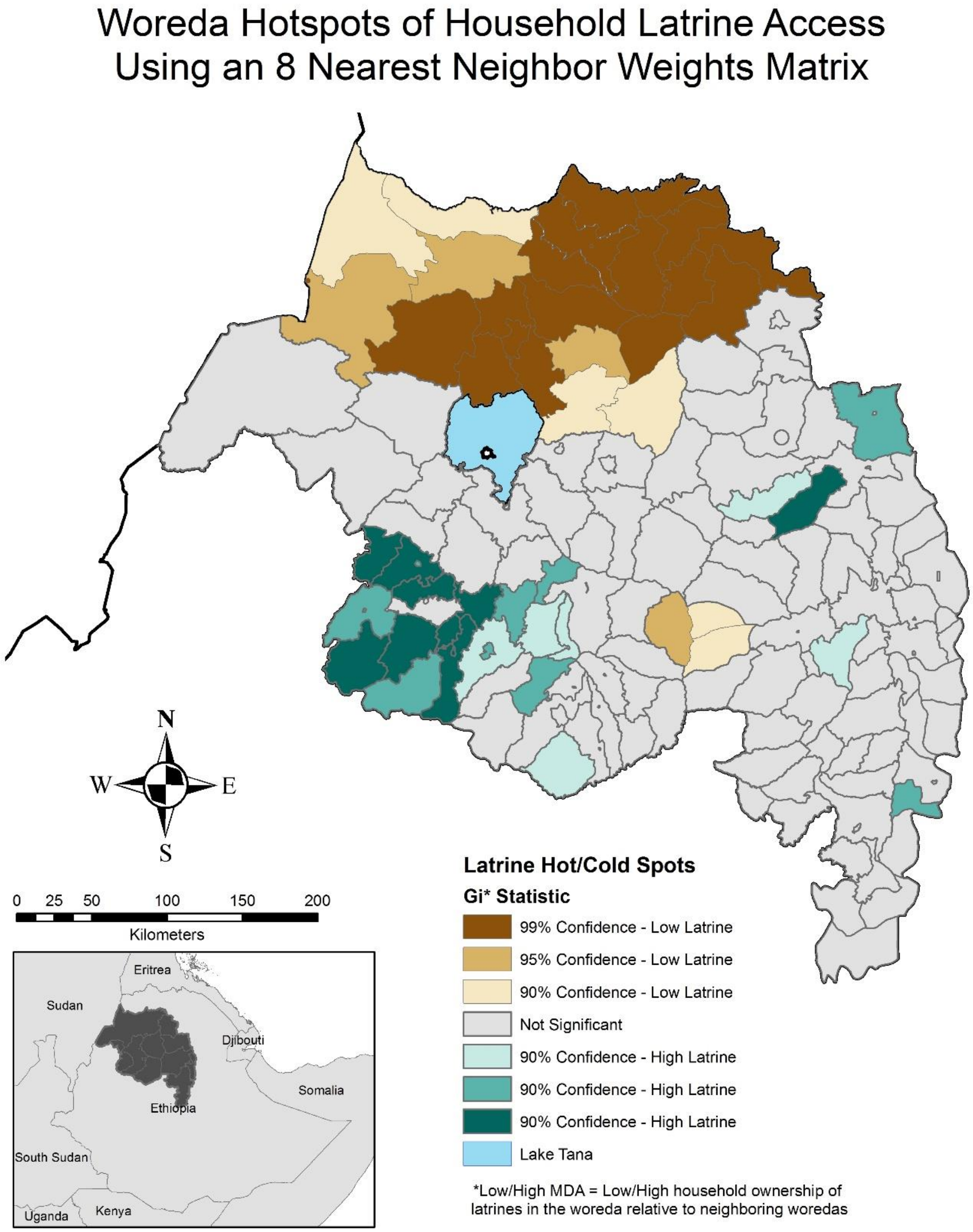


Figure 10. Clusters of high and low household access to any latrine throughout the woredas of Amhara

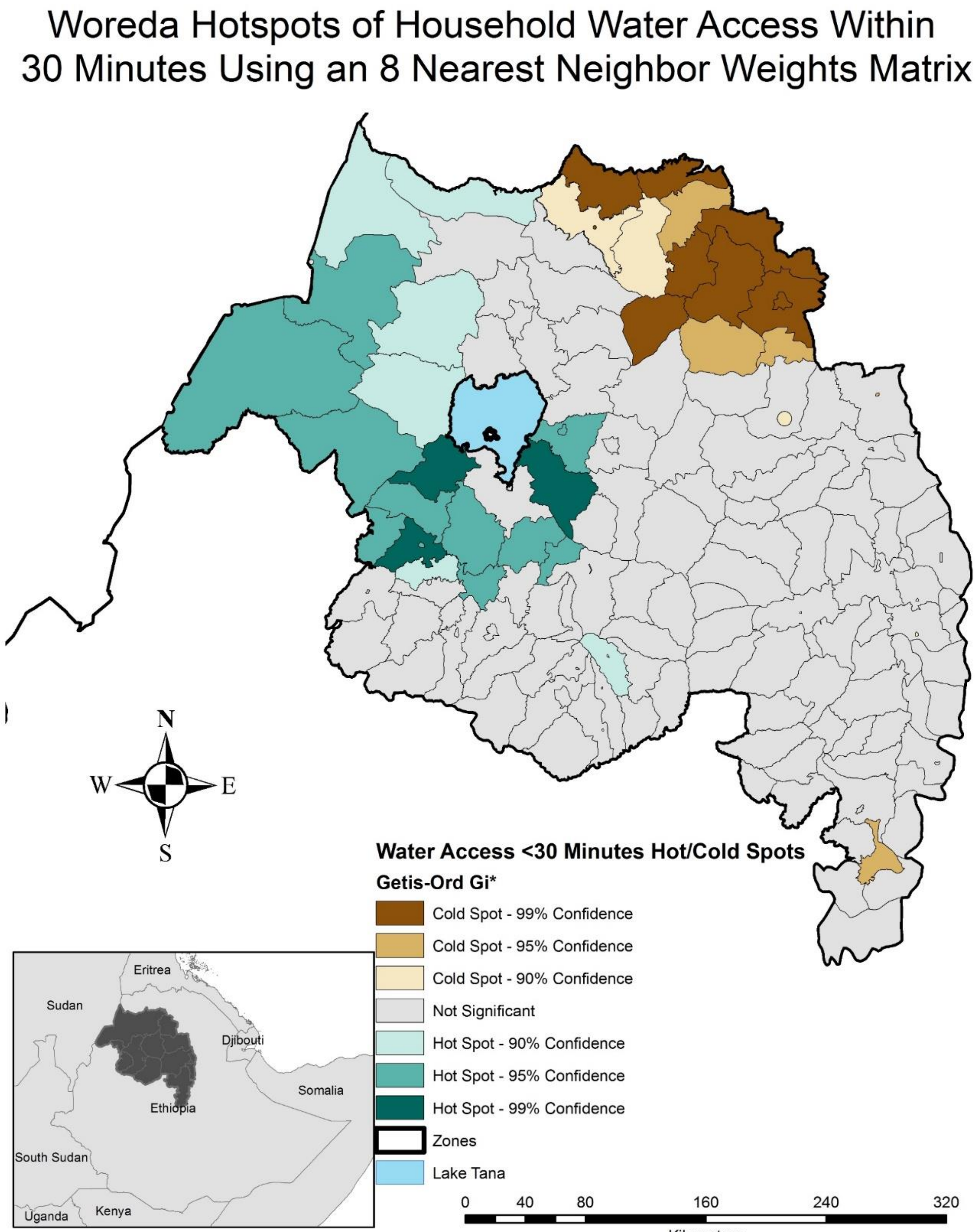


Figure 11. Clusters of high and low access to a water source within 30 minutes of a household across the woredas of Amhara. Cold spots are synonymous with high household access to water.

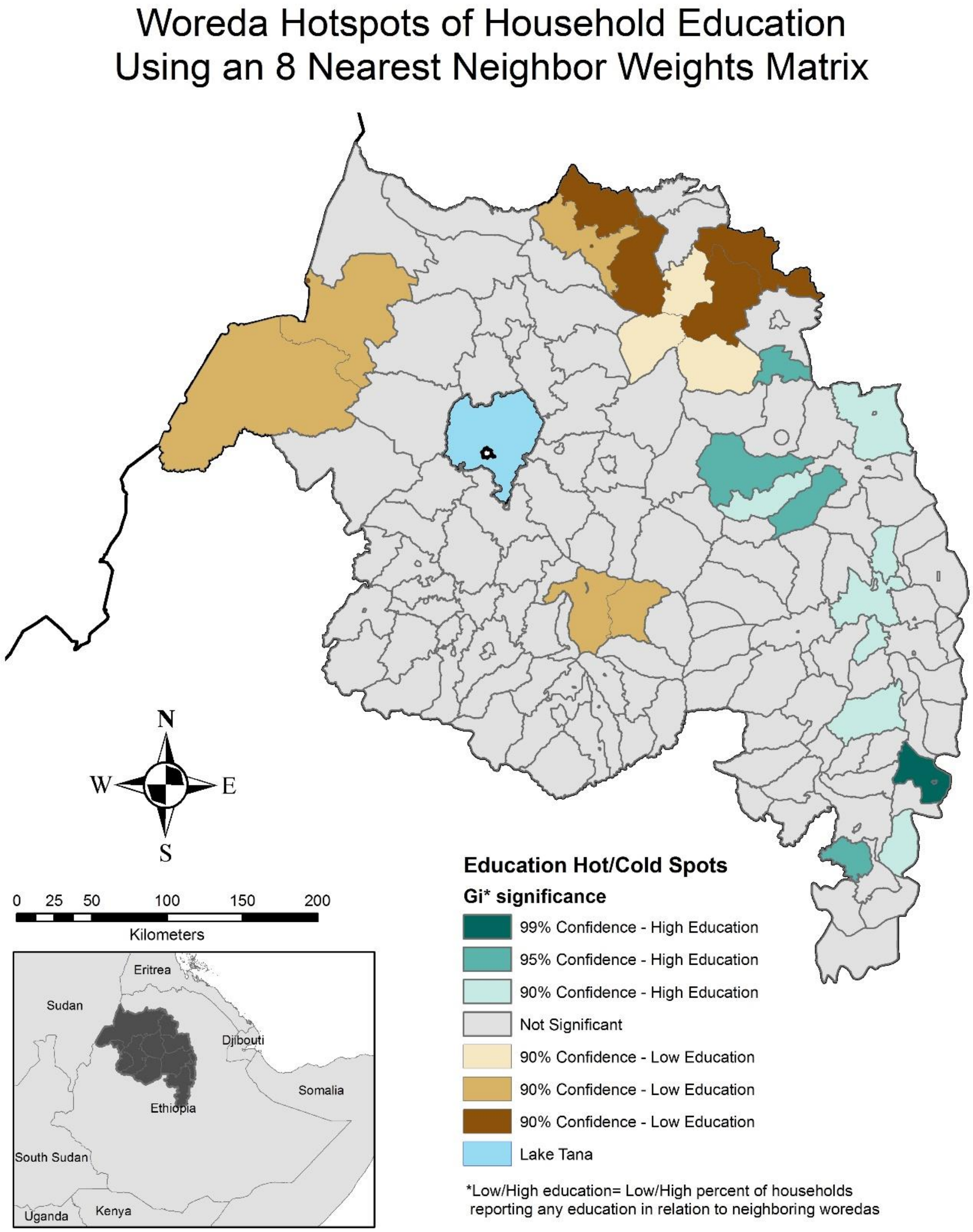


Figure 12. Clusters of high and low percentage of families having reported any formal education within the household

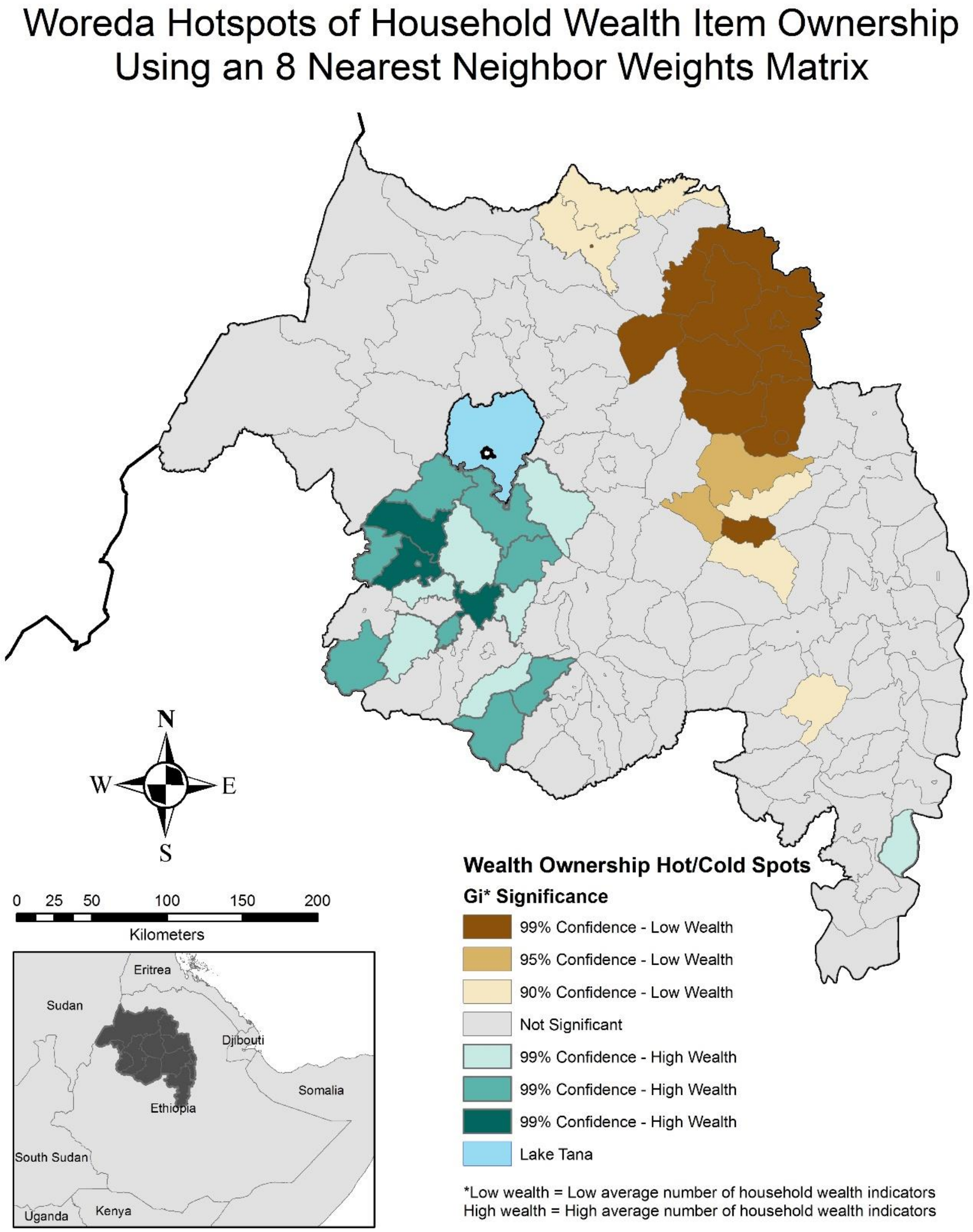


Figure 13. Clusters of high and low average numbers of wealth indicators within surveyed households in the Amhara region of Ethiopia

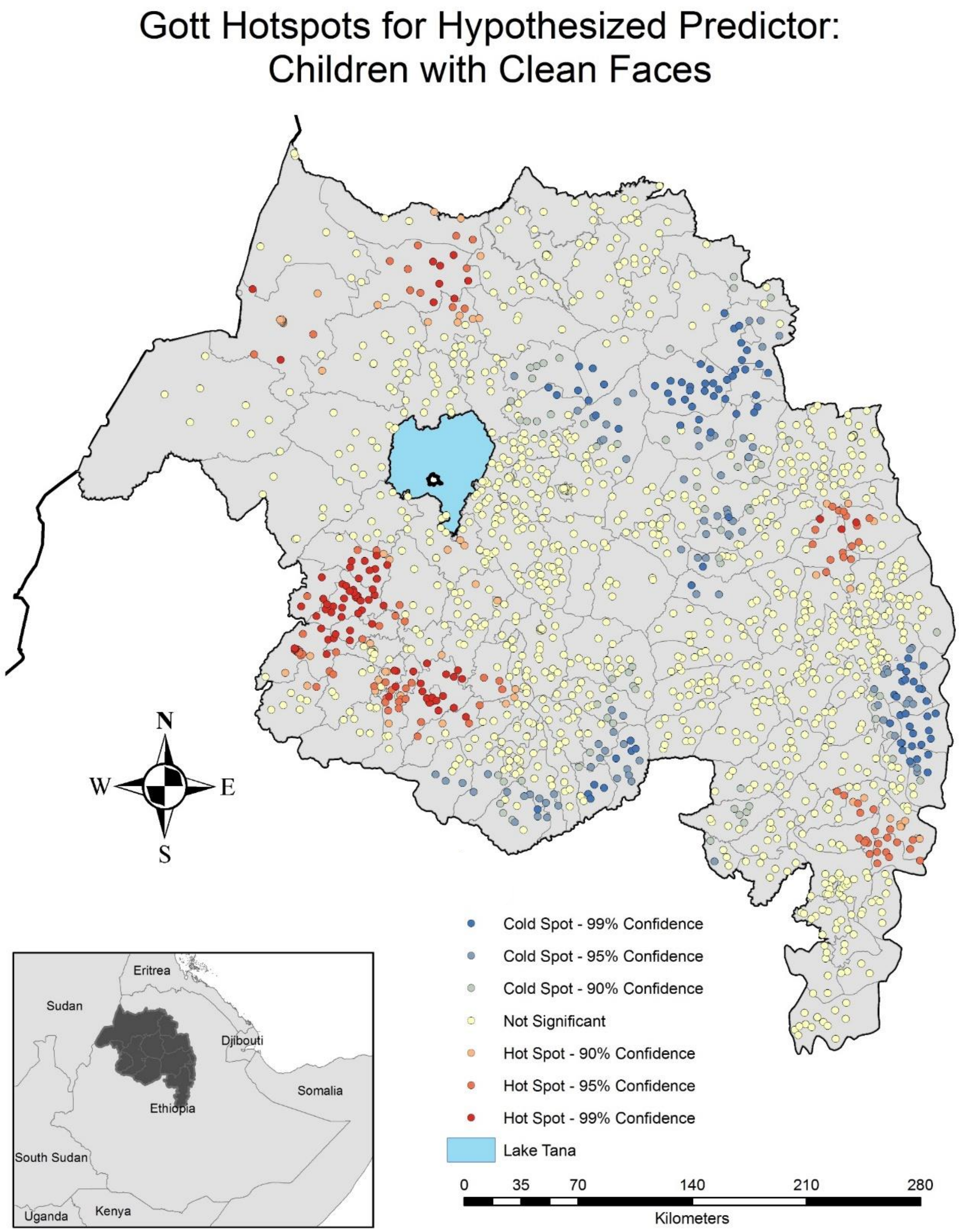


Figure 14. Village level clusters of high and low percentage of clean faces

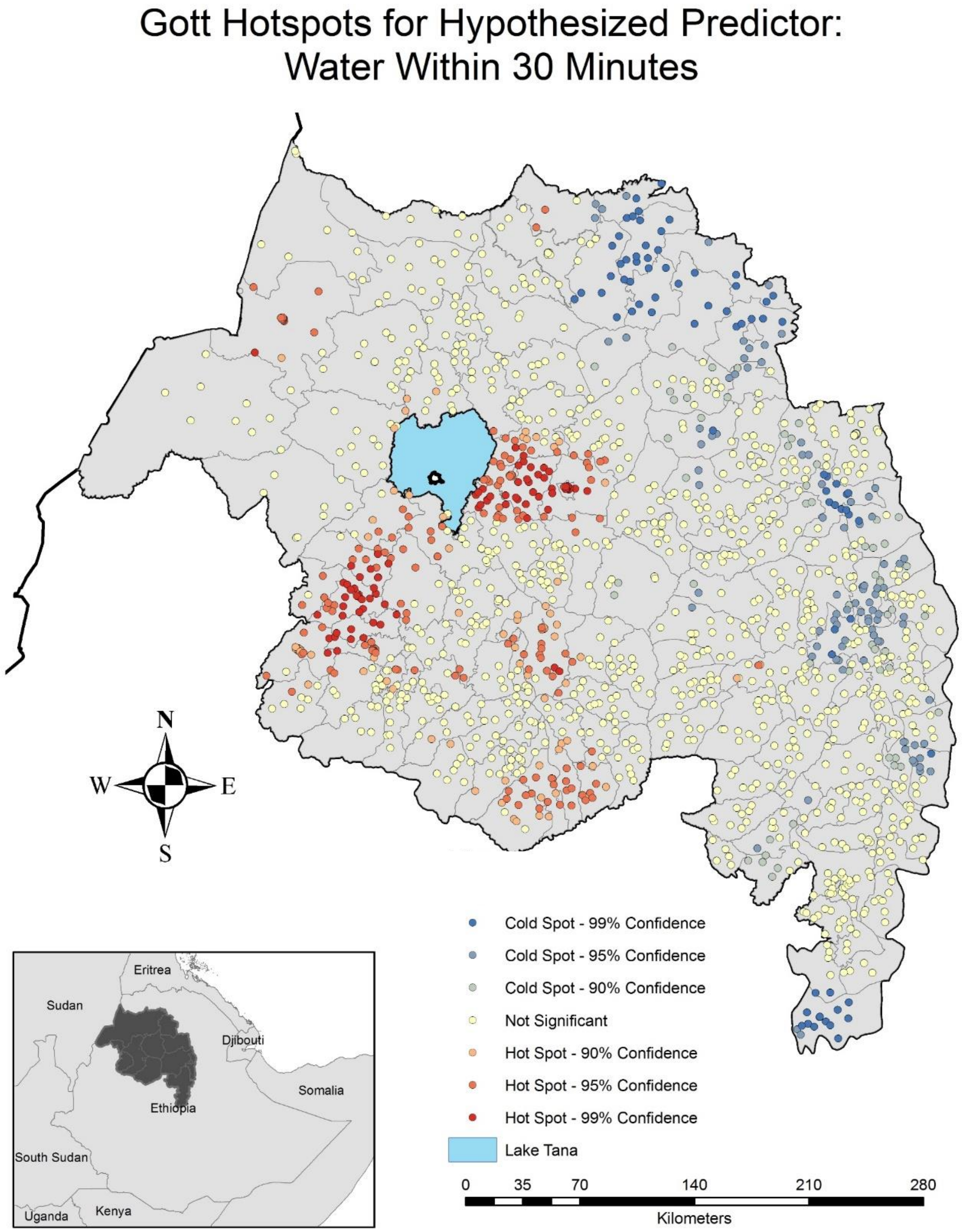


Figure 15. Gott level clusters of households with access to any water source within 30 minutes of the home

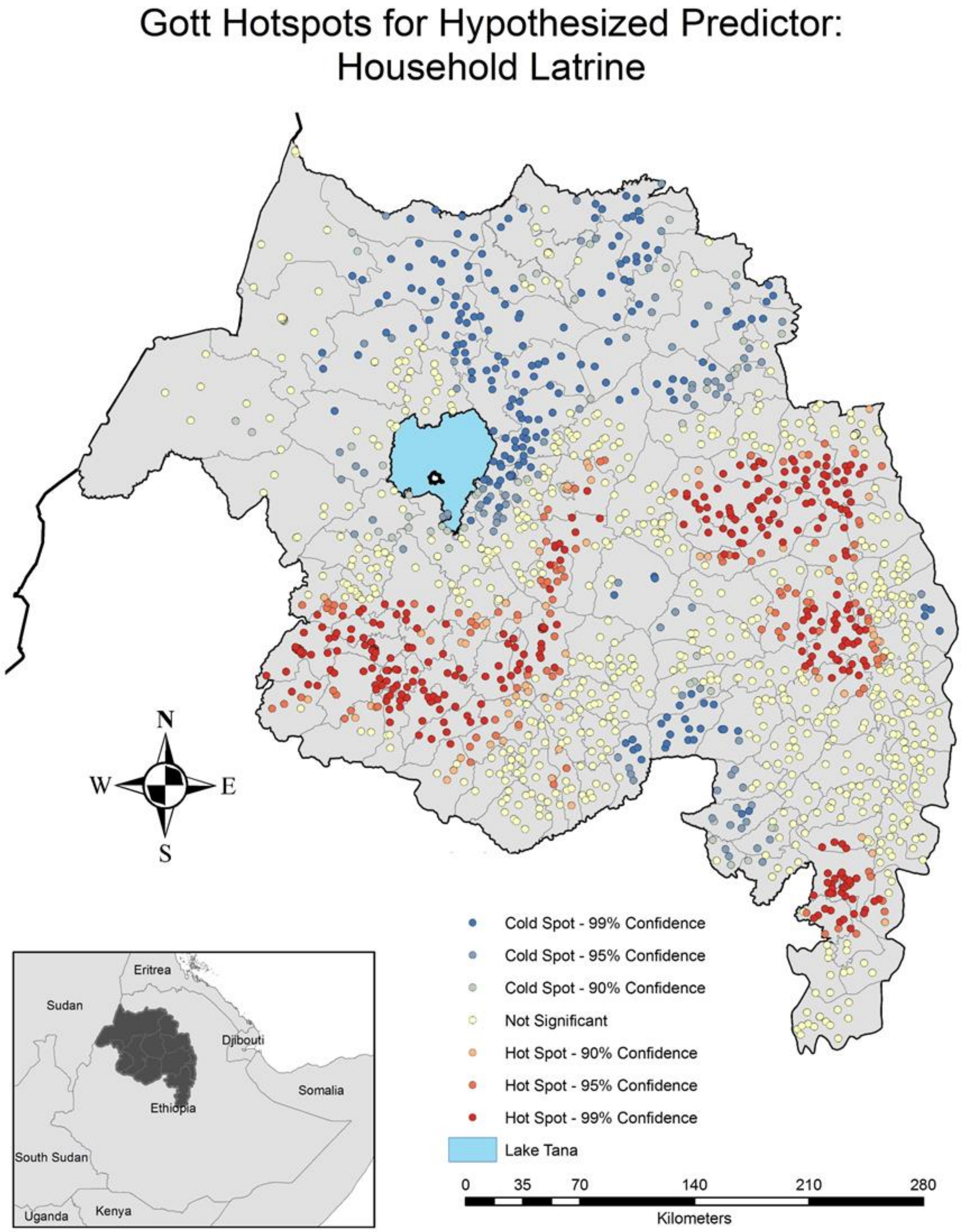


Figure 16. Gott level clusters of household access to any latrine throughout the Amhara Regional State of Ethiopia

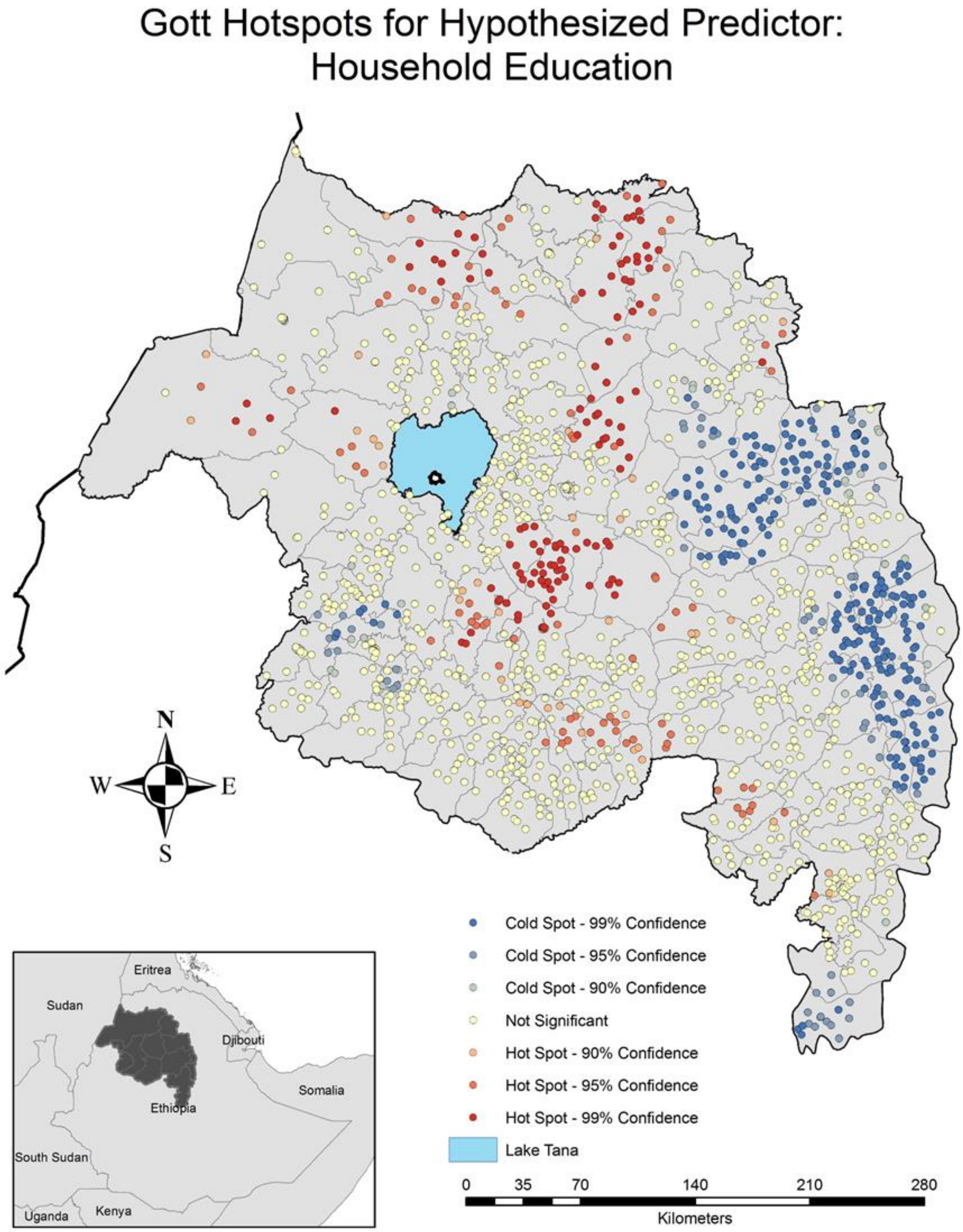


Figure 17. Gott level clusters of any household member reporting having received formal education

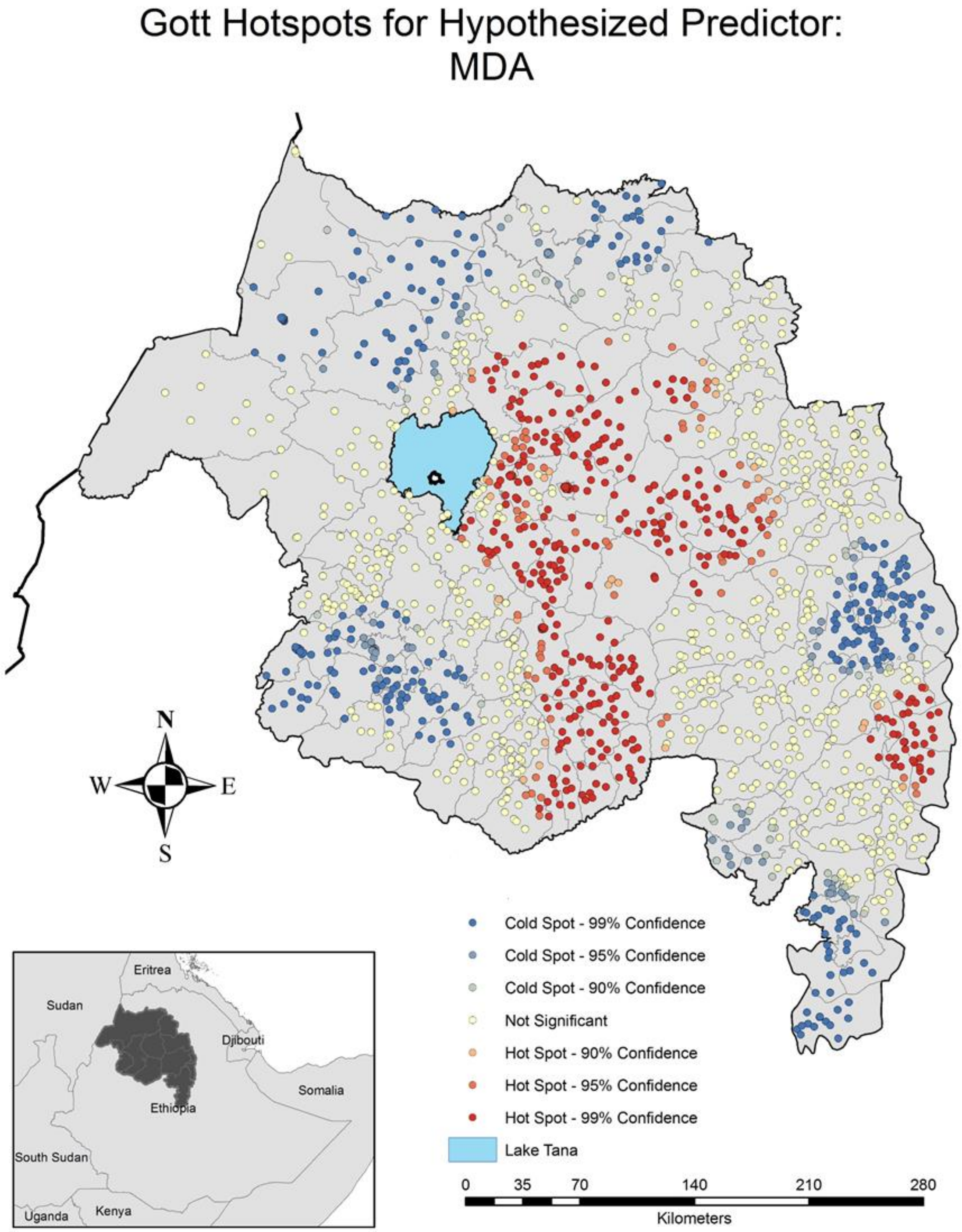


Figure 18. Gott level clusters of rounds of mass drug administration provided to the woredas

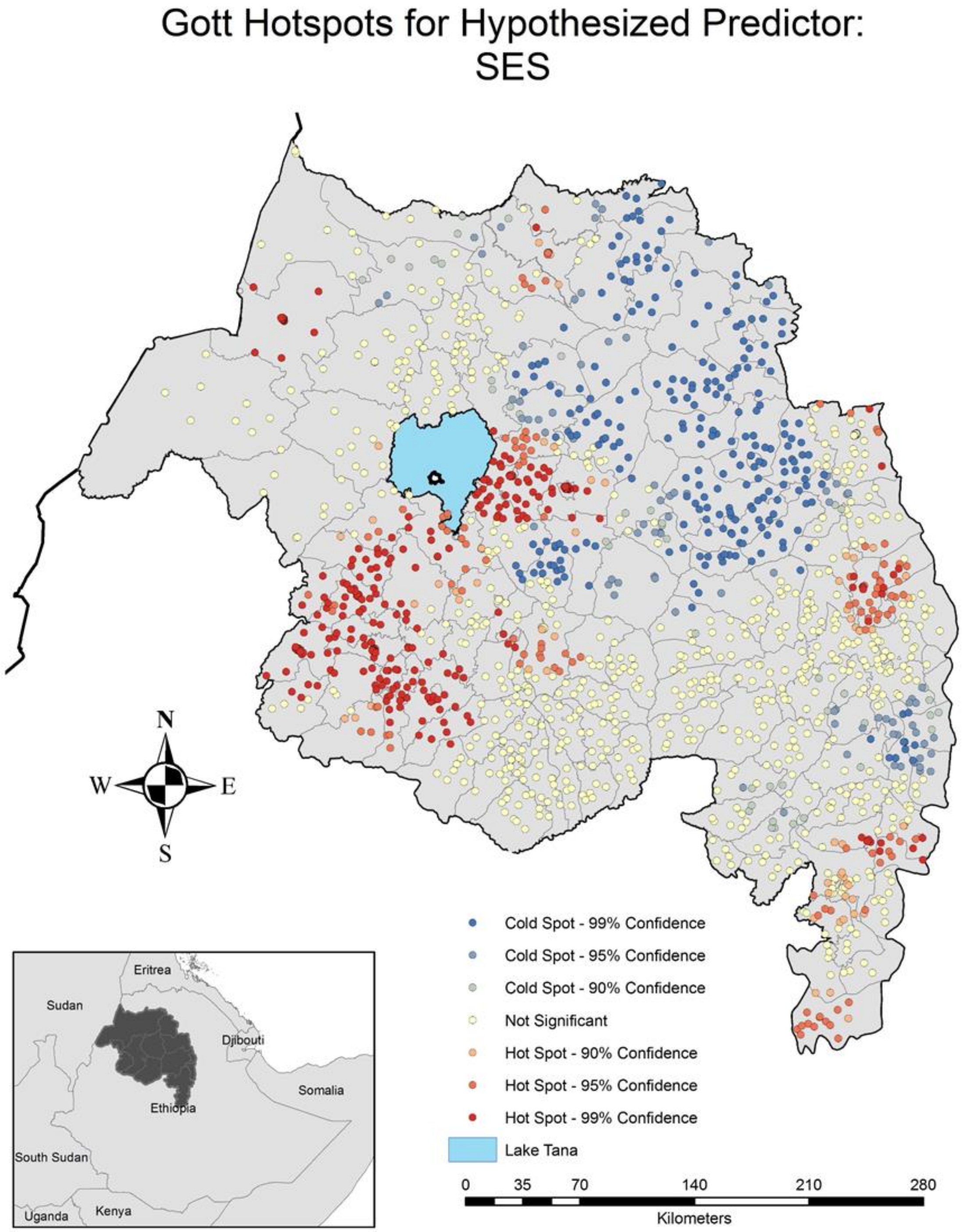


Figure 19. Gott level clusters of high and low average proportion of household wealth indicator items owned

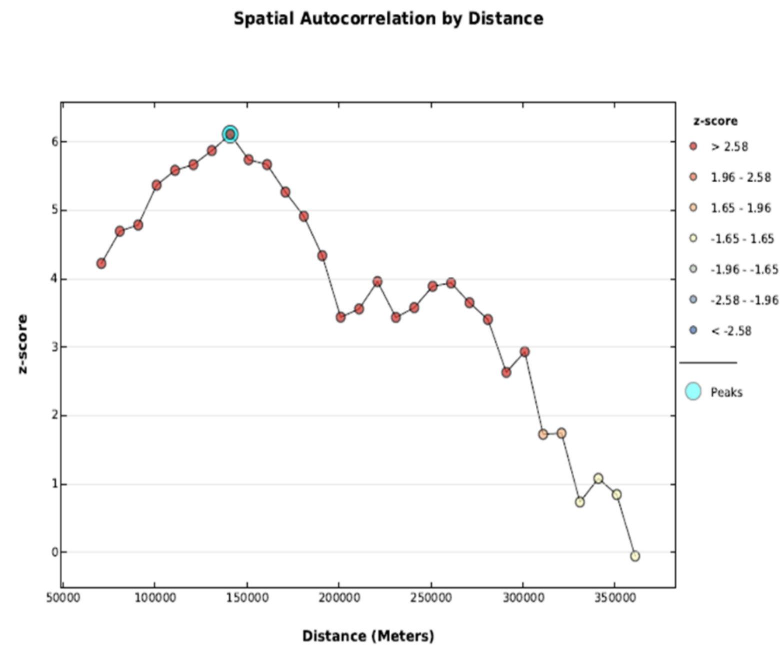


Figure 20. Incremental spatial autocorrelation of the residuals from the ordinary least squares regression model used to predict the prevalence of TF in children aged one to 9. The highlighted point at the peak of the curve indicates the highest Moran's I z-score

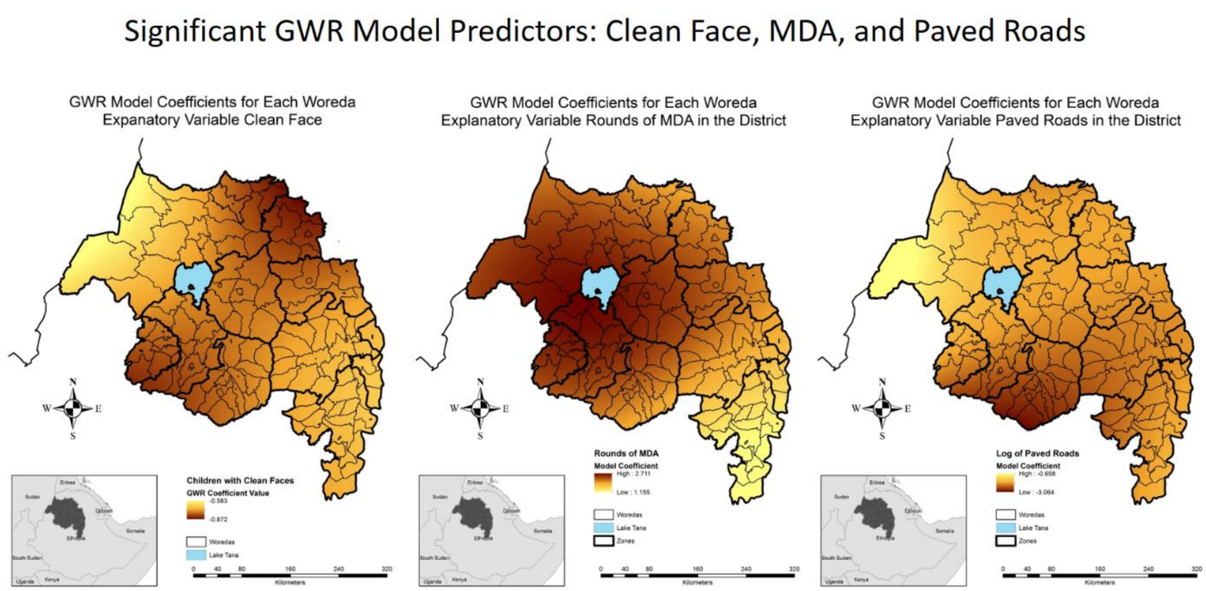


Figure 21. Significant explanatory factors from the geographically weighted regression model showing the regions with the strongest effect from the predictor in darker coloration.

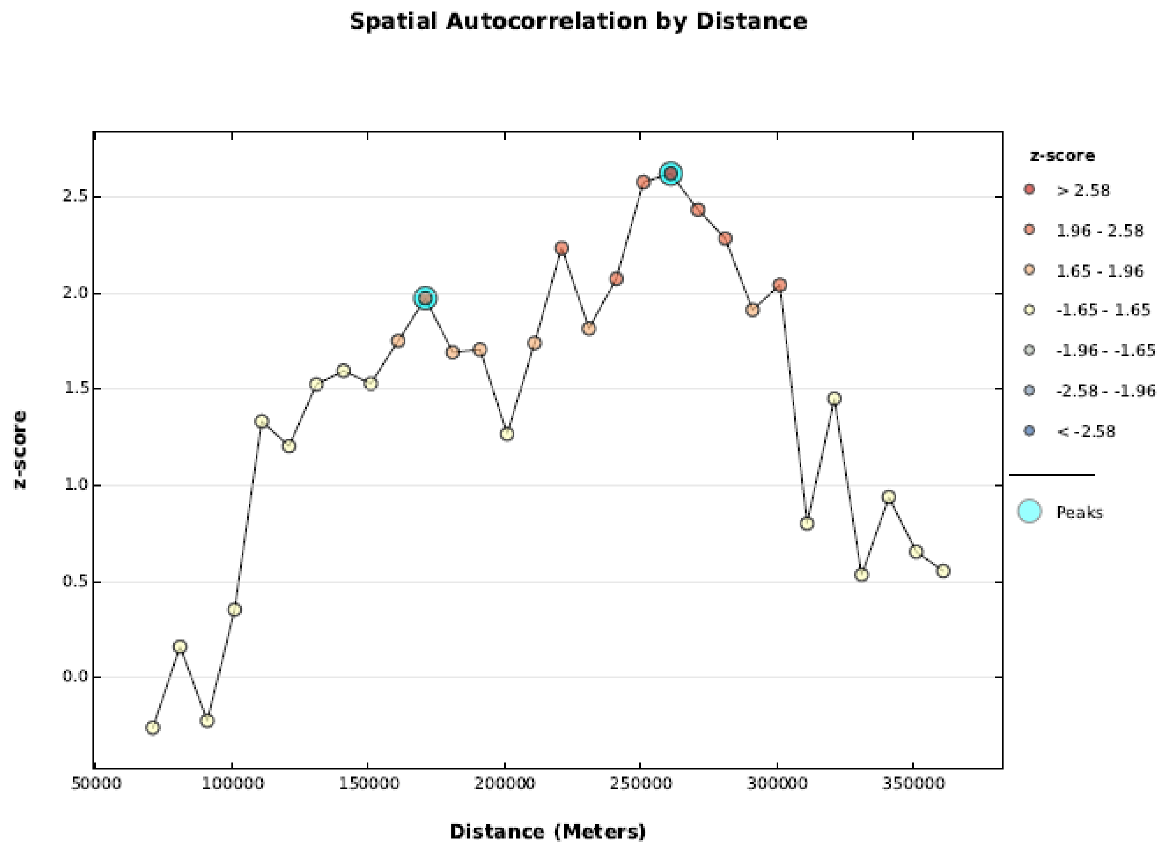


Figure 22. Spatial autocorrelation of the residuals for the spatial lag model that was identified as the best fit linear form of a model that explains TF in children aged 1 to 9 while accounting for spatial autocorrelation. The highlighted peaks in the figure indicate where the effect of spatial autocorrelation is most inherent.

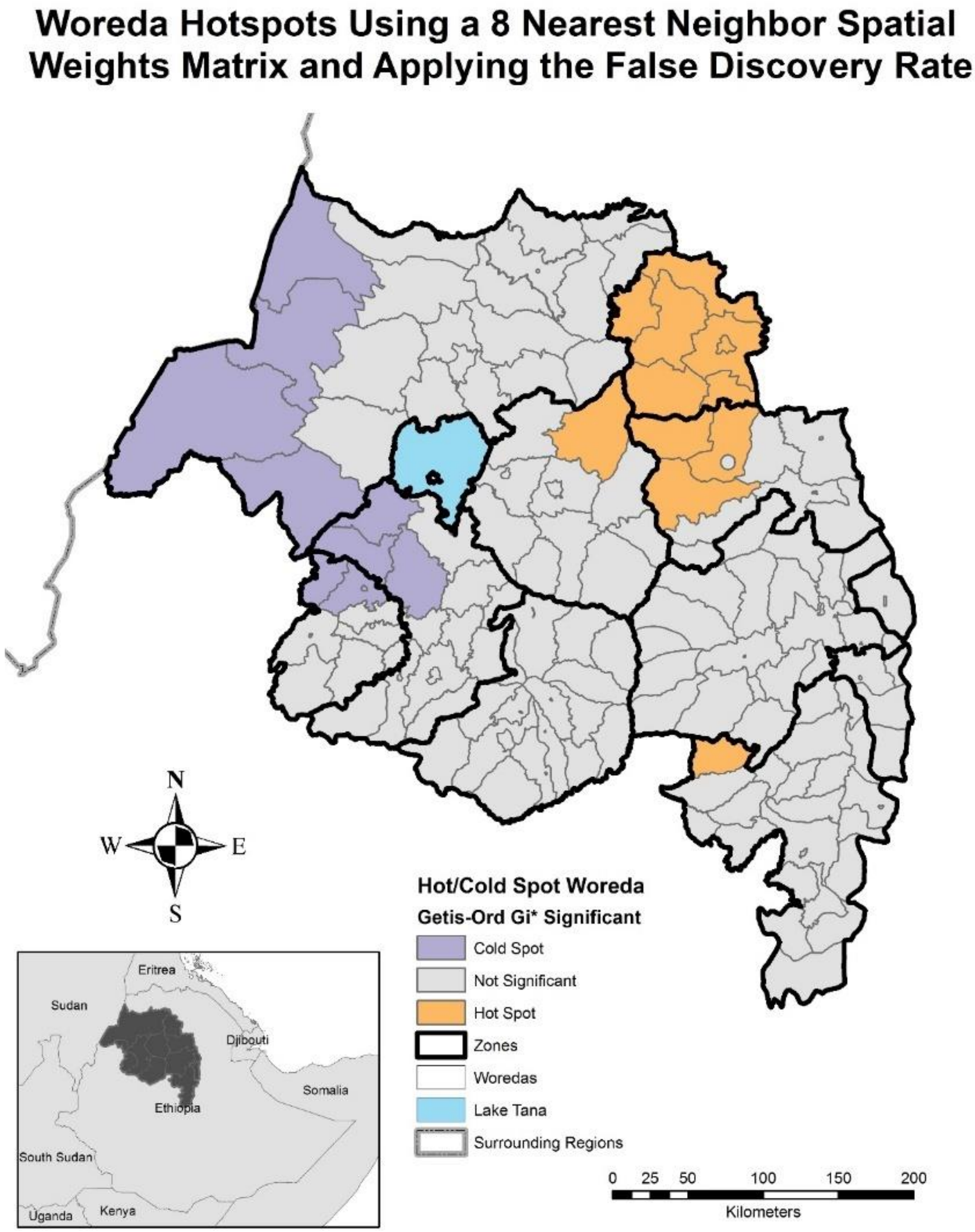


Figure 23. Woreda hot and cold spots of TF prevalence in children aged 1 to 9

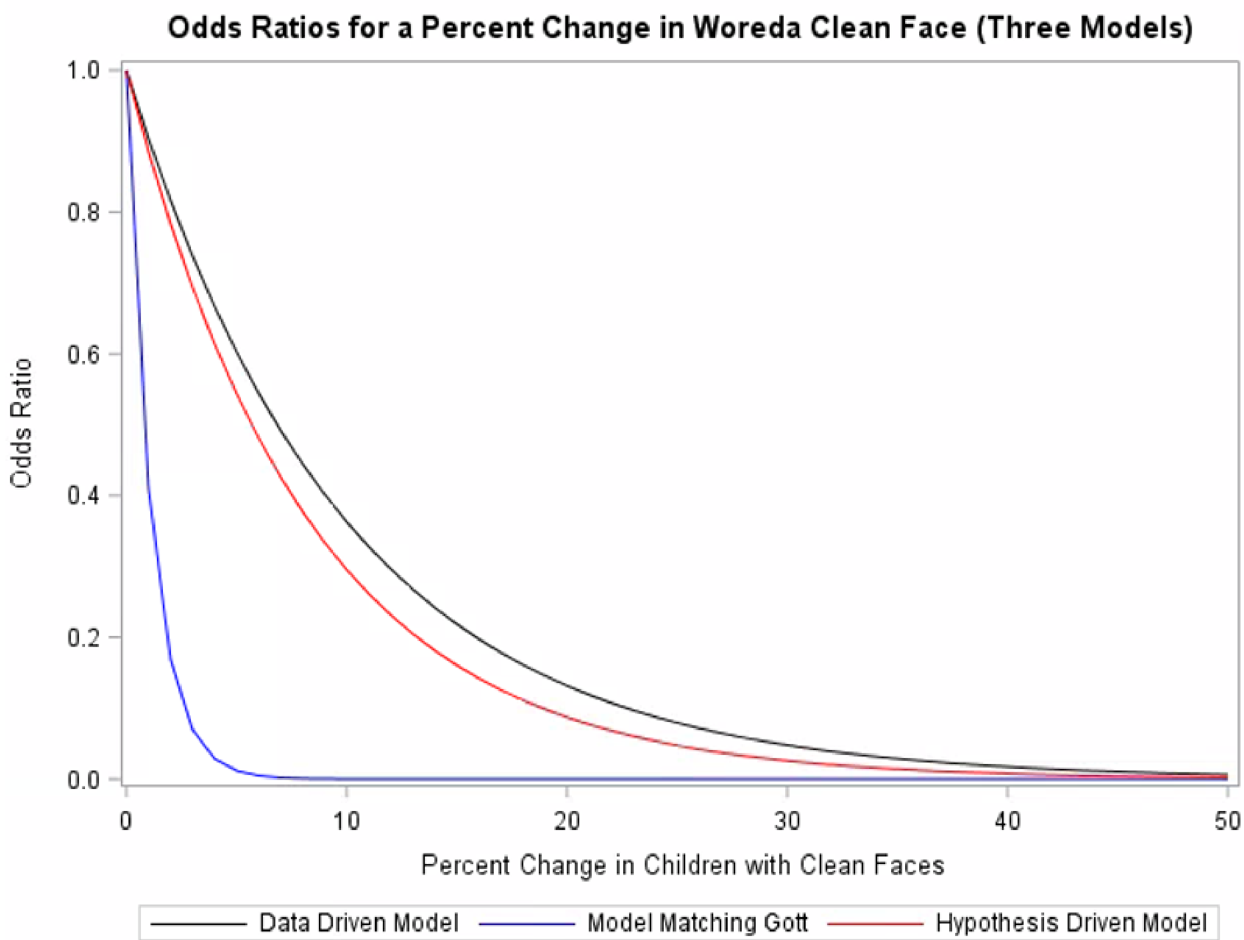


Figure 24. Odds ratios for a 1% increase in clean face at the woreda level controlling for confounders in three models

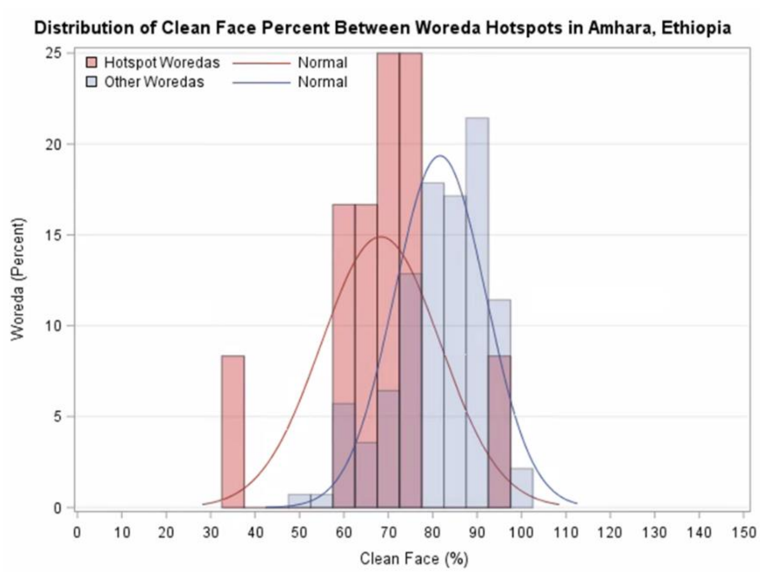


Figure 25. The distribution of the weighted percent of children presenting with clean faces in woredas that are statistically significant hotspots compared against all other woredas

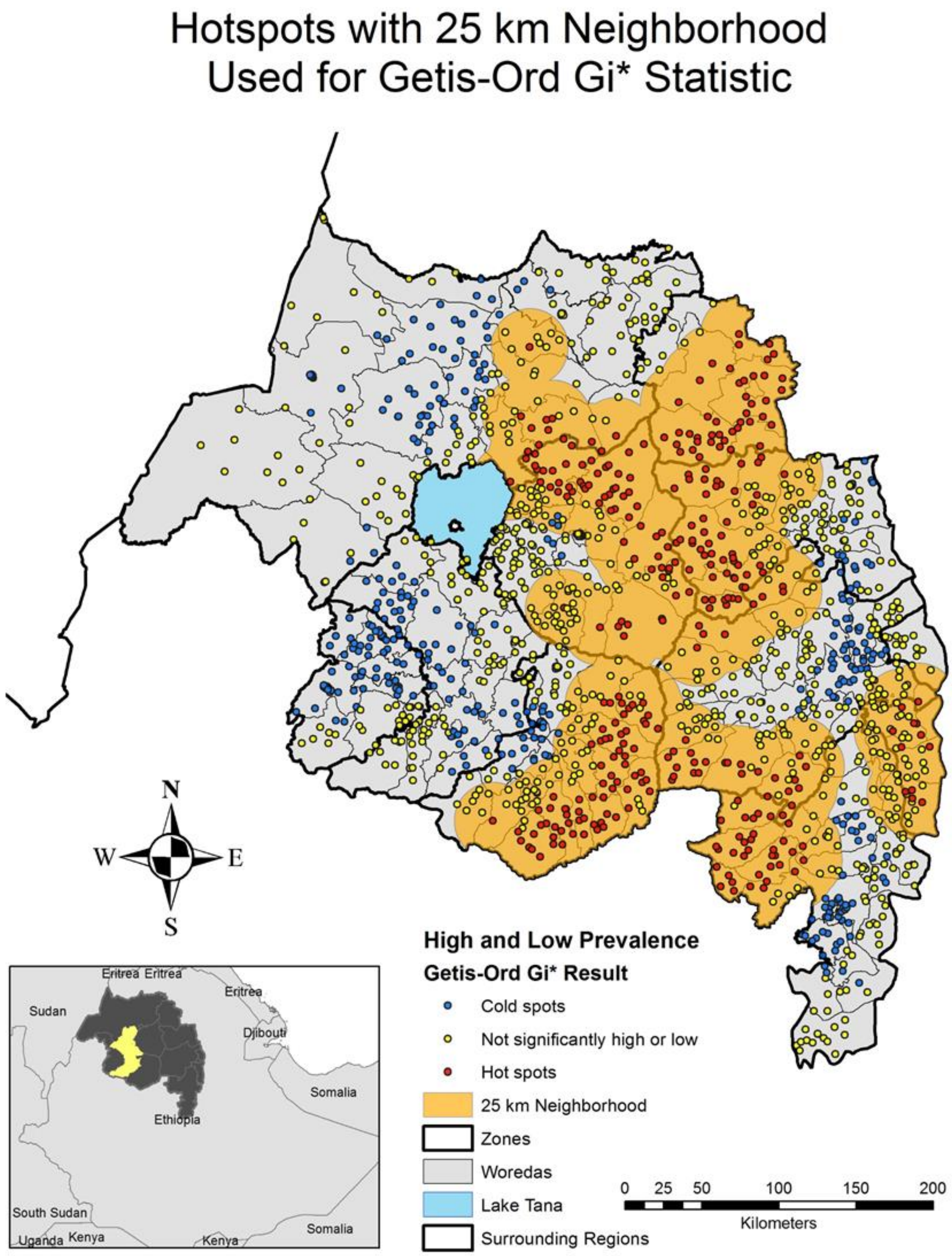


Figure 26. Geographic distribution of gotts and the neighborhood used to search for statistical significance of clustering similar TF prevalence values.

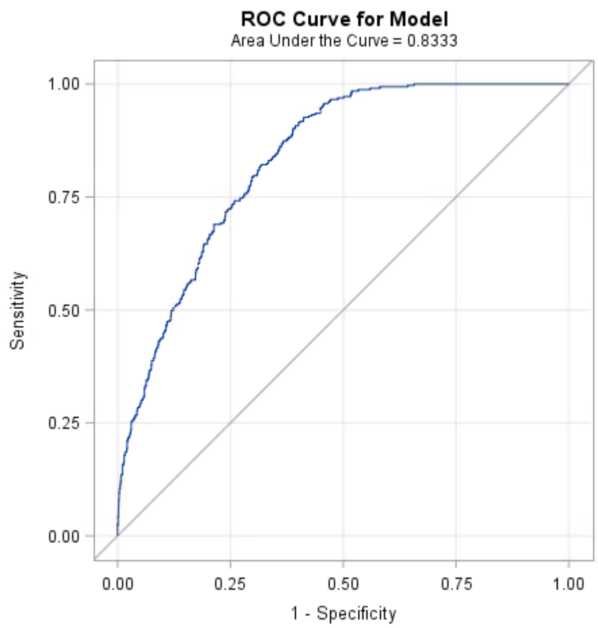


Figure 27. Receiver operator curve to test the discriminatory performance of the selected logistic model for the prediction of residence in a hotspot gott

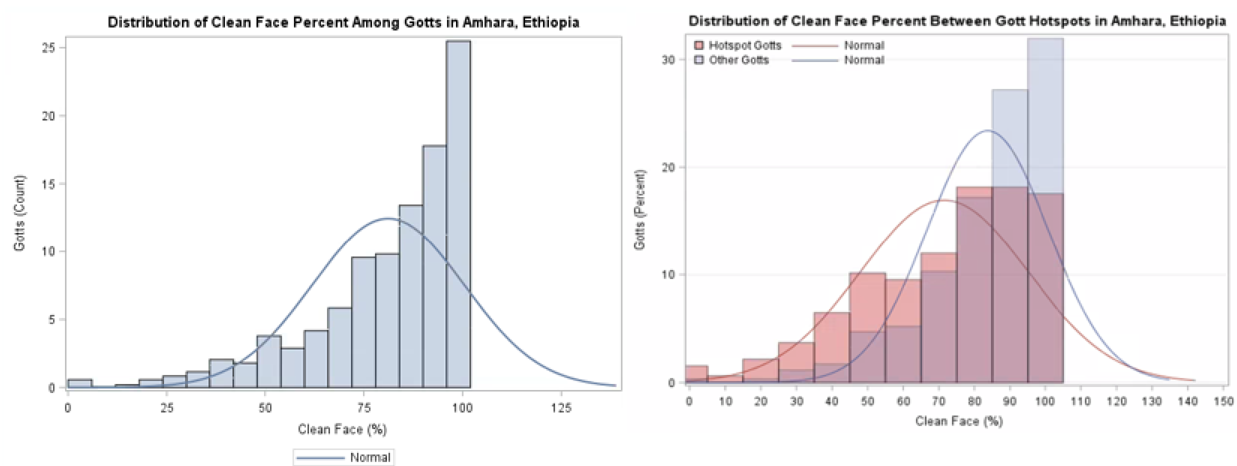


Figure 28. Distribution of the weighted percentage of children presenting with clean faces across the entire region and then stratified by gotts that were identified as hotspots and all other gotts.

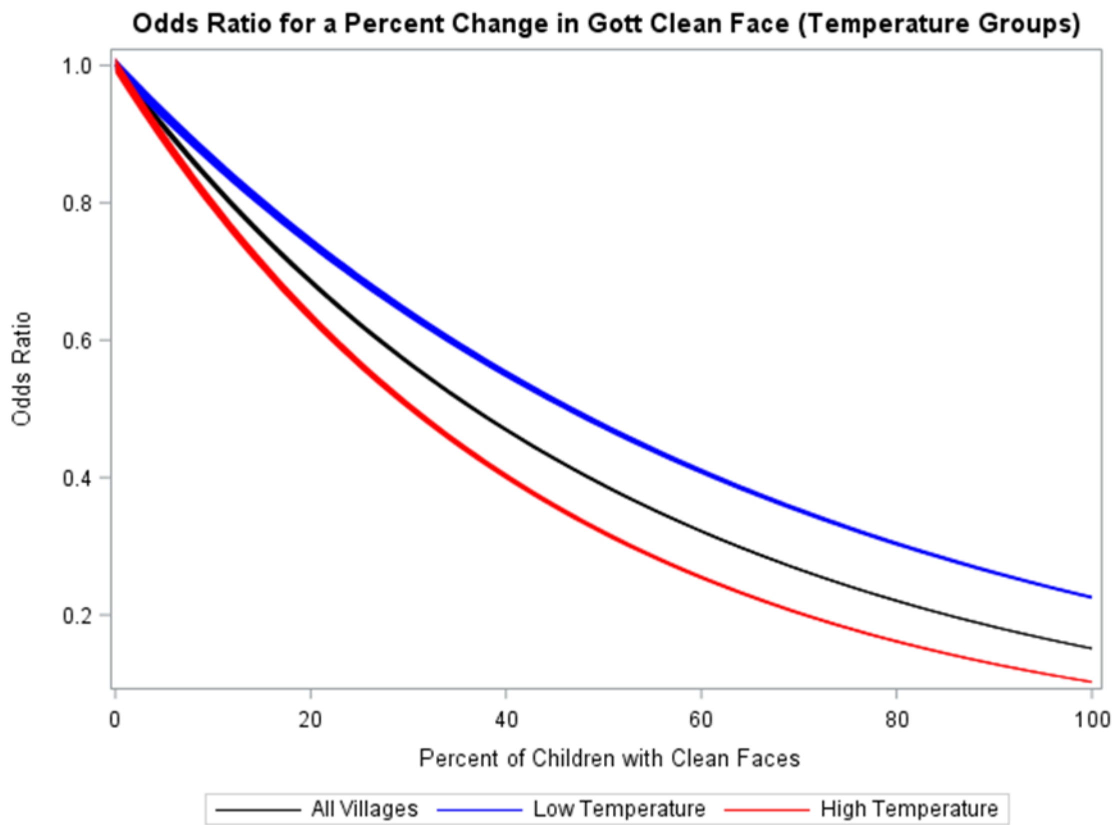
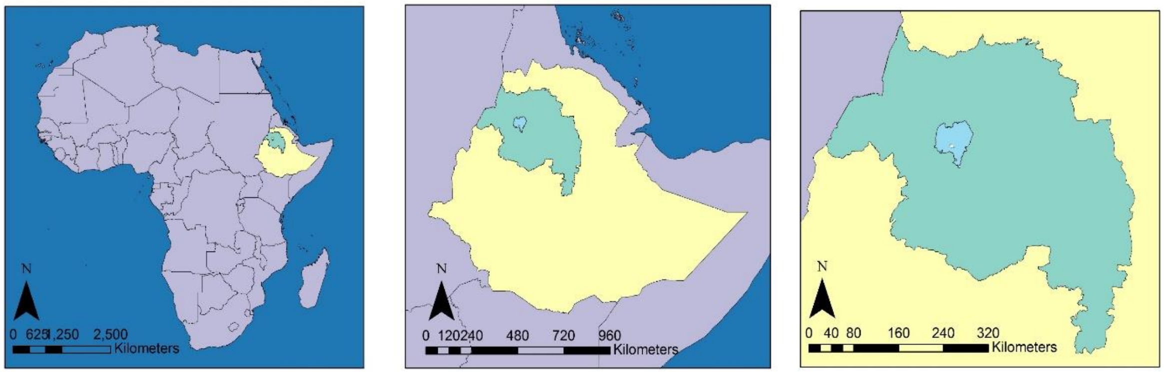


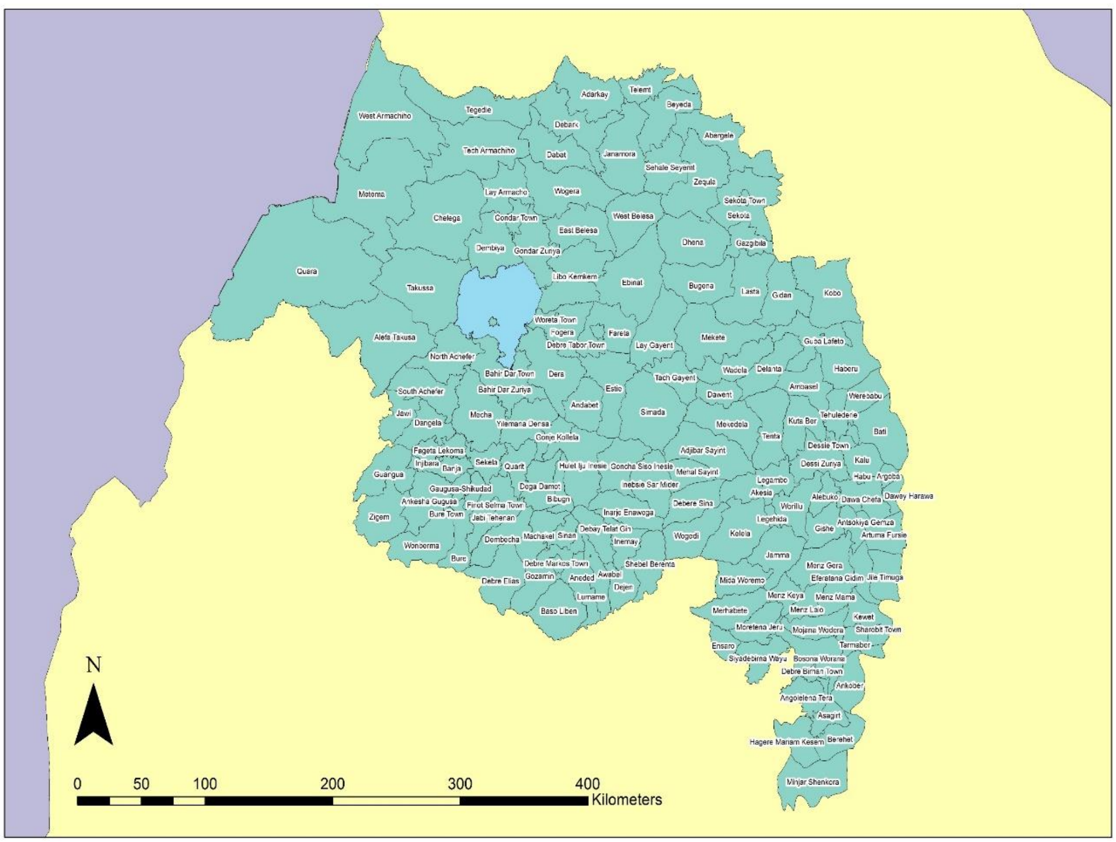
Figure 29. The change in the odds ratio stratified across temperature groups dichotomized at the median annual temperature for the Amhara National Regional State of Ethiopia.

Appendices

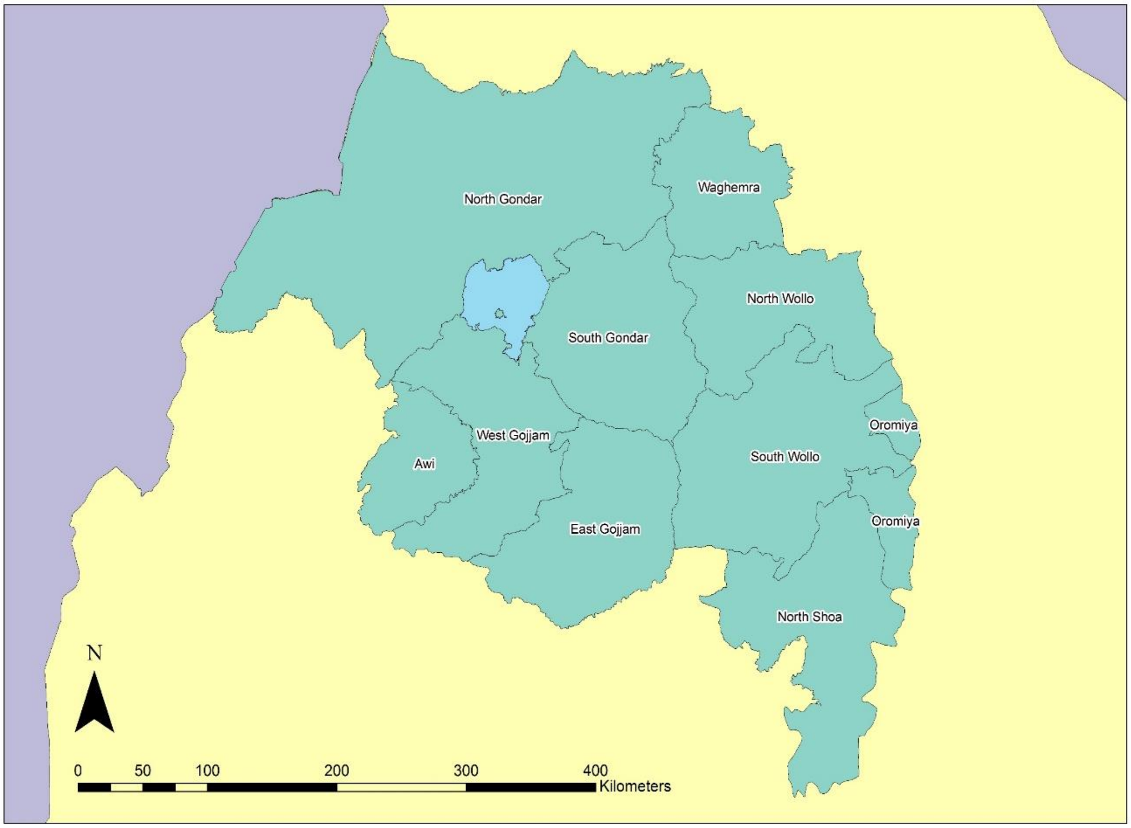
Appendix A. Maps showing the study location at increasingly smaller scale



Appendix B. Woredas throughout the Amhara Regional State of Ethiopia



Appendix C. Geographic range of zones throughout the Amhara National Regional State of Amhara



Appendix D. Studies evaluating clustering of trachoma at varied spatial scales

Author (Location)	Modeling Strategy	Disease Indicator	Clustering at Defined Level	Spatial Clustering
Broman et al. 2005. (Tanzania)	k-function analysis	PCR diagnostic test Children < 8	Household Level: Only observed clustering when using chlaymdia infection diagnosed with PCR as indicator. Clinical symptoms diagnosis revealed no significant clustering. Time following treatment and distance between clusters Baseline Clustering at < 2km 2 months No clustering 6 months Slight clustering at 0.5 km 12 months Clustering at < 1.3km	
Yohannan et al. 2014 (Tanzania)	k-function analysis	Clinical and PCR dignostics Children < 10	Houshold Level: Only observed clustering when using chlaymdia infection diagnosed with PCR as indicator. Clinical symptoms diagnosis revealed no significant clustering. Time following treatment and distance between clusters Baseline No clustering 6 months No clustering 12 months Boarderline clustering between 0 and 0.05 km 18 months Significant clustering between 0.05 and 0.10 km 24 months Clusters became larger with clustering between 0.05 and 0.20 km 30 months Clusters between 0.00 and 0.14 km 36 months Small clusters remained 42 months Small clusters remained	
Polack et al. 2005. (Tanzania)	Random effect in logistic regression analysis Kulldorf spatial scan statistic	PCR diagnostic test All ages	Household Level: Significant clustering by household (P <0.001) Bedroom Level: Significant clustering within bedrooms (P <0.001)	Most significant cluster (P <0.005) had a 0.283 km radius encompassing 24 households
Bailey et al. 1989. (Gambia)	Monte Carlo Simulation	Clinical diagnostic all grades All ages	Household Level: Significant clustering by compound (P<0.0001) Bedroom Level: Significant clustering within bedrooms (P<0.05)	No evidence of spatial clustering
Hagi et al. 2010. (Mali)	Bayesian hierarchical logistic models	Active Trachoma (TF/TI/both) Children <10	No clustering was observed between villages Significant clustering observed at the child (P<0.001), caretaker (P<0.001), household(P<0.001), and village levels (P<0.001)	

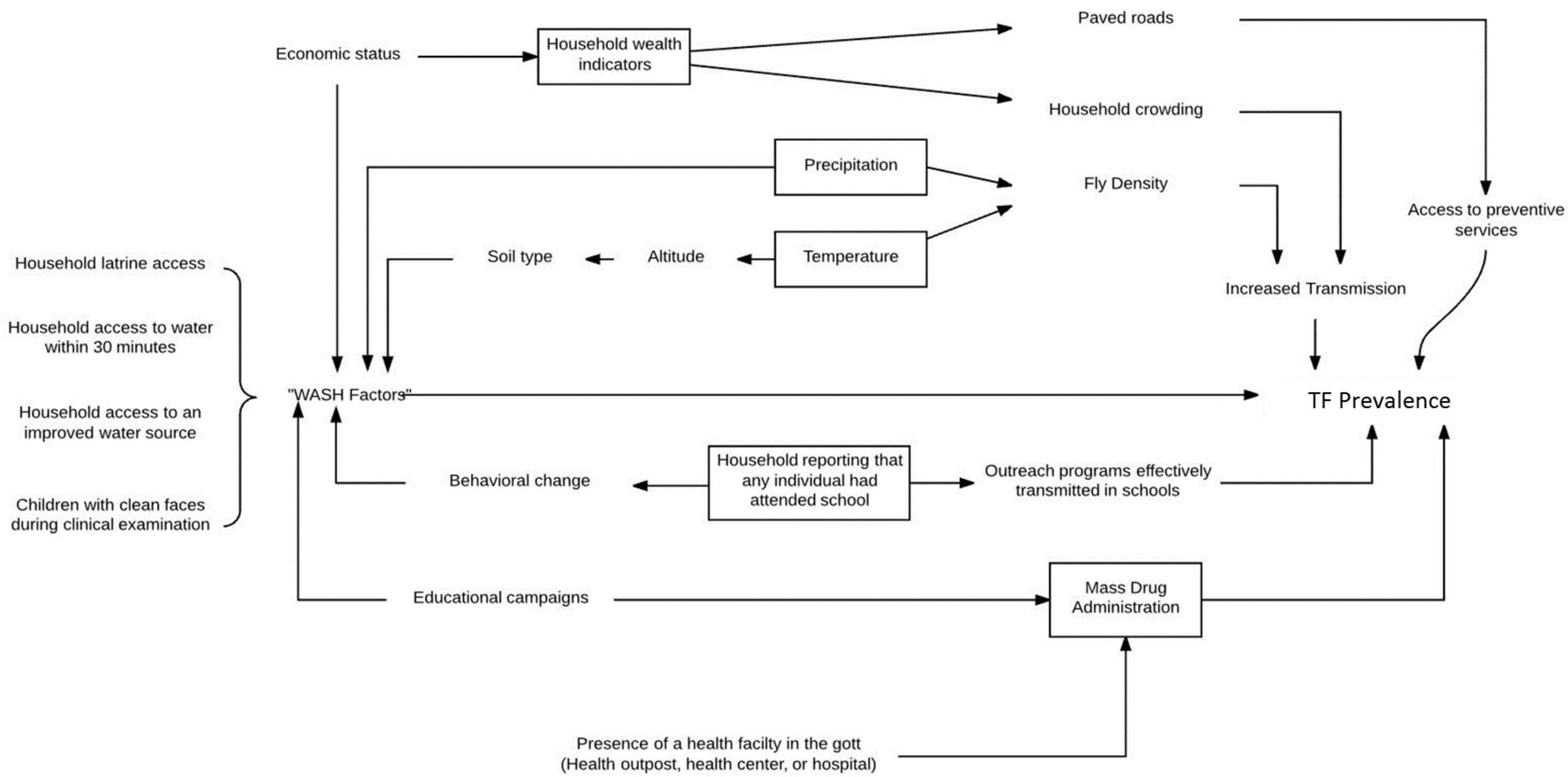
Appendix E. Studies using trachoma as an outcome that considered exposure as distance to a point of interest as a risk factor

Author (Location)	Modeling Strategy	Measure of Association	Distance/Time to Water	Distance to Latrine	Distance to Bar/Café	Distance to Religious Establishment	Distance to Commercial/Government Center
Shemann et al. 2007. (Mali)	Multivariate logistic regression	OR	< 100 m REF ≥ 100 m 1.26 (1.15-1.39)				
Baggaley et al. 2006. (Tanzania)	Logistic regression	OR	0.0- 79.0 m REF 79.1-190.5 m 1.67 (1.21-2.31) 190.6- 496.1 m 2.52 (1.78 -3.56) 496.2- 4855.7 m 2.99 (2.09-4.28) p value for trend <0.001				
Polack et al. 2005. (Tanzania)	Logistic regression	OR	<5 min REF 5-15 min 4.5 (1.0-19.0) 15-30 min 3.2 (0.7-13.9) >30 min 4.9 (1.1-21.7)	Null association (Unspecified OR)			
Montgomery et al. 2011. (Tanzania)	Logistic regression	OR			<700 m REF >700 to ≤1400 m 1.3 (0.5-3.2) >1400 m 2.3 (1.2-4.3)	<700 m REF >700 to ≤1400 m 1.6 (1.1-2.3) >1400 m 1.9 (1.5-2.4)	<700 m REF >700 to ≤1400 m 1.3 (0.6-2.8) >1400 m 1.8 (0.8-4.1)
Smith et al. 2015. (Nigeria)	Logistic regression	OR	Simple logistic model - Distance to surface water (km) 1.18 (<0.0001)				
Edwards et al. 2012. (South Sudan)	Logistic regression	OR	Simple logistic model - Time to water <15 min REF 15-30 min 0.86 (0.62-1.20) >30 min 0.96 (0.69-1.33)				
Bailey et al. 1989. (Gambia)	Monte Carlo Simulation	NA		Null association (Unspecified measure of association)			

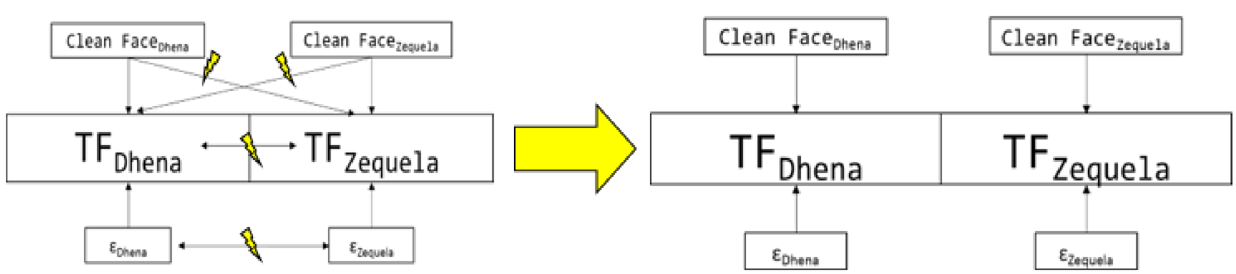
Appendix F. Studies that used trachoma as the primary outcome and considered exposure to geoclimatic variables as risk factors

Author (Location)	Modeling Strategy	Disease Indicator	Measure of Association	Altitude (m)	Temperature	Precipitation	Land cover/NVDI	Other Variables
Shemann et al. 2007 (Mali)	Simple and multiple logistic regression	Active Trachoma (TF/TI/both in either eye) Children <10	OR	Simple logistic model 0-260 REF 260-350 0.68 (0.62-0.74) >350 0.98 (0.88-1.01)	Simple logistic model - Avg daily <31°C REF >31°C 1.19 (1.10-1.29) Multiple logistic model <31°C REF >31°C 1.17 (1.02-1.34)	Simple logistic model - Avg annual 0-300 mm REF 300-650 mm 0.71 (0.65-0.77) 650 mm 0.48 (0.43-0.53)		Latitude Simple logistic model 15-21° REF 13-15° 0.73 (0.67-0.78) 10-13° 0.65 (0.59-0.71) Multiple logistic model 15-21° REF 13-15° 0.62 (0.54-0.73) 10-13° 0.58 (0.49-0.70) Longitude Simple logistic model 0-4° REF 4-7° 1.10 (0.97-1.24) >7 ° 2.00 (1.79-2.24) Relative humidity Simple logistic model 0-25% REF 25-40% 1.05 (0.96-1.14) >40% 0.64 (0.59-0.70)
Clements et al. 2010 (South Sudan)	Logistic regression model estimates Model 1: Fixed effects Model 2: Fixed effects plus geostatistical location level random effects	Active Trachoma (TF/TI) Children 1-9	OR			Per 100ml (Avg annual) Model 1: 0.55 (0.49-0.62) Model 2: 0.21 (0.08-0.46)	Model 1: Savanna 1.77 (1.48-2.11) Forest/Wetland 1.20 (0.98-1.46) Grass/Cropland 0.62 (0.50-0.75) Model 2: Savanna 3.44 (0.5-13.98) Forest/Wetland 2.48 (0.21-12.14) Grass/Cropland 0.81 (0.09-3.65)	
Hagi et al. 2010 (Mali)	Bayesian hierarchal logistic model	Active Trachoma (TF/TI) Children 1-10	OR		Max temperature (Avg monthly) <34.6°C REF 34.6-38.7°C 0.51 (0.29-0.90) >38.7°C 1.03 (0.51-2.05) Mean temperature (°C) (Avg annual) < 27.3°C REF 27.3-28.1°C 0.49 (0.32-0.74) > 28.1°C 0.57 (0.36-0.90)	Number of rainy days (Avg monthly) 0 REF ≥1 0.57 (0.31-1.07)		Sunshine fraction % (Avg monthly) < 62.8 REF 62.8-69.9 0.61 (0.41-0.91) > 69.9 0.50 (0.32-0.79)
Koukounari et al. 2011. (Burkina Faso)	Simple and multivariable hierarchal logistic regression models	Active Trachoma (TF/TI) Children 1-9	OR	Per 50 meters above sea level (MSL) Simple logistic model OR 1.730 (1.189-2.433) Multiple logistic model Dropped by BWE	Minimum temperature per 0.5°C Simple logistic model 0.570 (0.546-0.591) Multiple logistic model 0.746 (0.717-0.768) Maximum temperature per 1°C Simple logistic model 0.465 (0.436-0.484) Average temperature per 0.5°C 0.580(0.564-0.591)	Per 0.5 millimeters Simple logistic model 1.690 (1.336-2.016)		Air Pressure Simple logistic model Per 5 millibars 0.606 (0.602-0.609) Multiple logistic model 0.616 (0.608-0.622)
Smith et al. 2015 (Nigeria)	Bayesian hierarchal logistic model	TT/CO Adults >40	OR	Simple logistic model Meters above sea level (m) <200 REF 200-499 14.41 (p value <0.0001) >500 5.64 (p value <0.0001)	Random effects model- Residual LST 2.95 (1.36-6.85) Random effects model - Mean annual temperature 0.89 (0.69-0.87) Simple logistic model - Mean monthly LST 1.358 (p value <0.0001) Simple logistic model - Mean annual temperature °C 1.01 (p value 0.59)	Random effects model - Avg annual (mm) 0.17 (0.10-0.26) Simple logistic model - Avg annual (mm) 0.997 (p value <0.0001))	Simple logistic model - Enhanced vegetation index (EVI) ≥0.35 REF 0.25-0.34 6.66 (p value 0.001) 0.15-0.24 26.00 (p value <0.0001) <0.15 4.53 (p value 0.01) Simple logistic model - Land cover type Other REF Savannah/Grasslands 2.70 (p value <0.0001)	Urbanization Simple logistic model Rural REF Urban 0.27 (p value <0.0001) Random effects model Urban 0.27 (0.13-0.52)
Baggaley et al. 2006. (Tanzania)	Multiple logistic regression	TF/TI Children 1-9	OR	Meters above sea level (MSL) 822.2-1337.3 REF 1337.4-1514.8 0.40-(0.28-0.56) 1514.9-1703.8 0.42 (0.30-0.60) 1703.9-2268.5 0.56 (0.41-0.76) p value for trend (p<0.001)				

Appendix F. Conceptual diagram explaining the modelled relationship and the influence of hypothesized confounders



Appendix G. Conceptual diagram showing how spatial autocorrelation affects the assumption of uncorrelated error terms and intendent observations. The right side of the diagram indicates how adding a lag parameter can control for the diffusion process of geographic neighbors.



Appendix H. Survey Tools

Gott Level Assessment Form (to be filled out prior to random selection of segment)

GPS Latitude: (N) |___|___| . |___|___|___|___|___| Longitude: (E) |___|___|___| .|___|___|___|___| ___| Elevation: : |___|___|___|___| m Accuracy: |___|___| m		
Cluster Number :	|___|___|___|	(DT_total) Total number of development teams: ________ (DT_size) Average number of HH per development team _____ (segments) Total number of segments created ______________
Questions:		
D5	Is there an organized community wide environmental clean up?	No=0 ; Yes=1
D6	Does the water point management committee meet regularly?	No=0 ; Yes=1; No Committee=88
EC1	Is there a community savings system in this area?	No=0; Yes=1
EC2	Is there a market in this Gott?	No=0; Yes=1
H2	Antibiotics were distributed during MalTra in this Gott by which method?	1 =Central point 2 =House to house 3=both methods
cls	Has your gott held any type of event to discuss sanitation?	No=0; Yes=1; Don't know=88
cls2	Is your community open defecation free?	No=0; Yes=1; Don't know=88
Observations to be made by team:		
		(Circle response)
D1	Is there a paved road?	No=0; Yes=1
D2	Is there functioning electricity?	No=0; Yes, at least partial coverage=1
D3	Is there mobile phone network coverage?	No=0 ; Yes=1
D7	What is the setting of this Gott?	town=0 ; rural=1
D8	How long did it take to access the Gott by foot from the vehicle?	Less than 1 hour=0 ; 1 to 3 hours =1; more than 3 hours=2
H1	What type of health facility is located in this Gott?	None =0 Health post =1 Health center =2 Hospital =3

Trachoma Impact Evaluation Survey – Amhara, Ethiopia: Jan. 2014

HOUSEHOLD QUESTIONNAIRE

<table>
<tr><td>Team number:</td><td>|__|__|</td><td>Cluster Number:</td><td>|__|__|__|</td><td>Household number:</td><td>|__|__|</td></tr>
<tr><td>Serial Number:</td><td colspan="2">|__|__||__|__|__||__|__|
Team **Cluster** **Household**</td><td>Gott name:</td><td colspan="2"></td></tr>
<tr><td>Woreda name:</td><td colspan="2"></td><td>Health Cluster Name:</td><td colspan="2"></td></tr>
<tr><td>Kebele name:</td><td colspan="2"></td><td>Development team name:</td><td colspan="2"></td></tr>
<tr><td colspan="2">Survey Date (DD/MM/YYYY)
|__|__|/|__|__|/|__|__|__|__|</td><td rowspan="3">GPS</td><td colspan="3">Latitude: (N) |__|__| .
|__|__|__|__|__|</td></tr>
<tr><td colspan="2">(consent) Household consent given? No= 0 Yes= 1</td><td colspan="3">Longitude: (E) |__|__|__|
.|__|__|__|__|__| __|</td></tr>
<tr><td colspan="2">(chldcnst) Are there children living in this house under 16 years of age? No= 0 (skip to M1) Yes= 1</td><td colspan="3">Elevation: |__|__|__|__|__| m
Accuracy: |__|__| m</td></tr>
</table>

<table>
<tr><td colspan="4">**Respondents’ demographics (Adult females: women and mothers of children are the preferred respondents)**</td></tr>
<tr><td>rd1</td><td>Description of respondent</td><td>head of household=1
wife of head of household=2
Other =99</td><td></td></tr>
<tr><td>rd2</td><td>Gender of respondent</td><td>Male=1 Female=2</td><td></td></tr>
<tr><td></td><td>Does your household have any of the following?
es1 Functioning radio set
es2 Functioning television
es3 Working electricity
es5 Functioning mobile phone</td><td>No=0; Yes=1
No=0; Yes=1
No=0 ; Yes=1; Yes, generator=2
No=0 ; Yes=1</td><td></td></tr>
<tr><td>es7</td><td>**Observation:** What is the main construction material for the roof of this household?
(One response only)</td><td>Thatch=1
Stick and mud =2
Corrugated iron/metal =3
Other=99</td><td></td></tr>
<tr><td>es8</td><td>**Observation:** are cattle kept *beyond* 10 meters of the living space of the household?</td><td>No=0; Yes=1; No cattle=88</td><td></td></tr>
</table>

s4	What is the highest level of school any *adult* in this household has attended? *(If "none," ask "has any adult had any non-formal education?")*	None=0 Religious=1 Primary school (grade 1-6) =2 Junior secondary=3 Senior secondary=4 College/university=5 Non-formal education=6	
K3	Do you know what trachoma is?	No=0 Yes=1	If no, skip to FW2
HE2	From where did you hear about trachoma?	trachoma volunteer=1 health extension worker=2 mass media=3 health facility=4 community gathering=5 school or school child=6 Other=99	
k5	How can someone protect him/herself from trachoma? *(multiple response)* *(After each response ask 'anything else?' Do not read choices. Mark all responses given.)*	Face washing/hygiene =1 Take antibiotics or medicine =2 Trichiasis surgery =3 Keep environment clean =4 Using pit latrines =5 I don't know =6 Other =99	
fw2	If you have children under 6 years of age, how often are their faces washed? *(only 1 answer)*	Never=0 Every other day=1 Once a day=2 Twice a day=3 3 or more times a day=4 No children under6 years =88	
WS0	Is there a water source within the concession grounds?	No=0 Yes=1	
ws1	Is the water source you use for drinking the same water source you use for bathing?	No=0 Yes	If yes, skip to WS3
ws2	If no, how long does a round trip take for you to collect water used for bathing – including time to walk there, collect water and return home?	less than 30 minutes=1 30 minutes to 1 hour=2 more than1 hour=3	

ws3	What is the main source of water your household uses for drinking? *(Only one response)*	Unprotected spring=1 Protected spring=2 Unprotected dug well=3 Hand pump/tube well / borehole=4 Surface water (river, dam, lake, stream, canal)=5 Public piped water/ tap/standpipe=6 Private piped into yard/dwelling=7 Rainwater collection=8 Other=99	
ws4	How long does it take to collect water from the source of water used for drinking –including time to walk there, collect water, and return?	less than30 minutes=1 30 minutes to 1 hour=2 more than 1 hour=3	
hw2	Ask and observe: where do members of this house most often wash their hands	Observed in dwelling=1 Observed by latrine=2 Observed by dwelling and latrine=3 Not observed in dwelling/plot=4	
hw3	Observation: presence of water at specific place for handwashing	No water present=0 Water available from container=1 Water available from piped tap=2	
hw4	Observation: presence of soap (or cleaning agent) at the specific handwashing place	No cleaning agent available=0 Ash, mud, sand=1 Soap or detergent=2	
m1	Was this household registered for MalTra?	No=0; Yes=1; I don't know=88	
PL4	**OBSERVATION**: currently, what type of toilet facility is available at this household? *A solid platform has no openings other than the drop-hole.* *A rudimentary platform does not seal the pit (e.g. wood branches or poorly plastered).*	No facility/bush/field=0 Pit latrine with rudimentary platform (doesn't seal pit)=1 Pit latrine with solid platform of wood logs and mud plaster=2 Pit latrine with concrete slab=3 Flush or pour-flush toilet=4 Other=99	If "no facility," skip to G1
lo1	**Observation:** Evidence of latrine usage (*feces in pit)*	No=0; Yes=1	
lo2	**Observation:** hand washing container by the latrine	No=0; Yes=1	If no, skip to G1

LO3	**Observation:** Is there water in the handwashing container?	No=0 Yes=1	
pl2	Do you share this latrine with another household?	No=0 Yes=1	
pl7	Why do you not currently have a latrine in your house?	Not enough land to construct a latrine=1 Soil is loose=2 Did not have adequate construction materials=3 No one to build it=4 Too expensive to build=5 I do not know how to build=6 Not a priority =7 No reason specified=8 Use a latrine in another household=9 Other=99	Skip if PL4 $\neq 0$
g1	Has your gott held any type of event to discuss sanitation?	No=0 Yes, community meeting=1 Yes, walk of shame=2 Yes, other=3 Don't know=88	
G3	Has a HEW ever visited your house to talk about sanitation and hygiene?	No=0 Yes=1 Don't know=88	

Household Census

Household Member ID		**01**	**02**	**03**	**04**	**05**	**06**
Name (hh_name)							
Age (hh_age)							
Sex (M/F) (hh_sex)							
Absent? (hh_absent) Present=0, Out=1, Travel=2, Refuse=3, At school=4 (If at school, fill following question –child attends school)							
Currently attend school? (hh_school) No=0, Yes=1, Not eligible=88							
Antibiotics during MDA	**Have you ever taken antibiotics for trachoma during MalTra?** No=0, Yes=1 (hh_ever)						
	How many times have you taken the antibiotics? (hh_times)						
Discharge on child's face? (1-9 years) (hh_disch) None=0, Ocular=1, Nasal=2, Not eligible (<1 year; >9years) =88							
Right Eye (check signs present; if no trachoma sign, check None) (hh_right)	None=0						
	TF=1						
	TI=2						
	TS=3						
	TT=4						
	CO=5						
Left Eye (check signs present; if no trachoma sign, check None) (hh_left)	None=0						
	TF=1						
	TI=2						
	TS=3						

	TT=4						
	CO=5						
TT present? (TT) No=0, Yes=1							
Have you ever been operated for trichiasis? (surgery) No=0 Yes=1							
If yes, on which eye? (TT_eye) Left=1 Right=2 Both=3							
Why have you never been operated? (multiple responses allowed) (never_op) Refused surgery=0 Never been offered surgery/did not know about surgery for TT=1, Did not know where service was offered=2, Can't leave family activities/work=3, Surgery site too far/no transportation=4, No caretaker=5, Other=99							

Appendix I. Institutional Review Board Exemption Letter

EMORY UNIVERSITY

Institutional Review Board

January 29, 2015

Kelly Callahan, MPH
Assistant Director
The Carter Center
453 Freedom Pkwy, NE
Atlanta, GA 30307

RE: Determination: No IRB Review Required
IRB#: Paper Study IRB# 079-2006
Title: *Evaluation of national trachoma control programs supported by the Carter Center in Sudan, South Sudan, Ethiopia, Mali, Niger, and Nigeria (former PI – Emerson)*
PI: Kelly Callahan, MPH

Dear Ms. Callahan:

Thank you for requesting a re-determination from our office about the above-referenced project. Based on our additional review of the materials you provided via email, we have determined that this project still does not require IRB review because it does not meet the definition of "research" or "clinical investigation" as set forth in Emory policies and procedures and federal rules, if applicable. Specifically, this project was previously classified as "exempt" but upon further review of the protocol on 05/01/2011, was determined that the study no longer required IRB review. Specifically, this project involves public health practice program evaluations. In recent communications from the project team, it was noted that the program evaluations would no longer occur in Ghana. Also, since the project's initial review and determination, Sudan has become two countries, Sudan and South Sudan. These site changes do not affect the IRB's determination about the type of review needed for this project.

Please note that this determination does not mean that you cannot publish the results. This determination could be affected by substantive changes in the study design. If the project changes in any substantive way, please contact our office for clarification.

Thank you for consulting the IRB.

Sincerely,

Carol Corkran

Carol Corkran, MPH, CIP
Team Lead

Emory University
1599 Clifton Road, 5th Floor - Atlanta, Georgia 30322
Tel: 404.712.0720 - Fax: 404.727.1358 - Email: irb@emory.edu - Web: http://www.irb.emory.edu
An equal opportunity, affirmative action university