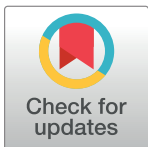


RESEARCH ARTICLE

Progression of a large syphilis outbreak in rural North Carolina through space and time: Application of a Bayesian Maximum Entropy graphical user interface

Lani C. Fox^{1,2}, William C. Miller^{3,4,5}, Dionne Gesink⁶, Irene Doherty^{4,7}, Kristen H. Hampton³, Peter A. Leone^{3,4,8}, Delbert E. Williams⁸, Yasuyuki Akita¹, Molly Dunn⁹, Marc L. Serre^{1*}



1 Department of Environmental Sciences and Engineering, Gillings School of Global Public Health, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, United States of America, **2** Lani Fox Geostatistical Consulting, Claremont, California, United States of America, **3** Department of Epidemiology, Gillings School of Global Public Health, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, United States of America, **4** School of Medicine, Division of Infectious Diseases University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, United States of America, **5** Division of Epidemiology, College of Public Health, The Ohio State University, Columbus, Ohio, United States of America, **6** Epidemiology Division, Dalla Lana School of Public Health, University of Toronto, Toronto, Ontario, Canada, **7** Julius L. Chambers Biomedical/Biotechnology Research Institute, North Carolina Central University, Durham, North Carolina, United States of America, **8** Division of Public Health, Communicable Disease Branch, North Carolina Department of Health and Human Services, Raleigh, North Carolina, United States of America, **9** Department of Maternal and Child Health, Gillings School of Global Public Health, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, United States of America

* marc_serre@unc.edu

OPEN ACCESS

Citation: Fox LC, Miller WC, Gesink D, Doherty I, Hampton KH, Leone PA, et al. (2023) Progression of a large syphilis outbreak in rural North Carolina through space and time: Application of a Bayesian Maximum Entropy graphical user interface. PLOS Glob Public Health 3(5): e0001714. <https://doi.org/10.1371/journal.pgph.0001714>

Editor: Sanghyuk S. Shin, University of California Irvine, UNITED STATES

Received: April 14, 2022

Accepted: February 16, 2023

Published: May 4, 2023

Copyright: © 2023 Fox et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: The data are the property of the North Carolina Department of Health and Human Services. Due to HIPAA restrictions and laws they are not routinely available. This data is directly covered under HIPAA's protected health information's privacy rule. This rule protects all "individually identifiable health information" (<https://www.hhs.gov/hipaa/for-professionals/privacy/laws-regulations/index.html>). The NCDHHS reviews data requests after a formal application for the data. Access to the data

Abstract

In 2001, the primary and secondary syphilis incidence rate in rural Columbus County, North Carolina was the highest in the nation. To understand the development of syphilis outbreaks in rural areas, we developed and used the Bayesian Maximum Entropy Graphical User Interface (BMEGUI) to map syphilis incidence rates from 1999–2004 in seven adjacent counties in North Carolina. Using BMEGUI, incidence rate maps were constructed for two aggregation scales (ZIP code and census tract) with two approaches (Poisson and simple kriging). The BME maps revealed the outbreak was initially localized in Robeson County and possibly connected to more urban endemic cases in adjacent Cumberland County. The outbreak spread to rural Columbus County in a leapfrog pattern with the subsequent development of a visible low incidence spatial corridor linking Roberson County with the rural areas of Columbus County. Though the data are from the early 2000s, they remain pertinent, as the combination of spatial data with the extensive sexual network analyses, particularly in rural areas gives thorough insights which have not been replicated in the past two decades. These observations support an important role for the connection of micropolitan areas with neighboring rural areas in the spread of syphilis. Public health interventions focusing on urban and micropolitan areas may effectively limit syphilis indirectly in nearby rural areas.

is not guaranteed. To apply to access the data, please contact Erika Samoff, HIV/STD Surveillance Manager, North Carolina Division of Public Health, Communicable Disease Branch; 225 N. McDowell Street Raleigh, NC USA 27699 (erika.samoff@dhhs.nc.gov).

Funding: This work was supported by the National Institutes of Health and was awarded to the following authors, WCM, ID, DG, PL, MS. The grant number is R01 AI067913. The funders had no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript. Additionally, WCM receives support from the American Sexually Transmitted Diseases Association, which is awarded via his institution, Ohio State University.

Competing interests: The authors have declared that no competing interests exist.

Introduction

Pronounced disparities for syphilis persist in the southeastern US, where discrimination against minorities, economic disadvantage, pervasive poverty, low quality education, limited access to health care, and residential and social segregation shape the social context [1–3]. Unlike other regions, rural areas of the Southeast have sustained high syphilis rates comparable to, or exceeding, rates in urban areas [4].

In 1999, the Centers for Disease Control and Prevention (CDC) launched the Syphilis Elimination program (SEP) to address the high burden of syphilis in a number of geographic areas including the southeastern US, targeting several states, including North Carolina [5]. In North Carolina, funding focused on the counties with the highest syphilis incidence rates (Mecklenburg, Wake, Durham, Guilford, Forsyth, and Robeson). Although the efforts of the elimination program initially showed success, in 2000, Robeson County, a largely rural county in southeastern North Carolina experienced a sharp rise in county-level rates of primary and secondary (P&S) syphilis, peaking at 71.7 cases per 100,000 person-years [6]. This incidence rate was more than 32 times higher than the national rate (2.2 cases per 100,000 of P&S in 2001), 21 times higher than the rate reported for the South (3.4/100,000 yearly person-time (p-y)), and 13 times higher than the overall rate reported for North Carolina (5.5/100,000) [6]. The next year, in 2001, Columbus County, a rural county with low population density bordering Robeson to the southeast, had the highest incidence of syphilis in the nation with 74.1 cases per 100,000 [2].

Similar to other syphilis outbreaks in the 1990's, the syphilis outbreak in Columbus and Robeson counties was predominantly among heterosexual African-Americans [2] and was attributed, at least in part to crack cocaine use and exchange of sex for drugs [2, 7–9]. Sexual network analysis showed strong cohesion of the network structures, with dense interconnection and substantial cyclicity, or complete circular linkages among a set of sex partners [2, 8]. The sexual network analysis, however, did not reveal substantial connections between Robeson County and its rural neighbor to the south, Columbus County [8].

Hence, how the outbreak progressed to rural Columbus County is unclear, given the absence of defined sexual network connections between persons in Robeson and Columbus counties. This question is particularly critical because the Columbus outbreak occurred concurrently with the implementation of CDC's Syphilis Elimination Program. Insight regarding the development and progression of this outbreak in a rural county in the midst of an active syphilis elimination program is useful to inform syphilis interventions today, given the persistence of syphilis epidemics in the rural south. Furthermore, previous network analyses showed limited connections across counties. When these connections were assessed temporally they were unlikely to have been the specific connection responsible for spread. This deficit likely reflects incomplete network analyses, given the inability even with strong contact tracing to determine every linkage in the communities.

Future technology that permits real-time visualization of network and outbreak progression on spatial maps, coupled with other details about cases, could expedite contact tracing and disease control. Our objective was to characterize the spatiotemporal development of the syphilis outbreak and observe the spread of infection through Robeson County and into Columbus County. We sought to determine whether the outbreak spread continuously across the two adjacent counties or occurred in distinct spatiotemporal clusters within each county. Although these data are now 20 years old, they remain highly relevant today, as this outbreak is one of the best characterized in terms of the underlying sexual networks. These data have served as the foundation for several of our studies as the contact tracing is in greater depth than most other data and thus the findings are highly valuable [9].

In addition, the critical questions of syphilis transmission mechanisms in rural areas remain. We hypothesized that the space/time mapping of incidence rates at the ZIP code and census tract level would provide information which, when combined with our previously reported sexual network analysis [8] would allow us to better understand the mechanism of propagation of the syphilis outbreak within a predominantly rural region. Our study addresses this hypothesis through the development of a series of Bayesian Maximum Entropy (BME) analyses of syphilis incidence rates.

While the mathematical framework of BME has been described [10–13], its implementation necessitates the use of a programming platform requiring specialized technical skills that may not be widely available at public health agencies. To address this issue we developed, and here demonstrate the use of a freely-available, shareware graphical user interface, which automates the BME map development and facilitates a disease mapping analysis. Disease incidence rates reflect uncertainties dependent on multiple factors including the data collection method (manual vs automated), number of observations available, population size, and level of aggregation. Additionally, incidence rates express high variability in space and time. This variability is especially relevant for Sexually Transmitted Disease (STI) outbreaks as outbreaks are often isolated to small geographic areas and specific time periods. Therefore the aggregation level and number of observations can have a large effect on the resulting incidence rates. In public health analyses, BME techniques can account for variabilities in rates resulting in more accurate results than other methods of geographical surveillance [14–17].

Materials and methods

This study was reviewed by the University of North Carolina Committee on the Protection of the Rights of Human Subjects and was determined to be exempt, based on exemption criterion 4 (IRB 05–3080). Criterion 4 is an IRB exemption for secondary research, which may be done without consent if certain criteria are met. The data included all primary, secondary, and early latent stage syphilis cases reported to the North Carolina Division of Public Health between January 1, 1999 and December 31, 2004 for seven adjacent counties: Robeson, Columbus, Bladen, Brunswick, Cumberland, Hoke, and Scotland. Columbus is a rural county with a low population density (58.4 people/mile², 54,749 total in 2000 [18]). Robeson County (130 people/mile², 123,339 total [18] is the closest county with a small urban town (Lumberton; 20,795 people in 2000) [18], considered micropolitan by the US census. Cumberland County (464.2 people/mile², 302,963 total [18]) is the closest county categorized by the US Census with a metropolitan statistical area (Fayetteville, 121,015 people in 2000 [18]).

We back-estimated the date of syphilis exposure and infection (i.e. transmission date) for each syphilis case by subtracting the median latency period for the specific syphilis disease stage from the diagnosis date [19]. Primary syphilis cases were back dated 45 days, secondary syphilis cases were back dated 90 days, and early latent syphilis cases were back dated 183 days.

Case locations were geocoded with ArcGIS 9.3.1 [20] using self-reported residential addresses, and geomasked by the North Carolina Division of Public Health to preserve patient confidentiality. Satori Bulk Mailer software [21] was used to reformat and correct street addresses before geocoding to optimize the match rate of case addresses to current maps. Cases with a post office box address were spatially assigned to the post office address. Cleaned addresses were then matched to a location using address matching maps from: 1) the North Carolina Emergency Response System (E911), which contains point locations for all North Carolina households including rural routes; 2) the North Carolina Department of Transportation, which contains street-level geographic data; or 3) ESRI's 2006 Street Map [22], which is primarily used for locating residences with outdated street names, prisons and military bases.

Geomasking was performed using the donut method [16, 23]. The donut method relocates each geocoded data point within a random distance to ensure patient privacy and maintain spatial resolution necessary for cluster and outbreak detection.

Syphilis rates were estimated by spatially aggregating cases by ZIP code and census tract and temporally aggregating over rolling 6 month intervals to allow for the finest temporal aggregation with the sparsity of the dataset. Rates were then assigned to areal centroids. We denote Y_{ij} as the number of new syphilis cases observed in an area i with centroid location s_i over a duration T (0.5 years) centered at time t_j . We calculate the 6-month areal incidence rate observed at location s_i and time t_j as $R_{ij} = Y_{ij}/(n_{ij}T)$ (cases per person-years), where n_{ij} is the population residing in area i at time t_j . As all cases reported to the state were used in the analysis, space-time periods without syphilis records were given a rate of 0. ZIP code, census tract boundaries and population estimates were obtained from the US Census Bureau for the years 2000 and 2006. Populations for years between 1999 and 2004 were estimated by linearly interpolating/extrapolating the 2000 and 2006 census population estimates.

The mathematical framework of BME (see S1 Text) used in our mapping analysis has already been described in detail [11, 15, 16, 24]. In summary, BME uses principles of *maximum entropy* to process general knowledge about a Space/Time Random Field (S/TRF), and *Bayesian* epistemic data integration to blend site specific knowledge characterized by a variety of distributions including truncated log normal distributions [25–28], interval and normal distributions [16, 29], normal distributions [30, 31], log normal distributions [32], triangular distributions [33], or only hard data [34]. In the context of BME disease mapping (see S1 Text), BME methods are reduced to space/time simple kriging [2, 35] when observed incidence rates are treated as hard data (no uncertainties), and Poisson kriging when observed incidence rate are treated as Gaussian soft data (with uncertainties) such as the small number problem (population variability) [16].

The BME framework is implemented as follows: Let $F06D(s, t)$ be a global mean trend function calculated for any spatial location s and time t in the mapping domain to capture global trends in disease rates. In short, the global mean trend is an exponential kernel smoothing function selected using an exploratory data analysis based on principles described in Nazelle et al. [36] to smooth local variability in the data and retain only global trends (see S1 Text for details). The global mean trend for this dataset has a spatial smoothing range set to 20 km and a temporal smoothing range set to 7 months.

Distinct covariance models were created for the ZIP code and census tract analyses. We define the transformation of incidence rates z_d observed at locations (s_i, t_j) (e.g. centroids of ZIP codes or census tracts) as $x_d = z_d - F06D(s_i, t_j)$. We then define $X(s, t)$ as a homogenous/stationary space/time random field for which the transformed data x_d is a realization, and we let $Z(s, t) = X(s, t) + F06D(s, t)$ represent the space/time random field of the incidence rate. We model the spatiotemporal autocorrelation of $X(s, t)$ using the covariance function $c_x(r, F074)$, where r and $F074$ are the spatial and temporal lags, respectively (See S1 Text for details on Syphilis Covariance models). The covariance function characterizes the autocorrelation in the transformed incidence rate data [37] x_d used by the simple and Poisson kriging methods to produce estimates of $X(s, t)$ across the continuous mapping domain. By adding back the global mean trend $F06D(s, t)$ to the estimated field $X(s, t)$, we obtain maps of the incidence rate $Z(s, t)$. Additional BME disease mapping approaches have been used in the past and are described in S1 Text.

Poisson kriging corrects the “small number problem”, whereby in administrative areas with small populations moderate changes in the number of cases lead to high variations in rates that visually dominate maps and may mask more relevant spatial patterns [38–40]. If the small number problem is an issue, Poisson kriging will smooth incidence rates observed on small populations (i.e. for small n_{ij} values). Conversely, if the small number problem is not an issue, Poisson kriging will not smooth observed incidence rates and the Poisson kriging maps will be

the same as the simple kriging maps. Implementing both simple and Poisson kriging is useful to evaluate if the small number problem affects a specific mapping situation.

Syphilis rates were modeled and mapped at both the ZIP code and census tract levels, and using both simple and Poisson kriging, yielding four sets of spatiotemporal analyses. For each analysis, the steps included (1) an exploratory analysis of the data, (2) modeling the mean trend function to de-trend the data, (3) modeling the covariance function of the mean trend removed data, and finally (4) creating maps at different times to visualize the evolution of outbreaks. The outbreak duration was identified by qualitative review of epidemic curves for each county. While separate, repeated outbreaks are possible, the NC DHHS treated the event as a single interconnected outbreak. The health departments and disease intervention specialists working in the field recognized the outbreak was ongoing before the data could be processed.

To reduce the programming requirements of BME analyses, we implemented and tested the Bayesian Maximum Entropy geographical user interface software (BMEGUI version 2) (see [S1 Text](#) on BMEGUI implementation and https://mserre.sph.unc.edu/BMElab_web/index.htm). BMEGUI was designed by co-authors MS and YA, developed by YA, and tested in this study in the context of disease mapping, with all co-authors contributing to the critical review of disease estimates obtained. Results provided by a BMEGUI analysis are identical to BME analyses performed on other platforms such as MATLAB without necessitating knowledge of a programming language. BMEGUI's interface is fully described in the user manual provided on the website for BMEGUI version 2 (https://mserre.sph.unc.edu/BMElab_web/index.htm). The syphilis rate estimations created by BMEGUI were then imported into Arc-Map 9.3.1 [20] for a more aesthetic display of the estimated syphilis rates through space and time.

Results

Geocoding results

Of the 7,347 syphilis cases reported to the state during the outbreak between January 1, 1999 and December 31, 2004, 65% (4,773) were diagnosed as primary, secondary or early latent syphilis. These five years of data capture syphilis incidence rates across the state before, during and after the Robeson-Columbus outbreak. One fifth (21.6%, $n = 1,032$) of the cases originate from the seven counties included in this study. Although the state was experiencing an overall decline in syphilis incidence during the study period, the counties included in this analysis demonstrated rising incidence from June, 1999 through December, 2001 followed by a rapid decline. Approximately 75% (771) of the syphilis cases in the study area and time period were geocoded and matched to a location; 25% (261) did not match to a location. The most common reasons for failing to match to a location included: missing street data, military addresses, unhoused or out of state persons, rural routes with unavailable E911 addresses, and false addresses.

Spatiotemporal mean trend and covariance models

The ZIP code and census tract aggregations displayed comparable spatial and temporal mean trends ([Fig 1](#)). The covariance analyses show syphilis rates are most correlated for ZIP codes less than 10km apart and for census tracts less than 5km apart. This difference in distances is likely due to the differences between average ZIP code and census tract areas.

Mapping results

As seen with simple kriging maps of ZIP code level syphilis incidence rates for six consecutive non-overlapping 6-month periods from April 1999 to March 2002, the outbreak began

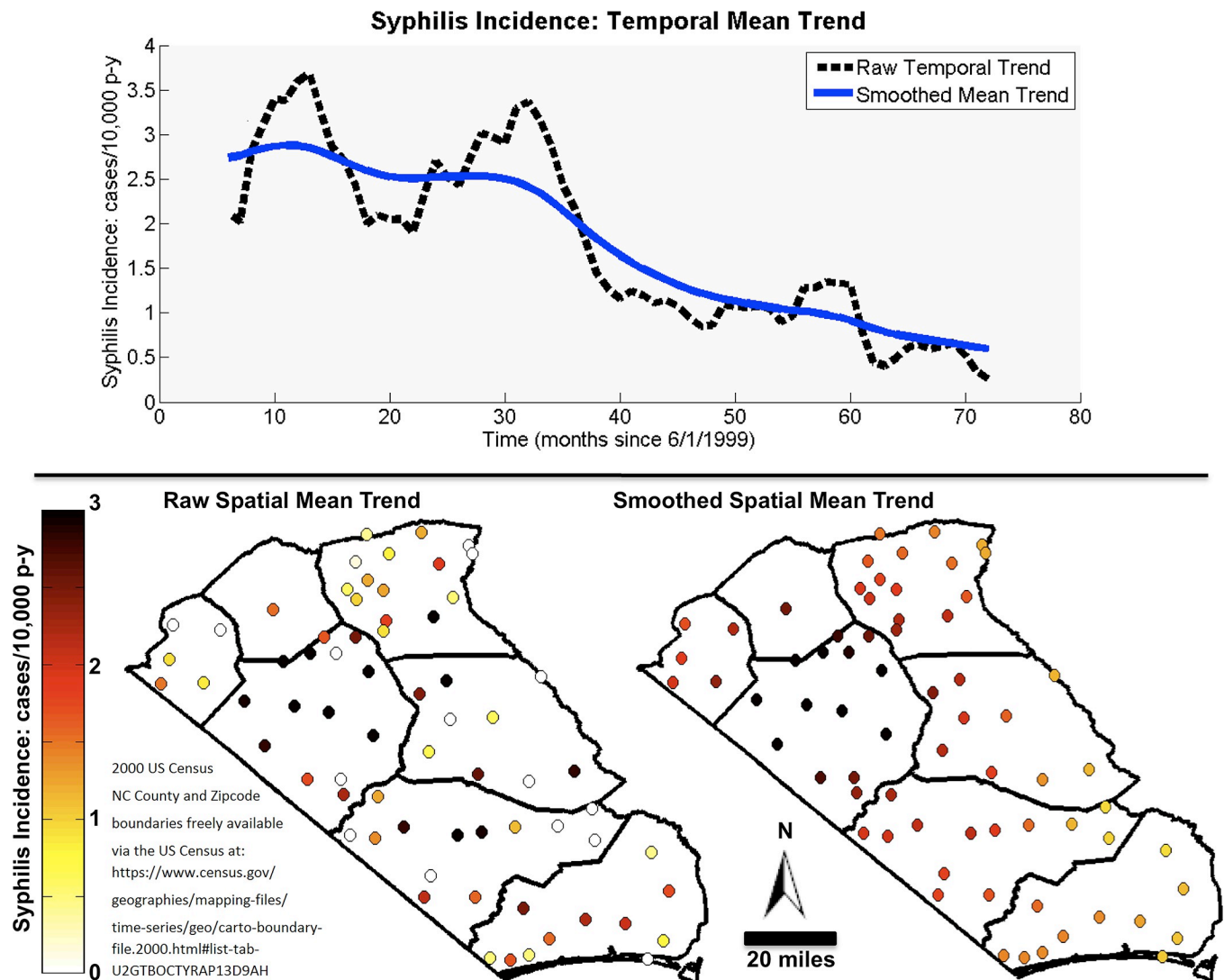


Fig 1. Spatial and temporal components of the mean trend of syphilis 6-month incidence rate (cases/person-years) over the Robeson Columbus outbreak area in 1999–2004. 2000 US Census North Carolina County, Census Tract and Zipcode boundaries freely available via the US Census at: <https://www.census.gov/geographies/mapping-files/time-series/geo/carto-boundary-file.2000.html#list-tab-U2GTBOCTYRAP13D9AH>.

<https://doi.org/10.1371/journal.pgph.0001714.g001>

emerging in Robeson County in 1999 and persisted until 2002 (Fig 2). Cumberland County, to the northeast, had moderate syphilis rates throughout the period consistent with its endemic high case rate. Syphilis incidence rates were low in Columbus County initially (April 1999–September 1999 and October 1999–March 2000) followed by an outbreak in the third 6-month period (April 2000–September 2000). A spatial “corridor” of moderate rates (5–10 cases/10,000 person-years) connected Robeson and Columbus counties in the following two 6-month periods (October 2000–March 2001 and April 2001–September 2001), and dissipated in the last 6-month period (October 2001–March 2002). In the census tract analyses, areas with high rates had less connectivity than the zip code analysis, presumably due to the census tract’s finer geographic resolution. However, the similar spatial and temporal outbreak development patterns were seen in both the zip code and census aggregations and the Simple and Poisson kriging results (see S1–S3 Figs).

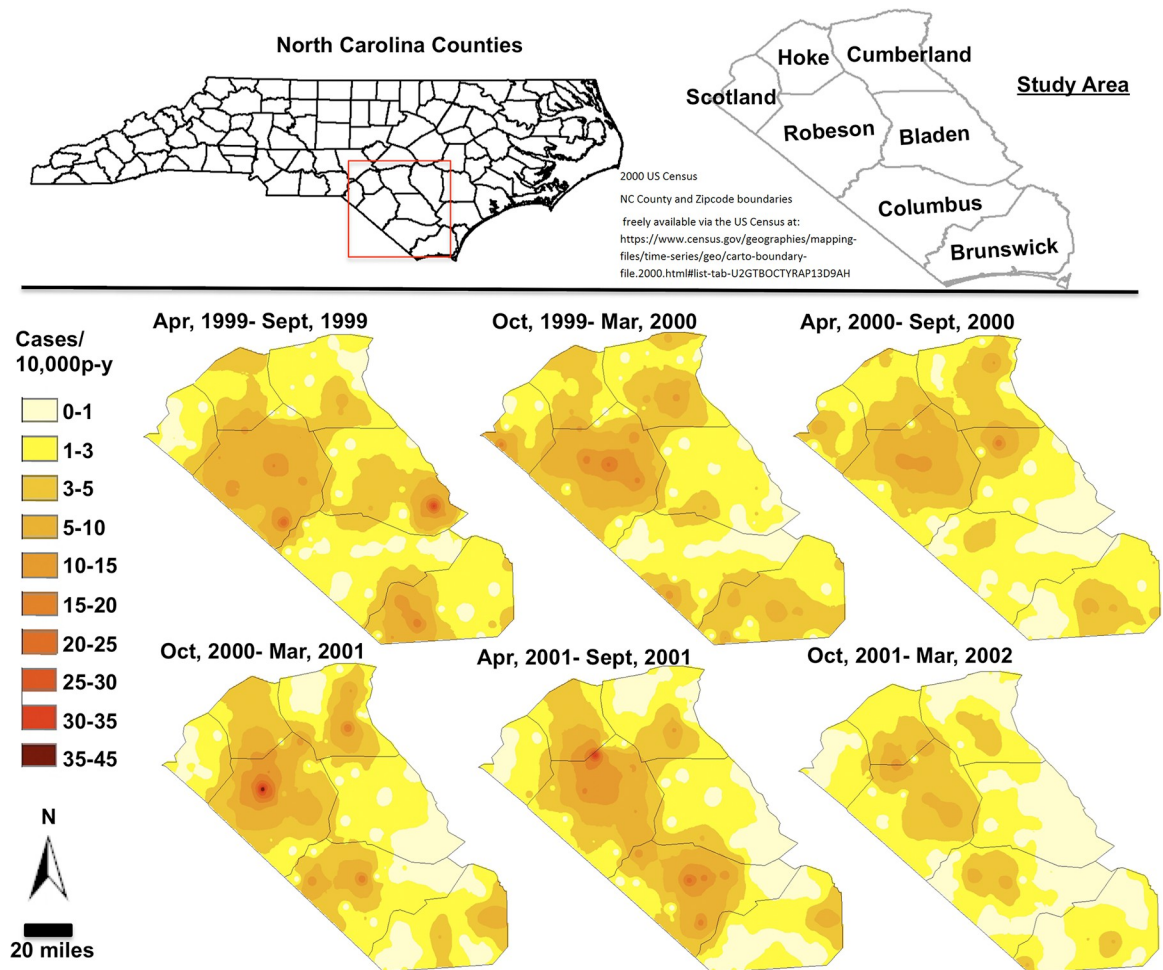


Fig 2. The location of the study area in North Carolina and the BME-simple kriging estimates of ZIP code-level syphilis incidence rates during the Robeson/Columbus outbreak from April 1999 to March 2002. 2000 US Census North Carolina County, Census Tract and Zipcode boundaries freely available via the US Census at: <https://www.census.gov/geographies/mapping-files/time-series/geo/carto-boundary-file.2000.html#list-tab-U2GTBOCTYRAP13D9AH>.

<https://doi.org/10.1371/journal.pgph.0001714.g002>

Discussion

Publicly available surveillance data on syphilis are typically reported at a highly aggregated space/time scale (e.g. by county and year) to protect patient confidentiality. By taking advantage of geomasked data, our work refined the spatial resolution of publicly available maps to ZIP code and census tract and added a temporal component. The geostatistical analyses we conducted would typically require substantial programming expertise, particularly in the selection of the mean trend model, the covariance function, and the estimation methods. However, the BMEGUI tool greatly facilitated the implementation of the BME analysis without requiring programming skills. This tool allowed us to perform the BME analysis at two aggregation scales ensuring comparable conclusions.

We found the ZIP code and census tract aggregations had slightly different results but displayed similar patterns. As expected, for both aggregations the modeled global mean trend, $F06D(s,t) = F06D_s(s) + \mu_t(t)$ displayed smoother spatial and temporal variability than that of the observed local incidence rate data, z_{dt} , because the spatial smoothing range (20 km) and temporal smoothing range (7 months) were greater than the aggregation areas (ZIP code or census

tract) and incidence time period (6 months). Consequently, the residual data $x_d = z_d - F06D(s_i, t_j)$ for the residual S/TRF $X(s, t)$ exhibited noticeable residual space/time variability modeled by the covariance function $c_x(r, \tau)$. The temporal covariance showed cases are correlated over relatively long periods of time and reflects the persistence of disease transmission over time.

We found indistinguishable results using simple and Poisson kriging, supporting the conclusion that the small number problem was not an important issue for our study area. Generally, we recommend both methods be performed as results may vary between mapping situations. Although approximately 25% of the data could not be geocoded the missing data appeared to be missing at random without any noticeable patterns. We also believe the missing data had a minimal effect on the final results, adding only noise rather than a specific bias.

The data analysis took place much later than the outbreak occurred for several reasons. First, at the time of the outbreak the health department used an antiquated data management system created prior to relational databases. Furthermore, the technology (hardware and software) available at the time was also limited. Extensive efforts were expended to geocode the data in a secure data environment, securely geomask and finally aggregate the data in administrative boundaries requiring institutional review and approval of each step by the NC DHHS. The tracking software and data mining technology available today would have undoubtedly streamlined the process of creating an analytical dataset. Furthermore, the spatiotemporal methods were developed after the outbreak for targeted research questions applied here. The BMEGUI framework was also developed and adapted to disease mapping for this specific data set. As it was the first application of BME on these data, and given the sensitive nature of the data, extensive validation had to be conducted.

The recognition of the outbreak by the NC DHHS was accompanied by an extensive effort of contact tracing and social contact referrals with testing of those contacted. This effort, in combination with exhaustion of a susceptible population within the sexual networks, may have dissipated the outbreak. The duration of the outbreak may have been lengthened by the density and size of the sexual network [8].

Our results are consistent with a study of a syphilis outbreak in Baltimore [10], which reported a leapfrog pattern of high incidence areas. In our context, extremely high and concentrated incidence areas in Robeson County were followed by high incidence areas in new, non-contiguous locations (e.g. Columbus County). Combining these results with our previous network analysis [8] provides a more comprehensive picture of the outbreak trajectory. First, our high resolution maps show the leapfrog jump of the outbreak from one county to its neighbor is followed by the establishment of a visible spatial corridor between these counties. Second, we learned from the socio-sexual network analysis that sexual network bridges linked individuals across some of these counties. Though our previous network analyses lacked temporal information, the combination of these findings provides a nuanced understanding of the mechanisms and transmission dynamics that propagated the outbreak: densely connected members of the sexual network geographically localized in Robeson County had sporadic bridges linking them with other densely connected persons in distinct geographical areas. In outbreaks driven by social networks where geographical proximity plays an important role, these bridges may predominantly link socio-sexual networks that are geographically close. These bridges may also lead to emergence of an outbreak in the newly connected sexual networks, leading to increasing rates in neighboring core areas and a visible corridor appears on the map.

This leapfrog-preceding-spatial-corridor mechanism explains how an outbreak can move to a rural area. In a previous syphilis cluster analysis in North Carolina we hypothesized the main pathway of STIs into rural areas may be through the interconnectedness of urban, metropolitan, small town, and rural areas [41]. Our findings support this hypothesis. Fayetteville, a

major city in Cumberland County with a population of 121,015 in 2000 [21], had the highest counts of syphilis cases (and low rates) throughout the study suggesting it was a likely source of the outbreak. Infection moved from Cumberland to Robeson County where rates remained high throughout the study period (Fig 2). The epicenter of the Robeson outbreak appears to be an extension of the high endemic syphilis cases in Cumberland. Sexually connected persons in Robeson continued to transmit locally, and this group was geographically and sexually distinct from networks in other counties [8].

After the initiation of the outbreak in Robeson County, its micropolitan center, Lumberton, was critical in the outbreak expansion. Lumberton is the closest urban area easily accessible by rural residents of nearby Columbus County. This connection between a small town and proximal rural areas allows introduction of the outbreak into the rural sexual network. Consequently, the outbreak leapfrogged spatially into Columbus with subsequent development of geographical bridges as the outbreak expanded. Had the outbreak not been contained, these bridges may have coalesced into a larger inter-county outbreak, and then possibly a regional epidemic. Critically, this series of events implies that addressing STIs in urban and micropolitan areas will also indirectly help address STI rates in rural areas, as was suggested in our previous work [41].

We also believe the visible leapfrog jumps can act as an early warning sign an outbreak is about to expand, potentially spanning multiple counties. BME incidence rate maps can be used to identify these leapfrog jumps and subsequent spatial corridors to create more targeted and potentially cost-effective interventions to contain localized outbreaks before they expand into regional epidemics.

As with most studies using interpolation methods, our ability to visualize spatial corridors was impaired by a well-known kriging artifact referred to as “kriging islands”. Kriging is an exact interpolator; thus, at a data point the estimator has the same value as the data point. Kriging islands are seen in our maps as spots at aggregation (i.e. census tracts) centroids. At these kriging islands, the rates are either higher or lower than the surrounding areas. Furthermore, since ZIP codes and census tracts have arbitrary boundaries with greatly differing sizes, shapes, and populations, high incidence areas can either be displaced to the centroid or completely removed, if the cases occur on the border of two or more aggregation areas. Methods to overcome this limitation are needed. For example, geocoded cases aggregated on a regularly spaced, overlapping grid may provide better representations.

Our observation of the outbreak expanding through a geographical front coupled with temporal outward leapfrog jumps suggests novel techniques are needed to track fronts. An effective system to track fronts could act as an early alarm to inform public health officials that an outbreak is growing from a localized occurrence to a regional concern. Tracking fronts will require the development of methods to smooth disease rates and detect ramp ups preceding the emergence of an outbreak.

Although the data used for this analysis was twenty years old, the results remain relevant today. The combination of these spatial data with the previous extensive sexual network analyses, which are some of the most extensive syphilis network analyses, especially in rural areas, provides rich insights that have not been replicated in the past two decades. Our disease mapping approaches and analyses strengthen and expand findings from our previous sexual network analysis. However the previous analyses had slightly different data sources and substantial missing data. Furthermore, the previous sexual network analysis did not incorporate temporal information or fine scale geographical data. Finally, the methods used here provide useful examples that can be applied to other conditions, especially within rural areas. Future work should consider integrating space/time disease mapping with temporal analyses of sexual and social networks, while addressing the impact of missing data. Technology that

permits real-time visualization of network and outbreak progression on spatial maps, coupled with other details about cases, could expedite contact tracing and disease control.

In conclusion, we observed a leapfrog pattern of syphilis transmission followed by development of local geographical bridges in a large outbreak in North Carolina. This outbreak with a large rural component appeared to be connected to both urban and micropolitan sources. When high disease incidence is seen in a relatively constrained geographic boundary and another geographically non-contiguous hot spot appears, it may suggest an outbreak is in progress as spatial linkages often do not appear immediately. Concentrated efforts to control an outbreak should be placed in densely populated areas as these resources will also substantially affect progression of an outbreak to less dense population centers. Future work must address the changing dynamics of syphilis, methods to monitor the leapfrog and bridging patterns of outbreaks, and the combination of social/sexual network and geospatial analyses.

Supporting information

S1 Fig. Spatiotemporal covariance $c_X(r,t)$ for the rolling 6-month syphilis incidence rate obtained using census tract aggregated cases (left) and zip code aggregated cases (right). The spatial plots (top) show $c_X(r,t=0)$ as a function of spatial lag r , and the temporal plots (bottom) show $c_X(r=0,t)$ as a function of temporal lag t . Circles depict experimental covariance values, while the line depicts the covariance model.
(TIF)

S2 Fig. Poisson kriging results for October, 2000- March, 2001 of the Zip code aggregated data set.
(GIF)

S3 Fig. Poisson kriging results for October, 2000- March, 2001 of the census tract aggregated data set.
(GIF)

S1 Text.
(DOCX)

Acknowledgments

We wish to acknowledge and thank the North Carolina Division of Public Health for their willingness to share their data and accommodate us with office space during the completion of this project. This work would not have been possible without their support.

Author Contributions

Conceptualization: William C. Miller, Dionne Gesink, Irene Doherty, Peter A. Leone, Marc L. Serre.

Data curation: William C. Miller, Peter A. Leone, Delbert E. Williams, Molly Dunn.

Formal analysis: Lani C. Fox, Molly Dunn, Marc L. Serre.

Funding acquisition: William C. Miller.

Investigation: Lani C. Fox, Molly Dunn, Marc L. Serre.

Methodology: Lani C. Fox, Molly Dunn, Marc L. Serre.

Project administration: Delbert E. Williams.

Resources: William C. Miller, Peter A. Leone, Delbert E. Williams.

Software: Yasuyuki Akita, Marc L. Serre.

Supervision: William C. Miller, Marc L. Serre.

Validation: Lani C. Fox, Marc L. Serre.

Visualization: Lani C. Fox, Marc L. Serre.

Writing – original draft: Lani C. Fox, William C. Miller, Dionne Gesink, Irene Doherty, Kristen H. Hampton, Molly Dunn, Marc L. Serre.

Writing – review & editing: Lani C. Fox, William C. Miller, Dionne Gesink, Irene Doherty, Kristen H. Hampton, Peter A. Leone, Delbert E. Williams, Yasuyuki Akita, Molly Dunn, Marc L. Serre.

References

1. Rothenberg R, Sterk C, Toomey K, Potterat J, Johnson D, Schrader M, et al. Using social network and ethnographic tools to evaluate syphilis transmission. *Sexually transmitted diseases*. 1998 Mar 1;154–60. <https://doi.org/10.1097/00007435-199803000-00009> PMID: 9524994
2. Seña A, Muth S, Heffelfinger J, O'DOWD J, Foust E, Leone P. Factors and the sociosexual network associated with a syphilis outbreak in rural North Carolina. *Sexually transmitted diseases*. 2007 May 1;280–7. <https://doi.org/10.1097/01.olq.0000237776.15870.c3> PMID: 17139235
3. Farley T. Sexually transmitted diseases in the Southeastern United States: location, race, and social context. *Sexually transmitted diseases*. 2006 Jul 1;S58–64. <https://doi.org/10.1097/01.olq.0000175378.20009.5a> PMID: 16432486
4. Thomas J, Kulik A, Schoenbach V. Syphilis in the South: rural rates surpass urban rates in North Carolina. *American Journal of Public Health*. 1995 Aug; 85(8_Pt_1):1119–22. https://doi.org/10.2105/ajph.85.8_pt_1.1119 PMID: 7625508
5. Division of STD Prevention. Sexually Transmitted Disease Surveillance 1999 [Internet]. Centers for Disease Control and Prevention; [cited 2022Mar30]. Available from: [chrome-extension://oemmnadbld-boiebfnladdacbfmadadm/https://www.cdc.gov/std/stats/archive/Surv-1999-All.pdf](https://www.cdc.gov/std/stats/archive/Surv-1999-All.pdf)
6. Centers for Disease Control and Prevention (CDC). Primary and secondary syphilis—United states, 2000–2001. *MMWR. Morbidity and mortality weekly report*. 2002 Nov 1; 51(43):971–3. PMID: 12433021
7. Cook R, Royce R, Thomas J, Hanusa B. What's driving an epidemic? The spread of syphilis along an interstate highway in rural North Carolina. *American Journal of Public Health*. 1999 Mar; 89(3):369–73. <https://doi.org/10.2105/ajph.89.3.369> PMID: 10076487
8. Doherty I, Serre M, Gesink D, Adimora A, Muth S, Leone P, et al. Sexual networks, surveillance, and geographical space during syphilis outbreaks in rural North Carolina. *Epidemiology (Cambridge, Mass.)*. 2012 Nov; 23(6):845. <https://doi.org/10.1097/EDE.0b013e31826c2b7e> PMID: 23007041
9. Hampton K. Spatial Dynamics of Syphilis and Gonorrhea Transmission Among Sexual Networks and Core Groups in North Carolina.
10. Christakos G. Random field models in earth sciences. Courier Corporation; 2012 Aug 9.
11. Christakos G, Li X. Bayesian maximum entropy analysis and mapping: a farewell to kriging estimators?. *Mathematical Geology*. 1998 May; 30(4):435–62.
12. Christakos G, Olea R, Serre M, Wang L, Yu H. Interdisciplinary public health reasoning and epidemic modelling: the case of black death. New York: Springer; 2005 Jun 24.
13. Serre M, Kolovos A, Christakos G, Modis K. An application of the holistochastic human exposure methodology to naturally occurring arsenic in Bangladesh drinking water. *Risk Analysis: An International Journal*. 2003 Jun; 23(3):515–28. <https://doi.org/10.1111/1539-6924.t01-1-00332> PMID: 12836844
14. Law D, Bernstein K, Serre M, Schumacher C, Leone P, Zenilman J, et al. Modeling a syphilis outbreak through space and time using the Bayesian maximum entropy approach. *Annals of Epidemiology*. 2006 Nov 1; 16(11):797–804. <https://doi.org/10.1016/j.annepidem.2006.05.003> PMID: 16882466
15. Choi K, Serre M, Christakos G. Efficient mapping of California mortality fields at different spatial scales. *Journal of Exposure Science & Environmental Epidemiology*. 2003 Mar; 13(2):120–33.

16. Hampton K, Serre M, Gesink D, Pilcher C, Miller W. Adjusting for sampling variability in sparse data: geostatistical approaches to disease mapping. *International Journal of Health Geographics*. 2011 Dec; 10(1):1–7. <https://doi.org/10.1186/1476-072X-10-54> PMID: 21978359
17. Lee S, Yeatts K, Serre M. A Bayesian Maximum Entropy approach to address the change of support problem in the spatial analysis of childhood asthma prevalence across North Carolina. *Spatial and spatio-temporal epidemiology*. 2009 Oct 1; 1(1):49–60. <https://doi.org/10.1016/j.sste.2009.07.005> PMID: 20300553
18. Bureau USC. American Fact Finder Profile of General Demographic Characteristics: 2000 Census 2000 Summary File 1 (SF 1) 100-Percent Data. [Internet]. Census.gov. 2021 [cited 2022Mar30]. Available from: <https://www.census.gov/data/datasets/2000/dec/summary-file-1.html>
19. Schumacher C, Bernstein K, Zenilman J, Rompalo A. Reassessing a large-scale syphilis epidemic using an estimated infection date. *Sexually transmitted diseases*. 2005 Nov 1:659–64. <https://doi.org/10.1097/01.olq.0000175400.40084.3e> PMID: 16254539
20. ESRI. ArcGIS 9. Redlands, CA: Environmental Systems Research Institute; 2006.
21. Fazio P. Satori Acquisition Resource Page [Internet]. Satori Acquisition | BCC Software. [cited 2022Mar30]. Available from: <https://bccsoftware.com/satori/>
22. ESRI. Esri U.S. Street Database. Redlands, CA: Environmental Systems Research Institute; 2006.
23. Allshouse W, Fitch M, Hampton K, Gesink D, Doherty I, Leone P, et al. Geomasking sensitive health data and privacy protection: an evaluation using an E911 database. *Geocarto international*. 2010 Oct 1; 25(6):443–52. <https://doi.org/10.1080/10106049.2010.496496> PMID: 20953360
24. Serre M, Christakos G. Modern geostatistics: Computational BME analysis in the light of uncertain physical knowledge—the Equus Beds study. *Stochastic Environmental Research and Risk Assessment*. 1999 Apr; 13(1):1–26.
25. Messier K, Akita Y, Serre M. Integrating address geocoding, land use regression, and spatiotemporal geostatistical estimation for groundwater tetrachloroethylene. *Environmental science & technology*. 2012 Mar 6; 46(5):2772–80. <https://doi.org/10.1021/es203152a> PMID: 22264162
26. Messier K, Kane E, Bolich R, Serre M. Nitrate variability in groundwater of North Carolina using monitoring and private well data models. *Environmental science & technology*. 2014 Sep 16; 48(18):10804–12. <https://doi.org/10.1021/es502725f> PMID: 25148521
27. Messier K, Campbell T, Bradley P, Serre M. Estimation of groundwater Radon in North Carolina using land use regression and Bayesian maximum entropy. *Environmental Science & Technology*. 2015 Aug 18; 49(16):9817–25. <https://doi.org/10.1021/acs.est.5b01503> PMID: 26191968
28. Jat P, Serre M. Bayesian Maximum Entropy space/time estimation of surface water chloride in Maryland using river distances. *Environmental pollution*. 2016 Dec 1; 219:1148–55. <https://doi.org/10.1016/j.envpol.2016.09.020> PMID: 27616646
29. Fox L, Serre M, Lippmann SJ, Rodríguez D, Bangdiwala S, Gutiérrez M, et al. Spatiotemporal approaches to analyzing pedestrian fatalities: the case of Cali, Colombia. *Traffic injury prevention*. 2015 Aug 18; 16(6):571–7. <https://doi.org/10.1080/15389588.2014.976336> PMID: 25551356
30. Xu Y, Serre M, Reyes J, Vizuete W. Bayesian Maximum Entropy integration of ozone observations and model predictions: A national application. *Environmental Science & Technology*. 2016 Apr 19; 50(8):4393–400. <https://doi.org/10.1021/acs.est.6b00096> PMID: 26998937
31. Xu Y, Serre M, Reyes J, Vizuete W. Impact of temporal upscaling and chemical transport model horizontal resolution on reducing ozone exposure misclassification. *Atmospheric Environment*. 2017 Oct 1; 166:374–82.32.
32. Cleland S, West J, Jia Y, Reid S, Raffuse S, O'Neill S, et al. Estimating wildfire smoke concentrations during the October 2017 California fires through BME space/time data fusion of observed, modeled, and satellite-derived PM2.5. *Environmental science & technology*. 2020 Oct 16; 54(21):13439–47.
33. Gómez A, Serre M, Wise E, Pavelsky T. Integrating Community Science Research and Space-Time Mapping to Determine Depth to Groundwater in a Remote Rural Region. *Water Resources Research*. 2021 Jun; 57(6):e2020WR029519.34.
34. DeLang M, Becker J, Chang K, Serre M, Cooper O, Schultz M, et al. Mapping yearly fine resolution global surface ozone through the bayesian maximum entropy data fusion of observations and model output for 1990–2017. *Environmental science & technology*. 2021 Mar 8; 55(8):4389–98. <https://doi.org/10.1021/acs.est.0c07742> PMID: 33682412
35. Olea RA. *Geostatistics for engineers and earth scientists*. Springer Science & Business Media; 2012 Dec 6.
36. Nazelle A, Arunachalam S, Serre M. Bayesian maximum entropy integration of ozone observations and model predictions: An application for attainment demonstration in North Carolina. *Environmental science & technology*. 2010 Aug 1; 44(15):5707–13. <https://doi.org/10.1021/es100228w> PMID: 20590110

37. Law D, Serre M, Christakos G, Leone P, Miller W. Spatial analysis and mapping of sexually transmitted diseases to optimise intervention and prevention strategies. *Sexually Transmitted Infections*. 2004 Aug 1; 80(4):294–9. <https://doi.org/10.1136/sti.2003.006700> PMID: 15295129
38. Goovaerts P. Geostatistical analysis of disease data: accounting for spatial support and population density in the isopleth mapping of cancer mortality risk using area-to-point Poisson kriging. *International Journal of Health Geographics*. 2006 Dec; 5(1):1–31. <https://doi.org/10.1186/1476-072X-5-52> PMID: 17137504
39. Kennedy S. The small number problem and the accuracy of spatial databases. *Accuracy of spatial databases*. 1989 Dec 6:187–96.
40. Kerry R, Goovaerts P, Haining RP, Ceccato V. Applying Geostatistical Analysis to Crime Data: Car-Related Thefts in the Baltic States. *Geographical analysis*. 2010. <https://doi.org/10.1111/j.1538-4632.2010.00782.x> PMID: 22190762
41. Gesink D, Sullivan A, Norwood T, Serre M, Miller W. Does core area theory apply to STIs in rural environments?. *Sexually transmitted diseases*. 2013 Jan; 40(1):32.