Spring 5-15-2015

Characteristics of Disease Transmission, Geography and Risk in an Urban Population with Endemic HIV

Evelyn J. Olansky
Georgia State University

CHARACTERISTICS OF DISEASE TRANSMISSION, GEOGRAPHY AND RISK IN AN URBAN POPULATION WITH ENDEMIC HIV

By

EVELYN JOLENE OLANSKY

B.A., Philosophy-Neuroscience-Psychology

WASHINGTON UNIVERSITY IN ST. LOUIS

A Thesis Submitted to the Graduate Faculty

of Georgia State University in Partial Fulfillment

of the

Requirements for the Degree

MASTER OF PUBLIC HEALTH

ATLANTA, GEORGIA

30303

TABLE OF CONTENTS

Acknowledgements

I would like to acknowledge the guidance and direction provided by Dr. Richard Rothenberg, my advisor and thesis chair, without whom this thesis would certainly have been another boilerplate analysis of an NHANES data scrape. I would also like to acknowledge the direction and assistance of Dr. Dajun Dai of the GSU Geosciences department, without whom the most finicky technical aspects of this thesis may have proven insurmountable. I would also like to thank my family and loved ones who I will not name here, as this document will be analyzed by search engines and it would be a tacky homage to their support to muddy their current SEO profiles.

List of Tables

Faculty Approval Page

OVERLAPPING PERSONAL GEOGRAPHIES AND EXTRAPOLATED HIV TRANSMISSION RISK

by

EVELYN JOLENE OLANSKY

Approved by:

Richard Rothenberg 3-5-2015

Dr. Richard Rothenberg - Committee Chair Date

Dajun Dai 3-5-2015

Dr. Dajun Dai - Committee Member Date

ABSTRACT

INTRODUCTION : Prior research suggests that sexually transmitted disease is not uniformly distributed throughout populations and geographic areas. Several studies of the geography of STI and HIV revealed a consistent core-like distribution of certain infections, such as gonorrhea, syphilis and HIV/AIDS. These studies further theorized that the conditions that precipitate endemic STI and HIV are not bound to populations, but strongly influenced by network-level features of social groups.

METHODS : Two geographic areas in Atlanta, GA were compared – one set of 5 zipcodes in which HIV was highly endemic, and another set of 5 zipcodes in which HIV was only moderately endemic. Two hypotheses were tested in the study area. First, risk variables were selected and composited into a variable representing compound risk, or the presence of multiple risk factors in a single individual, and the distribution of compound risk across the two geographic areas was compared. Second, the correlation between social distance (as geodesic length) and geographic distance (as distance between the centroid of connected individuals) was compared across the two geographic areas.

RESULTS : Compound risk was far more prevalent in the high HIV area than in the moderate HIV area (OR: 3.549; 95% CI: 2.438 -- 5.165), even after controlling for potential confounders. A breakdown of the individual risk variables indicates that involvement in sex work (OR: 2.279; 95% CI: 1.549 – 3.354), history of injection drug use (OR: 4.377; 95% CI: 2.35 – 8.152), and having any disease status disparity (OR: 1.511; 95% CI: 1.113 – 2.086) were each significantly more prevalent in the high HIV area than the moderate HIV area, even when stratifying by gender. The examination of the correlation between social distance and geographic distance revealed markedly different correlations in the two geographic areas. For residents of the high HIV area, the Pearson's correlation score (CC: 0.17175; 95% CI: 0.154887 – 0.188492) was significantly higher than in the moderate HIV area (CC: 0.07021; 95% CI: 0.050822 – 0.08954).

CONCLUSION : Areas of high HIV endemicity are associated with at least two of the characteristics described by Rothenberg (2005) : a higher prevalence of individual compound risk than observed in low or moderate HIV areas, and a higher correlation between geodesic and geographic distance than observed in low or moderate HIV areas. If the observed higher correlation is true and can be replicated in other study locations and with other demographic groups, then it may be useful to examine whether areas exhibiting a similar correlation are host to higher than expected rates of HIV. The compound risk finding is in line with the kinds of behavior-oriented HIV/STI risk studies that have been historically emphasized, while the difference correlation between geographic and geodesic distance suggests that behavioral factors do not provide a complete explanation for observed differences in endemicity.

Key words : HIV/AIDS, Social Networking, GIS, Transmission Risk

Author's Statement Page

In presenting this thesis as a partial fulfillment of the requirements for an advanced degree from Georgia State University, I agree that the Library of the University shall make it available for inspection and circulation in accordance with its regulations governing materials of this type. I agree that permission to quote from, to copy from, or to publish this thesis may be granted by the author or, in his/her absence, by the professor
under whose direction it was written, or in his/her absence, by the Associate Dean, College of Health and Human Sciences. Such quoting, copying, or publishing must be solely for scholarly purposes and will not involve potential financial gain. It is understood that any copying from or publication of this dissertation which involves potential financial gain will not be allowed without written permission of the author.

Evelyn J. Crowley

Signature of Author

3/6/15

Date

Notice to Borrowers Page

All theses deposited in the Georgia State University Library must be used in accordance with the stipulations prescribed by the author in the preceding statement.

The author of this thesis is :

Evelyn Jolene Olansky

37 Wiltshire Drive

Avondale Estates, GA 30002

The Chair of the committee for this thesis is :

Dr. Richard Rothenberg

College : School of Public Health

Georgia State University

P.O. Box 3995

Atlanta, Georgia 30302-3995

Users of this thesis who not regularly enrolled as students at Georgia State University are required to attest acceptance of the preceding stipulation by signing below. Libraries borrowing this thesis for the use of their patrons are required to see that each user records here the information requested.

NAME OF USER	**ADDRESS**	**DATE**	**TYPE OF USE *(EXAMINATION ONLY or COPY)***

Introduction

The dynamics of disease transmission are multi-faceted and highly variable. As our understanding of the biomechanics of disease improves, the role of social mechanics in disease transmission becomes more apparent. It is not enough to describe only the physiological pathways of transmission. The social organization of those who might transmit disease exerts as much influence, if not more, over patterns of transmission. Disease prevalence, which is often reported for large administrative areas like counties, states and nations, may be more fruitfully analyzed as interpersonal, street-level phenomena. Often, a disproportionate share of a large area's disease may be found in a relatively small area, among small social groups within which those diseases are highly endemic. This thesis aims to explore the interpersonal and geographic characteristics that contribute to disease endemicity within small social groups.

Over 30 years worth of epidemiological research suggests that sexually transmitted infections (STIs), such as gonorrhea, syphilis or HIV, are not evenly distributed throughout population centers, but rather disproportionately prevalent in a small number of locations while exhibiting sporadic prevalence elsewhere (Rothenberg, 1983; Potterat et al., 1985). Some traditional risk factors, such as injection drug use or high concentrations of men who have sex with men, can be found in such areas of disproportionate disease burden, yet the presence of these and other factors does not fully account for the maintenance of endemicity observed (Rothenberg, 1983). In pursuit of an explanation for this phenomenon, a number of studies have been conducted to explore the potential factors that contribute to these observed patterns of relative prevalence.

One factor thought to contribute to disease endemicity within social groups is geographic distance and proximity to groups of varying characteristics. While incident STIs can and do occur in individuals who live at a great distance from the source of transmission,

not only does the risk of singly-occurring infection increase with geographic proximity to areas of higher prevalence, but repeat infections are also more likely to occur in close geographic proximity to other subjects of repeat infection (Bernstein et al., 2004). This proximity effect is thought to underlie the observation that distinct administrative zones with markedly high prevalence are often geographically contiguous (Kerani et al., 2005; Zenilman et al., 2002).

Literature Review

Geography **-** Previous research into the geographic distribution of STI prevalence revealed a core-like distribution for some infections, in which the area of highest prevalence is the core, and an area's disease prevalence decreases with increased distance from the core area (Rothenberg 1983, Alvarez-Dardet 1985, Zenilman 1988, Zenilman 2002, Law 2004). The core-like distribution was found to occur in areas of endemic gonorrhea, syphilis, chlamydia, and HIV (Rothenberg 1983, Alvarez-Dardet 1985, Zenilman 1988, Heimer 2008) though later research suggests that the distribution of chlamydia is not as cleanly core-like as the distributions of syphilis and gonorrhea (Zenilman 2002, Kerani 2005). If a large enough area is studied, multiple disease cores within an area may be identified (Jennings 2005, Kerani 2005). Beyond the distribution of simple disease prevalence, it was also found that repeat infections are distributed in a core-like pattern (Bernstein 2004). Jennings and colleagues found that the core-like distribution persists even when controlling for known demographic confounders of the disease in question (Jennings 2005). It was later demonstrated that core-like distributions are not restricted to single types of infections, but can be applied to infection with STIs in general. Proximity to areas of high prevalence of any STI markedly increases the odds of being infected with any STI (Jennings 2010).

Hixson and colleagues identified a core-like geographic distribution of HIV cases in Atlanta, Georgia, as well as significant associations between living in a high-HIV geographic

cluster and higher rates of participation in specific risk factors (e.g. injection drug users, and men who have sex with men) (Hixson 2011). When an area of endemic infection is discovered through analysis, new contacts with the infection can be readily located by recruiting in the endemic area (Goswami 2012). More recent investigations into the limits of core theory suggest that core-like distributions are not found in rural settings of high endemicity (Gesink 2012), nor necessarily in urban areas of high endemicity where cultural factors overwhelmingly affect the distribution of risk factors (Ross 2012). A longitudinal study of HIV and IDU distribution in the San Francisco bay area showed that an urban core of disease endemicity may persist for 2 decades without markedly changing shape (Martinez 2014).

Networks - Existing research on the effect of network characteristics on disease transmission provides multiple avenues for exploration. Among the most prominent network features thought to relate to transmission are assortativity, concurrency, and the presence of compound risk. Assortativity, or within-group partner selection, limits a group's force of infectivity in the population as a whole, (Rothenberg & Potterat, 1987) while somewhat increasing the disease endemicity within highly assortative groups (Rothenberg et al. 2001) (Newman, 2003). Low burden groups in high risk areas were characterized by the absence of within-group assortativity over time, wherein partners were not shared among close friends or other partners, and few individuals maintained multiple consistent sexual partners over time (Rothenberg et al., 1998). More cohesive social groups seem to provide a more efficient platform for disease transmission and endemicity than less cohesive social groups, even which risk behaviors are equal across groups (Potterat et al., 1999). High assortativity also has the effect of making connected components more robust, such that the removal of a single node is less likely to have an impact on the overall connectedness of a given connected component (Newman 2003). These findings suggest that high network

assortativity, including areas with geographic characteristics that promote high assortativity, may play a large role in establishing and maintaining disease endemicity.

Some individual risk factors that are known to be associated with individual risk are similarly associated with risk within a network, even among group members who do not directly engage in those risk behaviors. Partner concurrency greatly increases transmission speed and breadth, both in simulation and reality. The greater the extent of concurrency within a social network, the higher the risk of disease transmission, and the higher the speed of that transmission (Morris and Kretzschmar, 1997; Rothenberg et al., 2000; 2001). The presence of individuals in networks who exhibit compound risk factors, injection drug use, involvement with sex work, or engagement in same-sex sexual activity, is associated with higher rates of disease in those networks (Rothenberg et al., 2000; 2001).

The social customs or habits within individual social groups exert influence over the rate of disease transmission in those groups, as well. For instance, social factors, such as the desire to maintain an extramarital affair in secret or other possible motivations for disguising a sexual relationship, often prevail over health risk concerns when selecting partners and sexual behaviors (Hirsch et al., 2007). Among drug users, the particular use habits and “customs” of intravenous drug users are highly dependent on the “customs” or habits of their friends who also use intravenous drugs, suggesting that norms pertaining to risk behaviors may be defined (or redefined) at the level of social group (Latkin et al. 2009). Beyond social customs, sometimes economic conditions serve as the prevailing determinant of the extent of risk behaviors within a particular social group. The use of condoms or avoiding any involvement with sex work are less frequent among populations in which multiple socio-economic hardships are present, including intravenous drug use, recent homelessness or general financial hardship (Gorbach et al., 2009).

Geography and Network - While the effects of geographic and network characteristics have been extensively studied independently, far fewer studies have examined both in concert. In a town identified for its high rate of gonorrhea, researchers found that the disease was largely isolated to a specific social network. Within that network were observed multiple, overlapping patterns of social behavior, each of which contributes to the establishment and maintenance of high disease endemicity among a disproportionately small share of that city's population, concentrated in only 4 census tracts. The group was highly self-assortative, with nearly half of all reported sexual contacts occurring between individuals who are familiar in the long-term (Potterat, et al., 1985). Another study suggests that couples in areas of higher STI prevalence exhibit a lower mean distance between partners than areas of lower STI prevalence. Couples within the core disease areas lived significantly closer together than couples outside of those core areas (Zenilman et al., 1999). Rothenberg and colleagues observed an increased association of geographic distance with social distance in areas of high HIV endemicity. Their findings suggest that disease endemicity in a group might be maintained in part by a strong association between social distance and geographic distance, which increases assortativity by increasing the likelihood that group members will make contact with strangers who are nonetheless members of the same social network, and thus share the same within-group transmission risks (Rothenberg et al., 2005). Lastly, participants living in areas of high HIV-prevalence are more likely to select spatially assortative partners than those living in areas of low-HIV prevalence (Gindi et al., 2011). While the emerging model could be more coherent, the overarching suggestion is that disease transmission is determined, at least in part, by the interrelationship of geographic and network characteristics. Higher geographic density and within-group assortativity are each independently associated with transmission

risk. Geographic density and within-group assortativity also exhibit an association with each other that is stronger in areas of high HIV and disease endemicity.

Methods

Background - The following analysis was conducted on secondary data, gathered by Rothenberg and colleagues at Emory University and Georgia State University between 2005 and 2011, under the auspices of a grant from the National Institute on Drug Abuse. The data were collected in Atlanta, GA, as part of an examination of sexually transmitted disease transmission risk and geography within inner-city social networks.

The stated aims of the Geography project were fourfold : 1.1. To examine the extent to which participants are exposed to disease transmission risk factors within their immediate social networks--chiefly patterns of drug use, needle-sharing and sexual activity with same- and opposite-sex partners. 1.2. To describe the geographic bounds of the participants' social networks. 1.3. To investigate whether these social groups exhibit network characteristics thought to facilitate transmission. 2.0. To determine the overall interaction of participants' behavior, patterns of interrelation, and geographic scope with regard to the prevalence of 7 infections : Human Immunosuppressive Virus, Syphilis, Herpes Simplex Virus Type 2, Chlamydia, Trichomoniasis, Gonorrhea and Hepatitis C Virus.

To these ends, the analyses contained herein were conducted to test the following related hypotheses about the co-contributions of geographic features, transmission risk factors and social network characteristics to the occurrence and maintenance of disease endemicity :

Hyp. 1. Persons with high compound risk will be more prevalent in areas of high endemicity than in areas of lower endemicity.

Hyp. 2. Study participants in areas of high endemicity will exhibit stronger correlation

between geographic distance and social distance (as measured by minimum geodesic path length) than areas of moderate endemicity.

Study Area - Participants were recruited from an inner city area of Atlanta composed of 10 urban zipcodes areas (figure 1). The 10 zipcodes were selected based on their relative burden of HIV and STI cases, with 5 zipcodes classified as “High Burden” (30308, 30310, 30314, 30315 and 30318) and 5 zipcodes classified as “Intermediate Burden” (30311, 30331, 30337, 30344 and 30349). Taken together, these 10 zipcodes form a conterminous study area of 240 square miles. The study area is contiguous with the exception of zipcode 30330, a non-civilian zipcode known as Fort McPherson, which occupies 1.1 square miles on the border between the High burden and Intermediate burden areas. Considered high and intermediate areas separately, the high burden zipcodes have an area of 67 square miles, while the intermediate regions have an area of 173 square miles. The high and intermediate regions meet along a border which is 16.7 miles in length.

Participant Recruitment - The study participants were recruited using a chain-link design. Rothenberg and colleagues conducted a 6-month ethnographic study to select 30 “seed” participants over age 18 who did not know each other. For each zipcode area, 3 “seeds” were selected on the basis of some ethnographic demonstration of experience with either sexual activity or drug use. Each “seed” was interviewed and asked to provide a list of contacts, from which the “seed” nominated a single contact to serve as the next source of new contacts. The secondary contacts were each asked to nominate a tertiary contact to be the final source of new contacts. In this way, each group contained the aggregated contacts of a chain of 3 core participants: the “seed”, the secondary contact, and the tertiary contact. Contacts could be listed in multiple groups, but each “seed” was unique to its respective group.

Survey Contents and Methods - Each respondent was interviewed using a standard questionnaire with items pertaining to their behavior, methods and extent of travel, medical history, socio-demographics, as well as sexual and drug use histories. The questionnaire was drafted by the researchers, then fine-tuned using observations gathered during the ethnographic period. For each contact named, the respondents were asked to describe the nature of the relationship (strength, duration, shared activities and activity locations). Respondents were each offered testing for HIV, Hepatitis C, Herpes-2, Syphilis, Gonorrhea, Chlamydia, and female respondents were also offered testing for Trichomoniasis, and the results of those tests were recorded. Interviews were repeated annually until each respondent had provided 3 interviews, or else were lost to attrition. No additional respondents were added after the primary round of interviews was completed.

Variables used **:**

Sociodemographic variables - From the researchers' original set of sociodemographic variables, the present analysis includes educational attainment, age at interview, race/ethnicity, gender, sexual orientation, religion and employment type. Though the survey included 9 educational strata, these were recoded into 3 strata : 1. no high school diploma or equivalent; 2. high school diploma or equivalent only; 3. some college or more. The survey recorded race/ethnicity as one of 9 options, which were recoded into 4 groups for the purposes of this analysis : 1. Black (African-American and Caribbean); 2. Hispanic (Black and White); 3. White; 4. Other (Native American Indian/Alaskan Native, Asian/Pacific Islander, Mixed, and Other).

The survey recorded sexual orientation as one of 8 options, which were recoded as 4 groups for the purposes of this analysis : 1. LGB (Gay, Lesbian, Bisexual or Homosexual); 2. Transgender; 3. Heterosexual; 4. Other. The survey recorded age at screening as a continuous variable, which was recoded into quartiles for the purposes of this analysis : 1.

18-23; 2. 24-34; 3. 35-46; 4. 47 and older. The survey recorded gender as one of 3 options, which were not recoded : 1. Female; 2. Male; 3. Other. All respondents who selected "Other" specified their gender as "Transgender Female".

Risk Variables – *Contact-based risk variables* - Participants were asked to indicate some features of their drug use and sexual history with each named contact, including whether they had ever had sex or shared drugs with each contact, as well as more detailed questions about the dates on which those activities last occurred. By calculating the difference between the stated dates of contact and the interview date, we calculated the total number of days since each dated encounter. If a respondent was found to have engaged in sexual activity with a named contact within 182.5 days, or half a year's time, that relationship was coded as a sexual relationship. Similarly, if a respondent indicated drug use with a named contact within the same duration, that relationship was coded as a drug relationship. For each contact, respondents were asked to indicate whether they had shared needles within the last 3 months, and affirmative responses were coded as needle relationships. Any contact listed who did not meet the criteria for sexual, drug or needle-type relationships was coded as a social relationship. A relationship could be coded as simultaneously "sexual", "drug" and "needle" in nature, but all "social" relationships definitionally lacked sexual, drug or needle characteristics.

For cases in which two contact reports conflicted--ie. one respondent declared a relationship type with a second respondent, while the second respondent either described the same relationship differently or not at all--the report reflecting greater interpersonal risk was favored over the conservative report.

Individual risk variables - HIV status was also included as a potential confounding variable, as previous studies indicate that it also may have some bearing on geographic

distance. Some self-reported characteristics of sex and drug history were used as individual-level risk variables. Characteristics of sex history used as risk variables were reporting multiple female sex partners within the last 6 months, reporting multiple male sex partners within the last 6 months, reporting ever having a sex partner who was an intravenous drug user, reporting any involvement with prostitution (reporting prostitution as one's occupation, or reporting ever accepting or giving money or drugs in exchange for sex), as well as reporting any non-use of condoms for anal or vaginal sex. Characteristics of drug history used as risk variables were reporting any use of intravenous drugs, as well as initiation of hard drugs (crack, cocaine, heroin and amphetamine) as a minor.

Disease Status Disparities - A comparison was made between each participant's self-reported disease status and the results of the lab testing performed as part of this study in order to determine the prevalence of disease status disparities. Status disparities of the false-negative type--cases where participants report no disease while testing revealed the presence of disease--were summed into a single variable. Notably, a small handful of false-positive type status disparities were revealed, but these were excluded from analysis due to being too few in number.

Geographic Variables : *Respondent midpoints* - Each respondent furnished a list of their frequented locations, as well as a list of locations where they met with each contact they listed, resulting in an aggregated list of locations roughly describing each respondent's personal geography. All identified locations found to be outside of the Atlanta Regional Commission's planning area (a 10-county area roughly 18 times larger than the core study zipcode areas) were excluded from analysis, such that the remaining locations represented each respondent's travels within the metro Atlanta area only. GPS coordinates were gathered for each listed location. For each respondent, the geographic midpoint of their coordinates was calculated and used for all subsequent calculations of geographic distance

between respondents. The geographic midpoint of each participant was calculated as the weighted centroid of every location they identified as a meeting point. Any location identified multiple times by the same participant was included that number of times in the centroid calculation. Centroids were calculated using a version of the Haversine formula modified for finding the midpoint of a group of greater than 2 points. The coordinates of each location were converted into Cartesian coordinates (X,Y,Z), and a weighted average was calculated for each of the three dimensions of the Cartesian coordinates – X, Y and Z respectively – producing the individual's geographic midpoint as a single Cartesian coordinate, which was then converted back to geodetic latitude and longitude.

Group midpoints - For each of the 30 distinct social groups in the study, a corresponding list of all locations named by its members was created, redundantly including locations named by multiple group members. The frequency of each location was calculated, and locations with a frequency of 1 (that is, only one group member identified that location) were discarded. The geographic midpoint of the remaining locations was calculated, with each location weighted by the total number of group members that reported that location. The resulting coordinates represent the point of greatest overlap for each group's members.

Social Distance – For a hypothetical Participant X, there are N contacts, N_{DIRECT} of whom are direct contacts, and $N_{INDIRECT}$ of whom are indirect contacts. There are 2 types of direct contact : 1. Study participants identified by Participant X as direct contacts; 2. Study participants who identified Participant X as a direct contact. An indirect contact is defined here as a participant that is not a direct contact of Participant X, but nonetheless linked socially. If Participant X is not a direct contact of Participant Z, but an uninterrupted line of direct contacts can be traced from Participant X and Participant Z, then they are indirect

contacts. For each participant, indirect contacts were identified by multiply layering direct contacts over increasing depths of social distance until no novel dyads were revealed. Each subsequent set of connections was calculated in toto, such that the set of all connections of geodesic length 2 contained the original set of all connections of geodesic length 1, and the set of all connections of geodesic length 3 contained within in the complete set of geodesic length 2 connections, and so on. Any paths passing through the same respondent more than once were excluded. For each pairing of two participants, *X* and *Y*, the shortest observed geodesic path was selected and all longer paths ultimately discarded, creating a master set of every unique connected pair and their corresponding minimum geodesic path length.

Informed Consent

The informed consent form, study protocols and other materials were approved by the Emory University Institutional Review Board at the study's outset, and approved again by Georgia State University's IRB when the study was moved to that institution. Interviewers administered a screener survey, and all those deemed eligible were asked to read and sign an informed consent form for the study. The informed consent included details about the study's HIPPA compliance, measures to secure and de-identify physical specimens collected, in addition to other relevant study details. For potentially vulnerable participants, such as parolees and pregnant women, additional information was provided to ensure that the specific risks and benefits of participation were fully understood. Consenting participants each provided a list of their contacts, and each subsequent contact was asked to sign their own informed consent form prior to participating in the survey. When listing their contacts, participants were given the option of anonymity. Where the option of anonymity was waived, interviewers had the choice of telling subsequent study recruits the names of those who had listed them as a contact. Where anonymity was requested, interviewers were instructed

simply to approach the new potential participant and ask them to participate in the study, without reference to the previous participant that named them as a contact.

Statistical Analyses

Statistical analyses were performed using SAS version 9.3 (SAS Institute, Inc.) Separate analyses were conducted to test each of the related hypotheses described above.

Univariate Analyses :

Hypothesis #1 : Prevalence of Compound Risk - first, risk variables were selected through individual univariate analysis using Chi-squares to determine the association between each potential variable and overall confirmed disease burden in each participant. Variables were selected if their presence significantly increased the odds of diagnosis with 1 or more infection or disease, and all those which had no significant relationship to overall disease burden were discarded, leaving the following set of seven risk variables : reporting 5 or more male partners within the last 6 months, reporting 10 or more sexual partners of both sexes in the last 6 months, ever being an injection drug user, ever being the partner of an injection drug user, having any number of disease status disparities, or being a sex worker of any kind. These risk factors were assigned binary values (0 or 1) and summed to create a compound risk value for each participant. A Chi-Square analysis was performed with odds ratio and confidence intervals calculated. The compound risk was treated as binary, with any summed risk score over 1 counted as the presence of compound risk.

Hypothesis #2. Correlation of Geographic and Social Distance - as previously described, network analysis was used to produce a list of every participant's distal social connections, as well as the minimum geodesic path length for each connection, to complement the list of direct connections provided by each participant. For every linked pair of participants, the geographic distance between their respective midpoints was calculated

using the Haversine formula for calculating the distance between two points on a great circle. Using the geographic distance between pair members and the minimum geodesic path length, a Pearson's correlation coefficient was calculated for both high and moderately endemic areas. A second set of Pearson's correlation coefficients were calculated after excluding all hybrid pairs, or pairs for whom one participant is from the highly endemic area and the other is from the intermediate endemicity area.

Multivariate analyses :

Hypothesis #1 : Prevalence of Compound Risk - Multivariate analyses were conducted using logistic regression to control for the effect of the five significant sociodemographic variables -- age quartile, gender, education, race and sexual orientation -- upon the association between compound risk and residence in the area of high HIV endemicity. Several logistic regressions were performed in total. To analyze the correlations with the predictor variable, residence in the high HIV area, two regressions were performed : one which took the independent variables as separate variables, and another using the composite variable, Compound Risk, as the sole independent variable. To analyze the correlations with the outcome variable, HIV status, the same two regressions were performed : one which took the independent variables as separate variables, and another using the composite variable, Compound Risk, as the sole independent variable. Each of these four regressions was performed again after stratifying by male and female gender, in each case excluding the gender demographic variable. Associations were examined as odds ratios with 95% confidence intervals, with significance defined as a confidence interval that did not intersect 1.00. For each demographic variable included in the regression analysis, the stratum with the highest N was analyzed as the reference group against which the other strata were compared.

Results

Basic Characteristics (Table 1) :

A total of 904 adults were included in the present analysis. Responses indicate that 420 (46.5%) were female, 469 (51.9%) were male, and 15 (1.66%) were transgender women. Racial classification revealed a largely black study population of 885 (97.9%), with 6 (0.7%) hispanic respondents, 9 (1.0%) non-hispanic white respondents, and 4 (0.4%) respondents of other racial backgrounds. Mean participant age was 36.2 years with a standard deviation of 12.8. The cutoffs for age quartiles were selected for groups of optimally equal size, with 18 to 23 year olds numbering 215 (23.8%), 24 to 34 year olds numbering 220 (24.3%), 35 to 46 year olds numbering 234 (25.9%) and those over 47 numbering 235 (26.0%). Regarding sexual orientation, 100 (11.1%) were either lesbian women, gay men or bisexual, while 17 (1.9%) did not identify with any of the options provided, and 786 (87.0%) of participants identified as heterosexual.

The majority of the participants, 417 (46.2%), reported educational attainment levels below high school equivalency, while 354 (39.2%) earned a high school diploma or GED, and the remaining 132 (14.6%) pursued education beyond high school. The majority of the participants, 466 (51.5%) were unemployed, while another 187 (20.7%) were employed. Among those neither definitionally unemployed or employed were 18 (2.0%) students, 12 (1.3%) for whom home duties and/or child care were their primary occupation, 9 (1.0%) retirees, 86 (9.5%) not employed due to disability, 36 (4.0%) with illegal occupations (drug dealer or prostitute), 68 (7.5%) earning money via other means, and 22 (2.4%) presumably unemployed participants providing no indication of their occupational status. Of the 904 participants, 49 (5.4%) were HIV positive and 855 (94.6%) were HIV negative. 482 (53.3%) of participants resided in zipcodes with a high HIV burden, and while 422 (46.7%) of participants resided in zipcodes with a moderate HIV burden.

Regarding the independent variables, 58 (6.42%) participants reported having sex with 6 or more male partners in the past 6 months, while 846 (93.6%) reported fewer than 6 male partners in that time. Sixty-nine (7.6%) participants reported having 10 or more total sexual partners in the past 6 months, while 835 (92.4%) reported fewer than 10 sexual partners of any gender. 101 (11.2%) participants reported ever having used injection drugs, while 803 (88.8%) reported no history of injection drug use. 23 (2.5%) participants reported having ever been the partner of an injection drug user (IDU), while 881 (97.5%) participants reported that they had never knowingly been a partner of an IDU. 467 (51.7%) participants were revealed to have 1 or more status disparities, or a positive diagnosis for at least one disease for which they report a negative status, while 437 (48.3%) had no status disparities. Two hundred and twenty-seven (25.11%) participants reported having any history of sex work, while 677 (75.9%) reported no such history. 387 (42.8%) participants tested negative for every infection covered within the study, while 517 (57.19%) participants tested positive for at least one of the infections under examination.

Hypothesis #1 : Compound Risk - Univariate analysis - Each independent risk variable was tested for association with overall disease/infection load. Odds ratios were calculated for each of the independent risk variables and their association with positive infection diagnoses (Table 2). Having 5 or more male partners within the last 6 months (OR : 2.034, 95% CI : 1.038 -- 3.985, p = .0004) significantly increased the odds of having one or more positive diagnoses. Being a sex worker significantly increased the odds of having one or more positive disease diagnoses (OR : 2.084, 95% CI : 1.511 -- 2.874, p < .0001). Reporting 10 or more partners of any sex was significantly associated with having one or more positive disease diagnoses (OR : 2.136, 95% CI : 1.164 -- 3.919, p = 0.0123). Participants reporting any history of injection drug use were significantly more likely to have one or more positive disease diagnoses (OR : 1.709, 95% CI : 1.099 -- 2.658, p = .0165).

Participants reporting ever having a relationship with an injection drug user showed increased odds of having two or more positive diagnoses, though with only marginal significance (OR : 1.879, 95% CI : 0.994 -- 3.552, p = .0491). Lastly, participants who were found to have at least one disease status disparity were more likely to have two or more positive diagnoses than participants with no disease status disparities (OR : 51.493, 95% CI : 22.535 -- 117.663, p < .0001). Among the potential risk variables tested and discarded due to not reaching significance were having multiple female partners (many threshold values were tested as the definition of "multiple", with 9 partners coming closest to significance without reaching it), any history of sex worker solicitation, and the use of non-injection drugs within 6 months or 30 days.

After combining the selected risk variables into the compound risk sum, a Chi-Square analysis was performed to test the extent to which the compound value was associated with odds of one or more disease diagnoses. Participants exhibiting compound risk, exhibiting two or more of the independent risk factors just described, were nine times more likely to have one or more disease diagnoses than participants with no compound risk (OR: 9.061, 95% CI : 5.822 -- 14.102, p < .0001).

Chi-Square analyses of the compound risk variable revealed significantly higher prevalence of compound risk in the highly endemic area when compared to the area of intermediate endemicity (OR : 4.785, 95% CI : 3.346 -- 6.843, p < .0001). This result stands as strong initial confirmation of the hypothesis that compound risk is more prevalent in the area of high endemicity than in the area of intermediate endemicity, though multivariate analysis must be carried out to determine whether this effect is maintained when controlling for possible confounders.

Hypothesis #1 : Multivariate analysis - Several logistic regressions were performed as described in the section describing statistical analyses, with the subsequent results

presented in the appendix (Table 5). In the model including the Compound Risk composite variable (Table 5.1), logistic regression analysis revealed a significant association between Compound Risk and residence in the study's high HIV area (OR : 3.549, 95% CI : 2.438 -- 5.165). Of the demographic and confounding variables included in the model, only age had a significant association with the outcome, residence in the high HIV area. More specifically, being in the two youngest age quartiles -- 18-23 years old (OR: 0.443, 95% CI : 0.294 -- 0.667) and 24-34 years old (OR: 0.498, 95% CI : 0.333 -- 0.746) -- appears to be a predictor against living in the high HIV area, when compared against 35 to 46 year olds as a reference group. No other potential confounders were significantly associated with the composite Compound Risk variable. The association of Compound Risk to residence in the high HIV area remained significant even after stratifying by gender, with a slightly stronger association appearing for women (OR: 3.852, 95% CI: 2.281 -- 6.506) than for men (OR: 3.354, 95% CI: 1.914 -- 5.877). Notably, age remained a significant factor for men, but failed to reach significance among women. Women presented no significant confounders in their stratified analysis, while analysis of the men revealed being gay or bisexual appears to be a moderate predictor against being a resident in the high HIV area (OR: 0.235, 95% CI: 0.092 -- 0.603).

The secondary analysis, in which the compound risk variables were treated independently and not as a composite variable (Table 5.2), revealed that 4 of 6 risk variables are significantly associated with residence in the high HIV area when controlling for confounders: any sexwork (OR: 2.279, 95% CI: 1.549 -- 3.354), sex with 5 or more male partners in 6 months (OR: 3.926, 95% CI: 1.306 -- 11.8), any history of injection drug use (OR: 4.377, 95% CI: 2.35 -- 8.152), and having any number of disease status disparities (OR: 1.544, 95% CI: 1.143 -- 2.086). When stratifying by gender, 3 of these 4 variables remained significant, with "Sex with 5 or more male partners" failing to reach significance

among females. The association of sexwork with residence in the high HIV area was markedly stronger among women (OR: 2.682, 95% CI: 1.548 -- 4.647) than men (OR: 1.946, 95% CI: 1.105 -- 3.429). As in the previous analysis of the composite Compound Risk, the two youngest age quartiles were the only potential confounders found to be significantly associated with residence in the high HIV area, with both age groups again appearing as predictors against living in the high HIV area.

Further logistic regression analyses were performed to assess the association between the independent variables and being HIV positive (Table 6). Broadly, the results greatly resemble those obtained from the analysis of the association between independent variables and residence in the high HIV area. In the model including the Compound Risk composite variable (Table 6.1), logistic regression analysis revealed a significant association between Compound Risk and positive HIV status (OR : 3.78, 95% CI: 1.94 -- 7.364). The association persisted when stratifying by gender, albeit with reduced confidence, with both males (OR : 4.548, 95% CI: 1.572 -- 13.16) and females (OR : 4.332, 95% CI: 1.656 -- 11.331) exhibiting the association. The analysis revealed no significant association of HIV status and any of the confounding variables, with the exception of being gay, lesbian or bisexual (OR : 2.684, 95% CI : 1.222 -- 5.897), which showed a significant association with being HIV positive. This association persisted and intensified among males (OR : 7.158, 95% CI : 2.213 -- 23.154), but failed to reach significance among females.

The secondary analysis of the association with HIV positive status, in which the risk variables were treated independently and not as a composite Compound Risk variable, again revealed results that broadly resembled those obtained in the analysis of association with residence in the high HIV area (Table 6.2). Any injection drug use (OR : 2.848, 95% CI : 1.167 -- 6.951) and any disease status disparity (OR : 11.124, 95% CI : 3.665 -- 33.763) were both significantly associated with being HIV positive. For men, the association of

injection drug use (OR : 7.089, 95% CI : 1.737 -- 28.938) and being HIV positive, as well as that of having any disease status disparity (OR : 15.679, 95% CI : 2.952 -- 83.281) and being HIV positive were markedly stronger than for all genders taken together. The association of injection drug use and being HIV positive failed to reach significance for women, but the association of disease status disparity persisted among women (OR : 11.3, 95% CI : 1.461 -- 87.387). As before, in the composite variable regression, a significant association between being gay, bisexual or lesbian and being HIV positive was revealed for both genders (OR : 3.324, 95% CI : 1.466 -- 7.536). The association was stronger for men (OR : 9.18, 95% CI : 2.169 -- 38.847) and not significant for women.

Hypothesis #2 : Correlation of Geographic and Geodesic Distance - Pearson's correlation coefficients were calculated for the set of all participants connected by any length of geodesic (Table 7), yielding a base correlation coefficient of 0.11715 (95% CI : 0.102676 -- 0.131559, $p < .0001$) for the full study population. Pearson's correlation coefficients were then calculated for two sub-populations : first, the set of all connected pairs with one or more participant from the area of intermediate endemicity, yielding a correlation coefficient of 0.07021 (95% CI : 0.050822 -- 0.08954, $p < .0001$); second, the set of all connected pairs with one or more participant from the area of high endemicity, yielding a correlation coefficient of 0.17175 (95% CI : 0.154887 -- 0.188492, $p < .0001$). These coefficients support the 2nd hypothesis.

After excluding all hybrid pairs, these coefficients were recalculated. Under these conditions, the Pearson's coefficients and their confidence intervals were almost entirely overlapping, with a value of .12182 for the highly endemic area (95% CI : .099844 -- .143661) and a value of .1028 for the intermediate endemicity area (95% CI : .075556 -- .129864). A correlation coefficient was also calculated for the set of hybrid pairs alone,

which predictably resulted in a very low correlation between geographic distance and geodesic distance (Pearson's CC : 0.0445, 95% CI : 0.016927 -- 0.071987).

Discussion - The results of the analyses support the hypotheses described at the outset of this work. The presence of compound risk in the high HIV area is safely 3 to 6 times higher than that found in the intermediate HIV area. This higher prevalence of individuals with compound risk undoubtedly contributes to the overall high endemicity of disease of this group. Failure to adopt safer sex and drug use practices for an individual with compound risk has an outsized influence on transmission risk when compared to those with only singular risk factors. While having multiple partners in a 6-month period or having a disease status disparity are risky when considered separately, the combination of the two greatly increases the risk of transmission. This finding suggests that, rather than interventions targeting single risk factors, a greater impact may be achieved by targeting those individuals who exhibit compound risk and working to lower their overall risk profile.

The results were also largely supportive of the second hypothesis regarding the correlation between geodesic path length and geographic distance between participant midpoints. The observed distribution of this correlation suggests that individuals in the highly HIV area live significantly closer to their contacts than those in the moderate HIV area. For those from the high HIV area, this highly local pattern of association increases exposure to others within the high HIV area. For residents of the moderate HIV area, the relatively lower correlation of geodesic path length and geographic distance suggests a greater opportunity to escape these local effects, while also somewhat increasing the likelihood of associating with individuals from the high HIV area. In this way, a high correlation in the high HIV area increases risk exposure, while a high correlation in the moderate HIV area could have a protective effect by virtue of preventing association with those at higher risk.

Limitations

A potential weakness of the compound risk variable is the inclusion of "Any disease status disparity" as one of the considered risk factors. Each risk factor used as part of the compound risk composite variable was tested for independent association with the total number of positive diagnoses, and variables found to not be associated with disease outcomes were excluded from the compound risk variable. For each variable considered, the association with having 1 or more positive diagnoses was tested. However, given that the odds that a person with a disease status disparity will have 1 or more positive diagnoses is 100%, the disease status disparity variable was tested for independent association with having 2 or more positive diagnoses. This comparison still holds some risk of collinearity, but the adjusted comparison mitigates this risk somewhat. "Any disease status disparity" is an undeniable risk factor for disease transmission. Any presence of disease within a social group represents an increased potential for transmission within that group, and disease status disparity represents the combination of disease presence alongside a lack of complete knowledge of disease presence. While behavior may ensure that the risk posed by a known disease presence can be avoided, an infected individual cannot plausibly act to avoid transmitting an infection they do not know they carry.

While polygon identification creates an approximation of the participants' geographic range based on points identified by each participant, there is significant risk that the boundaries of those polygons are artificial. A more complete analysis of each individual's range could be obtained through a more comprehensive survey that includes not only locations where each participant travels to, but the routes of their travel between each point. These routes could potentially be extrapolated as the most direct route via transit mode of choice, but the mixed preference and use of transit modes presents further methodological difficulties to such an approximation. Without significantly more intrusive recording of zones

of travel, a more exact representation of overlap within groups would be difficult, potentially relying on assumptions that could introduce further inaccuracy to a model that is already potentially inaccurate.

An additional dimension, time of travel, could also be considered to produce a more comprehensive model of participant travel and group member co-location. For instance, it's conceivable that two group members, each with areas of travel that appear to overlap without respect to time of travel, are actually occupying common geographical areas at different, non-overlapping times of day. A central assumption of the geographical overlay hypothesis is that geographic overlap represents an implicit opportunity for contact between participants. However, the temporal dimension weakens this assumption through the requirement that participants occupy the same geographic areas during the same periods of time.

Conclusions and Recommendations – Despite confirmation of the initial hypotheses, the work described herein must be replicated in order to adequately confirm its validity. If it is safe to assume that the results are valid, then it appears as if areas of high HIV endemicity are associated with at least two of the characteristics described by Rothenberg (2005) : a higher prevalence of individual compound risk than observed in low or moderate HIV areas, and a higher correlation between geodesic and geographic distance than observed in low or moderate HIV areas. Further research into the latter finding would do well to test the inverse of what is asserted here; that is, if the observed higher correlation is true and can be replicated in other study locations and with other demographic groups, then it may be useful to examine whether areas exhibiting a similar correlation are host to higher than expected rates of HIV. The finding that high HIV endemicity is associated with greater presence of compound risk behavior, while perhaps unsurprising, makes it clear that factors from many scopes contribute to an area's HIV level. The compound risk finding is in

line with the kinds of behavior-oriented HIV/STI risk studies that have been historically emphasized, while the difference correlation between geographic and geodesic distance suggests that behavioral factors do not provide a complete explanation for observed differences in endemicity.

Appendix 1 - Tables

Table 1. CORE VARIABLES		
Gender	N	%
Female	420	46.46
Male	469	51.88
Other	15	1.66
Race/Ethnicity	N	%
Non-Hispanic Black	885	97.90
Non-Hispanic White	9	1.00
Hispanic	6	0.66
Other	4	0.44
Age	Mean	STD
Mean	36.15	12.83
Age Quartiles	N	%
18-23	215	23.78
24-34	220	24.34
35-46	234	25.88
47 and older	235	26.00
Sexual Orientation	N	%
Lesbian	5	0.55
Gay	11	1.22
Bisexual	84	9.30
Heterosexual	786	87.04
Trans / Other	17	1.88
Education	N	%
Less than HS	417	46.18
HS Equivalent	354	39.20
Some college or more	132	14.62
Employment Type	N	%
unemployed	466	51.55
employed	187	20.69
student	18	1.99
illegal	36	3.98
home duties / child care	12	1.33
retired	9	1.00
disabled	86	9.51
other	68	7.52
no response	22	2.43

INDEPENDENT VARIABLES		
Male Partners w/i 6 months	N	%
0 - 5 male partners	846	93.58
6 or more male partners	58	6.42
All Gender Partners w/i 6 month	N	%
0 - 9 partners	835	92.37
10 or more partners	69	7.63
IDU	N	%
No history of injection drug use	803	88.83
Any history of injection drug use	101	11.17
Partner IDU	N	%
Never been partner of an IDU	881	97.46
Ever been partner of an IDU	23	2.54
Disease Status Disparities	N	%
No status disparities	437	48.34
1 or more status disparities	467	51.66
History of Sex Work	N	%
No history of sex work	677	74.89
Any history of sex work	227	25.11
Total Risk Factors	N	%
0	312	34.51
1	367	40.60
2	138	15.27
3	50	5.53
4	33	3.65
5	4	0.44
Compound Risk	N	%
0 or 1 total risk factors	679	75.11
2 or more total risk factors	225	24.89
Total # of current infections	N	%
0	387	42.81
1	316	34.96
2	149	16.48
3	43	4.76
4	8	0.88
5	1	0.11

>1 Current Infection	N	%
0 OR 1 infections	703	77.77
2 or more infections	201	22.23
HIV Status	N	%
HIV negative	855	94.58
HIV positive	49	5.42

Dependent Variable			
Zipcode Group	Zipcode	N	%
Zipcodes with High HIV burden	30308	77	8.52
	30310	79	8.74
	30314	87	9.62
	30315	115	12.72
	30318	124	13.72
	Total	***482***	***53.32***
Zipcodes with Low HIV burden	30311	87	9.62
	30331	68	7.52
	30337	82	9.07
	30344	119	13.16
	30349	66	7.30
	Total	***422***	***46.68***

Table 2 - Association of Risk Factors with Total Disease Diagnoses

Independent Association with 1 or more disease diagnoses

	OR	95% CI	Chi-Square	p-value
Compound Risk	8.689	5.620 -- 13.432	119.5178	p < .0001*
Sex Work	2.084	1.511 -- 2.874	20.4552	p < .0001*
>=10 Sex Partners	2.244	1.289 -- 3.906	8.5326	p = .0035**
>5 Male Sex Partners	3.052	1.594 -- 5.844	12.3861	p = .0004**
Ever Injected Drugs	1.709	1.099 -- 2.658	5.7495	p = .0165*
Independent Association with 2 or more disease diagnoses				
	OR	95% CI	Chi-Square	p value
Ever partner of IDU	2.307	0.984 -- 5.411	3.8965	p = .0484*
Any Status Disparity	51.498	22.536 -- 117.680	212.9169	p < .0001***

Table 3. Descriptive Statistics, stratified by HIV Burden Zone

3. 1 - Core Demographics	Zone HIV Burden				
Gender	Low	High	df	Chi-Square	p-value
Female	230	239	2	10.9424	0.0042**
Male	191	229			
Other	1	14			
Race/Ethnicity	Low	High	df	Chi-Square	p-value
Non-Hispanic Black	419	466	*3*	*10.0023*	*0.0185*
Non-Hispanic White	1	8			
Hispanic	0	6			
Other	2	2			
Age	Mean	STD	df	t-value	p-value
Age	33.1682	38.76	904	-6.44	<.0001***
95% CI	(31.96-34.36)				
Age Quartiles	Low	High	df	Chi-Square	p-value
18-23	131	84	3	44.2819	<.0001***
24-34	121	99			
35-46	83	151			
47 and older	87	148			

Sexual Orientation	Low	High	df	Chi-Square	p-value
Heterosexual	369	417	2	8.7919	0.0123*
Lesbian/Gay/Bisexual	50	50			
Transgender/Other	2	15			
Education	Low	High	df	Chi-Square	p-value
Less than HS	208	209	2	5.8396	0.0539
HS Equivalent	164	190			
Some college or more	50	82			
Employment Type	Low	High	df	Chi-Square	p-value
unemployed	217	249	8	27.8624	0.0005**
employed	103	84			
student	13	5			
illegal	6	30			
home duties / child care	6	6			
retired	5	4			
disabled	32	54			
other	28	40			
no response	12	10			
3.2 - INDEPENDENT VARIABLES					
Male Partners w/i 6 months	Low	High	df	Chi-Square	p-value
0 - 5 male partners	416	430	1	32.877	<.0001***
6 or more male partners	6	52			
All Gender Partners w/i 6 month	Low	High	df	Chi-Square	p-value
0 - 9 partners	409	426	1	23.2634	<.0001***
10 or more partners	13	56			
IDU	Low	High	df	Chi-Square	p-value
No history of injection drug use	406	397	1	43.4486	<.0001***
Any history of injection drug use	16	85			
Partner IDU	Low	High	df	Chi-Square	p-value
Never been partner of an IDU	417	464	1	5.8989	0.0152*
Ever been partner of an IDU	5	18			

Disease Status Disparities	Low	High	df	Chi-Square	p-value
No status disparities	240	197	1	23.0678	<.0001***
1 or more status disparities	182	285			
History of Sex Work	Low	High	df	Chi-Square	p-value
No history of sex work	363	314	1	52.1331	<.0001***
Any history of sex work	59	168			
Total Risk Factors	Low	High	df	Chi-Square	p-value
0	200	112	5	104.9138	<.0001***
1	175	192			
2	38	100			
3	6	44			
4	3	30			
5	0	4			
Compound Risk	Low	High	df	Chi-Square	p-value
No - 0 or 1 total risk factors	375	304	1	80.0657	<.0001***
Yes - 2 or more total risk factors	47	178			
Total # of current infections	Low	High	df	Chi-Square	p-value
0	217	170	3	31.845	<.0001***
1	139	177			
2	53	96			
3 or more	13	39			
>1 Current Infection	Low	High	df	Chi-Square	p-value
0 OR 1 infections	356	347	1	19.9072	<.0001***
2 or more infections	66	135			
HIV Status	Low	High	df	Chi-Square	p-value
HIV negative	414	441	1	19.1793	<.0001***
HIV positive	8	41			

Table 4 - Descriptive Statistics, stratified by HIV status

4.1 - Core

Demographics	**HIV Status**				
Gender	Negative	Positive	df	Chi-Square	p-value
Female	398	18	2	89.48	<.0001***
Male	451	22			
Other	6	9			
Race/Ethnicity	Negative	Positive	df	Chi-Square	p-value
Non-Hispanic Black	836	49	*3*	*1.1123*	*0.7741*
Non-Hispanic White	9	0			
Hispanic	6	0			
Other	4	0			
Age	Negative	Positive	df	t-value	p-value
Mean	35.8178	39.5102	904	-1.97	0.0494*
	(34.94 -- 36.69)	(36.62 -- 42.40)			
Age Quartiles	Negative	Positive	df	Chi-Square	p-value
18-23	213	2	3	11.7882	0.0081**
24-34	203	17			
35-46	218	16			
47 and older	221	14			
Sexual Orientation	Negative	Positive	df	Chi-Square	p-value
Heterosexual	757	29	2	85.4608	<.0001***
Lesbian/Gay/Bisexual	89	11			
Transgender/Other	8	9			
Education	Negative	Positive	df	Chi-Square	p-value
Less than HS	400	17	2	3.0971	0.2126
HS Equivalent	332	22			
Some college or more	122	10			
Employment Type	Negative	Positive	df	Chi-Square	p-value
unemployed	438	28	*8*	*24.4515*	*0.0019***
employed	185	2			
student	18	0			
illegal	31	5			
home duties / child care	12	0			
retired	9	0			
disabled	75	11			
other	66	2			
no response	21	1			

4.2 - INDEPENDENT VARIABLES

Male Partners w/i 6 months	Negative	Positive	df	Chi-Square	p-value
0 - 5 male partners	806	40	1	12.3247	0.0004**
6 or more male partners	49	9			

All Gender Partners w/i 6 month	Negative	Positive	df	Chi-Square	p-value
0 - 9 partners	794	41	1	5.5542	0.0184*
10 or more partners	61	8			

IDU	Negative	Positive	df	Chi-Square	p-value
No history of injection drug use	766	37	1	9.2582	0.0023**
Any history of injection drug use	89	12			

Partner IDU	Negative	Positive	df	Chi-Square	p-value
Never been partner of an IDU	835	46	*1*	*2.6752*	*0.1019*
Ever been partner of an IDU	20	3			

Disease Status Disparities	Negative	Positive	df	Chi-Square	p-value
No status disparities	432	5	1	30.1732	<.0001***
1 or more status disparities	423	44			

History of Sex Work	Negative	Positive	df	Chi-Square	p-value
No history of sex work	649	28	1	8.6766	0.0032**
Any history of sex work	206	21			

Total Risk Factors	Negative	Positive	df	Chi-Square	p-value
0	310	2	*5*	*47.661*	*<.0001****
1	250	17			
2	121	17			
3	44	6			
4	26	7			
5	4	0			

Compound Risk	Negative	Positive	df	Chi-Square	p-value
0 or 1 total risk factors	660	19	1	36.5878	<.0001***
2 or more total risk factors	195	30			

Total # of current infections	Negative	Positive	df	Chi-Square	p-value
0	387	0	3	150.9238	<.0001***
1	308	8			
2	127	22			
3 or more	33	19			
>1 Current Infection	Negative	Positive	df	Chi-Square	p-value
0 OR 1 infections	695	8	1	113.1023	<.0001***
2 or more infections	160	41			
Zipgroup	Negative	Positive	df	Chi-Square	p-value
Zipgroup 0	414	8	1	19.1793	<.0001***
Zipgroup 1	441	41			

Table 5.1 - Logistic Regression Analysis of the composite risk variable, stratified by gender, predicting odds of residing in High HIV Zipgroup Area

	Males (n=420)	**Females (n=469)**	**All (n=904)**
Compound Risk	**3.354 (1.914 -- 5.877)***	**3.852 (2.281 -- 6.506)***	**3.549 (2.438 -- 5.165)***
Demographic Variables			
Gender	**Males (n=420)**	**Females (n=469)**	**All (n=904)**
Male	---	---	1
Female	---	---	1.124 (0.841 -- 1.502)
Transgender	---	---	7.653 (0.354 -- 165.292)
Age Quartiles	**Males (n=420)**	**Females (n=469)**	**All (n=904)**
18-23 years old	**0.313 (0.173 -- 0.568)***	0.568 (0.316 -- 1.019)	**0.443 (0.294 -- 0.667)***
24-34 years old	**0.341 (0.188 -- 0.618)***	0.697 (0.393 -- 1.236)	**0.498 (0.333 -- 0.746)***
35-46 years old	1	1	1
47 and older	0.871 (0.512 -- 1.483)	1.005 (0.539 -- 1.873)	0.964 (0.648 -- 1.436)
Race/Ethnicity	**Males (n=420)**	**Females (n=469)**	**All (n=904)**
Non-Hispanic Black	1	1	1
Non-Hispanic White	---	2.547 (0.275 -- 23.555)	5.040 (0.585 -- 43.42)
Hispanic	---	---	---
Other	---	0.140 (0.010 -- 1.969)	0.522 (0.057 -- 4.778)

Education	**Males (n=420)**	**Females (n=469)**	**All (n=904)**
Less than HS	1	1	1
HS Equivalent	0.914 (0.591 -- 1.416)	1.094 (0.697 -- 1.717)	0.986 (0.725 -- 1.342)
Some college or more	0.939 (0.502 -- 1.756)	1.476 (0.788 -- 2.762)	1.203 (0.778 -- 1.860)
Sexual Orientation	**Males (n=420)**	**Females (n=469)**	**All (n=904)**
Heterosexual	1	1	1
Lesbian / Gay / Bi	**0.235** **(0.092 -- 0.603)***	1.188 (0.667 -- 2.115)	0.734 (0.461 -- 1.170)
Trans / Other	---	0.703 (0.031 -- 16.16)	1.042 (0.09 -- 12.086)
--- denotes an unavailable comparison	* denotes significance, α = .05	** denotes significance, α = .01	*** denotes α = .001

Table 5.2 - Logistic Regression Analysis of the independent risk variables, stratified by gender, predicting odds of residing in High HIV Zipgroup Area

Independent Variables	**Males (n=420)**	**Females (n=469)**	**All (n=904)**
Any sexwork	**1.946** **(1.105 -- 3.429)***	**2.682** **(1.548 -- 4.647)***	**2.279** **(1.549 -- 3.354)***
Sex with 5 or more male partners w/i 6 months	---	2.182 (0.427 -- 11.163)	**3.926** **(1.306 -- 11.8)***
Sex with 10 or more partners of any gender w/i 6 months	0.998 (0.376 -- 2.648)	1.752 (0.234 -- 13.135)	1.053 (0.452 -- 2.453)
Any Injection Drug Use	**4.571** **(2.106 -- 9.918)***	**4.368** **(1.401 -- 13.62)***	**4.377** **(2.35 -- 8.152)***
Ever been sexual partner of an Injection Drug User	1.307 (0.309 -- 5.53)	0.964 (0.155 -- 5.993)	1.129 (0.369 -- 3.46)
Any disease status disparity	**1.576** **(1.042 -- 2.386)***	**1.647** **(1.042 -- 2.603)***	**1.544** **(1.143 -- 2.086)***
Demographic Variables			
Gender	**Males (n=420)**	**Females (n=469)**	**All (n=904)**
Male	---	---	1
Female	---	---	1.046 (0.771 -- 1.420)
Transgender	---	---	6.172 (0.214 -- 177.904)

Age Quartiles	**Males (n=420)**	**Females (n=469)**	**All (n=904)**
18-23 years old	**0.412 (0.223 -- 0.763)***	0.713 (0.383 -- 1.327)	**0.554 (0.361 -- 0.85)***
24-34 years old	**0.435 (0.236 -- 0.799)***	0.790 (0.439 -- 1.422)	**0.587 (0.388 -- 0.886)***
35-46 years old	1	1	1
47 and older	0.82 (0.471 -- 1.427)	0.936 (0.489 -- 1.790)	0.881 (0.583 -- 1.333)
Race/Ethnicity	**Males (n=420)**	**Females (n=469)**	**All (n=904)**
Non-Hispanic Black	1	1	1
Non-Hispanic White	---	2.006 (0.18 -- 22.33)	4.428 (0.484 -- 40.488)
Hispanic	---	---	---
Other	---	**0.027 (0.001 -- 0.596)***	0.225 (0.016 -- 3.069)
Education	**Males (n=420)**	**Females (n=469)**	**All (n=904)**
Less than HS	1	1	1
HS Equivalent	0.952 (0.606 -- 1.497)	1.156 (0.726 -- 1.842)	1.016 (0.740 -- 1.393)
Some college or more	0.924 (0.486 -- 1.758)	1.536 (0.805 -- 2.929)	1.197 (0.767 -- 1.870)
Sexual Orientation	**Males (n=420)**	**Females (n=469)**	**All (n=904)**
Heterosexual	1	1	1
Lesbian / Gay / Bi	**0.134 (0.042 -- 0.43)***	1.09 (0.596 -- 1.990)	0.620 (0.379 -- 1.013)
Trans / Other	---	0.579 (0.016 -- 21.049)	0.808 (0.052 -- 12.598)
--- denotes an unavailable comparison	* denotes significance, α = .05	** denotes significance, α = .01	*** denotes α = .001

Table 6.1 - Logistic Regression Analysis of composite risk variable, stratified by gender, of the association between independent variables and HIV positive status			
Compound Risk	**4.548 (1.572 -- 13.16)***	**4.332 (1.656 -- 11.331)***	**3.78 (1.94 -- 7.364)***
Demographic Variables			
Gender			All
Male	1	1	1
Female	(--)	(--)	1.077 (0.545 -- 2.128)
Transgender	(--)	(--)	17.454 (0.933 -- 326.432)

Age Quartiles	Males	Females	All
18-23 years old	---	0.304 (0.034 -- 2.693)	0.129 (0.023 -- 0.716)
24-34 years old	1.497 (0.449 -- 4.986)	1.634 (0.478 -- 5.581)	1.386 (0.615 -- 3.12)
35-46 years old	1	1	1
47 and older	0.482 (0.129 -- 1.808)	2.417 (0.794 -- 7.355)	1.103 (0.492 -- 2.474)
Race/Ethnicity			
Non-Hispanic Black	1	1	1
Non-Hispanic White	---	---	---
Hispanic	---	---	---
Other	---	---	---
Education			
Less than HS	1	1	1
HS Equivalent	0.85 (0.263 -- 2.743)	1.39 (0.529 -- 3.651)	1.268 (0.619 -- 2.596)
Some college or more	1.757 (0.461 -- 6.701)	0.755 (0.189 -- 3.013)	1.267 (0.521 -- 3.081)
Sexual Orientation			
Heterosexual	1	1	1
Lesbian / Gay / Bi	**7.158** **(2.213 -- 23.154)***	1.454 (0.485 -- 4.362)	**2.684** **(1.222 -- 5.897)***
Trans / Other	---	---	1.925 (0.117 -- 31.639)
--- denotes an unavailable comparison	* denotes significance, α = .05	** denotes significance, α = .01	*** denotes α = .001

Table 6.2 - Logistic Regression Analysis of independent variables, stratified by gender, of the association between independent variables and HIV positive status

Independent Variables	**Males (n=420)**	**Females (n=469)**	**All (n=904)**
Any sexwork	0.68 (0.135 -- 3.42)	2.237 (0.827 -- 6.052)	1.406 (0.649 -- 3.043)
Sex with 5 or more male partners w/i 6 months	14.543 (0.272 -- 777.629)	---	1.484 (0.281 -- 7.841)
Sex with 10 or more partners of any gender w/i 6 months	---	---	0.471 (0.082 -- 2.722)
Any Injection Drug Use	**7.089** **(1.737 -- 28.938)**	1.892 (0.465 -- 7.698)	**2.848** **(1.167 -- 6.951)**
Ever been sexual partner of an Injection Drug User	3.694 (0.421 -- 32.381)	1.375 (0.114 -- 16.628)	3.094 (0.68 -- 14.087)
Any disease status disparity	**15.679** **(2.952 -- 83.281)**	**11.3** **(1.461 -- 87.387)**	**11.124** **(3.665 -- 33.763)**
Demographic Variables			
Gender			All
Male	**1**	1	1
Female	(--)	(--)	0.928 (0.459 -- 1.878)
Transgender	(--)	(--)	15.666 (0.302 -- 811.746)
Age Quartiles	Males	Females	All
18-23 years old	---	0.4 (0.044 -- 3.607)	0.321 (0.064 -- 1.612)
24-34 years old	2.268 (0.575 -- 8.944)	1.543 (0.452 -- 5.263)	1.613 (0.69 -- 3.772)
35-46 years old	1	1	1
47 and older	0.326 (0.075 -- 1.411)	2.494 (0.782 -- 7.948)	1.004 (0.429 -- 2.349)
Race/Ethnicity			
Non-Hispanic Black	1	1	1
Non-Hispanic White	---	---	---
Hispanic	---	---	---
Other	---	---	---

Education			
Less than HS	1	1	1
HS Equivalent	0.755 (0.206 -- 2.759)	1.541 (0.578 -- 4.108)	1.305 (0.621 -- 2.742)
Some college or more	1.926 (0.457 -- 8.113)	0.966 (0.239 -- 3.906)	1.426 (0.568 -- 3.58)
Sexual Orientation			
Heterosexual	1	1	1
Lesbian / Gay / Bi	**9.18** **(2.169 -- 38.847)***	1.953 (0.611 -- 6.243)	**3.324** **(1.466 -- 7.536)***
Trans / Other	---	---	7.034 (0.142 -- 349.159)
--- denotes an unavailable comparison	* denotes significance, α = .05	** denotes significance, α = .01	*** denotes α = .001

Table 7 - Pearson's Correlation Coefficients, correlation between geographic distance (km) and geodesic distance, compared among area HIV burden

All Social Pairs	N	Correlation Coeff	p-value	95% CI (lower)	95% CI (upper)
	17918	0.11715	<.0001	0.102676	0.131559
By Area Burden	N	Correlation Coeff	p-value	95% CI (lower)	95% CI (upper)
High burden	7768	0.12182	<.0001	0.099844	0.143661
Hybrid pairs	5049	0.0445	<.0001	0.016927	0.071987
Intermediate burden	5101	0.1028	<.0001	0.075556	0.129864
By Area Burden	N	Correlation Coeff	p-value	95% CI (lower)	95% CI (upper)
High burden	12817	0.17175	<.0001	0.154887	0.188492
Intermediate burden	10150	0.07021	<.0001	0.050822	0.08954
Excluding Hybrid Pairs					
All Social Pairs	N	Correlation Coeff	p-value	95% CI (lower)	95% CI (upper)
	12869	0.08312	<.0001	0.065938	0.100254

Bibliography

Alvarez-Dardet, C., Marquez, S., & Perea, E. J. (1985). Urban clusters of sexually transmitted diseases in the city of Seville, Spain. *Sexually Transmitted Diseases*, *12*(3), 166–168.

Bernstein, K. T., Curriero, F. C., Jennings, J. M., Olthoff, G., Erbelding, E. J., & Zenilman, J. (2004). Defining core gonorrhea transmission utilizing spatial data. *American Journal of Epidemiology*, *160*(1), 51–58. doi:10.1093/aje/kwh178

Blanchard, J. F., & Aral, S. O. (2010). Emergent properties and structural patterns in sexually transmitted infection and HIV research. *Sexually Transmitted Infections*, *86 Suppl 3*, iii4–9. doi:10.1136/sti.2010.046037

Cooper, H. L. F., Bossak, B., Tempalski, B., Des Jarlais, D. C., & Friedman, S. R. (2009). Geographic approaches to quantifying the risk environment: drug-related law enforcement and access to syringe exchange programmes. *The International Journal on Drug Policy*, *20*(3), 217–226. doi:10.1016/j.drugpo.2008.08.008

Doherty, I. A., Adimora, A. A., Muth, S. Q., Serre, M. L., Leone, P. A., & Miller, W. C. (2011). Comparison of sexual mixing patterns for syphilis in endemic and outbreak settings. *Sexually Transmitted Diseases*, *38*(5), 378–384. doi:10.1097/OLQ.0b013e318203e2ef

Doherty, I. A., Padian, N. S., Marlow, C., & Aral, S. O. (2005). Determinants and consequences of sexual networks as they affect the spread of sexually transmitted infections. *The Journal of Infectious Diseases*, *191 Suppl 1*, S42–54. doi:10.1086/425277

Gesink, D. C., Sullivan, A. B., Norwood, T. A., Serre, M. L., & Miller, W. C. (2013). Does core area theory apply to sexually transmitted diseases in rural environments? *Sexually Transmitted Diseases*, *40*(1), 32–40. doi:10.1097/OLQ.0b013e3182762524

Gindi, R. M., Sifakis, F., Sherman, S. G., Towe, V. L., Flynn, C., & Zenilman, J. M. (2011). The geography of heterosexual partnerships in Baltimore city adults. *Sexually Transmitted Diseases*, *38*(4), 260–266. doi:10.1097/OLQ.0b013e3181f7d7f4

Gorbach, P. M., Murphy, R., Weiss, R. E., Hucks-Ortiz, C., & Shoptaw, S. (2009). Bridging sexual boundaries: men who have sex with men and women in a street-based sample in Los Angeles.

Journal of Urban Health: Bulletin of the New York Academy of Medicine, *86 Suppl 1*, 63–76. doi:10.1007/s11524-009-9370-7

Goswami, N. D., Hecker, E. J., Vickery, C., Ahearn, M. A., Cox, G. M., Holland, D. P., … Stout, J. E. (2012). Geographic information system-based screening for TB, HIV, and syphilis (GIS-THIS): a cross-sectional study. *PloS One*, *7*(10), e46029. doi:10.1371/journal.pone.0046029

Heimer, R., Barbour, R., Shaboltas, A. V., Hoffman, I. F., & Kozlov, A. P. (2008). Spatial distribution of HIV prevalence and incidence among injection drugs users in St Petersburg: implications for HIV transmission. *AIDS (London, England)*, *22*(1), 123–130. doi:10.1097/QAD.0b013e3282f244ef

Hirsch, J. S., Meneses, S., Thompson, B., Negroni, M., Pelcastre, B., & del Rio, C. (2007). The inevitability of infidelity: sexual reputation, social geographies, and marital HIV risk in rural Mexico. *American Journal of Public Health*, *97*(6), 986–996. doi:10.2105/AJPH.2006.088492

Hixson, B. A., Omer, S. B., del Rio, C., & Frew, P. M. (2011). Spatial clustering of HIV prevalence in Atlanta, Georgia and population characteristics associated with case concentrations. *Journal of Urban Health: Bulletin of the New York Academy of Medicine*, *88*(1), 129–141. doi:10.1007/s11524-010-9510-0

Jenkins, W. D. (2009). Development and Evaluation of GIS-Based Chlamydia Trachomatis Intervention Policy in Illinois. *Online Journal of Public Health Informatics*, *1*(1). doi:10.5210/ojphi.v1i1.2771

Jennings, J. M., Curriero, F. C., Celentano, D., & Ellen, J. M. (2005). Geographic identification of high gonorrhea transmission areas in Baltimore, Maryland. *American Journal of Epidemiology*, *161*(1), 73–80. doi:10.1093/aje/kwi012

Jennings, J. M., Taylor, R., Iannacchione, V. G., Rogers, S. M., Chung, S.-E., Huettner, S., & Ellen, J. M. (2010). The available pool of sex partners and risk for a current bacterial sexually transmitted infection. *Annals of Epidemiology*, *20*(7), 532–538. doi:10.1016/j.annepidem.2010.03.016

Kerani, R. P., Handcock, M. S., Handsfield, H. H., & Holmes, K. K. (2005). Comparative geographic concentrations of 4 sexually transmitted infections. *American Journal of Public Health*, *95*(2), 324–330. doi:10.2105/AJPH.2003.029413

Latkin, C., Donnell, D., Celentano, D. D., Aramrattna, A., Liu, T.-Y., Vongchak, T., … Metzger, D. (2009). Relationships between social norms, social network characteristics, and HIV risk

behaviors in Thailand and the United States. *Health Psychology: Official Journal of the Division of Health Psychology, American Psychological Association*, *28*(3), 323–329. doi:10.1037/a0014707

Law, D. C. G., Serre, M. L., Christakos, G., Leone, P. A., & Miller, W. C. (2004). Spatial analysis and mapping of sexually transmitted diseases to optimise intervention and prevention strategies. *Sexually Transmitted Infections*, *80*(4), 294–299. doi:10.1136/sti.2003.006700

Martinez, A. N., Mobley, L. R., Lorvick, J., Novak, S. P., Lopez, A., & Kral, A. H. (2014). Spatial analysis of HIV positive injection drug users in San Francisco, 1987 to 2005. *International Journal of Environmental Research and Public Health*, *11*(4), 3937–3955. doi:10.3390/ijerph110403937

Morris, M., & Kretzschmar, M. (1997). Concurrent partnerships and the spread of HIV. *AIDS (London, England)*, *11*(5), 641–648.

Newman, M. E. J. (2003). Mixing patterns in networks. *Physical Review. E, Statistical, Nonlinear, and Soft Matter Physics*, *67*(2 Pt 2), 026126.

Potterat, J. J., Rothenberg, R. B., & Muth, S. Q. (1999). Network structural dynamics and infectious disease propagation. *International Journal of STD & AIDS*, *10*(3), 182–185.

Potterat, J. J., Rothenberg, R. B., Woodhouse, D. E., Muth, J. B., Pratts, C. I., & Fogle, J. S. (1985). Gonorrhea as a social disease. *Sexually Transmitted Diseases*, *12*(1), 25–32.

Ross, M. W., Nyoni, J., Bowen, A. M., Williams, M. L., & Kashiha, J. J. (2012). Sexual and geographic organisation of men who have sex with men in a large East African city: opportunities for outreach. *BMJ Open*, *2*(6). doi:10.1136/bmjopen-2012-001813

Rothenberg, R. B. (1983). The geography of gonorrhea. Empirical demonstration of core group transmission. *American Journal of Epidemiology*, *117*(6), 688–694.

Rothenberg, R. B., Long, D. M., Sterk, C. E., Pach, A., Potterat, J. J., Muth, S., … Trotter, R. T. (2000). The Atlanta Urban Networks Study: a blueprint for endemic transmission. *AIDS (London, England)*, *14*(14), 2191–2200.

Rothenberg, R. B., & Potterat, J. J. (1988). Temporal and social aspects of gonorrhea transmission: the force of infectivity. *Sexually Transmitted Diseases*, *15*(2), 88–92.

Rothenberg, R. B., Potterat, J. J., Woodhouse, D. E., Muth, S. Q., Darrow, W. W., & Klovdahl, A. S. (1998). Social network dynamics and HIV transmission. *AIDS (London, England)*, *12*(12), 1529–1536.

Rothenberg, R., Baldwin, J., Trotter, R., & Muth, S. (2001). The risk environment for HIV transmission: results from the Atlanta and Flagstaff network studies. *Journal of Urban Health: Bulletin of the New York Academy of Medicine*, *78*(3), 419–432. doi:10.1093/jurban/78.3.419

Rothenberg, R., Muth, S. Q., Malone, S., Potterat, J. J., & Woodhouse, D. E. (2005). Social and geographic distance in HIV risk. *Sexually Transmitted Diseases*, *32*(8), 506–512.

Woodhouse, D. E., Rothenberg, R. B., Potterat, J. J., Darrow, W. W., Muth, S. Q., Klovdahl, A. S., … Muth, J. B. (1994). Mapping a social network of heterosexuals at high risk for HIV infection. *AIDS (London, England)*, *8*(9), 1331–1336.

Wylie, J. L., Cabral, T., & Jolly, A. M. (2005). Identification of networks of sexually transmitted infection: a molecular, geographic, and social network analysis. *The Journal of Infectious Diseases*, *191*(6), 899–906. doi:10.1086/427661

Zenilman, J. M., Bonner, M., Sharp, K. L., Rabb, J. A., & Alexander, E. R. (1988). Penicillinase-producing Neisseria gonorrhoeae in Dade County, Florida: evidence of core-group transmitters and the impact of illicit antibiotics. *Sexually Transmitted Diseases*, *15*(1), 45–50.

Zenilman, J. M., Ellish, N., Fresia, A., & Glass, G. (1999). The geography of sexual partnerships in Baltimore: applications of core theory dynamics using a geographic information system. *Sexually Transmitted Diseases*, *26*(2), 75–81.

Zenilman, J. M., Glass, G., Shields, T., Jenkins, P. R., Gaydos, J. C., & McKee, K. T. (2002). Geographic epidemiology of gonorrhoea and chlamydia on a large military installation: application of a GIS system. *Sexually Transmitted Infections*, *78*(1), 40–44.