PLOS COMPUTATIONAL BIOLOGY

METHODS

Enhancing insights in sexually transmitted infection mapping: Syphilis in Forsyth County, North Carolina, a case study

Lani Fox[1,2], **William C. Miller**[3], **Dionne Gesink**[4], **Irene Doherty**[5], **Marc Serre**[1]*

1 Department of Environmental Sciences and Engineering, Gillings School of Global Public Health, University of North Carolina at 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, North Carolina, United States of America, **4** Epidemiology Division, Dalla Lana School of Public Health, University of Toronto, **5** Julius L. Chambers Biomedical/Biotechnology Research Institute / North Carolina Central University, North Carolina, United States of America

* marc_serre@unc.edu

OPEN ACCESS

Citation: Fox L, Miller WC, Gesink D, Doherty I, Serre M (2024) Enhancing insights in sexually transmitted infection mapping: Syphilis in Forsyth County, North Carolina, a case study. PLoS Comput Biol 20(10): e1012464. https://doi.org/10.1371/journal.pcbi.1012464

Editor: Samuel V. Scarpino, Northeastern University, UNITED STATES OF AMERICA

Received: February 29, 2024

Accepted: September 5, 2024

Published: October 31, 2024

Peer Review History: PLOS recognizes the benefits of transparency in the peer review process; therefore, we enable the publication of all of the content of peer review and author responses alongside final, published articles. The editorial history of this article is available here: https://doi.org/10.1371/journal.pcbi.1012464

Data Availability Statement: The data are the property of the North Carolina Department of Health and Human Services. They contain sensitive geographic location and demographic data and are

Abstract

In 2008–2011 Forsyth County, North Carolina (NC) experienced a four-fold increase in syphilis rising to over 35 cases per 100,000 mirroring the 2021 state syphilis rate. Our methodology extends current models with: 1) donut geomasking to enhance resolution while protecting patient privacy; 2) a moving window uniform grid to control the modifiable areal unit problem, edge effect and remove kriging islands; and 3) mitigating the "small number problem" with Uniform Model Bayesian Maximum Entropy (UMBME). Data is 2008–2011 early syphilis cases reported to the NC Department of Health and Human Services for Forsyth County. Results were assessed using latent rate theory cross validation. We show combining a moving window and a UMBME analysis with geomasked data effectively predicted the true or latent syphilis rate 5% to 26% more accurate than the traditional, geopolitical boundary method. It removed kriging islands, reduced background incidence rate to 0, relocated nine outbreak hotspots to more realistic locations, and elucidated hotspot connectivity producing more realistic geographical patterns for targeted insights. Using the Forsyth outbreak as a case study showed how the outbreak emerged from endemic areas spreading through sexual core transmitters and contextualizing the outbreak to current and past outbreaks. As the dynamics of sexually transmitted infections spread have changed to online partnership selection and demographically to include more women, partnership selection continues to remain highly localized. Furthermore, it is important to present methods to increase interpretability and accuracy of visual representations of data.

Author summary

From 2008 to 2011, Forsyth County, North Carolina saw a dramatic increase in syphilis cases reaching over 35 cases per 100,000, aligning with the state's highly elevated 2021 rate. Our study addresses the challenges of mapping such outbreaks by introducing

federally protected under the United States Government Health Insurance Portability and Accountability Act (HIPAA). They are not routinely available. The NC-DHHS reviews data requests after a formal application for the data. To request access to the data, contact the NC-DHHS Communicable Disease Branch, HIV/STD Prevention and Care Unit by phone at: 919- 733-3419 or mail at: 1905 Mail Service Center, Raleigh, NC 27699-1905; https://www.ncdhhs.gov/divisions/public-health. Access to the data is not guaranteed. The BME numerical processing of these data was performed using MATLAB 7.8 and the BMElib version 2.0c. The code created and used in the analysis can be found at: https://mserre.sph.unc.edu/BMElab_web/mappingStudies/Syp_ForsythNC/. The BMElib code is freely available at: https://mserre.sph.unc.edu/BMElib_web/. For those who do not have programming knowledge there is also a graphical user interface version of BME (BMEGUI) that guides a user through the BME process, download it at: https://mserre.sph.unc.edu/BMElab_web/; code and programs can also be accessed by contacting the University of North Carolina at Chapel Hill Gillings School of Global Public Health, Department of Environmental Sciences and Engineering: https://sph.unc.edu/envr/environmental-sciences-and-engineering-home/.

Funding: This study was supported by the National Institute of Allergy and Infectious Diseases, National Institutes of Health (R01 AI067913) to WCM, DG, ID and MS. Additionally, ID also received funding from National Institutes of Health (U54MD012392). All authors received salaries from the above grant to complete the research. The funders had no role in study design, data collection.

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

innovative methodologies that enhance spatial resolution while preserving patient privacy. Analyzing syphilis surveillance data from the North Carolina Department of Health and Human Services, our work finds the combination of these techniques resulted in more accurate predictions of true syphilis rates, improving accuracy by 5% to 26% over traditional geopolitical mapping methods. The results also identified localized hotspots more effectively, revealing a complex network of transmission emerging from urban endemic areas appearing to spread through sexual core transmitters.

Our study emphasizes the changing landscape of syphilis transmission including shifts in partnership dynamics, showing while online platforms have changed how individuals connect geographic proximity remains a key factor in partner selection. Our study underscores the importance of contextualizing outbreaks within historical and current trends and the continued need for precise visual representations of STI data to aid public health interventions. Overall, this work provides framework for future STI surveillance and response strategies.

1. Introduction

Increased understanding of syphilis outbreak development and progression is important as syphilis rates not only persist but are increasing in the United States (US) [1–2] and globally [3–4]. Since reaching a historic low in the US in 2000 and 2001, syphilis has increased nearly yearly and expanded 28.6% from 2020 to 2021 [1–2]. The southeastern region of the US continues to experience higher incidence rates of syphilis than elsewhere in the states [5–13]. Those with ulcerative sexually transmitted Infections (STI), such as syphilis, have an increased likelihood of transmitting or becoming infected with HIV leading to additional concerns about increases HIV morbidity [1–4,14–17]. Additionally, syphilis has continued to be understudied in both the United States and globally [12,18–27], with one article in 2002 titled, "Don't Forget syphilis. . ." [18]. Furthermore, the spread of syphilis is commonly described based on individual-level behaviors, rather than the spatiotemporal or community factors [27]. Thus, even fewer studies track the geographic progression of syphilis outbreaks in the United States [11–12,19–20,27] and additional research on this issue has been advocated for [6,12,19–20,27–28].

As syphilis rates rise in 2022 in North Carolina (NC) very little published research exists on NC outbreaks occurring after 2007. At the end of 2010 Forsyth County, North Carolina (NC) experienced a dramatic, over four-fold increase in syphilis cases and reported the second most cases of early syphilis in the state [14–17]. Similar to previous syphilis outbreaks [5–13,14–17,19–22], escalations in syphilis morbidity during the Forsyth County outbreak of 2008–2011 were present in nearly all demographic groups including those already co-infected with HIV [15–17]. The early syphilis rate in Forsyth County during this outbreak rose to over 35 cases per 100,000 [15–17] similar to the NC state average in 2021 (30 per 100,000) [14].

As shown in Fig 1, in 2022, early syphilis (primary, secondary and early latent syphilis) rates in North Carolina jumped 631% since 2018 [29]. Fig 1 also shows the rise of syphilis in both the US and NC over the past 40 years. Furthermore in 2022, the high-risk demographic groups continue to largely remain the same in North Carolina and the US [1–2,14,29]. Using syphilis surveillance data for Forsyth County, NC from 2008–2011 we analyze this outbreak and use it as a case study to introduce and build upon novel mapping methodologies to better understand STI outbreaks.

When creating STI outbreak maps, protecting patient privacy is essential. In public maps, privacy is commonly protected by aggregating to coarse incidence areas. An incidence area is

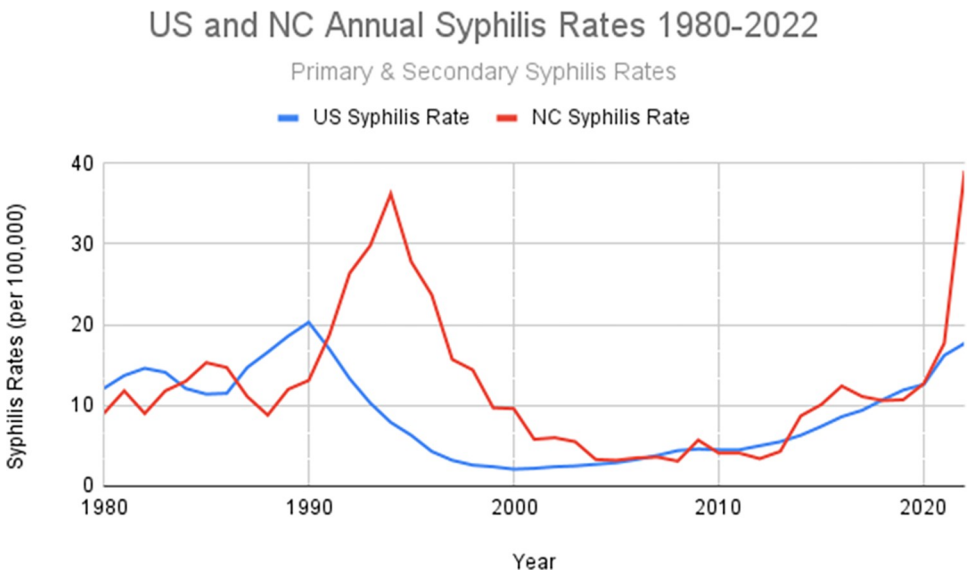

Fig 1. Annual Primary and Secondary Syphilis Rates from 1980–2022, data compiled from a variety of US Centers for Disease Control (CDC) and NC Department of Human Health Services (NC DHHS) public databases [29–36].

https://doi.org/10.1371/journal.pcbi.1012464.g001

the geographical foundation on which an incidence rate is defined; the number of new cases in an area over a given period. Most incidence areas are defined using geopolitical boundaries such as states, counties, or census block groups (CBGs). Incidence rates are then assigned to the incidence area's geographical centroid. Although most incidence areas have set geopolitical boundaries, they can also be defined with uniform boundaries with a set shape, such as a circle with centroids based on a regular grid.

Centroid assignment to geopolitical boundaries can obscure pertinent geographical information needed to conduct effective disease mapping [19,37] and cause multiple geostatistical issues leading to inappropriate interpretation of maps. These include the: 1) Modifiable Areal Unit Problem (MAUP); 2) edge effect [19,38–39]; 3) kriging islands of artificially higher and lower incidence at centroids [19]; 4) potentially masked hotspots (hotspots are localized geographical areas with elevated incidence rates) [19,37]; 5) a background rate greater than zero; and 6) the small number problem [19]. MAUP comprises two components: 1) aggregation zone size and 2) areal unit shape [20,38–42]. Even small administrative areas such as census block groups vary greatly in size, shape and underlying population creating a non-uniform distribution of population and geographic centroids. For example within US states, California vs Rhode Island have the same aggregation unit however they have massive differences in both size and population. The edge effect results from displacing an incidence rate to a geographical centroid causing ecological bias. In a map it is shown as hotspots not located in the correct geographic location, the complete loss of a hotspot and/or multiple hotspots incorrectly connected or disconnected to each other [19,38–40].

An increase in spatial resolution (i.e. from state to county to zip code) can reduce MAUP and the edge effect and clarify spatial patterns. With a resolution increase, aggregated rates are displaced less from their original, non-aggregated point sources [19–20,40–47]. Geomasking techniques such as donut geomasking used in this study can ensure patient privacy while increasing the spatial resolution necessary for cluster and outbreak detection [28, 40–47]. However with rare diseases such as syphilis, reducing the aggregation scale also introduces errors due to the scarcity of the data.

Scarcity in disease data can create maps dominated by locations with low populations. This produces challenges to interpret incidence rates and their maps because rates in low population areas can appear alarmingly elevated even when a single case is present, known as the "small number problem" [28,48–50]. Concurrently the size of many geopolitical areas such as census tracts or block groups are population dependent, therefore areas with large populations also have small areas and vice versa. This is generally seen as maps with a hotspot over a large area. Multiple smoothing algorithms have been developed to reduce the small number problem improving the interpretable accuracy of incidence rates estimates also referred to as the true or latent rate [19,48,28,51–54]. Uniform Model Bayesian Maximum Entropy (UMBME), an advanced kriging based smoothing method can produce incidence rate estimates with greater accuracy through mathematical penalizing of areas with small populations [19,28,44].

Past STI studies [19] found the edge effect is especially prevalent in the prediction method kriging where rates are highest at the prediction point. This phenomenon is known as "kriging islands". Kriging islands can make maps more difficult to interpret by obscuring the ability to fully understand the progression of an outbreak and how it moves between susceptible communities [19]. Using the Forsyth County 2008–2011 syphilis surveillance data, we advance STI outbreak analysis methods by increasing spatial resolution with donut geomasking and introducing a global moving window approach to minimize MAUP, edge effect and kriging islands. Finally, employing the Uniform Model Bayesian Maximum Entropy (UMBME) [28,38] method moderates the small number problem. We hypothesized combining these methods would create more easily interpretable maps better delineating the geographical extent and connection of clusters [19]. Using the 2008–2011 Forsyth outbreak as a case study for these methods, our study continues to fill in the gap in our geographic understanding of syphilis and other STIs contextualizing it to other outbreaks.

2. Methods

Study population & data preparation

The study used incidence data from the North Carolina Department of Health and Human Services (DHHS) and US Census data. North Carolina health care providers and laboratories are required to report each newly diagnosed case of syphilis to the county health department including diagnosis date, date of disease or symptoms onset, current residence, syphilis disease stage and limited demographic information. Data preparation, geocoding and geomasking were performed on site at the NC-DHHS. Our first step removed all patient demographic information. 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).

This work focuses on the spatiotemporal analysis of incidence rates in Forsyth County, NC from 2008–2011. We used syphilis diagnosis stage to estimate when a patient acquired the disease. To support the work of the DHHS who both provided the data for the project and supported it, the early syphilis categories (primary, secondary and early latent) were also used in our analysis. Early syphilis is the stages when the syphilis pathogen has highest infectivity for sexual partner transmission. Late latent syphilis is not considered infectious and excluded from the analysis [20,44]. We estimated the date of infection based on the diagnosis date and median latency period for each syphilis disease stage. Primary syphilis, secondary syphilis and early latent syphilis were back-estimated by 45, 90, and 183 days, respectively [60].

Geographically, Forsyth County is one of the smallest of the 100 counties in North Carolina spanning approximately 1,070 km. In 2010 it was also the 5th most population dense with approximately 351,000 residents. Forsyth has 2 major cities, Winston-Salem and High Point, the fourth

and seventh most populous cities in NC. The remainder of the county is primarily rural [55]. Interstate 40 (I-40) also runs through the county. I-40 spans the entirety of the state and is one of the most critical US east-west highways connecting North Carolina to California [56].

Self-reported residential addresses were corrected with Satori Bulk Mailer software [57] to optimize the geocoding match rate of addresses. Case residences were geocoded using ESRI's ArcGIS 9.3.1 [58] and matched to three geographical locators used by the NC-DHHS for geocoding. The first geocoding round used our highest quality and primary locator created by the North Carolina Department of Transportation containing street-level geographical data. Cases not matched using our first locator were then matched using our secondary locator created by the North Carolina Emergency Response System which contains GPS point locations for households. Cases not matched to the first or second locator were then matched to our tertiary locator created using ESRI's 2006 Street Map shapefile [59], which is primarily used for locating residences with outdated street names, prisons, and military bases. Finally, cases with a post office box address were spatially assigned to that post office address. Matching in this method allows for the highest quality geographic data to be linked to the cases while maximizing the number of cases matched to a geolocator.

Approximately 83% of records were successfully geocoded to a location (n = 497 of 601). Cases that were not geocoded (n = 104 of 601) were excluded from the analysis. Of the cases not geocoded, 36 were in 2009 (6%) and 16 in 2010 (3%). Failure to geocode was usually because: 1) no address is reported; 2) non-existent/false addresses reported; 3) street segments are missing locators- most often occurs on rural routes with unavailable E911 addresses; or 4) incorrect/incomplete addresses (misspellings, abbreviated street names and improper use of rural routes). Failure to match may also occur with persons unhoused, out of state, or with military addresses. Data that could not be geocoded appeared to be missing at random without any noticeable patterns thus we believe it only added noise rather than an inherent bias. After geocoding, data were geomasked using the donut method [28,44–45]. Donut geomasking randomly relocates each geocoded data point within a user defined minimum and maximum radii [28].

Syphilis incidence rates

Incidence areas. Two datasets of syphilis incidence rates were created and compared, US census block groups (CBG) and a uniform grid with a moving window (Fig 2). Our first dataset, composed of the traditional, geopolitical boundaries was created in ESRI's ArcGIS 10 [58] using the US 2000 Census populations and census block group boundaries packaged in a shapefile [58]. This file was also used to calculate the CBG centroids. Fig 2 clearly shows the irregular locations of the CBG centroids, size and shape demonstrating a strong potential for MAUP and the edge effect. The uniform grid approach changes the area of aggregation from a geopolitical boundary to a standardized area with a consistent centroid location allowing aggregation without the variability caused by size and shape [38–40]. Furthermore, overlapping the boundaries of the uniform grid area creates a continuous surface, reducing the edge effect [39–41]. We hypothesize kriging islands, MAUP and edge effect created when aggregating to geopolitical incidence groups can be corrected by creating a uniform grid with a moving window of overlapping perfect circles (shown as turquoise dots with blue circles in Fig 2).

To calculate incidence rates, CBG cases were aggregated spatially by census block group boundaries, and temporally with a rolling period of 6 months duration to lessen the small number problem. Population growth was also incorporated into the crude rate calculation through a linear interpolation of the census block group population for all 12-64-year-olds in 2000 and 2007 assuming positive population growth over the time period.

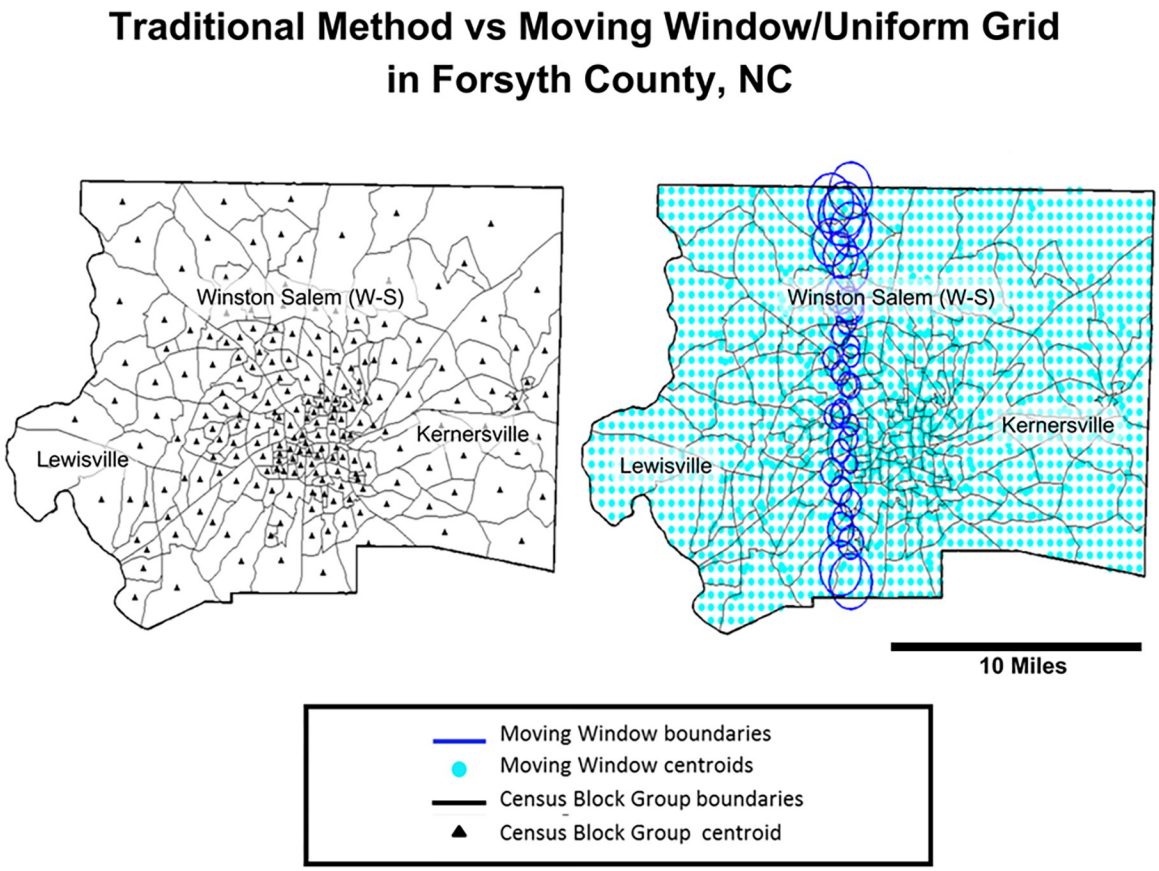

Fig 2. Geopolitical Census Block group aggregation vs standard grid with a moving window aggregation for incidence areas in Forsyth County, NC. The CBG boundaries show the non-uniform distribution of geographical centroids whereas the uniform grid/moving window show an even distribution of centroids and incidence areas size to reducing the edge effect and MAUP. Census Block Group Boundaries are publically available by the US Census at: census.gov/geographies/mapping-files/time-seriesgeo/carto-boundaryfile.2000.html#list-tab-U2GTBOCTYRAP13D9AH.

https://doi.org/10.1371/journal.pcbi.1012464.g002

The crude incidence is denoted as R_{ij} and calculated as $R_{ij} = y_{ij}/(n_{ij}T)$. Incidence period of duration T expressed in years (i.e. $T = 0.5$yr for a 6 month incidence), centered at time t_j, $j = 1$ denote the number of rolling time periods (39 rolling 6-month aggregated time periods with monthly estimates). $i = 1$ denotes the incidence area (5259 census block group centroids, 10376 ubiquitous group centroids), y_{ij} is the number of incident syphilis cases and n_{ij} is the population at time t_j. Time periods that did not contain incident syphilis reports were assumed to have a rate of 0 cases per 10,000. The CBG dataset is composed of these crude rates.

To construct rates for the uniform grid based method, the CBG R_{ij} rates are enriched with a grid-based incidence rate field. Cases located within each of these grid-based incidence areas are identified and the underlying population within the ubiquitous area is systematically calculated and used to construct a rate assigned to each grid centroid. The uniform grid is a series of grid points constructed solely within Forsyth County.

The distance between grid points, D was set using the following equation:

$D = f * \frac{1}{n}\left(\sum_{a=1}^{n} \frac{\sqrt{A_a}}{\pi}\right)$, where n is the user selected number of CBGs (we used 5), A_a is the area of the ath CBG (i.e. 1 of 5), $\frac{1}{n}\left(\sum_{a=1}^{n} \frac{\sqrt{A_a}}{\pi}\right)$ calculates the average CBG area, and f is the factor

used to increase or reduce the spatial resolution of the grid. For this study we used $f = 0.65$ to create a fine grid lattice of grid points throughout the study area.

Each uniform grid (UG) boundary is a perfect circle with a radius r_i, where i is the spatial location of the uniform grid centroid/"point". r_i is calculated using an inverse weighted distance area average of the five closest CBGs to r_i. It is shown as, $r_i = \sum_{a=1}^{5} d_{ia}^{-1} * \frac{\sqrt{A_a}}{\pi} / \sum_{a=1}^{5} d_{ia}^{-1}$ where d_{ia} is the distance between the grid point and its a^{th} CBG neighbor. This approach allows for a gradual change in the size of the uniform grid areas in relation to their local CBGs, as seen in Fig 2. The boundaries of the uniform grid overlap producing a continuous rate field that smooths the edge effect.

The population of the uniform incidence area is a summation of the proportion of populations from each census block groups it overlaps. Calculated as: $n_{ij} = \sum_{a=1}^{n} \frac{\|\mathrm{CBG}_a \cap \mathrm{UG}_i\|}{\|\mathrm{CBG}_a\|} \mathrm{n}_{aj}$. The percentage of every census block group within the uniform grid incidence area is calculated, where $\|\mathrm{CBG}_a \cap \mathrm{UG}_i\|$ is the area of the intersection between CBG_a and uniform grid boundary i *(UG$_i$), and* $\frac{\|\mathrm{CBG}_a \cap \mathrm{UG}_i\|}{\|\mathrm{CBG}_a\|}$ is the proportion of the area of *CBG$_a$* in $\mathrm{UG_i}$. This value is then multiplied by n_{aj}, the CBG population, n in *CBG$_a$* at time t_j. These population values were then summed to calculate the total population for the uniform grid incidence area, n_{ij}. Although this approach assumes a uniform population distribution within the CBG, the regularity of the grid centroid placement pulls populations from multiple centroids, creating an enriched, modeled population at each point smoothing some of the issues with the population homogeneity assumption.

To continue to increase the spatial accuracy of the dataset, we reverse calculated the donut geomask for each case. The number of cases within each uniform incidence area was calculated as a function of the probability a geomasked case is within the uniform grid. The following is known about the geomasked cases: 1) each case is geomasked within the CBG they are located in and 2) there is a maximum and minimum distance a case can be moved from its original location [28; 44–45] (these values can only be accessed on secure computers onsite at the NC-DHHS by selected researchers). From this we can infer the probability w_{li} of the l^{th} geomasked case at time t_j in area UG_i is $w_{li} \approx \frac{\|\mathrm{RDCBG}_l \cap \mathrm{UG}_i\|}{\|\mathrm{RDCBG}_l\|}$. *RDCBG$_l$* is geographic area of original donut area case l was initially located in its CBG before geomasking. More specifically, $\mathrm{RDCBG}_l = \mathrm{RD}_1 \cap \mathrm{CBG}_1$, where *RD$_l$* is the reverse geomasked donut centered on the geomasked location of case l, and *CBG$_l$* is the Census Block Group area case l is located in. Then, the uniform grid incidence rate over area *UG$_i$* at time t_j is calculated as $R_{ij} \approx \frac{\sum_{l=1}^{N_j} w_{li}}{n_{ij} T}$, where N_j is the total number of cases at time t_j, and $\sum_{l=1}^{N_j} w_{li}$ is the sum of the probabilities that each case at time t_j is located in *UG$_i$*.

Spatiotemporal analysis and incidence mapping

Bayesian Maximum Entropy (BME) is a highly developed approach of contemporary geostatistics using a spatiotemporal analysis structure [61–69]. BME techniques based on kriging have been successfully applied in wide range of public health and environmental applications [19,28,44,61–69]. These methods generate a continuous surface of exposure (such as syphilis incidence) and can incorporate soft data modeled by a distribution (population/ small number problem) as done in the Uniform Model BME (UMBME) method. In public health analysis UMBME methods have fundamental benefits over other methods including the ability to comprehensively minimize the small number problem.

UMBME introduces a measure of variance based on population allowing the method increase the variance (i.e. uncertainty) in areas with low populations, effectively penalizing their rates and categorizing them as more unreliable in the BME estimation model. This results

in both increased visual map interpretation and more accurate mathematical predictions [28,44]. Though kriging based interpolation methods such as BME are highly accurate [19,28,44,61–69], one downside is the creation of "kriging islands" [19]. This is seen at actual data point locations within the map (i.e. CBG centroids) as higher or lower rates than the surrounding areas [19].

UMBME rates were calculated for both the CBG and uniform grid data sets to allow for model comparison. BME and UMBME methods are coupled space-time calculations that provide estimates for each location at each time period using both space and time data to populate the model. The inputs for the BME spatiotemporal model include: the data points (CBG and UG) containing rate values, the mean trend for the data (must be removed to ensure model assumption of field homogeneity), a covariance model, and the UBME rates constructed from the localized population values. The output for the BME analysis returns a space-time rate field of the calculated point estimates and their associated estimation error. These point estimates can then create time series maps of the Forsyth County syphilis outbreak [28,54,62–64].

The BME numerical processing was performed using MATLAB 7.8 [70] and the BMElib version 2.0c [71]. The code created and used in the analysis can be found at: https://mserre.sph.unc.edu/BMElab_web/mappingStudies/Syp_ForsythNC/. The BMElib code is freely available at: https://mserre.sph.unc.edu/BMElib_web/. For those who do not have programming knowledge there is also a graphical user interface version of BME (BMEGUI) that guides a user through the BME process, download it at: https://mserre.sph.unc.edu/BMElab_web/ [19].

Cross-Validation of BME Methods

A cross-validation of the CBG and uniform grid methods was conducted to identify the most accurate method of modeling syphilis incidence rates. For an estimation method of interest, cross-validation comprises removing each observed value incrementally and allowing the model estimate the value of that point using the remaining data. An estimation error value is also calculated for each point and represents the difference between the actual data value and its estimation. To calculate the mean square error (MSE), the cross-validation errors for each point in each method are averaged. The MSE is always positive and a MSE of zero demonstrates the estimator perfectly predicts an observation. The MSE is effective for comparing the prediction ability of methods, where the method with the smallest MSE is considered the best predictor for the data set. To compare the MSE between the methods, only the CBG data points contained within the uniform grid and CBG datasets are evaluated (n = 28,290 points).

The MSE formulas are:

$$MSE_{CBG} = \frac{1}{n}\sum\left(\hat{X}_{ij}^{(CBG_i)} - \frac{y_{ij}}{n_{ij}T}\right)^2$$

$$MSE_{UG} = \frac{1}{n}\sum\left(\hat{X}_{ij}^{(UG_i)} - \frac{y_{ij}}{n_{ij}T}\right)^2$$

where y_{ij} is the number of incidence cases, n_{ij} is the population in CBG_i at time t_j, T is the incidence duration (i.e. 6 months), $\hat{X}_{ij}^{(CBG_i)}$ and $\hat{X}_{ij}^{(UG_i)}$ are the incidence rates estimated using the CBG and uniform grid methods. A mean squared error rate is calculated for each point in the dataset. These are summed and n is the number of observed rates across the CBGs and observation times.

However because of the small number problem, latent rate theory is introduced and used to more effectively compare model MSEs [28,54]. Latent rate theory states as a location's population approaches infinity, the observed rate reaches the latent also referred to as the true rate.

This can also be interpreted as the rate if the small number problem was not present [28,53–54]. The latent incidence rate of disease in a given region can be defined as: $X_{ij} = \lim_{n_{ij}\to\infty}\left(\frac{y_{ij}}{n_{ij}T}\right)$ where X_{ij} is the latent disease rate, y_{ij} are the incidence cases and n_{ij} is the population. In practice, you can very simply estimate the latent rate by stratifying the MSE result for each point in a model by its population. To calculate each point's MSE values and their corresponding populations are categorized into population percentiles and an average MSE is calculated for each population percentile (i.e. 10, 20. . .) [28]. Finally, the difference between the average MSE for the two models at the population percentiles are calculated, this quantifier is referred as the percent change in the mean square error (PCMSE) [28,53–54].

With PCMSE, the method that more accurately predicts the latent rate will result in both a negative PCMSE and its negative PCMSE will remain stable or become increasingly negative as population percentile increases [28; 53–54]. The PCMSE is calculated by: $PCMSE_{UG} = 100 * \frac{MSE_{UG} - MSE_{CBG}}{MSE_{CBG}}$

3. Results

Our cross-validation demonstrated the uniform grid model performed noticeably better in predicting the latent rate than the CBG model minimizing the mean squared error of the predictions. Fig 3 graphs the PCMSE between the two methods showing a clear decrease in the PCMSEUG as the population percentile increases. This demonstrates the uniform grid method is a better predictor of the latent rate, predicting syphilis rates 5% to 26% more accurate than the traditional, geopolitical based, CBG method.

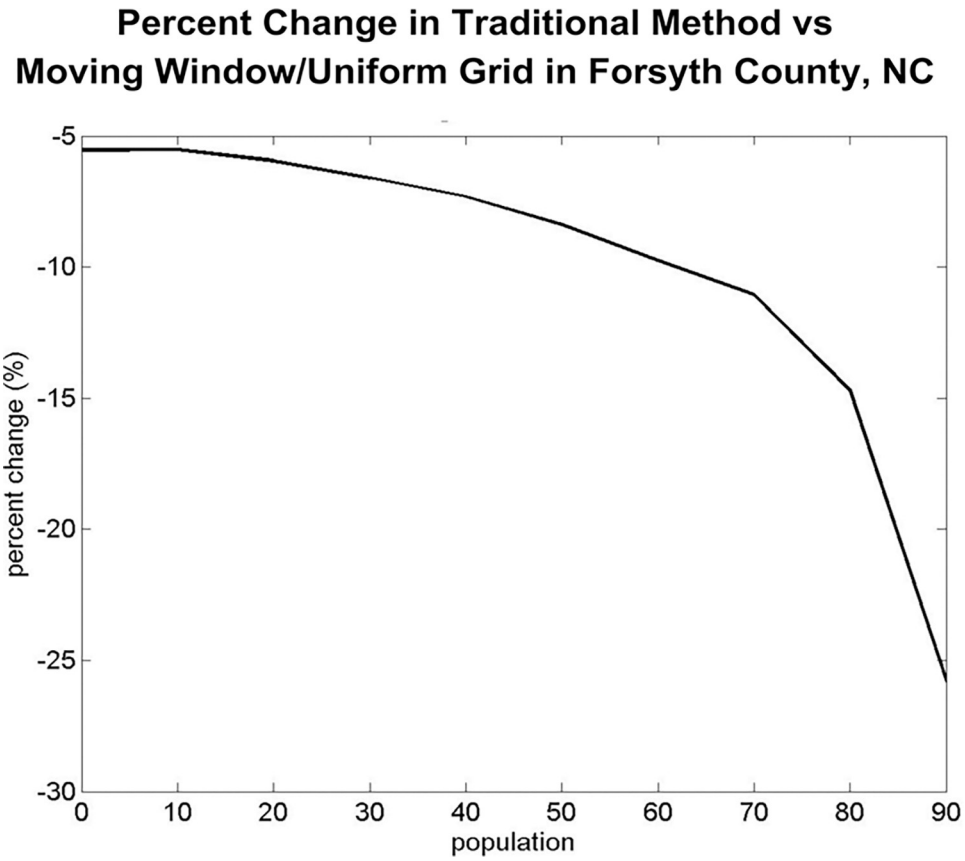

Fig 3. The percent change in MSE from CBG to uniform grid as a function of the population percentile.

https://doi.org/10.1371/journal.pcbi.1012464.g003

In addition to the mathematical increases in prediction accuracy differences in the methods are visible in the maps of the two approaches (Fig 4). The uniform grid method shows increased connectivity between hotspots compared to the CBG method (Fig 4, Boxes 1, 2, 3 and 5). The uniform grid method also led to identification of an additional hotspot (Fig 4 and Box 5). In the CBG map, Box 5, a small hotspot is misplaced outside of the Winston-Salem city limits, likely produced as a kriging island. In contrast, in Box 5 the uniform grid method reduces the kriging island effect showing instead a pronounced hotspot within the Winston-Salem city limits. New hotspots also appear in the uniform grid map and are shown in all the Boxes in Fig 4, particularly in areas between spatial aggregations. Furthermore the interconnectedness of hotspots is changed in both maps demonstrating the edge effect in the CBG map.

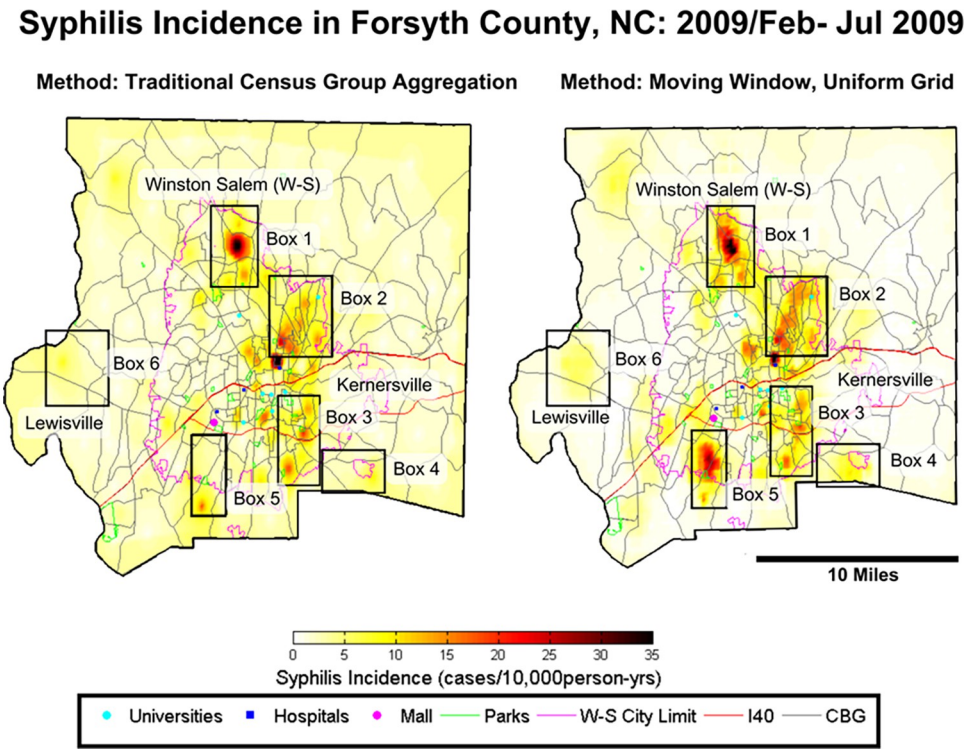

Fig 4. BME maps of the CBG & uniform grid methods in February-July, 2009 (the peak of the outbreak). The boxes highlight the differences in the maps showing increased connectivity, hotspot locations moved, and a reduced background rate in the Moving Window, Uniform Grid Map. Census Block Group Boundaries are publically available by the US Census at: census.gov/geographies/mapping-files/time-seriesgeo/carto-boundaryfile.2000.html#list-tab-U2GTBOCTYRAP13D9AH.

https://doi.org/10.1371/journal.pcbi.1012464.g004

Additionally, the background rate in the CBG map is approximately 5–9 cases/10,000person-years, indicated by the overall darker yellow color for the whole county and likely the result of the kriging islands effect. The uniform grid method has reduced the kriging island effect allowing a truer background rate of 0–2 cases to be shown. This permits more minor hotspots to appear in the rural areas of the map highlighting rural areas of concern shown in Boxes 4 and 6.

Outbreak information can also be gained from the covariance model produced in the BME analysis. The covariance model describes the geographic and temporal features of the dataset. The spatial component of the covariance model and it's the experimental covariances values (top panel of Fig 5), indicate syphilis outbreak hotspots are generally concentrated to small areas (less than 10% of the study area). Temporally, the covariance shows hotspots persisting for more than 6 months up to the duration of the Forsyth outbreak. This shows that resources could potentially be targeted to small areas that persist for relatively long time periods.

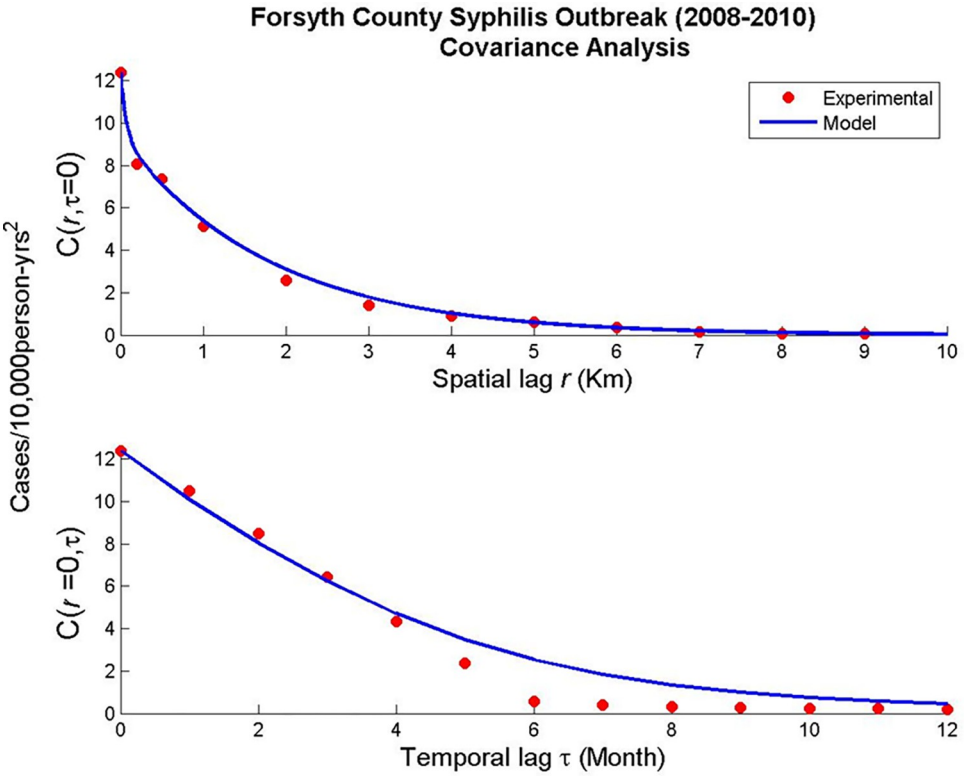

Fig 5. Spatial and Temporal Covariance Models.

https://doi.org/10.1371/journal.pcbi.1012464.g005

BME maps show the pattern of outbreak development (Fig 6; movie also available at https://mserre.sph.unc.edu/BMElab_web/mappingStudies/Syp_ForsythNC/RevisedUMBMEMovieForsythGrid.gif). Across Forsyth County, areas with an incidence of approximately 15 cases/10,000person-years are identifiable throughout the southeastern region of Winston-Salem and lower rates of 2-5cases/10,000person-years are found within Lewisville and Kernersville.

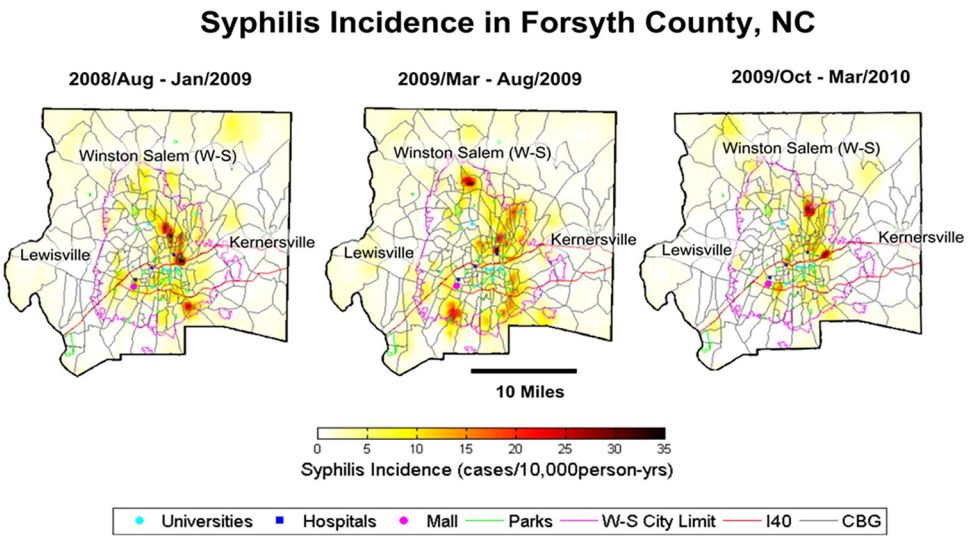

Fig 6. Uniform grid method time series of the Forysth County Outbreak (2008–2010). Census Block Group Boundaries are publically available by the US Census at: census.gov/geographies/mapping-files/time-seriesgeo/carto-boundaryfile.2000.html#list-tab-U2GTBOCTYRAP13D9AH.

https://doi.org/10.1371/journal.pcbi.1012464.g006

Rapid incidence increases to ≥20 cases/100,000 person-years within small geographical boundaries cumulatively lead to an outbreak that began to spread. In the aggregated time period of June-November, 2008 small hotspots begin to appear and increase; some portions exhibit rates at or above 35 cases/10,000person-years. In September, 2008-February, 2009 the hotspots continue to spread throughout Winston-Salem, increasing in intensity and connection in the north central and north eastern portions of the city. As the outbreak progresses, new hotspots appear in the eastern region of Winston-Salem and begin to connect (Fig 6 and video). The peak of the outbreak is February-July, 2009. As the outbreak wanes, the hotspot in the north-central region of the city is reduced, while a hotspot in northeast increases. This hotspot shows increased connection in April-September, 2009 specifically in central-east Winston-Salem near I-40 and in the northeast regions. These hotspots continue to grow and connect until May-October, 2009 then wane and disconnect.

In November, 2009-April 2010, the rates begin to return to endemic levels. In 2010–2011 the outbreak continues to decrease with increasing disconnection of the hotspots. By February-July, 2010 the Winston-Salem incidence has returned to endemic levels of 3–7 cases per 100,000 and the outbreak has subsided. Endemic is when the incidence is low and stable, and was defined in a study by Hook in the Lancet in 1998 as an area with 4 or more syphilis cases per 100,000 [72]. Furthermore, the Kernersville and Lewisville syphilis rate returns to 1–5 cases 10,000person-years. Fig 6 and the movie display the progression of the syphilis epidemic across Forsyth County between 2008–2010 with its three largest towns labeled (Winston-Salem, Lewisville, and Kernersville).

4. Discussion

The methods presented in this article advance outbreak analysis by addressing common issues in sexually transmitted infection (STI) mapping creating more accurate and more easily interpretable maps. Conducting a spatiotemporal case-study review of the syphilis outbreak in Forsyth County, North Carolina in 2008–2011 allowed insight into how the outbreak emerged from endemic areas, developed and then subsided. We show combining a moving window

and a UMBME analysis with geomasked data produced more realistic geographical patterns for more accurate and targeted insights into a syphilis outbreak. The improved visualization in the maps was accompanied by less apparent error, with the uniform grid model performing 5–26% better in predicting the true rate compared to the traditional, geopolitical CBG based model.

Our approach is particularly novel because we introduce the ability to remove what is commonly referred to as "kriging islands", where rates are higher and lower at centroids [19] providing greater sensitivity for hotspot detection. This enhanced visualization is especially important when hotspots are located on or across the boundary of two or more aggregation areas which traditionally can result in the complete loss of a hotspot. Furthermore, removing kriging islands reduces the background rate which unmasks rural hotspot areas of concern. Thus, removing kriging islands resulted in an increased ability to visualize spatial corridors of incidence and a more accurate representation of interconnection and spread. Throughout the outbreak, our methods placed incidence hotspots within the Winston-Salem city limits, rather than at administrative centroids which are often outside the city limits or in underdeveloped, sparsely populated areas.

This Forsyth outbreak in 2008–2011 is similar to the 2001 Robeson/Columbus, North Carolina syphilis outbreak [19], and began in areas with endemic syphilis. As the endemic hotspots intensified they also spread and leap-frogged to connect into non-endemic areas. In both outbreaks, the hotspots were highly localized and generally persisted for one year [19]. This endemic leap-frogging is also similar to what was seen in the Baltimore, Maryland syphilis outbreak in 2002 [67]. The Forsyth outbreak slightly contrasted to the Robeson/Columbus outbreak in 2000–2002, because in Forsyth the cases were primarily contained to the urban, endemic areas of Winston Salem; whereas in the Robeson/Columbus outbreak the cases leaped from the cities to the rural areas [19]. Additionally, we found many of the hotspots in the Forsyth outbreak were located near or in between popular meet up areas, such as parks, highways, and popular malls in Winston Salem. This is similar to past outbreaks which have moved along the Interstate Highway 95 [13].

Other work has also shown cases with partners from many locations can act as spatial spread/network bridges, geographically distributing the disease [12]; this bridging effect appeared to be occurring in this outbreak as well. In a geographic STI partner study in Toronto in 2016, 78% of their urban-outbreak sexual network participants remained in the outbreak/hotspot areas to find new partners and rarely travelled to the suburban or rural areas for partner selection. However, 81% of those living in the rural areas and suburbs travelled into the urban outbreak areas to meet new partners [73]. This may have been occurring in this outbreak as well, as the bulk of the cases in the 2008–2011 Forsyth outbreak were isolated to the urban areas, however there are some isolated cases in the more suburban and rural areas near Lewisville and outside of Winston-Salem. Removing the kriging islands allowed the identification of these rural hotspot locations, which otherwise would have been masked within a homogenous background rate.

Some of the dynamics of syphilis transmission have changed since 2008. First, in the US overall the rate of primary and secondary syphilis for men who have sex with men (MSM) has decreased from 72% in 2010 [74] to 45% in MSM with an additional 20% in men who have sex with unknown partners (MSU) in 2021 [2]. However, some of the 2021 MSU data likely could be categorized within MSM [75–76]. Conversely, in North Carolina in 2022 the early syphilis rate for gay, bisexual and other men who have sex with men was estimated to be 1,054 per 100,000 whereas the syphilis rates of: 1) men who only have sex with women and 2) women were both 17 per 100,000 [29]. As 85% of syphilis cases in 2008 were in men [16], in North Carolina the primary high risk population may not have significantly changed. Additionally many of

the rates seen in the 2008–2011 Forsyth outbreak [16] are similar to 2021 syphilis rates in the North Carolina, where the county rates in the 20 most affected counties ranged from 54.8–23.5 syphilis cases per 100,000 [14]. The gender change in syphilis transmission in the US in general from 2011 to 2021 could affect the application of the results of geographical transmission shown in this study to outbreaks outside of North Carolina.

Additionally methods for partnership selection have changed substantially since 2011. In 2008-

2011 people often "cruised", i.e. meeting partners at physical environments (e.g., bath-houses, parks, parks, pools, malls). With the wide-spread adoption of internet based phones, most people now rely on apps and online dating websites for partner selection changing the community level risk space for syphilis [73,77–80]. However studies have also shown participants who have a more challenging time with dating apps still frequent common activity spaces to meet partners [73,80].

Yet even in the dating app era, localized geographic proximity is still a key factor partner selection for both men and women [73,80–81]. In practice, distance travelled to meet sexual partners may not have drastically changed over time. Interviews on partner selection have shown participants preferentially selected partners for whom they had to travel the shortest distance possible to connect [73,80]. Research continues to support that partnership selection is a localized event whether this selection occurs online or offline [73,80–83]. Therefore many of the assumptions around geographical partner selection dynamics shown in this work may remain in 2024.

There are a few other limitations of our study. First, we focused on a single county and single outbreak. Primary and secondary syphilis cases are heavily geographically concentrated [19–20,27,84], more than other STIs such as chlamydial infection and gonorrhea [84–85]. Although our findings mirror other syphilis outbreaks [13,19–20,27,67], use of a different geographical area, such as an entire state or country, may have led to different conclusions [19–20,67,27]. Advances in BME techniques, big data methods and optimization have recently occurred to allow BME analysis to be performed on large domains, such as Kenya [86] with 2 billion data points and the continental United States [87]. In the past, BME studies required a relatively small spatiotemporal dataset for computation. Second, the combination of the UMBME method with a moving window and regular grid is ideally suited for relatively rare conditions, such as syphilis. The potential benefits of this approach for more common infections, such as chlamydial infection and gonorrhea, are uncertain [84–85].

Our approach can benefit health authorities by providing greater resolution of ongoing outbreaks, allowing for targeted intervention and resource allocation [12,19,27,44,65,73]. Even if the spatial dynamics of STI transmission have changed it is still important to have highly accurate and easily interpretable visual representations of STI data. Understanding the spatial progression of outbreaks can help to contextualize current and past outbreaks and help describe the spread within and outside sexual core transmitters and their sexual network [19,73]. These insights can provide public health officials a clearer view of the sexual network location and their potential activity spaces (where the sexual behaviors are actually occurring) [12,19–20,27,73] to isolate areas of need and inform intervention strategy selection and development [73].

Author Contributions

Conceptualization: Lani Fox, William C. Miller, Dionne Gesink, Marc Serre.

Data curation: Lani Fox, William C. Miller.

Formal analysis: Lani Fox, Marc Serre.

Funding acquisition: William C. Miller, Dionne Gesink, Irene Doherty, Marc Serre.

Investigation: Lani Fox, Marc Serre.

Methodology: Lani Fox, William C. Miller, Marc Serre.

Project administration: William C. Miller.

Resources: William C. Miller, Marc Serre.

Supervision: William C. Miller, Marc Serre.

Validation: Lani Fox.

Visualization: Lani Fox.

Writing – original draft: Lani Fox, William C. Miller, Marc Serre.

Writing – review & editing: Lani Fox, William C. Miller, Dionne Gesink, Irene Doherty, Marc Serre.

References

1. Reported STDs in the United States, 2021 [Internet]. www.cdc.gov. 2023. Available from: https://www.cdc.gov/nchhstp/newsroom/fact-sheets/std/std-us-2021.html.
2. Syphilis Surveillance Supplemental Slides, 2018–2022 [Internet]. www.cdc.gov. 2023. Available from: https://www.cdc.gov/std/statistics/syphilis-supplement/default.htm.
3. Kojima N, Klausner JD. An update on the global epidemiology of syphilis. Current epidemiology reports. 2018 Mar; 5:24–38. https://doi.org/10.1007/s40471-018-0138-z PMID: 30116697
4. Tsuboi M, Evans J, Davies EP, Rowley J, Korenromp EL, Clayton T, Taylor MM, Mabey D, Chico RM. Prevalence of syphilis among men who have sex with men: a global systematic review and meta-analysis from 2000–20. The Lancet Global Health. 2021 Aug 1; 9(8):e1110–8. https://doi.org/10.1016/S2214-109X(21)00221-7 PMID: 34246332
5. Farley TA. 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
6. Baffi CW, Aban I, Willig JH, Agrawal M, Mugavero MJ, Bachmann LH. New syphilis cases and concurrent STI screening in a southeastern US HIV clinic: a call to action. AIDS patient care and STDs. 2010 Jan 1; 24(1):23–9.
7. McNeil CJ, Bachmann LH. Syphilis: an old disease with present-day implications. North Carolina Medical Journal. 2016 Sep 1; 77(5):365–8. https://doi.org/10.18043/ncm.77.5.365 PMID: 27621352
8. Seña AC, Muth SQ, Heffelfinger JD, O'DOWD JO, 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
9. Rosenberg D, Moseley K, Kahn R, Kissinger P, Rice J, Kendall C, Coughlin S, Farley TA. Networks of persons with syphilis and at risk for syphilis in Louisiana: evidence of core transmitters. Sexually transmitted diseases. 1999 Feb 1:108–14. https://doi.org/10.1097/00007435-199902000-00009 PMID: 10029986
10. Seña A.C., Torrone E.A., Leone P.A. Foust E. and Weidman L.H. Endemic early syphilis among young newly diagnosed HIV-positive men in a southeastern US state. AIDS Patient Care and STDs. December 2008, 22(12): 955–963. https://doi.org/10.1089/apc.2008.0077 PMID: 19072102
11. Doherty L, Fenton KA, Jones J, Paine TC, Higgins SP, Williams D, Palfreeman A. Syphilis: old problem, new strategy. Bmj. 2002 Jul 20; 325(7356):153–6. https://doi.org/10.1136/bmj.325.7356.153 PMID: 12130615
12. Doherty IA, Serre ML, Gesink D, Adimora AA, Muth SQ, Leone PA, Miller WC. 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

13. Cook RL, Royce RA, Thomas JC, Hanusa BH. 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
14. 2021 North Carolina STD Surveillance Report [Internet]. 2022. Available from: https://epi.dph.ncdhhs.gov/cd/stds/figures/2021-STD-AnnualReportFinal.pdf.
15. North C, Beverly E, Perdue, Cansler L, Engel. State of [Internet]. [cited 2023 Jul 29]. Available from: http://epi.publichealth.nc.gov/cd/syphilis/NCSyphilisMorbidity2009.pdf.
16. 2010 HIV/STD Surveillance Report [Internet]. [cited 2023 Jul 29]. Available from: http://epi.publichealth.nc.gov/cd/stds/figures/std10rpt.pdf.
17. Disease Branch C. North Carolina STD Surveillance Report: 2011 Communicable Disease Surveillance Unit [Internet]. [cited 2023 Jul 29]. Available from: http://epi.publichealth.nc.gov/cd/stds/figures/std_tables_2011.pdf.
18. Clark P, Cook PA, Lighton L, Syed Q, Bellis MA. Don't forget syphilis: Syphilis outbreak is twice as big as reported. BMJ: British Medical Journal. 2002 Oct 10; 325(7367):775.
19. Fox LC, Miller WC, Gesink D, Doherty I, Hampton KH, Leone PA, Williams DE, Akita Y, Dunn M, Serre ML. Progression of a large syphilis outbreak in rural North Carolina through space and time: Application of a Bayesian Maximum Entropy graphical user interface. PLOS Global Public Health. 2023 May 4; 3 (5):e0001714. https://doi.org/10.1371/journal.pgph.0001714 PMID: 37141185
20. Escamilla V, Hampton KH, Gesink DC, Serre ML, Emch M, Leone PA, Samoff E, Miller WC. Influence of detection method and study area scale on syphilis cluster identification in North Carolina. Sexually transmitted diseases. 2016 Apr; 43(4):216. https://doi.org/10.1097/OLQ.0000000000000421 PMID: 26967297
21. Mobley V, Cope A, Dzialowy N, Maxwell J, Foust E, Samoff E. A comparison of syphilis partner notification outcomes by reported use of internet-based apps to meet sex partners in North Carolina, 2013–2016. Sexually transmitted diseases. 2018 Dec; 45(12):823. https://doi.org/10.1097/OLQ.0000000000000884 PMID: 29944644
22. Oliver SE, Cope AB, Rinsky JL, Williams C, Liu G, Hawks S, Peterman TA, Markowitz L, Fleischauer AT, Samoff E, Ocular Syphilis Disease Investigation Specialists Workgroup Hall Jason Hough Victor Ivey Andre Hawks Stephanie Greene Samantha Taylor Dishonda Mercurio Mike Gipson Miraida. Increases in ocular syphilis—North Carolina, 2014–2015. Clinical Infectious Diseases. 2017 Oct 30; 65 (10):1676–82.
23. Bowen VB, Peterman TA, Calles DL, Thompson AR, Kirkcaldy RD, Taylor MM. Multistate syphilis outbreak among American Indians, 2013 to 2015. Sexually transmitted diseases. 2018 Oct 1; 45(10):690–5. https://doi.org/10.1097/OLQ.0000000000000809 PMID: 30204746
24. DeSilva M, Hedberg K, Robinson B, Toevs K, Neblett-Fanfair R, Petrosky E, Hariri S, Schafer S. A case–control study evaluating the role of internet meet-up sites and mobile telephone applications in influencing a syphilis outbreak: Multnomah County, Oregon, USA 2014. Sexually transmitted infections. 2016 Aug 1; 92(5):353–8. https://doi.org/10.1136/sextrans-2015-052509 PMID: 27188272
25. Brodsky JL, Samuel MC, Mohle-Boetani JC, Ng RC, Miller J, Gorman JM, Espain G, Bolan G. Syphilis Outbreak at a California Men's Prison, 2007–2008: Propagation by Lapses in Clinical Management, Case Management, and Public Health Surveillance. Journal of Correctional Health Care. 2013 Jan 1; 19(1):54–64. https://doi.org/10.1177/1078345812458088 PMID: 22989493
26. Schillinger JA, Slutsker JS, Pathela P, Klingler EJ, Hennessy RR, Toro B, Blank S. The epidemiology of syphilis in New York City: historic trends and the current outbreak among men who have sex with men, 2016. Sexually Transmitted Diseases. 2018 Sep 1; 45(9S):S48–54. https://doi.org/10.1097/OLQ.0000000000000796 PMID: 29465651
27. Leichliter JS, Grey JA, Cuffe KM, de Voux A, Cramer R, Hexem S, Chesson HW, Bernstein KT. Geographical correlates of primary and secondary syphilis among men who have sex with men in the United States. Annals of epidemiology. 2019 Apr 1; 32:14–9.
28. Hampton KH, Serre ML, Gesink DC, Pilcher CD, Miller WC. 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
29. Syphilis in North Carolina, 2022 Fact Sheet [Internet]. Available from: https://epi.dph.ncdhhs.gov/cd/stds/figures/2022-SyphilisFactsheet.pdf.
30. 2009 HIV/STD Surveillance Report [Internet]. [cited 2024 May 29]. Available from: https://epi.dph.ncdhhs.gov/CD/stds/figures/std09rpt.pdf.
31. 2013 HIV/STD Surveillance Report [Internet]. [cited 2024 May 29]. Available from: https://epi.dph.ncdhhs.gov/CD/stds/figures/std13rpt.pdf.

32. Sexually Transmitted Disease Surveillance 2021 [Internet]. Available from: https://www.cdc.gov/std/statistics/2022/2021-STD-Surveillance-Report-PDF_ARCHIVED-2-16-24.pdf.

33. Table 21. Primary and Secondary Syphilis—Reported Cases and Rates of Reported Cases by State, Ranked by Rates, United States, 2022 [Internet]. www.cdc.gov. 2023. Available from: https://www.cdc.gov/std/statistics/2022/tables/21.htm.

34. Sexually Transmitted Disease Surveillance 2018 [Internet]. Available from: https://www.cdc.gov/std/stats18/STDSurveillance2018-full-report.pdf.

35. 2022 North Carolina STD Surveillance Report [Internet]. 2023. Available from: https://epi.dph.ncdhhs.gov/cd/stds/figures/2022-STDSurveillanceReportSummary.pdf.

36. NC SCHS: Statistics and Reports: Health Atlas: Syphilis—Additional Information [Internet]. schs.dph.ncdhhs.gov. [cited 2024 May 29]. Available from: https://schs.dph.ncdhhs.gov/data/hsa/syphilisnote.htm.

37. Zhang S, Freundschuh SM, Lenzer K, Zandbergen PA. The location swapping method for geomasking. Cartography and Geographical Information Science. 2017 Jan 2; 44(1):22–34.

38. Boulos MN, Cai Q, Padget JA, Rushton G. Using software agents to preserve individual health data confidentiality in micro-scale geographical analyses. Journal of Biomedical Informatics. 2006 Apr 1; 39 (2):160–70. https://doi.org/10.1016/j.jbi.2005.06.003 PMID: 16098819

39. Bailey TC, Gatrell AC. Interactive spatial data analysis. Essex: Longman Scientific & Technical; 1995 Oct.

40. Swift Andrew, Liu Lin, and Uber James. "MAUP sensitivity analysis of ecological bias in health studies." *GeoJournal* 79 (2014): 137–153.

41. Ratcliffe JH, McCullagh MJ. Hotbeds of crime and the search for spatial accuracy. Journal of geographical systems. 1999 Dec 1; 1(4):385–98.

42. Openshaw S. The modifiable areal unit problem. Quantitative geography: A British view. 1981:60–9.

43. Zandbergen PA. Ensuring confidentiality of geocoded health data: assessing geographical masking strategies for individual-level data. Advances in medicine. 2014 Apr 29;2014.

44. Hampton KH, Fitch MK, Allshouse WB, Doherty IA, Gesink DC, Leone PA, Serre ML, Miller WC. Mapping health data: improved privacy protection with donut method geomasking. American journal of epidemiology. 2010 Nov 1; 172(9):1062–9. https://doi.org/10.1093/aje/kwq248 PMID: 20817785

45. Allshouse WB, Fitch MK, Hampton KH, Gesink DC, Doherty IA, Leone PA, Serre ML, Miller WC. 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

46. Kounadi O, Leitner M. Spatial information divergence: Using global and local indices to compare geographical masks applied to crime data. Transactions in GIS. 2015 Oct; 19(5):737–57.

47. Swanlund D, Schuurman N, Zandbergen P, Brussoni M. Street masking: a network-based geographical mask for easily protecting geoprivacy. International Journal of Health Geographics. 2020 Dec; 19:1–1.

48. 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

49. Choi KM, Serre ML, Christakos G. Efficient mapping of California mortality fields at different spatial scales. Journal of Exposure Science & Environmental Epidemiology. 2003 Mar; 13(2):120–33.

50. Goovaerts P, Jacquez GM, Greiling D. Exploring scale-dependent correlations between cancer mortality rates using factorial kriging and population-weighted semivariograms. Geographical Analysis. 2005 Apr; 37(2):152–82. https://doi.org/10.1111/j.1538-4632.2005.00634.x PMID: 16915345

51. Kennedy S. The small number problem and the accuracy of spatial databases. Accuracy of spatial databases. 1989 Dec 6:187–96.

52. 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

53. Fox L, Serre ML, Lippmann SJ, Rodríguez DA, Bangdiwala SI, Gutiérrez MI, Escobar G, Villaveces A. 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

54. Nazelle AD, Arunachalam S, Serre ML. 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

55. Population and Housing Unit Counts 2010 Census of Population and Housing [Internet]. 2012. Available from: https://www2.census.gov/library/publications/decennial/2010/cph-2/cph-2-35.pdf.

56. Skinner B, Brian P, Peeler D, Lewis D, Hawes P. _ [Internet]. [cited 2024 Feb 6]. Available from: https://www.ncdot.gov/projects/40-440-us-1-interchange/Documents/fs-1005a-2015-feasibility-study.pdf.

57. Fazio P. Satori Acquisition Resource Page [Internet]. Satori Acquisition | BCC Software. [cited 2022Mar30]. Available from: https://bccsoftware.com/satori/.

58. ESRI. ArcGIS 10. Redlands, CA: Environmental Systems Research Institute; 2006.

59. ESRI. Esri U.S. Street Database. Redlands, CA: Environmental Systems Research Institute; 2006.

60. Schumacher CM, Bernstein KT, Zenilman JM, Rompalo AM. Reassessing a large-scale syphilis epidemic using an estimated infection date. Sexually transmitted diseases. 2005 Nov 1; 32(11):659–64. https://doi.org/10.1097/01.olq.0000175400.40084.3e PMID: 16254539

61. He J, Christakos G, Wu J, Li M, Leng J. Spatiotemporal BME characterization and mapping of sea surface chlorophyll in Chesapeake Bay (USA) using auxiliary sea surface temperature data. Science of the Total Environment. 2021 Nov 10; 794:148670. https://doi.org/10.1016/j.scitotenv.2021.148670 PMID: 34225143

62. Serre ML, 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.

63. Christakos G, Bogaert P, Serre M. Temporal GIS: advanced functions for field-based applications. Springer Science & Business Media; 2002 Jan 11.

64. Akita Y, Carter G, Serre ML. Spatiotemporal nonattainment assessment of surface water tetrachloroethylene in New Jersey. Journal of Environmental Quality. 2007 Mar; 36(2):508–20. https://doi.org/10.2134/jeq2005.0426 PMID: 17332255

65. Allshouse WB, Pleil JD, Rappaport SM, Serre ML. Mass fraction spatiotemporal geostatistics and its application to map atmospheric polycyclic aromatic hydrocarbons after 9/11. Stochastic Environmental Research and Risk Assessment. 2009 Dec; 23(8):1213–23.

66. Serre ML, Christakos G, Lee SJ. Soft data space/time mapping of coarse particulate matter annual arithmetic average over the US. IngeoENV IV—Geostatistics for Environmental Applications 2004 (pp. 115–126). Springer, Dordrecht.

67. Law DC, Bernstein KT, Serre ML, Schumacher CM, Leone PA, Zenilman JM, Miller WC, Rompalo AM. Modeling a syphilis outbreak through space and time using the Bayesian maximum entropy approach. Annals of Epidemiology. 2006 Nov 1; 16(11):797–804.

68. Orton TG, Lark RM. The Bayesian maximum entropy method for lognormal variables. Stochastic Environmental Research and Risk Assessment. 2009 Mar; 23:319–28.

69. Lee SJ, Yeatts KB, Serre ML. 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

70. Mathworks.com. [cited 2022 Feb 2]. Available from: http://www.mathworks.com/.

71. Unc.edu. [cited 2022 Feb 2]. Available from: https://mserre.sph.unc.edu/BMElab_web/.

72. Hook EW. Is elimination of endemic syphilis transmission a realistic goal for the USA?. The Lancet. 1998 Jun 1; 351:S19–21. https://doi.org/10.1016/s0140-6736(98)90006-x PMID: 9652716

73. Gesink D, Salway T, Kimura L, Connell J, Widener M, Ferlatte O. The social geography of partner selection in toronto, canada: A qualitative description of "convection mixing". Archives of Sexual Behavior. 2020 Jul; 49:1839–51. https://doi.org/10.1007/s10508-019-01484-1 PMID: 31628629

74. Sexually Transmitted Disease Surveillance 2011 [Internet]. www.cdc.gov. 2012. Available fromhttps://www.cdc.gov/std/stats/archive/Surv2011.pdf.

75. Cope AB, Bernstein K, Matthias J, Rahman M, Diesel J, Pugsley RA, Schillinger JA, Ng RA, Sachdev D, Shaw R, Nguyen TQ. Unnamed partners from syphilis partner services interviews, 7 jurisdictions. Sexually transmitted diseases. 2020 Dec 1; 47(12):811–8. https://doi.org/10.1097/OLQ.0000000000001269 PMID: 32890335

76. Holbrook JM. Homelessness among substance-using minority men who have sex with men. InHealth Issues Confronting Minority Men Who Have Sex with Men 2008 (pp. 109–124). New York, NY: Springer New York.

77. Beymer MR, Weiss RE, Bolan RK, Rudy ET, Bourque LB, Rodriguez JP, Morisky DE. Sex on demand: geosocial networking phone apps and risk of sexually transmitted infections among a cross-sectional

sample of men who have sex with men in Los Angeles County. Sexually transmitted infections. 2014 Nov 1; 90(7):567–72. https://doi.org/10.1136/sextrans-2013-051494 PMID: 24926041

78. Grosskopf NA, LeVasseur MT, Glaser DB. Use of the internet and mobile-based "apps" for sex-seeking among men who have sex with men in New York City. American journal of men's health. 2014 Nov; 8 (6):510–20. https://doi.org/10.1177/1557988314527311 PMID: 24658284

79. White Hughto JM, Pachankis JE, Eldahan AI, Keene DE. "You can't just walk down the street and meet someone": The intersection of social–sexual networking technology, stigma, and health among gay and bisexual men in the small city. American journal of men's health. 2017 May; 11(3):726–36. https://doi.org/10.1177/1557988316679563 PMID: 27885147

80. Arnold T, Stopka TJ, Gomillia CE, Murphy M, Johnson K, Chan PA, Klasko-Foster L, Rogers B, Soler JH, Monger ML, Jacque E. Locating the risk: Using participatory mapping to contextualize perceived HIV risk across geography and social networks among men who have sex with men in the Deep South. The Journal of Sex Research. 2022 Sep 2; 59(7):931–8. https://doi.org/10.1080/00224499.2021.1906397 PMID: 33826434

81. Haandrikman K. Partner choice in Sweden: How distance still matters. Environment and Planning A: Economy and Space. 2019 Mar; 51(2):440–60.

82. Gesink D, Wang S, Guimond T, Kimura L, Connell J, Salway T, Gilbert M, Mishra S, Tan D, Burchell AN, Brennan DJ. Conceptualizing geosexual archetypes: mapping the sexual travels and egocentric sexual networks of gay and bisexual men in Toronto, Canada. Sexually transmitted diseases. 2018 Jun 1; 45(6):368–73. https://doi.org/10.1097/OLQ.0000000000000752 PMID: 29465690

83. Zenilman JM, Ellish N, Fresia A, Glass G. The geography of sexual partnerships in Baltimore: applications of core theory dynamics using a geographic information system. Sexually transmitted diseases. 1999 Feb 1; 26(2):75–81. https://doi.org/10.1097/00007435-199902000-00002 PMID: 10029979

84. Kerani RP, Handcock MS, Handsfield HH, Holmes KK. Comparative geographical concentrations of 4 sexually transmitted infections. American Journal of Public Health. 2005 Feb; 95(2):324–30.

85. Chesson HW, Sternberg M, Leichliter JS, Aral SO. The distribution of chlamydia, gonorrhoea and syphilis cases across states and counties in the USA, 2007. Sexually transmitted infections. 2010 Dec 1; 86 (Suppl 3):iii52–7.

86. Fox LC, Peters BG, Frake AN, Messina JP. A Bayesian Maximum Entropy Model for Predicting Tsetse Ecological Distributions. International Journal of Health Geographics. 2023 https://doi.org/10.1186/s12942-023-00349-0 PMID: 37974150

87. Xu Y, Serre ML, 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