Geographic Epidemiology of Gonorrhea in Baltimore, Maryland, Using a Geographic Information System

Karen M. Becker,1,2 Gregory E. Glass,2 Wayne Brathwaite,3 and Jonathan M. Zenilman 1,2

The epidemiology of gonorrhea is characterized by geographically defined hyperendemic areas, or "cores." Geographic information system (GIS) technology offers new opportunities to evaluate these patterns. The authors developed a GIS system linked to the disease surveillance database at the Baltimore Health Department and used this system to evaluate the geographic epidemiology of gonorrhea in Baltimore, Maryland, during 1994. There were 7,330 reported cases, of which 87.4% were in persons aged 15-39 years; 56.6% of the cases were in males; and 60.5% of the cases were reported from the nonsexually transmitted disease (STD) clinic sector. Valid residential addresses were available for 6,831 (93.5%) of cases. In the GIS system, gonorrhea cases were geocoded by reported address using digitized maps, and assigned to census tract. Census tract-specific rates for persons aged 15-39 years were calculated using 1990 census data. Gonorrhea was reported from 196/202 (97%) of census tracts, of which 90 census tracts had >30 cases. For these 90 census tracts, rates were ranked. The core was considered as the top rate quartile, consisting of 13 geographically contiguous census tracts with rates 4,370-6,370 per 100,000; adjacent areas were 19 census tracts in the second quartile (rates: 3,730-4,370 per 100,000). As radial distance from the core areas increased, incidence rates decreased and male/female ratio increased, which is consistent with previous definitions of the core theory of STD transmission. Mapping of cases by provider showed that cases reported from STD clinics had similar geographic distribution to those from the non-STD clinic sector. From an operational perspective, GIS can be effectively integrated with clinical data systems to provide epidemiologic analysis. Am J Epidemiol 1998;147:709-16.

In the developed world, gonorrhea incidence typically has a sociogeographic distribution of hyperendemic areas, or "core" areas (1-12). Core areas are defined geographically and are also characterized by poverty and poor health care access. A central tenet of the core theory is that the transmission of gonorrhea is not homogeneous (3), and that most cases occur in a small proportion of the population.

The epidemic reproductive rate ($R_0$) model of microparasites has been applied to sexually transmitted disease (STD) control (2, 3, 13, 14) as $R_0 = \beta cd$, where $\beta$ is disease transmission efficiency between sexual partners, $c$ is sexual partner turnover, and $d$ is the duration of infectivity. Within the core areas $R_0 > 1$, which could function as a reservoir for further disease spread to surrounding communities areas where $R_0 < 1$ (2, 15). From a disease control perspective, an intervention to reduce incidence within the core area should therefore impact on the community-wide disease incidence.

In the early 1980s, Rothenberg (6) empirically applied the core area concept models to health department STD program data from New York State. He found that gonorrhea incidence was inversely proportional to the distance from the core. These concepts were confirmed by investigations in the United States in Miami, Florida (7), Seattle, Washington (5), rural North Carolina (8), Colorado Springs, Colorado (9), and San Francisco, California (10), and in the United Kingdom in Leeds (11) and London (12). Studies in Colorado Springs (9, 16) showed that the duration of disease infectivity was longer in patients who resided within the core area, and that core area residents were 4.5 times more likely than people who lived outside of
tected directly to the Baltimore City Health Department and the Maryland State Department of Health and Mental Hygiene (figure 1). Cases in Baltimore, Maryland, have historically had high gonorrhea rates. In 1994, the crude citywide rate was 1,234/100,000, which was third highest in the nation (17). This paper describes a system of data management and analysis developed at the Baltimore City Health Department that provides the capability for real-time morbidity and geographic data analysis, and could form a basis to evaluate measures for STD interventions. We used the 1994 data set to develop the prototype for these studies, because it was in that year that the STD database was fully computerized. Compared with other STDs, gonorrhea was selected because the geography of gonorrhea has been most widely studied, incidence was high, and the disease is easily and reliably diagnosed. In addition, because of the disease’s short incubation period and absence of acquired immunity, gonorrhea incidence responds rapidly to epidemiologic changes, and does not exhibit the wide oscillations characteristic of some other sexually transmitted diseases such as syphilis.

MATERIALS AND METHODS

Cases were defined on the basis of isolation of *Neisseria gonorrhoeae* from a urethral, cervical, pharyngeal, or rectal clinical specimen, or a positive result of a licensed gonococcal nucleic acid probe test from a cervical or urethral specimen (18). Cases of gonorrhea were treated according to the 1993 Centers for Disease Control (CDC) “STD Treatment Guidelines” (19). Any patient who returned ≥ 30 days after treatment with a diagnosed case of gonorrhea was assumed to have acquired a new infection (20).

Reported cases of gonorrhea were obtained from STD clinic and non-STD clinic sources, following the CDC reporting scheme. The Baltimore City Health Department operates two public STD clinics which are located in high-incidence areas and serve a predominantly minority, indigent population. Data from all other sources are classified as non-STD clinic, and includes private physicians, community health centers, and emergency room settings. Integration of data from these sources relied on an automated Central Disease Registry, maintained by the Baltimore City Health Department and the Maryland State Department of Health and Mental Hygiene (figure 1).

At every patient clinical encounter, clinical data, medical records, laboratory specimens, and test results are linked using a relational database. Cases in Baltimore City residents from non-STD clinic providers are reported directly to the Baltimore City Health Department where they are entered into the morbidity system. The Maryland State Health Department provides reports of cases diagnosed in city residents seen by providers outside the city limits.

The database generated from the public and non-STD clinic providers was sorted and checked for duplicate records. Multiple cultures taken from different anatomic sites of any patient during one visit were deleted. In cases where multiple visits of a patient occurred within one month, medical records were evaluated to investigate why patients positive for gonorrhea returned. If patients changed providers before a case was treated, records were examined to determine the cause of clinic shift. Regardless of the reason, cases were regarded as non-STD clinic or public depending on the origin of the first reported source of information.

The reported residential street address for each case in the Central Disease Registry was used to characterize the geographic distribution of gonorrhea within Baltimore City. We constructed a geographic information system (GIS) which included data layers for street addresses, zip codes, 1990 census tracts, and demographic data linked to each census tract. The GIS software that we used was Mapinfo (Mapinfo Corporation, Troy, New York). The digital base maps used for constructing the GIS were commercially procured. These base maps were modified from the Maryland Department of Transportation road maps, which allowed more accurate spatial locations of streets, and linkages with the other data layers.

Addresses in the Central Disease Registry were geocoded to the digital base maps of roads and census tracts in Baltimore City (Baltimore Regional Commission, Baltimore, Maryland) using the Mapinfo software. The census tract for each case was then obtained by overlaying the database of residential addresses with the census tract database. The software automatically links addresses from the clinic database to census tract if there is an exact address match. Addresses that could not be automatically geocoded were matched interactively, using zip code as a secondary search criterion to reduce potential assignment errors. Cases that could not be geocoded but that listed the city as their addresses were assumed to have lived somewhere in the city and retained for comparative demographic analyses but were excluded from spatial comparisons.

After cases were assigned to appropriate census tracts, case counts per census tract were enumerated, and rates per census tract were calculated by importing the 1990 census data (21) into the software spreadsheet component. We then created thematic maps of annualized disease rates by census tract. We defined
FIGURE 1. Information flow of the data obtained from providers to the Central Disease Registry, Baltimore City, Maryland. Information from non-sexually transmitted disease (STD) clinic providers and Baltimore City Health Department (BCHD) STD clinics are merged within the Central Disease Registry. The patient data were geocoded to reported residential address, assigned to census tracts, and rates calculated.
"core," "adjacent," and "peripheral" areas of gonorrhea incidence using the modifications of definitions previously used by Rothenberg (6) and Thomas (8). In their original analysis of upstate New York gonorrhea, a decision rule was used that included only census tracts that had ≥10 cases. Because the Baltimore city-wide rate (1,234/100,000) was approximately three times the rate in the urban New York State counties (365/100,000, from table 1 in Rothenberg (6)), we adopted a decision rule to include only the 90 census tracts that had >30 cases.

We considered a definition of core based on determination of rates. Census tract-specific rates were calculated for gonorrhea in the 15-39 years age group for the 90 census tracts with >30 cases, ranked by rates (range, 1,140–6,370 per 100,000), and divided into quartiles. We defined the core as the 13 census tracts with rates in the upper quartile (rates: 4,370–6,370 per 100,000); the adjacent areas were defined as the 19 census tracts with rates in the second quartile (rates: 3,730–4,370 per 100,000), and the peripheral census tracts were the 58 census tracts with rates in the lower two quartiles (rates: 1,140–3,730 per 100,000).

We compared the geographic characteristics of patients who used non-STD clinic providers and public clinics. For each case with a valid residential address, we calculated the distance from the geographic centroid (NS,EW) of Baltimore City, expressed as a north-south and east-west vector. The mean distance was calculated as

\[
\sqrt{\sum (\text{NS}_i - \text{NS}_0)^2 + (\text{EW}_i - \text{EW}_0)^2}/n,
\]

where NS and EW are the north-south and east-west vectors calculated in meters from the geographic centroid (NS,EW), the square root is the linear distance calculated by the Pythagorean theorem, and n is the number of cases used in the calculation.

RESULTS

Descriptive epidemiology of gonorrhea morbidity

In 1994, 7,330 cases of gonorrhea were reported in Baltimore City residents (table 1). Of these, 2,895 (39.5 percent) were reported from the public STD clinics. One hundred and ten patients (1.5 percent) were first seen and tested by non-STD clinic providers, but also sought treatment at the STD clinics. Men with gonorrhea comprised 56 percent of cases and were significantly \((p < 0.001)\) older (mean age = 27.7 years) than women (mean age = 22.5 years). In the STD clinics, 2,895 cases (82.0 percent) were in males, whereas for cases diagnosed by the non-STD clinic providers, 4,416 cases (59.9 percent) were in females.

<table>
<thead>
<tr>
<th>Variable</th>
<th>STD* clinic</th>
<th>Non-STD clinic providers</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Sex</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>522</td>
<td>2,643</td>
<td>3,165</td>
</tr>
<tr>
<td>Male</td>
<td>2,373</td>
<td>1,773</td>
<td>4,146</td>
</tr>
<tr>
<td>Not reported</td>
<td>0</td>
<td>19</td>
<td>19</td>
</tr>
<tr>
<td><strong>Race</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Black</td>
<td>2,855</td>
<td>3,174</td>
<td>6,029</td>
</tr>
<tr>
<td>White</td>
<td>30</td>
<td>152</td>
<td>182</td>
</tr>
<tr>
<td>Other</td>
<td>10</td>
<td>15</td>
<td>25</td>
</tr>
<tr>
<td>Not available</td>
<td>0</td>
<td>1,094</td>
<td>1,094</td>
</tr>
<tr>
<td>Total</td>
<td>2,895</td>
<td>4,435</td>
<td>7,330</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Age (years)</th>
<th>Males</th>
<th>Females</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;15</td>
<td>88</td>
<td>215</td>
<td>303</td>
</tr>
<tr>
<td>15–39</td>
<td>3,534</td>
<td>2,866</td>
<td>6,410</td>
</tr>
<tr>
<td>&gt;39</td>
<td>524</td>
<td>82</td>
<td>606</td>
</tr>
<tr>
<td>Total</td>
<td>4,146</td>
<td>3,165‡</td>
<td>7,330§</td>
</tr>
</tbody>
</table>

* STD, sexually transmitted disease.
† Includes 10 individuals without information on sex.
‡ Includes two females without age information.
§ Includes 11 individuals with missing information for sex and/or age.

Of the 7,330 reported cases, 499 (6.5 percent) could not be located to residential address. Most \((n = 449)\) of these were cases from non-STD clinic providers, representing 10.1 percent of their records. Among STD clinic patients, 50 (1.7 percent) could not be mapped. Causes for failure to map addresses were: address information missing or unknown \((n = 306)\), invalid addresses or post office box numbers \((n = 185)\), or the patient was listed as homeless \((n = 6)\). Cases of gonorrhea were reported from 196 (97.0 percent) of the 202 census tracts (figure 2). The geographic distribution of the 6,831 cases with valid residential addresses showed that cases of gonorrhea tended to cluster along two principle axes, one running northwest out of the city and the other running essentially north-south along the eastern portion of the city (figure 3).

A total of 6,410 cases (87.4 percent) were diagnosed in individuals aged 15–39 years. Ethnicity was not reported for 1,094 (14.9 percent) of cases, all from the non-STD clinic sector. When ethnicity was reported, 6,029/6,236 (96.7 percent) were blacks. The citywide rates for blacks aged 15–39 years were 3,470/100,000, which was significantly \((p < 0.001)\) older (mean age = 27.7 years) than those for other ethnic groups (mean age = 22.5 years). In the STD clinics, 2,895 cases (82.0 percent) were in males, whereas for cases diagnosed by the non-STD clinic providers, 4,416 cases (59.9 percent) were in females.

For the analyses described here, we evaluated gonorrhea incidence in the population aged 15–39 years because this group comprised 87.4 percent of all reported gonorrhea, and was therefore more representative of the population-at-risk. Valid residential
addresses were available for 5,980 (92.3 percent) of these cases.

We initially considered defining the core based on ranking the census tract case counts as in Rothenberg (6). However, case counts did not accurately reflect the underlying rates, because the distribution of the population-at-risk, i.e., individuals aged 15–39 years, is not homogeneous. For example, two census tracts in northwestern Baltimore had high case counts, but also had large populations (figures 2 and 4), and therefore had disease rates (figure 5) which were substantially lower than many other areas.

Ninety census tracts had >30 cases and were included in the evaluation of the core areas. Rates were calculated for the 5,690/6,410 (93.2 percent) cases aged 15–39 years with valid residential addresses. Using our rate-based definitions described earlier (table 2), we defined 13 census tracts as core located in geographically contiguous areas in East and West Baltimore. These census tracts accounted for 15.5 percent of the morbidity in 6.5 percent of the total population. Nineteen census tracts were defined as adjacent, these contained 17.1 percent of the disease in 9.6 percent of the population. Fifty-eight census tracts were peripheral, containing 38 percent of the cases in 35 percent of the population. Adjacent areas tended to cluster around the edges of the core tracts with the exception of a region in the northwest portion of the city which was separated from the western adjacent area by an intervening city park (figure 5). Demographic characteristics of individuals with disease were similar in core, adjacent, and peripheral areas.

We then compared the geographic (spatial) distributions of residential addresses of patients from the two provider groups and found no significant difference in the spatial distributions of patients from BCHD STD clinic and non-STD clinic providers. The mean distances from the Baltimore City centroid differed by 299 m. The cumulative distributions of cases from the geographic centroid were not statistically different ($T_1 = 0.08; p > 0.20$). The maximum difference in the proportion of patients occurred at 4.5 km with 67.5 percent of BCHD STD clinic and 59.2 percent of non-STD clinic patients living within this distance.
DISCUSSION

Routinely collected STD program data coupled with an easily adaptable GIS system can define the geographic patterns of gonorrhea, including the delineation of core groups. Our results confirm the presence of geographically defined "core areas" for gonorrhea in Baltimore, Maryland. We also have demonstrated the operational utility of integrating a local health department disease management database with a GIS system that can facilitate sophisticated epidemiologic analysis and disease surveillance.

The gonorrhea core distribution observed in Baltimore is similar to that seen in other areas (5–12, 15). The large core in Baltimore may explain the relative stability of local gonorrhea rates. For example, the national crude national gonorrhea rates have decreased over 50 percent since 1975 (17). However, rates in subgroups, such as adolescents, have either not decreased, or have decreased minimally. Similar patterns have been observed in Baltimore, where the crude gonorrhea rate decreased 11.5 percent from 1988 to 1994 (from 1,382 to 1,234 per 100,000), whereas in the most severely impacted age groups, women aged 15–19 years and men aged 20–24 years, such decreases in rates were not observed (9, 17). In 1994, the rates in persons aged 15–39 years within the defined core areas were 6,821/100,000 for men and 4,343/100,000 for women, and were more than two times higher only when adolescent women (ages 15–19 years) and young men (ages 20–24 years) were considered. In four census tracts, the rates for black men aged 15–39 years were >8,000/100,000. Citywide, the rate for blacks aged 15–39 was 3,482 for men and 2,370 for women. Rates of this magnitude are more typical of resource-poor, developing countries (22), and demonstrate a critical need for public health commitment to develop and implement effective screening, diagnostic, and treatment services (23).

Males who had gonorrhea were significantly older than females who had the disease, and were more...
likely to get their care from the public STD clinics. This suggests that young women have increased access to reproductive health services, such as family planning clinics and adolescent clinics that are operated by the health department and a variety of private providers.

In contrast to previous studies, we used rates to define the core. In large urban areas, often with diverse population, we believe that this is a better representation of disease density. For example, in three northwest Baltimore census tracts, because of a large population denominator, several areas with high case counts actually had rates substantially lower than the inner-city areas.

Nevertheless, the core, peripheral, and adjacent definitions are empirical constructs, and the graphical appearances of these analyses can be manipulated by changing the definitions of the strata so the core could be made to appear to be smaller or larger. As these analyses are more widely used, there is a critical need to develop standardized procedures for descriptive analysis of geographic epidemiologic data, which would be applicable to individual localities. One approach is to use the rate ratio:

\[
\text{rate}_{(\text{census tract})}/\text{rate}_{(\text{total } = \text{ citywide})}
\]

which would, by definition, be proportional to the rate quartile approach that we used in this analysis. Rate ratios may provide a standardized method for determining the core, without needing to rely on stratification procedures or case counts. For example, the range of rate ratios was as follows in Baltimore City: core areas, 2.19–3.19; adjacent areas, 1.88–2.17; and peripheral areas, 0.66–1.85.

Our findings may be limited by the potential for reporting bias, especially from the non-STD clinic sector. Historically, underreporting of gonorrhea has been estimated to be 0–4 cases for each reported case (24–26), although recent studies suggest that reporting from private providers has improved (27). Underreporting from outside the core areas could artificially increase the rate gradient from the core to peripheral areas. Nevertheless, because the rates within the core are so high, the pattern would still be valid even if there was substantial underreporting. Our group is also currently performing a population-based STD surveillance assessment that should provide further insight into the reporting biases.

In STD morbidity data, the public sector is defined as the cases diagnosed in the health department clinics. The non-STD clinic health sector is defined by default as every other health care provider, including private physicians, hospital outpatient departments, emergency clinics, and family planning clinics. Cases with missing geographic data were predominantly from the non-STD clinic sector; this could in turn insert additional bias into our data. We also found no geographic difference in the distribution of public and non-public cases (figure 3), which suggests that bias due to provider source is not an issue.

We believe the geographically defined core distribution of disease has significant applications for disease control. Geographic analysis alone does not provide direct behavioral data—which is also a critical component of understanding STD epidemiology. More information on sexual partners and other factors that define the socioeconomic network are needed to explain the observed spatial and demographic differences. Because of the geographic disease distribution, partners recruited from within close geographic proximity are also more likely to have gonococcal infection. This in turn increases the risk of disease transmission.

Formal, controlled trials of targeted screening and interventions such as contact tracing approaches to gonorrhea control and overall impact on incidence still need to be conducted. If the core theory were operative, focused intervention, diagnosis, and treatment
efforts within the core would have a substantial impact on community-wide disease incidence. GIS is a useful and easily adaptable tool to define areas of highest incidence and people with the highest risks, and where resources may be targeted where needed most.

ACKNOWLEDGMENTS

This work was supported by grant no. 1U19 AI38533 from the National Institutes for Allergies and Infectious Diseases.

The authors thank Jim Kus and Jeremy Lifschutz for providing helpful comments and review.

REFERENCES


