Sexually Transmitted Disease Core Theory: Roles of Person, Place, and Time

Dionne C. Gesink*, Ashleigh B. Sullivan, William C. Miller, and Kyle T. Bernstein

* Correspondence to Dr. Dionne C. Gesink, Dalla Lana School of Public Health, University of Toronto, 155 College Street, 6th Floor, Toronto, Ontario, Canada M5T 3M7 (e-mail: dionne.gesink@utoronto.ca).

Initially submitted October 27, 2010; accepted for publication January 25, 2011.

The authors’ purpose was to expand sexually transmitted disease core theory by examining the roles of person, place, and time in differentiating geographic core areas from outbreak areas. The authors mapped yearly census-tract-level syphilis rates for San Francisco, California, based on new primary and secondary syphilis cases reported to the San Francisco City sexually transmitted disease surveillance program between January 1, 1985, and December 31, 2007. SaTScan software (Information Management Services, Inc., Silver Spring, Maryland) was used to identify geographic clusters of significantly elevated syphilis rates over space and time. The authors graphed epidemic curves for 1) core areas, 2) outbreak areas, 3) neither core nor outbreak areas, and 4) noncore areas, where noncore areas included outbreaks, and stratified these curves according to demographic characteristics. Five clusters of significantly elevated primary and secondary syphilis rates were identified. A 5-year threshold was useful for differentiating core clusters from outbreak clusters. Epidemic curves for core areas, outbreak areas, neither core nor outbreak areas, and noncore areas were perfectly synchronized in phase trends and wavelength over time, even when broken down by demographic characteristics. Between epidemics, the occurrence of syphilis affected all demographic groups equally. During an epidemic, a temporary disparity in syphilis occurrence arose and a homogeneous core group of cases could be defined.

disease outbreaks; epidemic model; geography; sexually transmitted diseases; spatial analysis; syphilis

Abbreviations: MSM, men who have sex with men; STD(s), sexually transmitted disease(s).

The “core group” has been a critical theoretical concept in sexually transmitted disease (STD) epidemiology since its initial description (1). However, the definition of the core group has varied (2). Common definitions have included persons with repeat infections (3), persons with a high number of sexual partners (5–7), and persons in high-risk occupations, such as commercial sex workers and truck drivers (6, 8–11). Although these definitions have varied, it is largely believed that the “core group” is disproportionately responsible for maintaining transmission in a population.

Shortly after the introduction of the core group concept, Rothenberg (12) recognized that persons with reported STDs were often concentrated geographically and added a spatial component to the definition of “core.” Although geographic concentrations of infection have been referred to as “core groups,” more appropriate terms include “geographic core,” “core areas,” or “risk space” (13). These terms distinguish the core group, based on groups of individuals with high-risk behavior, from the geographic areas where persons infected with STDs live or meet. Geographically defined core areas have been observed in several locations in the United States and Canada (2, 14–20). However, the social and epidemiologic mechanisms underlying the formation of geographically definable core areas are unclear. Core areas may reflect local sexual partner selection (20, 21), neighborhood-level sociocultural factors (15, 22–35), other spatially dependent co-occurring activities and behaviors, such as injection drug use (36), or a combination of these factors.

While significant attention has been given to defining core areas in space, little attention has been given to defining or characterizing core areas in both space and time. Time is an important aspect of the concept of core areas because it effectively differentiates a true geographic core cluster, persistent in time, from an outbreak cluster of limited duration. Changes in the size and location of the core over time can be
used to measure the stability of the core area. Knowing whether a geographic cluster of infection is a core area or an outbreak, and whether it is stable in space and time, directly affects public health decisions pertaining to intervention, prevention, and response strategies (37, 38).

Our purpose in this investigation was to expand core theory by examining the roles of person, place, and time in relation to the spatial distribution of syphilis. Syphilis provides a good model for investigating core theory because sexual contact tracing and, in some cases, social contact tracing enhances the completeness of surveillance data. Consequently, the observed spatiotemporal distribution of infection will be based on reasonably complete information. We investigated the concept of geographic core areas using syphilis data from San Francisco, California, which experienced 2 separate syphilis epidemics between 1985 and 2007. Here, we use the term “epidemic” to refer primarily to temporal periods in which observed rates of syphilis infection exceeded what was expected and remained elevated for a limited period of time. With this definition, space is implicit and refers to the entire city of San Francisco. The first San Francisco syphilis epidemic (1985–1991) occurred predominantly among young, black, heterosexual crack-cocaine users. The second epidemic (2001 to the present and ongoing) occurred predominantly among older white and Hispanic men who have sex with men (MSM).

Specifically, we addressed whether a distinct core area or areas could be identified in San Francisco that could be differentiated from the known outbreaks, both geographically and temporally. We sought to determine whether such a core area would be spatiotemporally stable. Finally, we sought to assess whether the characteristics of the persons with syphilis in the core area(s) were consistent over time and whether the core area(s) contributed to the emergence of the outbreaks. The answers to these questions should provide important insights into the epidemiology of syphilis and the relation of core areas, core groups, and outbreaks.

MATERIALS AND METHODS

Under California state public health law, all reactive syphilis serologic results are reported to the local health jurisdiction from laboratories, along with provider reports of syphilis. Syphilis case investigations are conducted by local health department staff to ensure treatment and partner services. Syphilis events that occur more than 30 days apart are considered separate morbid events; thus, repeat infections were included in our analysis, since we were calculating incident rates of infection. The 30-day period is standard and is used to remove likely duplicate reports. Syphilis morbidity from persons residing outside of the geographic boundaries of the City and County of San Francisco are not included in local morbidity statistics. We mapped yearly, census-tract-level syphilis incidence rates for San Francisco based on new primary and secondary syphilis cases reported to the San Francisco City STD surveillance program between January 1, 1985, and December 31, 2007.

Syphilis incidence rates were estimated using US Census Bureau census-tract-level population estimates for 2007 and population totals for 1990 and 2000. We interpolated population estimates for 1991–1999 and 2001–2007 and then back-extrapolated population estimates for 1985–1989 based on the 1990s interpolation. Rates were calculated using the year 2000 census tract delineations. Some census tracts were divided into multiple smaller census tracts between the 1990 and 2000 delineations. Population estimates for census tracts that changed boundaries prior to 2000 were calculated by multiplying the percentage of the population in each census tract in 2000 by the population estimate prior to 2000. For example, if a census tract had 1,000 people in 1990 and was split into 2 census tracts in 2000, one with 1,600 people and one with 2,400 people, the population estimates for the census tracts in 1990, using year 2000 census tract delineations, would be 400 (1,600/(1,600 + 2,400) × 1,000) and 600 (2,400/(1,600 + 2,400) × 1,000), respectively.

Basic demographic information was available for reported syphilis cases at the individual level, including gender, race/ethnicity, and age group. Hence, population estimates stratified by the basic demographic characteristics of gender, race/ethnicity, and age group were constructed for San Francisco for the study time frame so demography-specific rates of infection could be estimated. Demographic data were used to 1) examine the evolution and transition of the 2 outbreaks that occurred during the study period, 2) compare characteristics of core and noncore cases, and 3) look for changes in the characteristics of core and noncore areas during different phases of the epidemic and determine whether these changes could be used to predict or signal changes in the epidemic profile.

SaTScan (Information Management Services, Inc., Silver Spring, Maryland) was used to identify geographic clusters of significantly elevated syphilis rates over space and time (39). All 23 years of data were used, with a maximum cluster population size of 20%. Significant clusters were examined for persistence based on duration. A priori, we set a minimum threshold of 5 years for a cluster to be defined as a core area (26) but anticipated that a true core area would persist for much longer. Clusters lasting for less than 5 years were defined as outbreaks.

We developed 2 different categorization systems to classify census tracts. First, we developed a “retrospective” 3-level classification based on 1) core areas, 2) outbreak areas, and 3) neither core nor outbreak areas. This classification had the benefit of hindsight and knowing when and where an outbreak would occur. We then developed a 2-level classification based on 1) core areas and 2) noncore areas, where noncore areas included outbreaks. This 2-level classification system was used to determine whether trends in the core and noncore areas could still be informative if it was not known when an outbreak was going to occur, much the way a local health department might use the terms and observations prospectively, since outbreaks are often identified after they have peaked or waned. For the purposes of this analysis, we use the term “core” to mean geographic or spatial core area.

Similar to Rothenberg (12), we graphed epidemic curves for each of the 2 classification systems outlined above (core/outbreak/Neither core nor outbreak areas and core/noncore areas). We also stratified epidemic curves by age (<35 years
and ≥35 years), race/ethnicity (black, white, or Hispanic), and gender/sexuality (female/male, MSM). Epidemic curves were qualitatively assessed by visually comparing phase trends, wavelengths, and amplitudes of curve peaks and troughs over time. Pearson’s correlation coefficient was used to evaluate synchronicity of the epidemic curve trends (null hypothesis: no correlation between epidemic curves). By synchronicity, we mean whether or not the epidemic curves moved in the same direction at the same time. Epidemic curves were constructed, and Pearson’s correlation coefficients were calculated, in R (40).

The University of Toronto Research Ethics Board approved this study. In San Francisco, this study was considered exempt from human subjects considerations in accordance with the Code of Federal Regulations, Title 45, since these data were de-identified and were undergoing retrospective analysis.

RESULTS

The San Francisco City Health Department reported 4,104 primary and secondary syphilis cases between 1985 and 2007. Four percent (176 of 4,104) of the reported primary and secondary syphilis cases could not be geocoded to a location. Over the entire period, the incidence rate of primary and secondary syphilis ranged from 3.6 cases per 100,000 person-years to 44.9 cases per 100,000 person-years. The first epidemic, which occurred predominantly among young, black, heterosexual crack-cocaine users, lasted from 1985 through 1991. The period 1992–2000 was a stable period of low infection rates. The second epidemic, which occurred predominantly among older white and Hispanic MSM, started in 2001 and is ongoing as of 2010.

SaTScan analysis identified 5 clusters of significantly elevated primary and secondary syphilis rates for San Francisco (Figure 1). The risk of being infected with primary or secondary syphilis was 4.6 times higher for someone living in the most likely cluster (cluster 1) between 1985 and 2007 than for someone living outside of that area. This cluster persisted for the entire duration of the study period (23 years), even through the stable period of low infection during the 1990s, and thus was identified as a core area.

Although the following 3 lower-ranked significant clusters (clusters 2, 3, and 4) had higher relative risks than the most likely cluster (cluster 1), they persisted for less than 5 years each and thus were defined as outbreak clusters (Figure 1). These 3 clusters were associated with the first syphilis epidemic (1985–1991). The fifth, least likely, significant cluster (cluster 5) had the lowest relative risk and was associated with the second syphilis epidemic
For the purposes of discussion, we identified this cluster as an outbreak area. However, it is unclear whether this area was an outbreak or more of an expansion/extension of the original core resulting from spatial overflow of cases to adjacent areas.

Demographic characteristics of primary and secondary syphilis cases from core areas, noncore areas (including outbreaks), and neither core nor outbreak areas suggested that the profiles of these cases were reasonably similar (Table 1). However, the characteristics of cases from outbreak areas were quite different. At the community level (based on census data from the year 2000), core areas had a greater white population, more households headed by a single male, and a higher proportion of rented houses compared with outbreak, neither core nor outbreak, and noncore areas (Table 1).

Using the 3-level definition of areas (i.e., core, outbreak, and neither), visual comparison of the overall and phase trends of the epidemic curves showed that all groups tracked in near-perfect synchrony, with the core epidemic curve remaining elevated over the curve for areas that were neither core nor outbreak areas. Outbreak cases also appeared and peaked when there were peaks in the core and neither core nor outbreak curves (Figure 2A).

When the 2-level definition of areas (i.e., core and noncore, including outbreaks) was used, the core and noncore epidemic curves still tracked in near-perfect synchrony (Figure 2B; Pearson’s $R^2 = 0.92$). However, now the amplitude of the noncore curve surpassed the peak for the core curve during the first epidemic, and the core and noncore distributions did not appear to differ.

Core and noncore epidemic curves stratified by demographic characteristics were also synchronized and significantly correlated (Figure 3 and Figure 4; for more detail, see Web Figure 1 (http://aje.oxfordjournals.org/)). The demography-specific epidemic curves indicated that dominant demographic characteristics for a given epidemic (e.g., black race/ethnicity for the first epidemic) “went away” during the interepidemic period (Figures 3 and 4). This period of equality (no elevated characteristics) was followed by a sudden change in demographic characteristics of core and noncore cases with the emergence of the second epidemic. For example, the epidemic curve for men was high during the first epidemic among heterosexuals in the late 1980s, fell to near zero during the interepidemic quiet period of the 1990s, and then began to rise again with the emergence of the second epidemic among MSM in the early

| Table 1. Patient and Community Characteristics (%) of Incident Syphilis Cases Based on San Francisco, California, Sexually Transmitted Infection Surveillance Data From 1985–2007 and US Census Data From the Year 2000a |
|-----------------------------------------------|-------------------------------|-------------------------------|-------------------------------|-------------------------------|-------------------------------|-------------------------------|-------------------------------|-------------------------------|
| Characteristic                             | Primary and Secondary Syphilis Cases | Census Tract (Neighborhood) |
|                                            | Core Areas | Outbreak Areas | Neither Core nor Outbreak Areas | Noncore Areas (Including Outbreak Areas) | Core Areas | Outbreak Areas | Neither Core nor Outbreak Areas | Noncore Areas (Including Outbreak Areas) |
| Race/ethnicity                             |            |                |                                |                                   |            |                |                                |                                   |
| White                                      | 52         | 31.7           | 48.9                           | 43                                | 57.4       | 40.8           | 50                             | 47.8                              |
| Black                                      | 22.5       | 51.3           | 18                             | 29.4                              | 11.4       | 17.7           | 3.6                            | 6.9                               |
| Hispanic                                   | 18.1       | 9.6            | 22                             | 17.8                              | 17.3       | 9.8            | 14.5                           | 13.4                              |
| Age group, years                           |            |                |                                |                                   |            |                |                                |                                   |
| 15–24                                      | 10.7       | 21.5           | 14                             | 16.6                              | 11.4       | 11.7           | 11.5                           | 11.5                              |
| 25–34                                      | 37.4       | 33.8           | 36.6                           | 35.6                              | 28.9       | 19.4           | 22.7                           | 21.9                              |
| 35–49                                      | 44.7       | 36.6           | 39.8                           | 38.7                              | 25.5       | 24.2           | 24.3                           | 24.3                              |
| ≥50                                        | 7.2        | 8.1            | 9.5                            | 9.1                               | 25.8       | 28.5           | 29.6                           | 29.3                              |
| Gender                                     |            |                |                                |                                   |            |                |                                |                                   |
| Female                                     | 10.5       | 24.7           | 12.7                           | 16.8                              | 43.6       | 50.9           | 50.3                           | 50.5                              |
| Male                                       | 89.5       | 75.3           | 87.3                           | 83.2                              | 56.4       | 49.1           | 49.7                           | 49.5                              |
| Men who have sex with men                  | 49.9       | 33.5           | 45.8                           | 41.6                              |            |                |                                |                                   |
| Household characteristics                  |            |                |                                |                                   |            |                |                                |                                   |
| Single male living alone                   | 29.1       | 12.4           | 17.6                           | 16.5                              |            |                |                                |                                   |
| Single female living alone                 | 20.8       | 14.8           | 19.9                           | 18.8                              |            |                |                                |                                   |
| Female head of household with a child      | 2.9        | 6.3            | 2.8                            | 3.5                               |            |                |                                |                                   |
| Male head of household with a child        | 0.8        | 1.3            | 0.9                            | 1                                 |            |                |                                |                                   |
| Owner-occupied housing                     | 15.1       | 50.4           | 35.3                           | 38.6                              |            |                |                                |                                   |
| Renter-occupied housing                    | 80.1       | 46.2           | 59.3                           | 56.5                              |            |                |                                |                                   |
| Vacant housing                             | 4.8        | 3.3            | 5.3                            | 4.9                               |            |                |                                |                                   |

a Percentages in some sections do not total 100 because of the exclusion of some racial/ethnic and age groups.
2000s. Similarly, the epidemic curves for blacks and females were high during the first epidemic, returned to near zero during the interepidemic period, and remained low during the second epidemic among MSM.

**DISCUSSION**

Our investigation into the epidemiology (person, place, and time) of syphilis in San Francisco indicates that both space and time need to be used to define core areas and to differentiate core areas from outbreak areas. We identified 1 core area and several outbreak areas over the course of 2 very different syphilis epidemics that occurred in San Francisco between 1985 and 2007 (the second epidemic is ongoing). During this period, the spatiotemporal dynamics of the core area were distinctly different from those of the outbreak areas. That is, the core tended to cover a larger geographic area and persisted for a longer period of time (23 years) compared with the outbreaks (4 years). The risk of being a case was much higher for outbreak areas than for the core area (Figure 1), but this could have been due to the core estimate’s being diluted by the interepidemic period of low rates.

A priori, we used a threshold of 5 years to differentiate a core area from an outbreak area. This threshold was based on the identification of 2 core areas present before, during, and after a syphilis outbreak in Baltimore, Maryland, in the 1990s (26). In Baltimore, the core areas persisted for 8 years or longer. For San Francisco, 5 clusters were identified using SaTScan (Figure 1). Only cluster 1 lasted longer than 5 years, and the fact that it was present for the entire 23-year time frame of the study (and more likely longer) clearly identified it as a geographic core area. The remaining 4 clusters (clusters 2–5) lasted 1–4 years (Figure 1). These findings support the threshold of 5 years for differentiating core areas from outbreak areas for syphilis.

We found that epidemic curves for both our 2- and 3-level classification systems were perfectly synchronized in phase trends and wavelength of curve peaks and troughs over time, even when broken down by demographic characteristics (Figures 3 and 4). Like Rothenberg (12), we found attenuation of the epidemic pattern when core, outbreak, and neither core nor outbreak areas were compared (3-level classification; Figure 2A). However, when outbreaks were included in the definition of noncore areas (2-level classification: core/noncore), attenuation of the epidemic pattern from core to noncore was not observed (Figure 2B and Figure 4).

Looking at the epidemic curves for both total cases (Figure 2B) and cases broken down by demographic characteristics (Figure 4), we found that the noncore epidemic curve consistently rose before the core epidemic curve and consistently exceeded the core epidemic curve during the first epidemic (late 1980s), suggesting that this epidemic may have emerged from the noncore area rather than the core area. Conversely, comparing the epidemic curves during the 2000s, both curves rose simultaneously, with the core area curve remaining elevated over the noncore area curve. This suggests that the second epidemic may have been initiated within the core. The difference in the spatial origin of the 2 epidemics could explain why the second epidemic (potentially emerging from within the geographic core) is ongoing and has been so difficult to bring under control.
Detection of an outbreak initiated outside the geographic core area may be delayed initially, if the core area is monitored more intensely. However, these outbreaks may be more manageable, with a shorter duration, unless they are sustained by spread within the geographic core area. In contrast, outbreaks within the core may be difficult to detect, as the distinction between endemic core cases and outbreak cases within a core area may not be clear. A sustained rise in cases in the core areas may reflect a relatively mature outbreak, necessitating careful analysis of demographic factors and sexual networks to characterize the outbreak. Thus, the location where an epidemic is initiated or emerges may have important implications for public health response strategies. This further suggests that 1) both core and noncore areas...
should be monitored for STDs patterns, 2) comparison of core and noncore epidemic patterns can lend insights into where an epidemic is emerging (or has emerged), and 3) it may be more important to target responsive intervention resources toward the noncore areas, depending on where the epidemic is initiated (i.e., core or noncore area). Our data suggest that some efforts could be directed toward all affected areas and not just the geographic core area.

Research on the relative efficacy and cost-efficiency of public health interventions targeting geographic core areas as compared with all affected areas is warranted.

Geographically profiling the demographic characteristics of cases in core and noncore areas showed that between epidemics, the occurrence of syphilis was equally affecting all demographic groups (Figures 3 and 4). It was only during an epidemic that a temporary disparity in syphilis

**Figure 4.** Syphilis epidemic curves according to demographic characteristics for core areas (cluster 1 (–)) and noncore areas (census tracts that were never part of a significant cluster during the study period (---)), San Francisco, California, 1985–2007. MSM, men who have sex with men.
occurrence arose and a homogeneous core group of cases could be defined.

Underreporting is always a concern in epidemiologic studies of STDs. Not only did our study rely on syphilis events reported to the health department, but these events had to have a residential address that could be geocoded. Syphilis patients were interviewed face-to-face, so case addresses were considered reliable. San Francisco does not have the same rural route or post office box issues that many suburban or rural communities have, thus minimizing the potential for ungeocodable addresses. Only a small percentage of cases failed to match to a location (4%), and the primary reasons for geocoding failure were address errors and homelessness. If all of the cases that could not be geocoded came from the same area, and that area was outside of the core area, we could have missed an important cluster of infection. However, the active contact tracing for syphilis should have minimized the opportunity for missed cases, and it is unlikely that a missing cluster would have changed the interpretation of the clusters and relations between core, outbreak, and noncore areas.

Another potential limitation of this analysis is the lack of more current census-level population estimates for San Francisco after 2000. Given that local population estimates for San Francisco from the 2010 Census will not be available for some time, we chose to include more recent syphilis data and the estimates derived from the 2000 Census. It is likely that the characteristics of the San Francisco population have changed and that for the later years of our analysis (2005, 2006, and 2007), the estimates from the 2000 Census may be less valid. However, we believe these limitations had a minimal impact on our analysis, since the areas in San Francisco that have seen the largest increases in population are minimally affected by syphilis.

In terms of core theory, our findings suggest the following.

1. Core groups cluster in space, creating core areas of infection that persist over time. As a corollary, core areas comprise core group members who appear to initiate epidemics and drive the spread of infection.
2. Time is a necessary factor for defining core areas and differentiating core areas from outbreak areas. Five years is a reasonable threshold for differentiating core areas from outbreak areas for syphilis.
3. The incidence rate of syphilis may vary substantially over time within a core area and may increase in parallel with an outbreak in adjacent areas.
4. Characteristics of cases in the core reflect the characteristics of cases outside of the core in the general population but do not necessarily reflect the characteristics of cases involved in outbreaks.
5. A demographic shift in the core group may occur during an interepidemic period. We observed a shift from a similar risk for members of the general population to increased risk for a specific population subgroup. Furthermore, disparities in core group demographic characteristics may be temporary.

We found evidence stressing the importance of monitoring both the spatial distribution of and epidemic curves for core and noncore areas, retrospectively and prospectively. Shifts in both epidemic curves and demographic characteristics for core and noncore areas may be useful for identifying where an epidemic has emerged or could emerge (core or noncore) and what group may be most affected, both of which give cause to consider different intervention strategies for epidemic control.

ACKNOWLEDGMENTS

Author affiliations: Dalla Lana School of Public Health, University of Toronto, Toronto, Ontario, Canada (Dionne C. Gesink, Ashleigh B. Sullivan); Department of Epidemiology, School of Public Health, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina (William C. Miller); and Epidemiology, Research and Surveillance Branch, STD Prevention and Control Services Section, San Francisco Department of Public Health, San Francisco, California (Kyle T. Bernstein).

This work was supported by the National Institute of Allergy and Infectious Diseases (grant R01 AI067913). Ashleigh Sullivan was also supported by a Public Health Professional Masters Award from the Canadian Institute for Health Research.

The authors thank the San Francisco Department of Public Health.

Conflict of interest: none declared.

REFERENCES


