## **Original Paper**

## Using a Geolocation Social Networking Application to Calculate the Population Density of Sex-Seeking Gay Men for Research and Prevention Services

Kevin P Delaney<sup>1</sup>, MPH; Michael R Kramer<sup>1</sup>, PhD; Lance A Waller<sup>2</sup>, PhD; W Dana Flanders<sup>1</sup>, MD, PhD; Patrick S Sullivan<sup>1</sup>, DVM, PhD

<sup>1</sup>Department of Epidemiology, Laney Graduate School, Emory University, Atlanta, GA, United States

<sup>2</sup>Department of Biostatistics and Bioinformatics, Rollins School of Public Health, Emory University, Atlanta, GA, United States

#### **Corresponding Author:**

Kevin P Delaney, MPH Department of Epidemiology Laney Graduate School Emory University Mailstop: 1518-002-4AA (SPH: Epidemiology) 1518 Clifton Road NE Atlanta, GA, 30322 United States Phone: 1 404 727 8710 Fax: 1 404 727 8737 Email: kpdelan@emory.edu

## Abstract

**Background:** In the United States, human immunodeficiency virus/acquired immunodeficiency syndrome (HIV/AIDS) continues to have a heavy impact on men who have sex with men (MSM). Among MSM, black men under the age of 30 are at the most risk for being diagnosed with HIV. The US National HIV/AIDS strategy recommends intensifying efforts in communities that are most heavily impacted; to do so requires new methods for identifying and targeting prevention resources to young MSM, especially young MSM of color.

**Objective:** We piloted a methodology for using the geolocation features of social and sexual networking applications as a novel approach to calculating the local population density of sex-seeking MSM and to use self-reported age and race from profile postings to highlight areas with a high density of minority and young minority MSM in Atlanta, Georgia.

**Methods:** We collected data from a geographically systematic sample of points in Atlanta. We used a sexual network mobile phone app and collected application profile data, including age, race, and distance from each point, for either the 50 closest users or for all users within a 2-mile radius of sampled points. From these data, we developed estimates of the spatial density of application users in the entire city, stratified by race. We then compared the ratios and differences between the spatial densities of black and white users and developed an indicator of areas with the highest density of users of each race.

**Results:** We collected data from 2666 profiles at 79 sampled points covering 883 square miles; overlapping circles of data included the entire 132.4 square miles in Atlanta. Of the 2666 men whose profiles were observed, 1563 (58.63%) were white, 810 (30.38%) were black, 146 (5.48%) were another race, and 147 (5.51%) did not report a race in their profile. The mean age was 31.5 years, with 591 (22.17%) between the ages of 18-25, and 496 (18.60%) between the ages of 26-30. The mean spatial density of observed profiles was 33 per square mile, but the distribution of profiles observed across the 79 sampled points was highly skewed (median 17, range 1-208). Ratio, difference, and distribution outlier measures all provided similar information, highlighting areas with higher densities of minority and young minority MSM.

**Conclusions:** Using a limited number of sampled points, we developed a geospatial density map of MSM using a social-networking sex-seeking app. This approach provides a simple method to describe the density of specific MSM subpopulations (users of a particular app) for future HIV behavioral surveillance and allow targeting of prevention resources such as HIV testing to populations and areas of highest need.

## KEYWORDS

Internet; HIV; MSM; sampling, location services

## Introduction

In the United States, human immunodeficiency virus (HIV) continues to have a heavy impact on men who report having sex with men (MSM) [1,2]. Although HIV incidence is increasing among MSM overall, there are pronounced disparities in both prevalence and incidence in the United States within the MSM HIV epidemic by race/ethnicity. A Centers for Disease Control and Prevention surveillance study conducted in 2008 [3] found that black non-Hispanic MSM were significantly more likely to be living with HIV than were white non-Hispanic MSM (28% vs 18%), and among those living with HIV, blacks were also significantly more likely to be unaware of their HIV infection (59% vs 26%). The disparity in HIV prevalence is consistent with a marked difference in estimated incidence of new infections for young minority MSM. From 2006-2009, black MSM under age 30 experienced a 47% increase in the estimated annual number of new infections and in 2009, there were more new infections in black MSM under age 30 than in white MSM under age 39 and more than all Hispanic MSM [4].

As a result, there is renewed emphasis [5] on identifying reasons for these disparities [6,7] and developing and providing interventions specifically for young minority MSM. However, the number of HIV prevention interventions implemented and evaluated with young minority MSM remains relatively low [8,9]. One reason for the lack of interventions specifically targeted to black MSM may be difficulty identifying a sampling frame for this population [6,8]. Stigma experienced by black MSM [10-12] may pose particular challenges in enumerating and accessing these men for provision of services [11]. A variety of sampling methods have been developed to access hidden or marginalized populations [13-18], with varying degrees of success [17-24].

Social networking websites and apps represent novel means for individual communication. A variety of new social networking tools designed for MSM are now available for most smartphones [25-28], and combined, these apps have more than 6 million users and 10,000 new users added daily. Many of these apps build their services on the ability to use the geolocation features

available on most phones and other communication devices (iPods, iPads, and tablets) to provide location information for other app users, including their geographic proximity (in feet or miles) to the user's location. In this paper, we describe a methodology for using the geolocation features of one of these apps as a novel approach to calculating the population density of men using the app at given times and describe how to use this density measure to highlight areas with a high-density of minority and young minority MSM.

## Methods

#### Overview

To pilot the study methodology, we chose a sexual networking app and collected data from publicly available profiles at sampled locations around the city of Atlanta, Georgia. App profiles (see Figure 1 for a hypothetical example showing the types of data typically included in user profiles) include information on the linear distance from the user to each other member, in feet for distances less than one mile, and miles for larger distances. For example, the person whose profile is represented in the first panel in Figure 1 was 1258 feet from our sampling location when the profile was viewed. Although we piloted this approach with several of the available apps [25-28], data generated for this paper were from a single app whose name is not revealed at the request of the developer. App profiles indicate the distance but not the direction of the person in question. In order to develop measures of density of users, we began by establishing a grid over Atlanta and selecting points within the grid at which to collect information (Figure 2). Points were selected systematically with the following protocol: we selected a starting point near one author's (KPD) home and drove along major roads to sample at roughly 2-mile intervals through most of the city. In areas with a high density of app users, we used a sampling strategy designed to collect data more frequently at closer intervals (see below). At each point where profile data were observed, study staff used the "GeoLocation" app [29] to pinpoint the location of data collection to latitude and longitude.

Figure 1. Hypothetical example of social networking application profile data provided by application users (we used age, race, and distance in feet from our location in this analysis).

Approx. 1258 ft away Online 57 sec ago 23 years old Black Neg as of 8/14/14



Figure 2. Map of Atlanta (gray outline) including major interstates (black lines) and selected major roads (dark red lines); points in the figure represent 79 data collection locations.



#### Validation of Geolocation Data

In order to assess the accuracy of the geolocating app, we also recorded the global positioning system (GPS) location at a subset of the same points at several different days/times using both GeoLocation and a GPS unit (Garmin model GPSmap 60CS [30]). The GeoLocation app was found to be consistent with the GPS unit, with the mean of the difference between them of 144 feet (range 7-344 feet) over a total of 25 sampled points. The GeoLocation app was also used at the same 10 locations 6 months apart and found to give consistent results with a mean of the difference in location coordinates of 76 feet (range 0-232). Thus we found it sufficient to use the free GeoLocation latitude and longitude data available on the same device as the social networking app for our purposes, rather than using two different devices for data collection. See Figure 3 for a screenshot from the GeoLocation iPhone app, available from [29]. Similar tools available for Android devices [31] were not evaluated in this study.

Figure 3. Example of the GeoLocation app used in this study, which relies on cell towers and an Internet connection and provides latitude and longitude in decimal degrees.



#### **Data Collection**

At each sampling point, study staff collected screenshots of user profiles. These apps sort profiles based on distance from the user to other users. We collected profile data for either the 50 closest users or for all users within 2 miles of the sampling point, whichever was less. Profiles were saved on a password-protected iPod Touch. These data were entered into a database after field collection of the screenshots. Staff also recorded the day and time of data collection at each point. We calculated the total time spent collecting data as a process measure for this pilot study. For each profile recorded, we extracted self-reported race and age, and the reported distance from the sampled point (see Figure 1). Race was categorized as "white", "black", or "other", and age was recorded as a continuous variable. If a profile included no information on race or age, this was indicated with a missing value in the database. Because the main objective was to compare the distribution of persons reporting their race as white to those reporting their race as black, when either race or age were missing, we recorded missing race as "other" and missing age as missing. Individual profile data from each sampled point were aggregated as the number of users by self-reported race (grouped as white, black, and other) and self-reported age group (grouped as 18-24, 25-30, >30, or



#### Delaney et al

unknown), and summary measures comparing those reporting black or white race in their profiles (further described below) were calculated for Atlanta.

#### **Sampling Strategy**

At points where there were greater than 50 users within less than a 2-mile radius, we recorded the maximum distance to the  $50^{\text{th}}$  closest user (ordered by distance) and moved this same distance along city streets to establish the next sample point. Thus smaller radii were used in areas with a higher density of

users. Figure 4 shows the sampling radii for each point: the smaller circles represent the areas of Atlanta with the highest density of users, and thus larger numbers of individual profiles available within a given (eg, 2 mile) radius. Because we collected different numbers of users from circles of different radii, we chose to standardize these measures to a common area, for example, converting each observation into the number of users within 1 mile of the point (thus describing a circle with a mile radius and/or an area of  $\pi$  square miles), and stratifying these measures by race and age group.

Figure 4. Map of Atlanta showing 79 data collection points from profiles on a sex-seeking networking app; radii of yellow circles represent distance to user sample at the maximum distance from the sample point, and overlapping circles completely cover Atlanta, with smaller circular areas used for data collection where there were the largest numbers of application users.



#### Analysis

The data in this study provide a somewhat unique challenge to geospatial statistical methods because they combine the characteristics of point and area processes [32-37]. Data are collected at points on a grid, but the data at that point represent a density over an area of sampling in a concentric circle around that point. Still, the data are more analogous to point data, with the measure collected at each point representing an area rather than an individual data point. Thus we chose to treat these densities of users per square mile as the measure of interest but use point data statistics [32,38,39] to summarize over the entire study area. ArcGIS [39] performs kernel smoothing to estimate the density measured at each sample point where each sample point is weighted by the observed population density at that point. In our case, the Kernel Density smoother [34] counts every white and black user observed at that location. For example, a point at which we observed 12 profiles within 2 miles, including 8 white and 4 black users, would be counted 8 times in the white density measure and 4 times in the black

density measure. Next, these weighted values for each point are also averaged with other points within a specified radius [32,36,37], resulting in a smoothed surface representing the density of users, by race, in the sample space. The kernel approach may place non-zero density in areas where no data were collected, but only as a result of averaging between points separated by the area with no data. We also experimented with methods for interpolation of spatial data such as kriging [32,38] and found similar results. We focus on kernel density estimates here. As noted above, sampling was conducted at different times and days of the week over a 6-month period (see Multimedia Appendix 1 for documentation of days and times sampled). While an in-depth analysis of time of day and day of week variability is of interest for future research, to illustrate our approach, we present the kernel densities calculated here as averages over sampled days and times.

After estimating the population density, we used ArcGIS to compute the mean and standard deviation for the calculated density measure over the entire sample space. We compared

density surfaces through ratio and difference measures via the Map Algebra tool in ArcGIS, which solves standard algebraic equations at each point in a grid across the density surface and creates a new map displaying the results of these calculations. When comparing the density of users, the difference between surfaces for different races, for example, (density of black users - density of white users) has the property that its null value (no difference) is zero, and if positive, it identifies an area with a higher density of black users than white users. This represents an absolute difference in the densities of the two groups. When positive, this approach identifies areas where it might be easier to recruit black users because the density of black users is greater in absolute terms (ie, the number of excess individuals). We note that this example says nothing about the magnitude (size of the density of black and/or white users), only that one number is bigger than the other. To capture areas where there are relatively more black users than white users (ie, the ratio of black to white users is higher), we also calculated the ratio of the two density surfaces.

As a further exploration of the possibilities with the approach, we also considered a measure to highlight areas with the largest densities for each race and then compare these areas as follows. First, for each density surface (eg, the density of black users <25 years of age) we identified areas with the highest density values (density value > mean + 2 SD). For example, if the estimated mean density for white users was 14/square mile with standard deviation of 7, we would ask ArcGIS to select points with a density of white users greater than 28. We then used Map Algebra to calculate the difference between the surfaces including these highest density points for each race according to the following formula:

I(Density of black users > mean + 2SD of estimated kernel density distribution) - I(Density of white users > mean + 2SD of estimated kernel density distribution)

where I(statement) represents an indicator function with value 1 if the statement is true and zero otherwise. This equation takes only three values: zero when a point is greater than mean + 2SD of both distributions or neither is greater than mean + 2SD; 1 when a point is greater than the mean + 2SD for only the first distribution; and -1 when the point is only greater than the mean + 2SD of the second distribution. This measure identifies not only locations with more users of a given race, but also locations with the highest density areas overall. Similar measures can be constructed to highlight other features of interest, for example, comparing densities by age group or combinations of race and age. Finally, to provide some context to our results, we present them in relation to the location of recruitment sites seeking to enroll MSM for two ongoing HIV prevention studies in Atlanta.

## Results

Over a 2-week period, we spent a total of 21 hours traversing Atlanta, collecting data at the 79 sample points (Figure 2) covering 883 square miles of area (Figure 4) in order to collect overlapping circles of data and cover the entire 132.4 square miles in the city of Atlanta. The average radius of data collection at each sample point was 1.65 miles, with smaller radii resulting from the more densely populated areas.

```
http://www.jmir.org/2014/11/e249/
```

We extracted profile data (race and age) for 2666 user profiles. Of these, 1563 (58.63%) were white, 810 (30.38%) were black, 146 (5.48%) were some other race, and 147 (5.51%) did not report a race in their profile. The mean age was 31.5 years, with 591 (22.17%) between the ages of 18-25, and 496 (18.60%) between the ages of 26-30. Age was more likely than race to be missing from profile information with 593 (22.24%) of profiles sampled not providing age information. The remaining 37% of profiles reported ages greater than 30; whites were more likely to report being >30 years of age than blacks (46% vs 25%, P<.001). Black users were younger than white users (median 28 vs 33 years, P<.001 via the Wilcoxon Sign rank test).

Across the 79 sampled points, the mean number of users was 33 per square mile, but the distribution of users across points was highly skewed with median of 17 and range 0.86-208 (Figure 5).

Figure 6 shows the density of app users, smoothed using a kernel density function with a 2-mile radius, for white (A) and black (B) users. A 2-mile radius was chosen as the smoothing parameter because it was the next largest integer that covered the average radius of 1.6 miles in the sampled points and also was the maximum distance to which we sampled data when a sample point had fewer than 50 users. Multimedia Appendix 1 shows the analogs of Figures 6 and 7 with a 1-mile kernel density smoothing parameter for comparison; the results were not qualitatively different. The highest density of white users (the darkest blues in the first panel in Figure 6) concentrates in Midtown Atlanta (roughly bounded by the yellow rectangle on the map). While much of the highest density of black users also concentrates in this area, it is clear that there are areas with high densities of black users further south and to the west (to the lower left) of Midtown. The kernel approach smooths observations according to a two-dimensional distribution centered at the observed point and declining out to the radius used to define the search area, essentially "spreading" observations from sample points across the study area. For example, the density values for white users over the 79 sample points ranged from 0.3 to 154 profiles per square mile, but the range of values for the smoothed density shown in the first panel in Figure 6 was 0-57 profiles per square mile. For the 1-mile smoothed density (Multimedia Appendix 1) the range (0-138) was closer to the observed values, but with many more points with density estimates of zero (ie, observations were not "spread" as far).

There are several ways to compare surfaces to illustrate local differences between the densities of white and black users. Figure 7 shows two similar but nonidentical ways to compare these densities. Panel A in Figure 7 shows the difference between the two surfaces, colored so that areas with higher absolute density of white users are blue and areas with higher density of black users are red. Panel B in Figure 7 shows the relative difference, with areas where the ratio of black to white profile densities is higher than one as red and lower than one as blue.

The ratio measure shows that most of Southwest Atlanta has relatively more black user profiles observed than white profiles,

```
XSL•FO
RenderX
```

but when we compare the map with that of the overall number of black users, we find a much smaller region in which to focus efforts, that is, south and west of Midtown, shown with a yellow band in Figure 7.

A third way to visualize differences between the surfaces is to focus on the areas with extreme values. This provides a within-density comparison: over the entire surface of the density of black user profiles, where is the density the greatest? In Figure 8, we highlight the regions with density greater than the mean+2 standard deviations over the entire map, separately for all white (A), all black (B), and young black (<25 years old, C) users based on data in their observed profiles. This approach

again highlights Midtown Atlanta (yellow rectangle) as the region with the most users observed in each graph.

Figure 9 calculates the difference between the first two panels in Figure 8 and shows that black user profiles have high density much further south than white profiles.

The third figure included in Multimedia Appendix 1 compares the difference between the 1-mile smoothed densities for young black and all black users (an analog to Figure 9 but comparing panels B and C of Figure 8). Overall the results are similar, but there are a few additional areas (highlighted in Multimedia Appendix 1 figure) with extreme densities of young black users that did not appear in the 2-mile estimates shown in Figure 8c or 9).

**Figure 5.** Histogram showing distribution of observed density of social network application users per 1-mile circle for the 79 sampled locations in Atlanta (inset includes statistics for the distribution, which is highly skewed with SD estimated to be larger than the mean; numbers above bars are the number of sample points with density along the X axis and the Y axis representing the percent of all points with this density).



#### Delaney et al

Figure 6. Estimated density of white (A) and black (B) social network application users in Atlanta (gray outline), showing major highways (black lines) and roads (dark red lines) and highlighting the "Midtown" area of Atlanta (yellow rectangle); kernel densities estimated from sample data standardized to 1-mile circular radii and smoothed to 2 miles using a Gaussian smoother that concentrates the majority of the density at the sample point and averages over all adjacent data points within the smoothing radius.



**Figure 7.** Comparison of the density of black and white social networking application users in Atlanta. Panel A shows the absolute difference in users (Density of black users – Density of white users) color-coded so that areas with more black users appear red and those with more white users appear blue. Yellow regions are areas where the two densities are similar. Panel A highlights a small section of the city (the area shaded the darkest red) where there are many more black than white application users. Panel B shows a comparison of the relative size of the densities of black and white users (Density of black users/Density of white users). With this measure, Atlanta is divided nearly in half, with relatively more black users in the southwest and more white users to the north and east. The yellow band in Panel B shows the region with the highest absolute excess of black users for comparison purposes.



http://www.jmir.org/2014/11/e249/

XSL•FO RenderX J Med Internet Res 2014 | vol. 16 | iss. 11 | e249 | p. 9 (page number not for citation purposes)

#### Delaney et al

**Figure 8.** Density of social networking application users in Atlanta, highlighting points with values >95th percentile of estimated kernel densities for white (Panel A), black (Panel B), and young black (<25 years of age, Panel C) users. For Panel A, points with an estimated density >17.2 users/mile<sup>2</sup> are highlighted dark blue; for Panel B those >5.65/mile<sup>2</sup> are dark red, and for Panel C >2.8/mile<sup>2</sup> are dark green. The yellow rectangle highlights the midtown area of Atlanta for reference. The yellow oval in Panel B highlights an area with high density of black users but not white users. The yellow circle in Panel C highlights an area with a high density of young black users, but not black users overall (ie, an area highlighted in Panel C but not Panel B).



**Figure 9.** Difference between extreme values of estimated kernel densities of white and black users of a social networking application. We use the formula I(Density of black users > mean+2 standard deviations) – I(Density of white users > mean+2 standard deviations). Figure shows regions where the values of this equation are -1 (green shading, indicating areas with extremes of density for white but not black users), 1 (red shading, indicating areas with extremes of the density or black but not white users), and 0 (white shading indicating areas that are either not extremes of either density or are extremes for both races).





XSL•FU RenderX

## Discussion

#### **Principal Findings**

We sampled 2666 profiles from a mobile phone–based social networking app at 79 sites in Atlanta and, under our sampling protocol, observed a mean of 33 app users per square mile. We also identified areas where there were more black and young black user profiles observed compared to white user profiles, describing three different summary measures of the density of profiles in a sampling frame. Finally, we showed the impact of the choice of the kernel radius in construction and interpretation of such data.

#### **Application of the Proposed Methods**

The goal of this study was primarily descriptive, in that we sought to describe a method for calculating the density of user profiles by race and age in Atlanta, and to compare and contrast the information provided by different outcome measures that can be constructed from these data. In addition, the methods described here may have practical application in HIV prevention research. The results are promising and illustrate how the use of self-reported location data can provide information on the geographic distribution of users in time and space. The study methodology could provide a more efficient way to identify locations for recruitment of MSM in future studies. Significant time and effort is spent on formative research to develop sampling frames for studies of MSM [15,21]. The goal of such formative research is to identify locations for sampling MSM using time space sampling methods [15]. Our methodology, based on the geolocation data incorporated into popular social networking apps, allowed us to quickly describe the density of sex-seeking MSM in Atlanta. Furthermore, we were able to use profile information to stratify these density measures by race and age. This might allow for oversampling or exclusive sampling in areas of the city that are expected to yield a particular subset of the population, for example, young black MSM. As an example, Figure 10 illustrates how these data can inform study implementation in practice. Figure 10 shows Panel B of Figure 7 and a variation of Figure 9, along with recruitment venues currently in use for two HIV prevention studies in Atlanta (green triangles). Panel A in Figure 10 shows that, to date, there have not been very many sampling locations in the southwestern part of Atlanta, where, based on the ratio of the density of black to white app users, there are relatively more black users than white users. However, Panel B in Figure 10 shows the difference between extremes for the densities of young white and young black users of the social networking app, using a formula similar to that used to calculate Figure 9. Looking at this representation of the data, we see that we have identified recruitment venues in an area of the city where there are the most young black app users and not that many white users. In this case, while going further into the areas of higher relative densities of black users might yield additional recruitment sites, we seem to have covered the areas with the highest number of both black and white users. Also, we find that there are not many recruitment sites outside of the area with the highest densities of white users, black users, or both, confirming that past recruitment sites were located in parts of the city where there are the most app users overall. Further

```
http://www.jmir.org/2014/11/e249/
```

potential applications of this methodology include identification of areas with need for prevention services, for example, overlaying HIV testing locations on the density grid to identify local areas with greatest unmet need.

Since the early 2000s, there has been a significant rise in Internet usage by MSM [40-43] and young minorities [44-46]. Three different groups have found that gay men now report meeting the majority of their sex partners online [40,47-49], and many [43,47-49] but not all [50] studies of sex behavior have shown increased reports of behaviors associated with higher HIV risk among partners met online compared to offline. The most popular and well-studied of these location-based social networking apps is Grindr [51-53], which is currently being used by over 4 million men worldwide [25] and is likely to continue to grow in popularity. MSM use this app for a variety of purposes, but a survey of Grindr users in Los Angeles found that 76% have had sex with someone they met on Grindr [51], suggesting that Grindr users are using the app to help find sex partners. Many other similar apps exist such as Adam-4-Adam, Jack'd, and BoyAhoy [26-28], and our methodology can be applied to any such app that provides data on race and age as well as distance to the user within member profiles. In our research, we have found that users of these apps vary by race and less so by age, with, for example, a greater proportion of white men reporting using Grindr and more black men reporting using Jack'd (unpublished Emory University data). In this study, although we illustrate our approach using only one app (and have chosen not to identify the specific app used to generate these data), we did validate the methodology with more than one app. Any of the apps that report race, age, and other characteristics of interest (eg, HIV serostatus), as well as geographic distance from the user's present location, could be used to make density maps and calculate summary statistics using the methods we report in this paper. In some cases, it may be useful to calculate one or more density measures with more than one app to try to get a better overall picture of the spatial distribution of men using sex-seeking apps in a given location.

Because users of these apps make both their profile information and their location public, it was possible to simply observe these publically available data without contacting the users directly for this research. However, there is still an ethical requirement to protect individually identifying information when the information is collected for research purposes. In this study, we used screen captures to record profile information, storing these pictures on a password-protected iPod Touch until the data of interest (age, race, and location information) could be entered into a database with no identifiers. Because we were recording only publically available data from user profiles without identifiers, the Institutional Review Board at Emory University considered the study to be research exempt from review.

More generally, using social networking apps for HIV prevention is likely a key strategy for future research [52-54] but comes with new ethical and methodological questions. Our study only sought to summarize the data publically available within these apps, but social media apps may themselves serve as an important public health communications tool. Recently, public health agencies have sought to partner with Grindr and use its built-in advertisements as a medium for disseminating

XSL•FO RenderX

prevention information and recruit MSM for research studies [52,53]. Future research might adapt our methodology further to establish a sampling frame and then use the density information to sample app users and contact them to either conduct a cross-sectional survey or recruit them into a follow-up

study. At that time, one would have to develop mechanisms for consenting study participants, as well as a way to keep sensitive information, such as sex and drug use behavior, protected and ideally separate from any identifying online profile information.

Figure 10. Application of two density metrics to evaluate recruitment for HIV prevention studies in Atlanta, showing recruitment venues currently in use for two HIV prevention studies (green triangles). Panel A shows few recruitment locations in the southwestern part of Atlanta, where there are relatively more young black application users than white users. Panel B uses the formula (Density of young black users > mean+2 standard deviations) – (Density of young white users > mean+2 SD). Regions where the values of this equation are -1 (blue shading, indicating areas with extremes of density for young white but not young black users), 1 (red shading, indicating areas with extremes of the density for young black but not young white users), and 0 (white shading indicating areas that are either not extremes of either density or are extremes for both races) can then be compared to the locations of current recruitment venues.



# Limitations of the Current Work and Opportunities for Future Research

Piloting this methodology in Atlanta exposed other challenges as well. Atlanta is geographically large and contains both densely settled neighborhoods in the inner city along with a large amount of semi-urban and even rural areas with less dense populations. Atlanta also exhibits a large degree of geospatial segregation by race, both in the population overall [55] and in the relative measures of the distribution of social-network app users (Panel B in Figure 7). However, although the overall black population density is low in Midtown Atlanta [55,56], it still represented the area with the highest concentration of black users of the sex-seeking application (Figure 4). To obtain a picture of the distributions of both black and white app users, we therefore had to sample enough points with a sufficiently wide radius to cover the entire city. We also found the density of users to vary widely within the city, and we therefore had to adapt our sampling strategy. We chose to collect either the first 50 profiles and record the distance to the 50<sup>th</sup> user, or to sample out to a 2-mile radius if there were less than 50 profiles observed in that area. In areas with large numbers of users, we had to collect data at more closely sampled points. For example, if



there were 50 profiles within a half mile, we moved only that short distance before collecting more data. If there were only 13 users within the 2 mile radius, we moved the full 2 miles between sample points. This allowed us to cover the whole city, but despite collecting data at 79 points that represented an area equivalent to 882 square miles, there were still areas of the city where we did not directly sample any users.

This makes the choice of the smoothing parameter (radius) for the kernel smoothing algorithm important because it provides a balance between too much interpolation of data between sampling points and presuming that the data collected at a particular sampling point occur only at the point and do not represent an area defined by the radius of a circle based on the linear distance to the person whose profile is being observed. Using our sampling plan, we collected data from concentric circles with an average radius of 1.65 miles and then fit weighted kernel densities smoothed to 1 and 2 miles. Both of these smoothing parameters provided similar interpretations of density of black, white, and young black individuals, with the 1-mile radius leaving more areas of the city with no estimates for the density of app users. The 2-mile radius covers the whole city, but as a result it reduces the emphasis of several points which,

when using a 1-mile radius are considered to have a particularly high density of black users.

Some questions remain about the precise interpretation of the density of social-network app users. For example, are users simply a subset of all MSM seeking sex on the Internet? Is the population that uses any one of these apps different by important characteristics (race, age, sex behavior with persons met through a social-network app or with sex partners generally) from the underlying population? Are persons who use specific services (eg, Adam-4-Adam, Jack'd, Grindr) different by one or more of these characteristics than those that use other online apps [47-49]? Future studies [54] will seek to quantify the density and characteristics of men who use each of these apps and compare the characteristics of men who use each of the apps exclusively, while also capturing information about men who use more than one service to describe whether their behaviors vary when using different services.

It would be useful to test the methodology in other cities with significant minority MSM populations (eg, Washington, DC, or Los Angeles, CA) and also to assess the utility of the method in less densely populated areas (eg, in rural areas of Georgia), to describe the extent to which the utility of the methods vary by characteristics of the geography of the region. We have already identified that Atlanta is a challenging place to conduct this kind of study because of its racial distribution, which was borne out in the social-network app user density data. In areas with sparse numbers of users, our adaptive sampling methodology, which sampled a maximum of 50 users or to a 2-mile radius, might help to stabilize density estimates, but this needs further testing. Additionally, although we averaged over day and time of sampling in our current analysis, the method could be refined to capture spatiotemporal trends in density. For example, it would be possible to select points to be sampled multiple times over a grid of specific times and days [14,15]. This modification could provide a clear description of how the user profile's population density changes over the course of a week. This last component may identify trends in the spatial and temporal clustering of app users, for example on weekend nights, as compared to mid-day during the work week.

#### Conclusions

We have found that it is possible to use a limited number of sample points to develop a geospatial density of men using a social-networking app to seek sex in the city of Atlanta. Such a density could serve as a sampling frame for future cross-sectional or longitudinal research. We also describe several methods to compare two densities with a goal of identifying areas with a high density of a particular subset of the population. We hope that this novel methodology and its further adaptations will prove useful to future research and prevention efforts that can be tailored to areas of the community where they will be most effective.

#### Acknowledgments

This work was supported in part by the Emory Center for AIDS Research (P30 AI050409) National Institutes of Mental Health R01-MH00085600. Thanks to Dr Eli Rosenberg and Adam Vaughn for helpful comments during the development of the project and development of the paper.

#### **Conflicts of Interest**

None declared.

## **Multimedia Appendix 1**

Kernel density calculations and days/times of data collection.

[PPTX File, 992KB-Multimedia Appendix 1]

#### References

- Centers for Disease Control and Prevention. Diagnoses of HIV Infection in the United States and Dependent Areas, 2011. 2013 Feb. HIV Surveillance Report, 2011; vol 23 URL: <u>http://www.cdc.gov/hiv/library/reports/surveillance/2011/</u> <u>surveillance\_Report\_vol\_23.html</u> [accessed 2013-12-12] [WebCite Cache ID 6LodIMTIRIR]
- Centers for Disease Control and Prevention. HIV Surveillance Supplemental Report 17(4); 2012. 2012 Dec. Estimated HIV incidence among adults and adolescents in the United States, 2007-2010 URL: <u>http://www.cdc.gov/hiv/pdf/statistics hssr vol 17 no 4.pdf</u> [accessed 2013-12-12] [WebCite Cache ID 6LodWskUl]
- 3. Centers for Disease ControlPrevention (CDC). Prevalence and awareness of HIV infection among men who have sex with men --- 21 cities, United States, 2008. MMWR Morb Mortal Wkly Rep 2010 Sep 24;59(37):1201-1207 [FREE Full text] [Medline: 20864920]
- Prejean J, Song R, Hernandez A, Ziebell R, Green T, Walker F, HIV Incidence Surveillance Group. Estimated HIV incidence in the United States, 2006-2009. PLoS One 2011;6(8):e17502 [FREE Full text] [doi: 10.1371/journal.pone.0017502] [Medline: 21826193]
- 5. whitehouse.gov. 2010 Jul. National HIV/AIDS Strategy URL: <u>http://www.whitehouse.gov/sites/default/files/uploads/NHAS.</u> pdf [accessed 2013-12-12] [WebCite Cache ID 6LodbOGzf]

- Millett GA, Flores SA, Peterson JL, Bakeman R. Explaining disparities in HIV infection among black and white men who have sex with men: a meta-analysis of HIV risk behaviors. AIDS 2007 Oct 1;21(15):2083-2091. [doi: 10.1097/QAD.0b013e3282e9a64b] [Medline: 17885299]
- Oster AM, Wiegand RE, Sionean C, Miles IJ, Thomas PE, Melendez-Morales L, et al. Understanding disparities in HIV infection between black and white MSM in the United States. AIDS 2011 May 15;25(8):1103-1112. [doi: 10.1097/QAD.0b013e3283471efa] [Medline: 21505305]
- Maulsby C, Millett G, Lindsey K, Kelley R, Johnson K, Montoya D, et al. HIV among Black men who have sex with men (MSM) in the United States: a review of the literature. AIDS Behav 2014 Jan;18(1):10-25. [doi: 10.1007/s10461-013-0476-2] [Medline: 23620241]
- 9. Bing EG, Bingham T, Millett GA. Research needed to more effectively combat HIV among African-American men who have sex with men. J Natl Med Assoc 2008 Jan;100(1):52-56 [FREE Full text] [Medline: 18277808]
- Jeffries WL, Marks G, Lauby J, Murrill CS, Millett GA. Homophobia is associated with sexual behavior that increases risk of acquiring and transmitting HIV infection among black men who have sex with men. AIDS Behav 2013 May;17(4):1442-1453. [doi: 10.1007/s10461-012-0189-y] [Medline: 22569942]
- Millett GA, Jeffries WL, Peterson JL, Malebranche DJ, Lane T, Flores SA, et al. Common roots: a contextual review of HIV epidemics in black men who have sex with men across the African diaspora. Lancet 2012 Jul 28;380(9839):411-423. [doi: 10.1016/S0140-6736(12)60722-3] [Medline: 22819654]
- 12. Ayala G, Bingham T, Kim J, Wheeler DP, Millett GA. Modeling the impact of social discrimination and financial hardship on the sexual risk of HIV among Latino and Black men who have sex with men. Am J Public Health 2012 May;102 Suppl 2:S242-S249. [doi: 10.2105/AJPH.2011.300641] [Medline: 22401516]
- 13. Heckathorn DD. Respondent-driven sampling: a new approach to the study of hidden populations. Social Problems 1997;44(2):175-200.
- 14. MacKellar DL, Vallaroy L, Karon J, Lemp G, Janssen R. The young men's survey: methods for estimating HIV seroprevalence and risk factors among young men who have sex with men. Public Health Reports 1996;111(Suppl 1):138-144.
- 15. MacKellar DA, Gallagher KM, Finlayson T, Sanchez T, Lansky A, Sullivan PS. Surveillance of HIV risk and prevention behaviors of men who have sex with men--a national application of venue-based, time-space sampling. Public Health Rep 2007;122 Suppl 1:39-47 [FREE Full text] [Medline: <u>17354526</u>]
- Ompad DC, Galea S, Marshall G, Fuller CM, Weiss L, Beard JR, et al. Sampling and recruitment in multilevel studies among marginalized urban populations: the IMPACT studies. J Urban Health 2008 Mar;85(2):268-280 [FREE Full text] [doi: 10.1007/s11524-008-9256-0] [Medline: 18214686]
- Sanchez T, Smith A, Denson D, Dinenno E, Lansky A. Developing a Web-Based HIV Behavioral Surveillance Pilot Project Among Men Who Have Sex with Men. Open AIDS J 2012;6:224-231 [FREE Full text] [doi: 10.2174/1874613601206010224] [Medline: 23091580]
- Johnston LG, Prybylski D, Raymond HF, Mirzazadeh A, Manopaiboon C, McFarland W. Incorporating the service multiplier method in respondent-driven sampling surveys to estimate the size of hidden and hard-to-reach populations: case studies from around the world. Sex Transm Dis 2013 Apr;40(4):304-310. [doi: <u>10.1097/OLQ.0b013e31827fd650</u>] [Medline: <u>23486495</u>]
- Montealegre JR, Johnston LG, Murrill C, Monterroso E. Respondent driven sampling for HIV biological and behavioral surveillance in Latin America and the Caribbean. AIDS Behav 2013 Sep;17(7):2313-2340. [doi: <u>10.1007/s10461-013-0466-4</u>] [Medline: <u>23568227</u>]
- 20. Pollack LM, Osmond DH, Paul JP, Catania JA. Evaluation of the Center for Disease Control and Prevention's HIV behavioral surveillance of men who have sex with men: sampling issues. Sex Transm Dis 2005 Sep;32(9):581-589. [Medline: 16118608]
- Lansky A, MacKellar D, Gallagher KM, Lin LS, Sullivan PS, Onorato IM. Evaluation of the new National HIV Behavioral Surveillance System (NHBS) is imperative. Sex Transm Dis 2006 Apr;33(4):272-3; author reply 274. [doi: 10.1097/01.olq.0000215745.61542.d4] [Medline: 16565646]
- 22. Manopaiboon C, Prybylski D, Subhachaturas W, Tanpradech S, Suksripanich O, Siangphoe U, et al. Unexpectedly high HIV prevalence among female sex workers in Bangkok, Thailand in a respondent-driven sampling survey. Int J STD AIDS 2013 Jan;24(1):34-38. [doi: 10.1177/0956462412472300] [Medline: 23512512]
- 23. Jenkins RA. Recruiting substance-using men who have sex with men into HIV prevention research: current status and future directions. AIDS Behav 2012 Aug;16(6):1411-1419. [doi: <u>10.1007/s10461-011-0037-5</u>] [Medline: <u>22016329</u>]
- 24. Kendall C, Kerr LR, Gondim RC, Werneck GL, Macena RH, Pontes MK, et al. An empirical comparison of respondent-driven sampling, time location sampling, and snowball sampling for behavioral surveillance in men who have sex with men, Fortaleza, Brazil. AIDS Behav 2008 Jul;12(4 Suppl):S97-104. [doi: <u>10.1007/s10461-008-9390-4</u>] [Medline: <u>18389357</u>]
- 25. Grindr. URL: <u>http://grindr.com/</u> [accessed 2013-12-12] [WebCite Cache ID 6LodfWVwq]
- 26. Jack'd. URL: <u>http://www.jackd.mobi/</u> [accessed 2013-12-12] [WebCite Cache ID 6LoduUIDX]
- 27. Adam-4-Adam Radar. URL: http://radar.adam4adam.com/ [accessed 2013-12-12] [WebCite Cache ID 6Loe4v0I0]
- 28. BoyAhoy. URL: <u>http://www.boyahoy.skout.com/</u> [accessed 2013-12-12] [WebCite Cache ID 6LoePJ6e1]
- 29. GeoLocation. URL: <u>https://itunes.apple.com/us/app/geolocation/id376832615?mt=8</u> [accessed 2013-12-12] [WebCite Cache ID 6LoeUcVPP]

- 30. Garmin geolocation tools. URL: <u>https://buy.garmin.com/en-US/US/on-the-trail/discontinued/gpsmap-60cs/prod6447.html</u> [accessed 2013-12-12] [WebCite Cache ID 6LoecZCII]
- 31. Android GeoLocation apps. URL: <u>https://play.google.com/store/apps/details?id=com.exposure101.</u> geolocation&feature=search\_result#?t=W10 [accessed 2013-12-12] [WebCite Cache ID 6LoejGmQ7]
- 32. Waller LA, Gotway CA. Applied spatial statistics for public health data. Hoboken, NJ: John Wiley & Sons; 2004.
- 33. Bivand RS, Pebesma EJ, Gomez-Rubio V. Applied spatial data analysis with R. New York: Springer; 2008.
- 34. ESRI ArcGIS Help. ArcGIS Online Documentation. How Kernel Density Works URL: <u>http://resources.arcgis.com/en/help/</u> <u>main/10.1/index.html</u> [accessed 2013-12-12] [WebCite Cache ID 6Loevcer4]
- 35. Silverman B. Density estimation for statistics and data analysis. London: Chapman and Hall; 1986.
- 36. Kafadar K. Smoothing geographical data, particularly rates of disease. Stat Med 1996 Dec 15;15(23):2539-2560. [doi: 10.1002/(SICI)1097-0258(19961215)15:23<2539::AID-SIM379>3.0.CO;2-B] [Medline: <u>8961462</u>]
- 37. Bithell JF. A classification of disease mapping methods. Stat Med 2000;19(17-18):2203-2215. [Medline: 10960848]
- 38. ESRI ArcGIS Help. ArcGIS Online Documentation. Empirical Bayesian Kriging (Geostatistical Analyst) URL: <u>http://resources.arcgis.com/en/help/main/10.1/index.html</u> [accessed 2013-12-12] [WebCite Cache ID 6Lof78o32]
- 39. ArcGIS 10 (Software). URL: <u>http://www.esri.com/products</u> [accessed 2013-12-12] [WebCite Cache ID 6LofNmaCJ]
- Bolding G, Davis M, Sherr L, Hart G, Elford J. Use of gay Internet sites and views about online health promotion among men who have sex with men. AIDS Care 2004 Nov;16(8):993-1001. [doi: <u>10.1080/09540120412331292453</u>] [Medline: <u>15511731</u>]
- 41. Zhang D, Bi P, Hiller JE, Lv F. Web-based HIV/AIDS behavioral surveillance among men who have sex with men: potential and challenges. Int J Infect Dis 2008 Mar;12(2):126-131 [FREE Full text] [doi: 10.1016/j.ijid.2007.06.007] [Medline: 17884663]
- 42. Sullivan PS, Khosropour CM, Luisi N, Amsden M, Coggia T, Wingood GM, et al. Bias in online recruitment and retention of racial and ethnic minority men who have sex with men. J Med Internet Res 2011;13(2):e38 [FREE Full text] [doi: 10.2196/jmir.1797] [Medline: 21571632]
- 43. Kakietek J, Sullivan PS, Heffelfinger JD. You've got male: internet use, rural residence, and risky sex in men who have sex with men recruited in 12 U.S. cities. AIDS Educ Prev 2011 Apr;23(2):118-127 [FREE Full text] [doi: 10.1521/aeap.2011.23.2.118] [Medline: 21517661]
- 44. Whiteley LB, Brown LK, Swenson RR, Romer D, DiClemente RJ, Salazar LE, et al. African American adolescents and new media: associations with HIV/STI risk behavior and psychosocial variables. Ethn Dis 2011;21(2):216-222 [FREE Full text] [Medline: 21749027]
- 45. Whiteley LB, Brown LK, Swenson RR, Valois RF, Vanable PA, Carey MP, et al. African American adolescents meeting sex partners online: closing the digital research divide in STI/HIV prevention. J Prim Prev 2012 Feb;33(1):13-18 [FREE Full text] [doi: 10.1007/s10935-012-0262-3] [Medline: 22293979]
- Cameron KA, Salazar LF, Bernhardt JM, Burgess-Whitman N, Wingood GM, DiClemente RJ. Adolescents' experience with sex on the web: results from online focus groups. J Adolesc 2005 Aug;28(4):535-540. [doi: 10.1016/j.adolescence.2004.10.006] [Medline: 16022887]
- 47. Grov C. HIV risk and substance use in men who have sex with men surveyed in bathhouses, bars/clubs, and on Craigslist.org: venue of recruitment matters. AIDS Behav 2012 May;16(4):807-817. [doi: 10.1007/s10461-011-9999-6] [Medline: 21748276]
- 48. Grov C, Crow T. Attitudes about and HIV risk related to the "most common place" MSM meet their sex partners: comparing men from bathhouses, bars/clubs, and Craigslist.org. AIDS Educ Prev 2012 Apr;24(2):102-116. [doi: 10.1521/aeap.2012.24.2.102] [Medline: 22468972]
- 49. Mayer KH, Ducharme R, Zaller ND, Chan PA, Case P, Abbott D, et al. Unprotected sex, underestimated risk, undiagnosed HIV and sexually transmitted diseases among men who have sex with men accessing testing services in a New England bathhouse. J Acquir Immune Defic Syndr 2012 Feb 1;59(2):194-198 [FREE Full text] [doi: 10.1097/QAI.0b013e31823bbecf] [Medline: 22027871]
- 50. Bolding G, Davis M, Hart G, Sherr L, Elford J. Gay men who look for sex on the Internet: is there more HIV/STI risk with online partners? AIDS 2005 Jun 10;19(9):961-968. [Medline: <u>15905678</u>]
- 51. Rice E. Sex on Grindr: Young Men Who Have Sex with Men and Geo-Locating Social Networking Technology. 2012 Presented at: ISIS Annual Conference on New Media, Youth and Sexual Health, SexTech 2012; March 30-April 2, 2012; San Francisco, CA.
- 52. Burrell ER, Pines HA, Robbie E, Coleman L, Murphy RD, Hess KL, et al. Use of the location-based social networking application GRINDR as a recruitment tool in rectal microbicide development research. AIDS Behav 2012 Oct;16(7):1816-1820 [FREE Full text] [doi: 10.1007/s10461-012-0277-z] [Medline: 22851153]
- Usher D, Frye V, Shinnick J, Greene E, Baez E, Benitez J, et al. Recruitment by a Geospatial Networking Application for Research and Practice: The New York City Experience. J Acquir Immune Defic Syndr 2014 Sep 22:-. [doi: <u>10.1097/QAI.0000000000066</u>] [Medline: <u>25247432</u>]
- Sullivan PS, Grey JA, Simon Rosser BR. Emerging technologies for HIV prevention for MSM: what we have learned, and ways forward. J Acquir Immune Defic Syndr 2013 Jun 1;63 Suppl 1:S102-S107 [FREE Full text] [doi: 10.1097/QAI.0b013e3182949e85] [Medline: 23673879]

```
XSL•FO
RenderX
```

- 55. Bloch M, Carter S, McLean A. The New York Times. Mapping the 2010 US Census, Largest racial/ethnic groups in 2010 URL: <u>http://projects.nytimes.com/census/2010/map?view=MajorGroupView2010&lat=33.7615&lng=-84.285&l=11</u> [accessed 2014-09-03] [WebCite Cache ID 6Rx49628r]
- 56. Bloch M, Carter S, McLean A. The New York Times. Mapping the 2010 US Census, Distribution of racial/ethnic groups in 2010 URL: <u>http://projects.nytimes.com/census/2010/map?view=DotView2010&lat=33.7726&lng=-84.378&l=14</u> [accessed 2014-09-03] [WebCite Cache ID 6Rx44zPyU]

## Abbreviations

**HIV:** human immunodeficiency virus **MSM:** men who have sex with men

Edited by G Eysenbach; submitted 07.05.14; peer-reviewed by E Scrimshaw, HJ Rendina; comments to author 17.07.14; revised version received 04.09.14; accepted 22.09.14; published 18.11.14 <u>Please cite as:</u> Delaney KP, Kramer MR, Waller LA, Flanders WD, Sullivan PS Using a Geolocation Social Networking Application to Calculate the Population Density of Sex-Seeking Gay Men for Research and Prevention Services J Med Internet Res 2014;16(11):e249 URL: <u>http://www.jmir.org/2014/11/e249/</u> doi: 10.2196/jmir.3523 PMID: 25406722

©Kevin P Delaney, Michael R Kramer, Lance A Waller, W Dana Flanders, Patrick S Sullivan. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 18.11.2014. This is an open-access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/2.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

