5-13-2016

Comparing HIV Risk Among Individuals Living in High and Low Burden Zip Codes in Atlanta, GA Using Different Risk Assessment Models

Joshua Renfroe
Georgia State University

COMPARING HIV RISK AMONG INDIVIDUALS LIVING IN HIGH AND LOW BURDEN ZIP CODES IN ATLANTA USING DIFFERENT RISK ASSESSMENT MODELS

by

JOSHUA RENFROE

B.A., UNIVERSITY of CALIFORNIA, LOS ANGELES

M.S., GEORGIA STATE UNIVERSITY

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

APPROVAL PAGE

COMPARING HIV RISK AMONG INDIVIDUALS LIVING IN HIGH AND LOW BURDEN ZIP CODES IN ATLANTA USING DIFFERENT RISK ASSESSMENT MODELS

by

JOSHUA RENFROE

APPROVED:

_____

Committee Chair

_____

Committee Member

_____

Committee Member

_____

Date

ABSTRACT

JOSHUA RENFROE
Comparing HIV Risk Among Individuals Living in High and Low HIV Burden Zip Codes in Atlanta Using Different Risk Assessment Models
(Under the direction of Richard Rothenberg)

HIV risk assessment models use multiple risk factors to build composite index scores to evaluate population level HIV risk. In this report, four risk assessment models were applied to a dataset with demographic, biological, and behavioral risk factors from 927 individuals in high and low HIV burden zip code groups in metro Atlanta, GA. Predictive ability of the risk assessment models were evaluated by comparing their sensitivity and specificity, area under the ROC curve, and mean score difference between high-burden and low-burden zip code area. The results show that the proportion of study participants who scored high in the risk assessment method are significantly greater in high-HIV burden zip code area than in low-HIV burden zip code area in all four risk assessment models. The Clinical Decision Rule risk-scoring model showed the best predictive ability of HIV risk and Binary Risk Indicator model showed the best predictive ability in predicting the residence zip code area.

INDEX WORDS: HIV risk, composite score, risk index, risk assessment

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.

_____ Signature of Author

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:

Student's Name:_____

Street Address: _____

City, State, and Zip Code: _____

The Chair of the committee for this thesis is:

Professor's Name: _____

Department: _____

College: _____

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 COPYING)

Joshua Renfroe

139 Swanee Lane
jrenfroe5@gsu.edu
Woodstock, GA 30188
818-430-8099

Education

Doctorate of Philosophy (Ph.D.) **Expected 2019**
Georgia State University, Atlanta, GA
Concentration: Applied and Environmental Microbiology

Masters of Science in Biology (M.S.) **August 2015**
Georgia State University, Atlanta, GA
Concentration: Molecular Genetics and Biochemistry

Masters of Public Health (M.P.H.) **Expected May 2016**
Georgia State University, Atlanta, GA
Concentration: Epidemiology/Biostatistics

Post-Baccalaureate Biology **May 2011**
Humboldt State University, Arcata, CA

Bachelor of Arts in Philosophy (B.A.) **March 2009**
University of California Los Angeles, Los Angeles, CA

Research Experience

Graduate Research Assistant **May 2014 - Present**
Georgia State University, Atlanta, GA
Applied and Environmental Microbiology Lab

Teaching Experience

Critical Thinking Through Writing Instructor **January 2014 – Present**
Georgia State University, Atlanta, GA
Molecular Cell Biology Lab

Public Health Experience

Graduate Research Assistant **August 2011 – December 2013**
Georgia State University, School of Public Health,
Center of Excellence on Health Disparities
Atlanta, Georgia, USA

TABLE OF CONTENTS

Chapter I. Introduction

1.1 Background

In the United States, more than 1.2 million people are estimated to be living with HIV and 50,000 new infections occur every year.[1] Identification of persons and networks at highest HIV risk is a priority for resource-limited healthcare programs seeking to prevent transmission. Screening persons at highest HIV risk is important for cost-effective interventions, aimed to avoid excessive testing of low risk individuals and prioritizing targeted prevention and treatment.[2]

Accounting for multiple types of risk exposure may best characterize individual risk. How multiple HIV risk factors co-exist in dense social networks is a productive area of research, which may lead to improved public health interventions, endemic and small network outbreak characterization, and improved understanding of HIV transmission and acquisition dynamics. However, assessment methods to account for multiple HIV risk may vary in usefulness in different populations.

Composite risk scores may be used to characterize multiple risks and are defined as sums of risk factors converging into a single index, representing some risk of disease acquisition by an individual. Composite risk scores may act as a variable for use in clinical decisions, predict health outcomes, estimate groups at highest disease risk for public health interventions, or used to compare populations in epidemiological studies. Composite HIV risk scores allow characterization of at risk populations and have been used in serodiscordant couple studies[3] to characterize transmission and acquisition of HIV.

Metro Atlanta has one of the highest HIV burdens among US cities and its prevalence is disproportionately spread among Atlanta neighborhoods. The Geography Project, a study lead by Dr. Richard Rothenberg in the Center of Excellence for Health Disparities in the School of Public Health at Georgia State University, surveyed social networks and tested HIV outcomes. Between 2007 and 2010, 927 individuals from five high and five lower HIV-burdened Atlanta zip codes were screened for STIs/HIV and collected behavioral and demographic characteristics through a survey. The interview sought to capture detailed HIV risk factors beyond traditional categorizations such as MSM and IDU and determine individual HIV risk and social network relationships.

1.2. Purpose of the Study

Evaluating different risk assessment models using the Geography Project dataset may help develop an appropriate risk assessment tool for studying networks and risk prediction in populations similar to metro Atlanta. By identifying factors associated with HIV in study participants and HIV risk factor literature, risk score indices will be developed using four risk assessment models: Simple Unit-weighted (Burgess), Subject Matter Expert-weighted, Clinical Decision Rule, and Binary Risk Indicator.

Each risk assessment model will be evaluated by predicting two associations: 1) Individual HIV status and 2) Resident of high HIV-burdened zip codes. It is expected that high HIV-burdened zip code groups would have a larger proportion of high-risk individuals. Risk assessment models will be compared by three different diagnostic evaluation methods to determine for best predictive ability for both outcomes.

Chapter II. Literature Review

This review examines the current literature on individual risk factors (biological, behavioral, and demographic) associated with HIV acquisition and discusses the variety of methods used to calculate risk scores.

2.1 Biological Risk Factors

Biological risk factors for HIV acquisition include the presence of other STI infections, lack of circumcision, HIV viral load, and sexual partner stage of infection. Although not explored in this review, the biological risk factors with the greatest potential for HIV transmission are blood transfusions; having a sexual partner in the primary stage of HIV infection; and having a sexual partner with a CD4 count below 200.[4,5]

Herpes Simplex Virus 2 (HSV-2). Presence of HSV-2 is associated with HIV acquisition and transmission. A meta-analysis of nine cohort and case-control studies showed a preceding HSV-2 infection more than doubled the risk of HIV acquisition RR 2.1 (95% CI, 1.4 – 3.2).[6] HSV-2 may increase the risk of HIV acquisition through the presence of genital ulcers, which are accompanied by an increase innate immune system response and concentration of macrophages. The immune response fosters herpetic lesions, creating an influx of lymphocytes, which in the presence of HIV virions, increase the availability of HIV epitopes and uptake by immune cells.[7] In addition to increasing susceptibility, HIV-positive individuals who are co-infected with HSV-2 may have high

levels of HIV virions present in herpetic lesions during outbreaks, increasing HIV infectiousness.[8] Most individuals infected with HSV are asymptomatic, but can still infect sexual partners.[9] Serological HSV tests detect antibodies specific for HSV G-1 and G-2 glycoproteins, allowing for distinction between HSV1 and HSV2 infections.[10]

Hepatitis C. Hepatitis C (HCV) shares similar transmission routes with HIV. HCV is transmitted through unprotected sexual contact, injection drug use, and vertical transmission at birth.[11] Of all HIV-positive individuals in the US, 33% are co-infected with HCV.[12] Recombinant immunoblot assays (RIBA) can detect the presence of antibodies specific for HCV antigens. Presence of HCV antibodies does not distinguish between resolved HCV and current HCV infection and is not reliable in detecting an infection occurring within the previous three months. RT-PCR can detect and quantify viral RNA levels in the blood. A combination of antibody and RT-PCR testing can give an accurate HCV profile for individuals.[13]

Bacterial STIs. Bacterial pathogens such as *N. gonorrhoeae*, *C. trachomatis*, and *T. pallidum* increase the number of innate immune cells in the host genital tract.[8] HIV virions were more likely to be detected in the presence of mucosal or cervical discharge in a bacterial STI-infected host.[14] In HIV-positive women, the presence of inflammatory bacterial infections increased HIV shedding compared to HIV-positive women with no bacterial co-infection.[15] A male having an HIV-positive female sexual partner who has a bacterial co-infection may therefore be at greater risk of HIV infection.

Despite clinical evidence showing conditions consistent for increased HIV infectiousness, epidemiological studies have not shown a clear causal relationship

between HIV infection and presence of bacterial STIs.[8] Mayer and Venkatesh[8] suggest that many epidemiological studies have looked at STI and HIV transmission in African populations with high HSV-2 prevalence, leading to confounding when assessing bacterial STI and HIV co-infections. The main risk factor for HIV transmission with STIs may be overall genital tract inflammation, which may be residually present in either treated bacterial STI infections or HSV-2 infections. Although studies have not confirmed bacterial infections to have biological synergy for HIV transmission, the presence of bacterial STIs may be a useful indication for high-risk sexual activity and may be treated as a surrogate for behavioral risk in composite risk score calculations.

2.2 Demographic Risk Factors

Demographic risk factors may be indicators for HIV infection. Risk factors in this category include age, race, gender, socioeconomic status, education level, and sexual orientation.

Age. In 2006, individuals aged 29 and below represented the highest risk for HIV acquisition.[16] 38% of new infections occurred in this age group, followed by the 30 – 39 age group (30%), 40 – 49 age group (22%), and 50 – 99 age group (9%). In 2010, in Fulton County, GA, HIV prevalence for ages 13-24 was 0.27%, 25-34 1.00%, 35-44 1.75%, 45-54 2.70%, and 55+ was 0.97%.[17]

Race. African Americans are at highest risk for acquiring HIV in the US. HIV in the US disproportionately affects African Americans- 41% of people living with HIV in 2010 were African American. African Americans accounted for 44% of new HIV infections in

2010, despite representing only 12% of the US population, with an HIV infection risk that is 7.9 times greater than Caucasians.[18]

Sex/Gender. Those that identify as transgendered represent higher risk for HIV acquisition. According to a meta-analysis by Herbst, 27.7% (95% [CI], 24.8-30.6%) of male to female transgendered individuals tested positive for HIV.[21] Transgendered individuals are at higher HIV risk due to behavioral, social, and economic risk factors such as high rate of unprotected receptive anal intercourse with sex work clients (38.5%), increased prevalence of mental health disorders and lack of transgender-sensitive mental health services, increased substance abuse, social isolation, economic marginalization, and needle-sharing behaviors for purposes of hormone injections.[19]

In 2010, the rate of HIV infection among males was 4.2 times greater than females.[1] Many of the new HIV infections in males are dependent upon high-risk sexual behaviors involving other males. For females, the main transmission category is heterosexual contact.[1]

2.3 Behavioral Risk Factors

Behavioral risk factors that increase risk fall into the categories of sexual intercourse and substance use. Sexual intercourse can be further divided into subcategories: number of sexual partners, type of commercial sex work, use of condoms, and type of sexual intercourse (receptive/insertive anal, and vaginal). Substance use represents HIV risk and can be divided into several categories: alcohol, marijuana, crack, intravenous drug use (IDU), needle sharing, and level of substance dependence.

Anal intercourse. Unprotected anal intercourse is a high-risk sexual behavior for HIV transmission. A meta-analysis investigating heterosexual and homosexual sero-discordant couple transmission risk for HIV documented estimates of per-act and per-partner for several categories of sexual behavior involving anal intercourse. Chance of HIV transmission from unprotected receptive anal intercourse (URAI) was estimated to be 40.4% (95% CI 6.0-74.9) per partner and 1.4% (95% CI 0.2-2.5) per act. Risk of HIV transmission was lower in unprotected insertive anal intercourse (UIAI), with a per-partner estimate at 21.7% (95% CI 0.2-43.3) and 0.11% (95% CI 4-28) per act. Combined UIAI and URAI per-partner risk was estimated at 39.9% (95% CI 22.5-57.4).[20]

Vaginal intercourse. Unprotected vaginal sex represents a risk for HIV transmission. Receptive vaginal acquisition of HIV (male to female) was estimated to carry a risk of 0.08% (95% CI 0.06-0.11) per act. The HIV risk of insertive vaginal transmission (female to male) was estimated to be 0.04% (95% CI 0.01-0.14) per act.[21]

Alcohol. Alcohol affects the brain of an individual and lowers inhibitions. Lowered inhibitions allow a person to engage in more high-risk behavior. High blood alcohol concentrations have been associated with reduced intention of condom use.[22] Among injection drug users, use of alcohol has been associated with high-risk behaviors such as sharing needles and injection equipment.[23] It is difficult to isolate risk of HIV infection related to alcohol consumption from other behavioral or demographic risk factors, but a

meta-analysis estimated a 1.70-fold increase (95% CI 1.42 – 1.72) in testing positive for HIV in alcohol drinkers among 20 studies based in Africa.[24]

Crack cocaine. Crack cocaine use puts a person at higher risk of HIV infection through impaired judgment and exposure to high-risk social networks. Since crack is addictive, individuals are exploited and are inclined to exchange sex for money or drugs. Injection drug users who also smoke crack are more likely to be infected with HIV.[25] Use of crack cocaine among 18-29 year old non-injection drug users shows a 2.1-fold greater risk (CI: 99% 1.2-3.8) of HIV infection than non-injecting non-crack smokers.[26]

Injection Drug Use. Sharing needles between individuals injecting drugs represents an efficient way for HIV transmission. HIV transmission through sharing needles is dependent upon the viral load present in the infected individual and how much blood is present in the shared needle. Transmission can also occur through sharing needle-cleaning equipment such as filters or water. Chance of transmission per needle-sharing activity is estimated to be 0.33%.[27] Of all injection drug users screened in 2009 in the US, 9% tested positive for HIV.[28]

2.4 Risk assessment models

Selection of risk variables. The selection of the predictive variables depends upon the method of analysis being used and type of variables present in the survey. If epidemiological literature exists for a disease, then selection of risk factors may rely on accepted associations.

Unit weights. Unit weights, also known as raw score weights, standardized scores or unweighted scores, are the simplest method used to develop composite risk scores. This method assigns a single unit to an individual for each risk factor present and zero units if not present.

$Y = x_1 + x_2 + x_3 + x_4 + \ldots + x_i$

Y= composite risk

x_i = individual risk contribution

The earliest use of the unit-weight method was by E.W. Burgess in 1928 in his assessment of risk of re-incarceration of paroled prisoners.[29] Although differential weights are argued to have more validity than unit weights because each risk is individually evaluated to determine relative contribution of risk, unit weights have been demonstrated to be as useful as other weighting methods in many circumstances. In a literature review by Bobko, Roth, and Buster, a meta-analysis demonstrated the predictive validity of scores created by unit weights and were compared to scores created by differential weights.[30] However, as the number of risk variables increases, the less of an effect each extraneous variable has on the outcome. This method may best be used when large numbers of predictor variables are used to create a risk score, the regression model fits the data poorly, there is low observation to predictor variable ratio, highly correlated predictor variables, or when measurement error is present in predictor variables.[31]

Differential weights. Differential weights may be beneficial to HIV risk score calculations because some risks are more likely to result in HIV infection. Weights of the independent variable may be determined by their relative contribution through different methods. In one method of obtaining differential weights, subject matter experts (SMEs) grade the contribution of risk factors by assigning a numerical value to each risk with the total equaling 100.[30] The product of the weights and the presence of the risk factor are summed to give the composite risk score.

$Y = x_1(a) + x_2(b) + x_3(c) + x_4(d) + \ldots + x_i$

Y= composite risk

x_1 = individual risk contribution

a = differential weight

Binary Risk Indicator model. Binary Risk Indicator model was used to assess risk in Rothenberg, Baldwin, Trotter, and Muth study [32] evaluating risk environments and networks in Flagstaff and Atlanta. A risk score is assigned based on the following risk categories occupied by the individual: Low level, medium level, high level, and very high-level risk. Risk factors may be assigned to a category based on different methods, but attempt to put highest risk activities such as needle sharing and unprotected anal intercourse at very high risk and activities such as protected vaginal sex or past infection with gonorrhea or chlamydia in low risk category.

Clinical decision rules or model. Clinical Decision Rules (CDR) are evidence-based assessment tools used in clinical settings to quantify patient data leading to clinical decision-making.[33] CDRs can be applied to assessing risk for prevention interventions by stratifying individuals into risk categories. Menza TW, et al[34] and Kahle EM, et al[35] used this methodology to assess risk in men who have sex with men and heterosexual serodiscordant HIV couples. The predictors or risk variables were derived from epidemiological studies and using Cox proportional hazard model their associations with HIV infection risk were assessed. The coefficient of each risk factor was used to develop a risk score. In the Kahle EM, et al study, coefficients of each factor were divided by the lowest coefficient among all risk variables and rounded to the integer to get the value for that risk factor. The sum of risk factor values give the composite score of individual. This method develops the risk score from data present in the dataset and if used on the same dataset, is considered internally validated.

2.5 Diagnostic tool evaluation

Evaluation of diagnostic tools is used to compare diagnostic methods to determine usefulness of a test. McNemar chi-square test has been used to evaluate new diagnostic tools by measuring discordance between sensitivities and specificities between a new test and a reference test.[36] Receiver-Operating Characteristics (ROC) has been used to evaluate diagnostic tests to measure diagnostic accuracy over cutoff points by comparing Area Under the Curve (AUC) for each diagnostic tool.

Chapter III. Methods

This study is a secondary analysis of the data collected from the Geography Project to compare risk scores among the study population living in low and high HIV burden zip codes using different risk scoring models and to assess the model validity using SPSS Version 2.0.0. Some figures were generated by SAS version 9.3.

3.2 Study setting

Geography Project

The Geography Project was an observational study arranged by Dr. Richard Rothenberg at the Center of Excellence for Health Disparities at Georgia State University's School of Public Health. The survey collected network, geographic, and risk data for 927 individuals in the metro Atlanta area from 2007 to 2010. The purpose of this study was to understand interactions between compound risk, social environment, and geographic proximity for several STIs in high-risk social networks in higher-burden (30318, 30308, 30314, 30310, 30315) vs. lower-burden (30311, 30344, 30331, 30337, 30349) metro Atlanta zip codes (Fig. 1). Chain link sampling method was used to recruit participants from high-risk social networks in each zip code group. Epidemiologic, social and behavioral, geographic, and network variables were collected through surveys.

Eight STIs (HIV, HSV2, HCV, Syphilis, Gonorrhea, Chlamydia, and Trichomoniases) were tested through serological and urine samples from each participant. An Orasure ELISA HIV test was given at the time of interview for

preliminary HIV test results. A blood sample was taken from each participant and delivered to the CDC where a Western blot was used to confirm the presence of HIV, RIBA to detect presence of resolved or current HCV, and RT-PCR to quantify HCV RNA if RIBA tested positive. RPR and IgG antibody tests were used to detect presence of syphilis. Current chlamydia and gonorrhea infections were tested by urine sample.

Out of 927 participants, 185 were followed up 12 months after their initial interview and were surveyed and tested again. Out 185 of these participants, 12 were interviewed a third time. Of the 927 participants in the first iteration of interviews, 49 tested positive for HIV infection.

High-Burden vs. Low-Burden Atlanta Zip Codes in the Geography Project

10 Atlanta zip codes are organized by higher and lower burden of HIV infection. High-burden zip codes have high prevalence of HIV infection and low-burden zip codes have lower HIV prevalence.

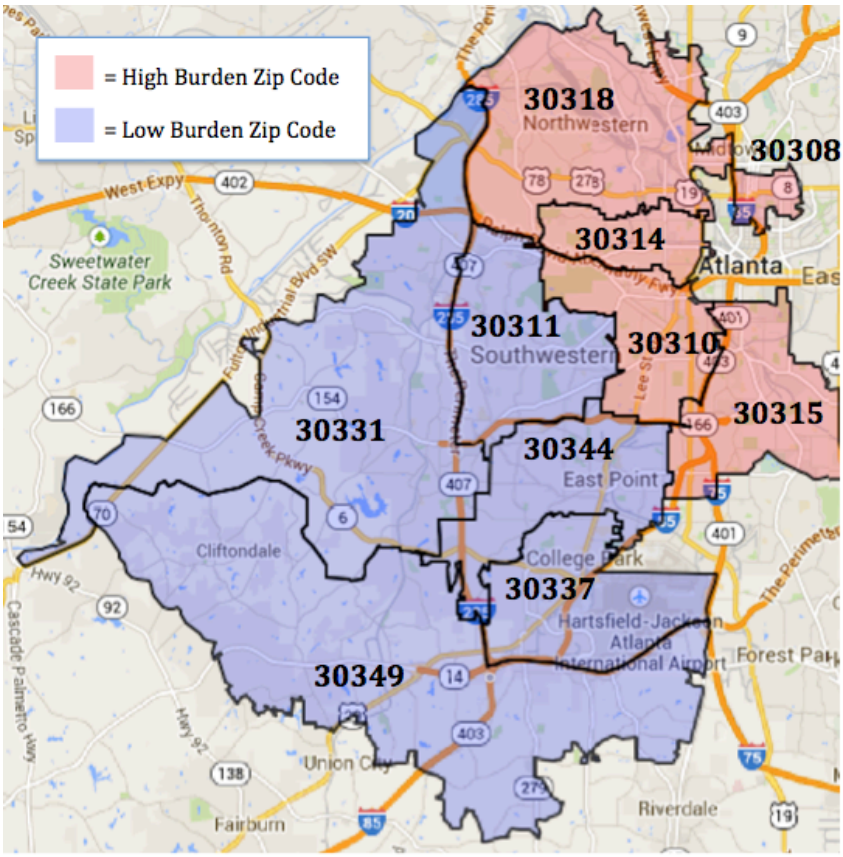

Figure 1. Atlanta zip codes screened in the Geography Project. AIDSVu HIV mapping tool used to show five high-burdened (30308, 30318, 30314, 30310, 30315) and five low-burdened metro Atlanta zip codes (30311, 30344, 30337, 30349, 30331).

3.3 Study population

From the Geography Project data, second and third iterations were removed, resulting in 927 participants from the first iteration. First iteration data was treated as a cross-sectional study and is shown in Table 1 as descriptive statistics. Twelve records missing conclusive HIV status were excluded from the risk scoring model and 33 records that were missing zip code data were removed when comparing risk scores between high- and low-burden zip code groups.

	HIV Status	Positive (n=49)		Negative (n=864)		Missing/Indeterminate (n=14)		Total (n=927)	
		No.	%	No.	%	No.	%	No.	%
Age	<20	0	0.00%	59	6.80%	1	8.30%	60	6.5
	20-29	12	24.50%	308	35.60%	6	42.90%	326	35.2
	30-39	14	28.60%	151	17.50%	2	14.30%	167	18
	40-49	17	34.70%	213	24.70%	3	21.40%	233	25.1
	50-59	6	12.20%	114	13.20%	1	7.10%	121	13.1
	60-80	0	0.00%	19	2.20%	1	7.10%	20	2.2
Race/ethnicity	Black (African American)	49	100.00%	840	97.20%	14	100.00%	903	97.4
Gender	Male	18	36.70%	454	52.50%	7	50.00%	479	51.7
	Female	22	44.90%	404	46.80%	7	50.00%	433	46.7
	Transgender	9	18.40%	6	0.70%	0	0.00%	15	1.6
Reside in high-burden vs. Low-burden	Low-burden zipcode	8	16.30%	408	49.00%	6	46.20%	422	45.5
	High-burden zipcode	41	83.70%	424	51.00%	7	53.80%	472	50.9
STDs	Herpes simplex 2 virus	37	75.50%	375	43.4%	0	0.00%	414	44.7
	Hepatitis C RTBA/RNA	4	8.20%	64	7.40%	0	0.00%	68	7.3
Previous STDs	Gonorrhea	2	4.10%	27	3.10%	0	0.00%	29	3.1
	Syphilis	8	16.30%	51	5.90%	0	0.00%	59	6.4
	Chlamydia	4	8.20%	67	7.80%	0	0.00%	71	7.7
	Trichmonas	9	18.40%	100	11.60%	1	8.30%	110	11.9
Sexual behavior	Had insertive anal sex a male in last 6 months	2	4.10%	7	0.80%	0	0.00%	9	1
	Had receptive anal sex in last 6 months	5	10.20%	10	1.20%	0	0.00%	15	1.6
	Had vaginal sex in last 6 months	29	59.20%	778	90.00%	11	78.60%	818	88.2
	Ever used crack	37	75.50%	387	44.80%	6	42.90%	430	46.4
	Sex partners who smoked crack rock	12	24.50%	125	14.50%	0	0.00%	137	14.8
	Ever injected any drug	11	22.40%	83	9.60%	1	7.10%	95	10.2
	Any sex partners ever inject drugs	3	6.10%	27	3.10%	0	0.00%	30	3.2
	given woman drugs to have sex in last 6 months	6	12.20%	95	11.00%	0	0.00%	101	10.9
	Woman paid respondent drugs for sex	2	4.10%	56	6.50%	0	0.00%	58	6.3
	Given man drugs to have sex in last 6 months	2	4.10%	13	1.50%	0	0.00%	15	1.6
	Man paid respondent drugs for sex in last 6 months	12	24.50%	63	7.30%	1	7.10%	76	8.2
	Paid man for sex	0	0.00%	11	1.30%	0	0.00%	11	1.2
	Sex worker	5	10.20%	23	2.70%	0	0.00%	28	3

Table. 1. Descriptive statistics of risk factors used in this study from Geography Project population. Risk factor frequency sorted by HIV status.

3.4 Risk assessment methods

Between May 5th 2013 and June 1st 2014, a literature search collected peer-reviewed publications of research related to HIV risk in Pubmed and the Cochrane Library. Priority was given to systematic reviews and meta-analyses evaluating HIV risk. Literature search fell under two main categories: per-act HIV transmission risk and behavioral or demographic risk factors for HIV. Only studies including human participants and published in the English language were considered. Preference was given to studies with participants in developed countries.

From several risk assessment methodologies present in the literature, four methods were selected and used to evaluate HIV risk in the Geography Project. Each risk score model was calculated in SPSS Version 2.0.0.

3.4.1. Burgess Unit-weighted method. Risk scores were developed for each individual by assigning a value of one if the known risk factor associated with HIV was present, and zero if absent. Units were summed to give a composite risk score for each individual.

3.4.2. Subject Matter Expert (SME) differential-weighted method. Differential weights were obtained by consulting two SMEs and instructing each to distribute 100 points among a list of 20 identified risk factors (Table 2) determined by a search of published literature. Factors considered high-risk by SMEs were given a greater proportion of 100 points. The two SME scores were combined and averaged to give a representative weight for each factor. Each unit was multiplied by the weight to give a weighted product and summed to give a composite score for each individual.

	Risk Factor	Score
1	Black (African American)	2
2	Herpes simplex 2 virus	5
3	Hepatitis C RIBA/RNA	5
4	Gonorrhea	7
5	Syphilis	7
6	Chlamydia	7
7	Trichmonas	7
8	had insertive anal sex a male, 6 mos	6
9	had receptive anal sex, 6 mos	6
10	had vaginal sex, 6 mos	5
11	ever used crack	5
12	sex partners who smoked crack rock	4
13	ever injected any drug	7
14	any sex partners ever inject drugs?	5
15	given woman drugs to have sex, 6 mos	3
16	Woman paid respondent drugs for sex, 6 mos	3
17	given man drugs to have sex, 6 mos	3
18	man paid respondent drugs for sex, 6 mos	3
19	paid man for sex, 6 mos	3
20	Sex worker	7
	Total	100

Table 2. SME differential weights. Assigned differential weights for each risk score for use in the SME-weighted method. SMEs were given 100 points to distribute among 20 risk factors. Final weight is the result of an average between the two SMEs.

3.4.3. Clinical Decision Rule. A model similar to that used by Kahle EM, et al, except multivariate regression model was used in place of Cox proportional hazard model because the data used to derive the risk variable was from a cross-sectional study. From the dataset, risk factors associated with positive HIV status in univariate analysis were evaluated in multivariate analysis (Table 3). Risk factors identified as having negative or no association were assigned zero. Risk factors that showed association in multivariate model were assigned a risk score by dividing the lowest coefficient into all other

coefficients and rounded to the nearest integer. The sum of these values was used to determine the composite risk score for each individual.

3.4.4. Binary Risk Indicator

Risk variables were categorized into "Very High", "High", "Medium", and "Low" risk (Table 4). Individuals possessing one or more risk in each category were assigned units of 8 for "very high risk", 4 for "high risk", 2 for "medium risk", and 1 for "low risk". Units were summed for all categories to give a composite score for each individual.

Very High – 8 points
had receptive anal sex, 6 mos
Paid man for sex, 6 mos
Woman paid resp. drugs for sex, 6 mos
man paid resp. drugs for sex, 6 mos
High – 4 points
paid man for sex, 6 mos
had insertive anal sex with male, 6 mos
man paid resp. drugs for sex, 6 mos
given man drugs to have sex, 6 mos
Paid woman for sex, 6 mos
Medium – 2 points
HSV-2 Infection
Had vaginal sex (receptive or insertive)
Had insertive anal intercou with female
Sex partner IDU
Non-injecting crack use
Low – 1 point
Current gonor, chlamy, HCV infection
Tried crack at any time

Table 4. Binary Risk Indicator risk factor categorization. Risk factors sorted into Binary Risk Indicator group categories.

	Univariate analysis				Multivariate analysis					Risk score assigned
	OR	95% CI		p-value	OR	95% CI		coefficient	p-value	
Black (African American)	1.058	1.042	1.075	>0.05						0
Herpes simplex 2 virus	4.021	2.068	7.817	<0.0001	3.195	1.584	6.443	1.162	< 0.01	**3**
Hepatitis C RTBA/RNA	1.111	0.387	3.187	>0.05						0
Gonorrhea	1.319	0.305	5.715	>0.05						0
Syphilis	3.11	1.386	6.983	>0.05						0
Chlamydia	1.057	0.369	3.029	>0.05						0
Trichmonas	1.719	0.81	3.648	>0.05						0
Had insertive anal sex a male, 6 mos	5.21	1.053	25.769	>0.05						0
Had receptive anal sex, 6 mos	9.705	3.181	29.608	<0.01	2.369	0.623	9.007	.863	>0.05	**2**
Had vaginal sex, 6 mos	0.16	0.087	0.295	<0.0001	0.209	0.106	0.412	-1.564	<0.001	0
Ever used crack	3.8	1.955	7.388	<0.0001	1.993	0.967	4.106	.690	>0.05	**2**
Sex partners who smoked crack	1.917	0.973	3.778	>0.05						0
Ever injected any drug	1.129	0.468	2.724	>0.05						0
Any sex partners ever inject drugs	2.022	0.591	6.911	>0.05						0
Given woman drugs to have sex, 6 mos	1.129	0.468	2.724	>0.05						0
Woman paid respondent drugs for sex, 6 mos	0.614	0.145	2.593	>0.05						0
Given man drugs to have sex, 6 mos	2.786	0.611	12.703	>0.05						0
Man paid respondent drugs for sex, 6 mos	4.124	2.048	8.302	<0.0001	2.326	0.985	5.493	0.844	>0.05	**2**
Paid man for sex, 6 mos	0.946	0.931	0.961	>0.05						0
Sex worker	4.155	1.508	11.448	<0.05	1.478	0.431	5.07	0.390	>0.05	**1**

Table 3. Clinical Decision Rule risk factor identification. Risk factors were assessed in univariate analysis and evaluated in multivariate analysis if found significantly associated with having HIV. Lowest coefficient generated from multivariate analysis was divided into all other coefficients of risk factors to assign risk score.

3.5. Analysis of risk assessment models to predict HIV-positive status. Each risk assessment method was used to test how well a high score predicted HIV-positive individuals. Sensitivity and specificity was determined for each cut off point in SPSS to generate an ROC curve. Youden-index determined the optimal specificity and sensitivity cutoff point, allowing categorization of "high-risk" and "low-risk" individuals. Pearson chi-square test was used to test the significance of difference in proportion of high-risk scored individuals in high and low-burdened zip code groups.

3.6 Analysis of risk score predicting high-risk individuals in high-burdened zip code groups. Each risk assessment method was used predict high-risk individuals living in high-burden zip group. Sensitivity and specificity was determined for each cut off point in SPSS.

3.7. Evaluation of risk assessment methods for HIV-positive status.

Predictive ability of each risk assessment model using the Geography Project dataset was evaluated using the area under the ROC curve. Area under the ROC curve was calculated by SPSS and compared and evaluated qualitatively by shape.

3.8. Evaluation of risk assessment methods for predicting high-risk individuals in high-burdened-zip groups. Predictive ability of each risk assessment model using the Geography Project dataset was evaluated using three methods: 1) Sensitivity of each risk assessment method was compared to sensitivity of Binary Risk Indicator by McNemar's chi-square test. 2) For each risk assessment method, the difference of means of

distributed risk scores among high and low burdened zip codes were standardized and compared. 3) Area under the ROC curve was determined for each risk assessment model and compared.

Chapter IV. Results

4.1. Characteristics of study population

Of 927 individuals that participated in the Geography Project, almost all (97%) were African American and most were between 20 to 29 and 40 to 49 years of age. Forty six percent were female and 1.6 % identified as transgendered. Half lived in high HIV burden zip code, Forty four percent were infected with HSV-2 virus, and 7% were infected with HCV. Only 3% reported having a previous case of gonorrhea. Other reported sexual transmitted diseases were syphilis (6.4%), chlamydia (7.7%) and trichomonas (11.9%). Sexual and drug use behaviors are presented in Table 1. Most reported (88%) having vaginal sex within the last 6 months. Nearly half used crack at least once in their life and 3% of the participants reported they engaged in sex work.

4.2. Components of composite risk

Twenty risk factors were included in Burgess, SME and CDR risk assessment methods and 16 were used in the Binary Risk Indicator method. Frequency and percentage of participants having each risk factor are presented in Table 5. Participants with no HIV status data or inconclusive HIV results were excluded

	Risk Factor	N	%
1	Black (African American)	889	97.4
2	Herpes simplex 2 virus	412	45.1
3	Hepatitis C RTBI/RNA	68	7.4
4	Gonorrhea	29	3.2
5	Syphilis	59	6.5
6	Chlamydia	71	7.8
7	Trichmonas	109	11.9
8	had insertive anal sex with a male, 6 mos	9	1
9	had receptive anal sex, 6 mos	15	1.6
10	had vaginal sex, 6 mos	807	88.4
11	ever used crack	424	46.4
12	sex partners who smoked crack rock	137	15
13	ever injected any drug	94	10.3
14	any sex partners ever inject drugs	30	3.3
15	given woman drugs to have sex, 6 mos	101	11.1
16	Woman paid respondent drugs for sex, 6 mos	58	6.4
17	given man drugs to have sex, 6 mos	15	1.6
18	man paid respondent drugs for sex, 6 mos	75	8.2
19	paid man for sex, 6 mos	11	1.2
20	Sex worker	28	3.1

Table 5. Frequency distribution of risk factors used in the Burgess Method (Total N = 894)

4.3. High-burden vs. Low-burden Zip Code Risk Score Comparison Using Different Risk Assessment Models

4.3.1 Burgess Unit-Weighted Method

The distribution of risk scores using Burgess Unit-weighted methods among participants is shown in Table 6. Half of participants had a risk score of more than three. ROC and Youden-index show the optimal cutoff point for both HIV-status and high-burden zip code group prediction to be 4, categorizing “High-HIV risk” to 4 and above for both outcomes. Area under the ROC for HIV-status prediction was 0.70.

The Burgess Method showed an increase in risk scores among HIV-positive participants as well as among high-burden zip code groups. Mean (SD) risk score of HIV positive and negative participants is 4.8 (1.6) and 3.7 (1.7) respectively, and median is 5 and 3 respectively (Fig. 2(a)). The mean difference is 1.17. Mean risk score of the high-burden zip codes is significantly higher than that of low-burden zip code. High burden population had a mean (SD) risk score of 4.25 (1.8) and median of 4. Low-burden zip group had a mean (SD) of 3.18 (1.4) with a median of 3 (Fig. 2(b)) with a mean difference between the low and high burden zip code group at 1.1 units. Area under the ROC for prediction of HIV status by using this risk scoring methods is 0.7 and prediction of high-risk participants living in high-risk zip code group was 0.68. The proportion of participants with a risk score of four or more who are living in high-burden zip code are significantly higher than the proportion of those living in low burden zip code (p-value <0.0001) (Table 7).

Risk Score	N	%
1	17	1.9
2	252	27.6
3	194	21.2
4	169	18.5
5	127	13.9
6	87	9.5
7	42	4.6
8	16	1.8
9	7	0.8
10	2	0.2
Total	913	100

Table 6. Distribution of score frequency using Burgess Unit Weighted Method in total population

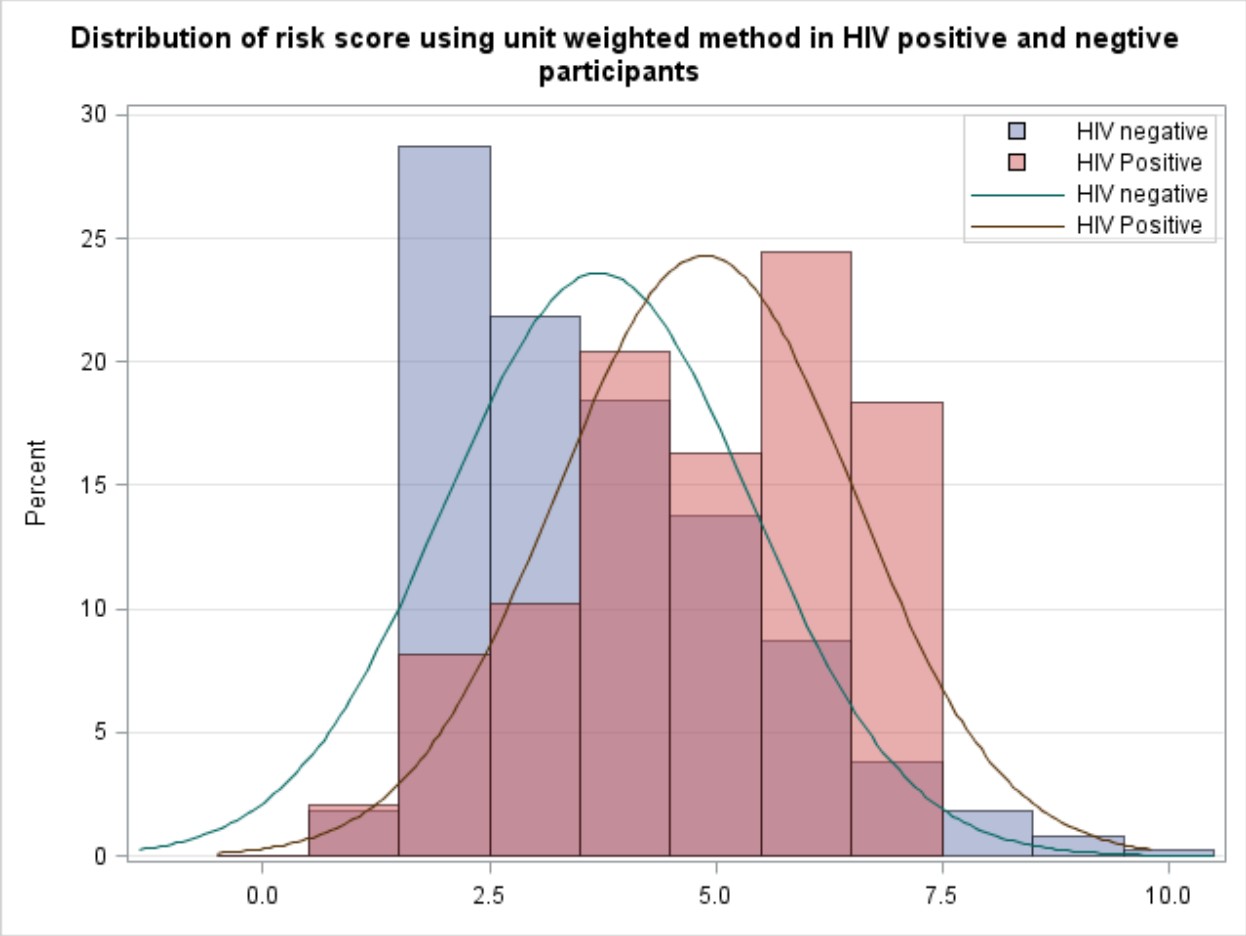

Fig. 2(a). Distribution of individual risk scores assigned using the Burgess Unit-Weighted method among HIV-positive and negative participants

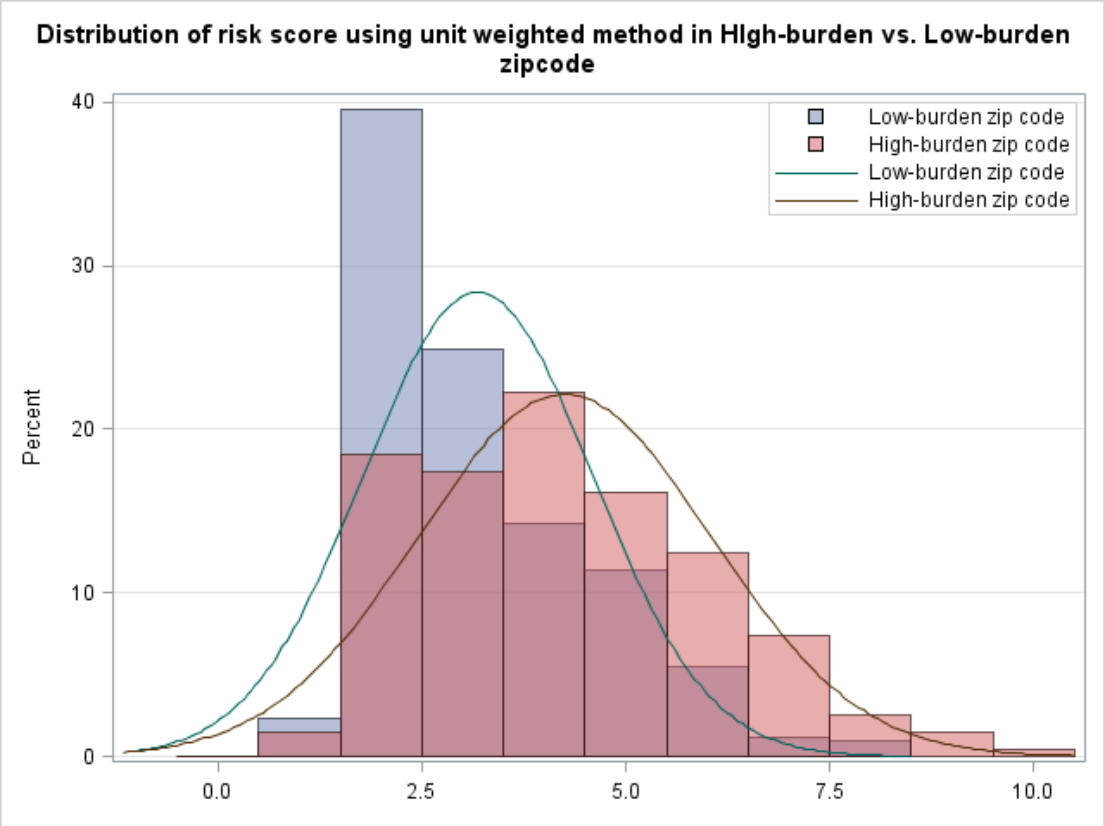

Fig. 2(b). Distribution of individual risk scores assigned using the Burgess Unit-Weighted method by low and high zip code groups.

	N	%
Lower Risk Area	140	33.20%
Higher Risk Area	296	62.70%
Areas combined	436	48.80%

z = -8.8202
p <0.0001

Table 7. Comparison of compound risk in low and high burden zip codes using Burgess Unit-weighted method

4.3.2. Subject matter expert (SME) differential-weighted method

The individual scores derived from SME are presented in Table 2. The distribution of risk scores using SME-weighted methods among participants is shown in Table 8. Half of the participants had risk scores more than 15. ROC and Youden-index show the optimal cutoff point for both HIV-status and high-burden zip code group prediction to be 15,

categorizing “High-HIV risk” to 16 and above for both outcomes. Area under the ROC for HIV-status prediction was 0.71 (Fig. 2).

Using this risk scoring method, there was an increase in risk scores among HIV-positive participants as well as among high-burden zip code groups. Mean (SD) risk score of HIV positive and negative participants is 22 (8.4) and 15.8 (8.6) respectively, and median is 23 and 14 respectively (Fig. 3(a)). The mean difference is 6.23. . Mean risk score of the high-burden zip codes is also significantly higher than that of low-burden zip code. High-burdened population had a mean (SD) risk score of 18.48 (9.1) and median of 19. Low-burden had a 13.14 (7.2) with a median of 12 (Fig. 3). The mean difference between low and high burden zip group was 5.3.

Area under the ROC for prediction of HIV status is 0.71 and prediction for high-risk participants living in high-risk zip code group was 0.67. The proportion of participants with a risk score of 16 or more living in high-burden zip code are significantly higher than the those living in low burden zip code (p-value <0.0001) (Table 9).

	Score	N	%
1-5	1	17	1.9
6-10	2	256	28
11-15	3	210	23
16-20	4	174	19.1
21-25	5	117	12.8
26-30	6	80	8.8
31-35	7	32	3.5
36-40	8	17	1.9
41-45	9	9	1
46-50	10	1	0.1
	Total	913	100

Table 8. Distribution of risk scores assigned using Subject Matter Expert-weighted method

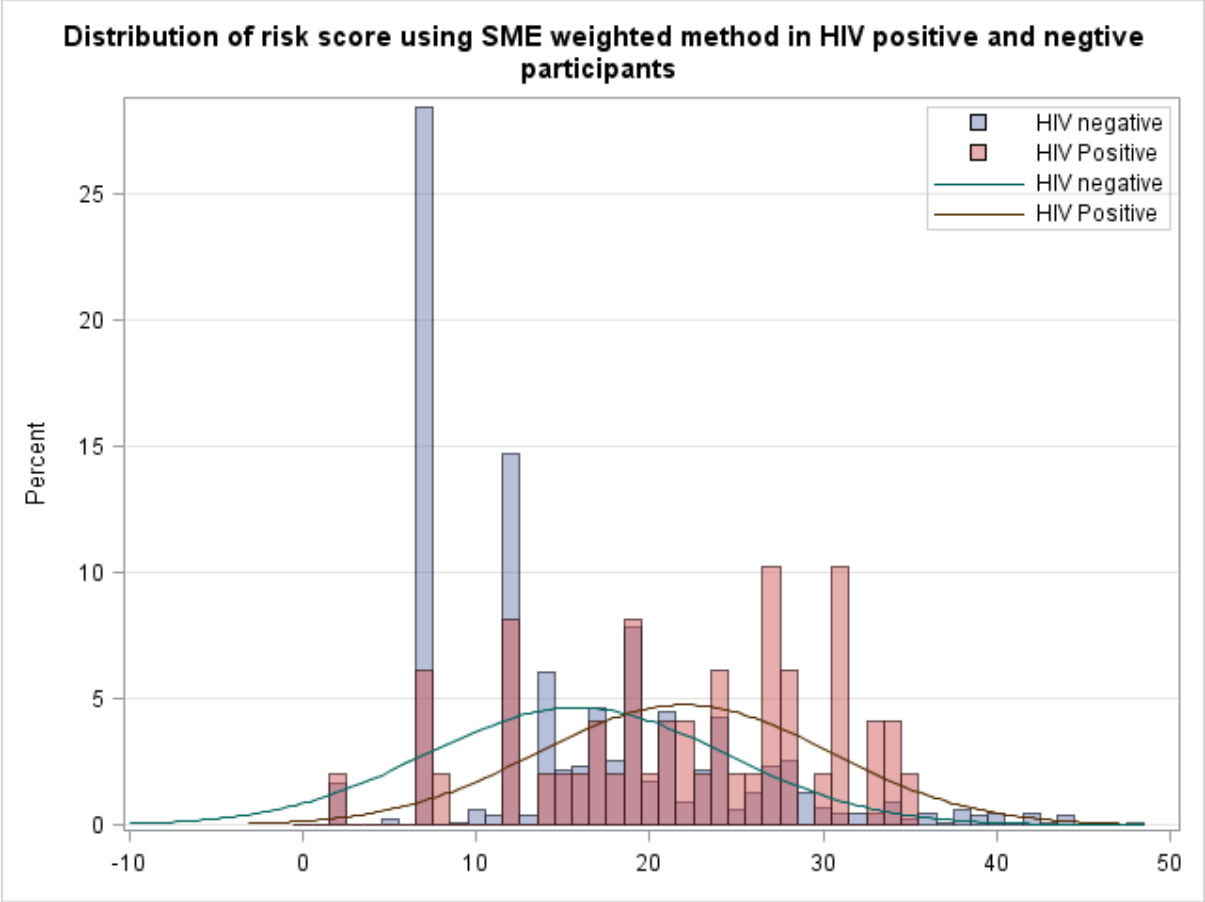

Fig. 3(a). Distribution of individual risk scores assigned using the SME Unit-Weighted method among HIV positive and negative participants

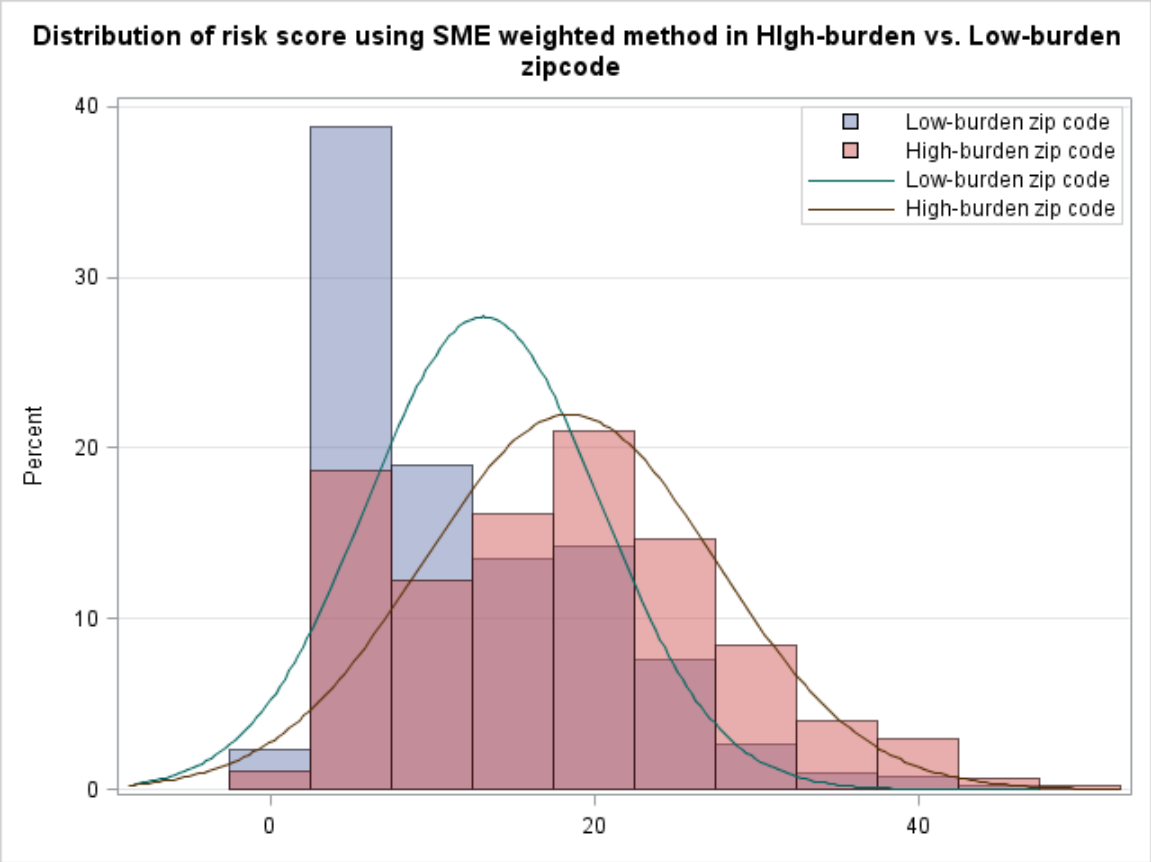

Fig. 3(b). Distribution of individual risk scores assigned using the SME Unit-Weighted method by low and high zip code groups.

	N	%
Lower Risk Area	142	33.60%
Higher Risk Area	294	62.30%
Areas combined	436	48.80%

z = -8.5522
p <0.0001

Table 9. Comparison of compound risk using SME in high and low risk zip code groups (Score of 16 or above)

4.3.3 Clinical Decision Rule risk scoring method

Univariate analysis identified the following five variables as being associated with HIV-positive status in the Geography Project dataset: HSV-2 infection, having had receptive anal intercourse, past crack use, been paid by a man for sex, and having engaged in sex work. (Table 3)

Risk score distribution using CDR method is shown in Table 10. Half of the participants had a risk score of more than 2. ROC and Youden-index show the optimal cutoff point for both HIV-status and high-burden zip code group prediction to be 4, categorizing "High-HIV risk" to 4 and above for both HIV-status and zip code group prediction. Area under the ROC for HIV-status prediction was 0.75.

Using this risk scoring method showed an increase in risk among HIV positive participants as well as among high-burden zip code groups. Mean (SD) risk score of HIV positive and negative participants is 4.6 (2.3) and 2.3 (2.1) respectively, and median is 5 and 2 respectively (Fig. 6(a)). The mean difference is 2.2. Mean risk score of the high-burden zip codes is significantly higher than that of the low-burden zip code group. High-burden population had a mean (SD) risk score of 3.1 (2.3) and median of 3. Low-burden had a 1.9 (1.9) with a median of 2 (Fig. 6(b)). The mean difference between low burden and high burden zip group is 1.21.

ROC Curve (Fig. 8) and Youden-Index show the best cutoff to be at 4 for both prediction for HIV-status and living in high burden zip codes and the AUC for prediction of HIV-status is 0.75 and for living in High burden zip codes is 0.64. Using this method, the proportion of participants who had a risk score >4 is significantly higher than the proportion of those living in low burden zip code (p-value <0.0001) (table 11).

Score	N	%
0	317	34.7
2	150	16.4
3	170	18.6
4	24	2.6
5	196	21.5
6	5	0.5
7	36	3.9
8	12	1.3
9	2	0.2
10	1	0.1
Total	913	100

Table 10. Distribution of individuals assigned risk scores from Clinical Decision Rule method

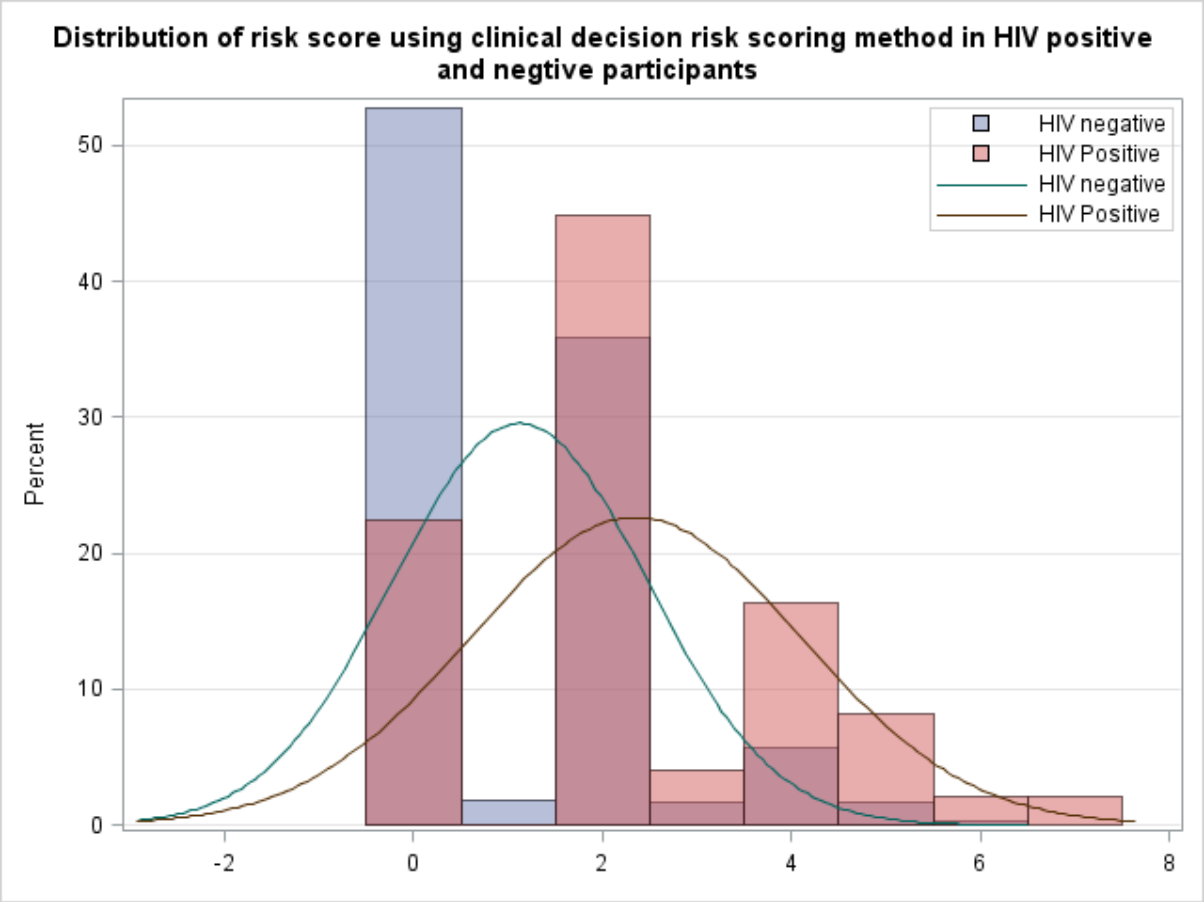

Fig. 6(a). Distribution of individual risk scores assigned using the Clinical Decision Rule method among HIV- positive and negative participants

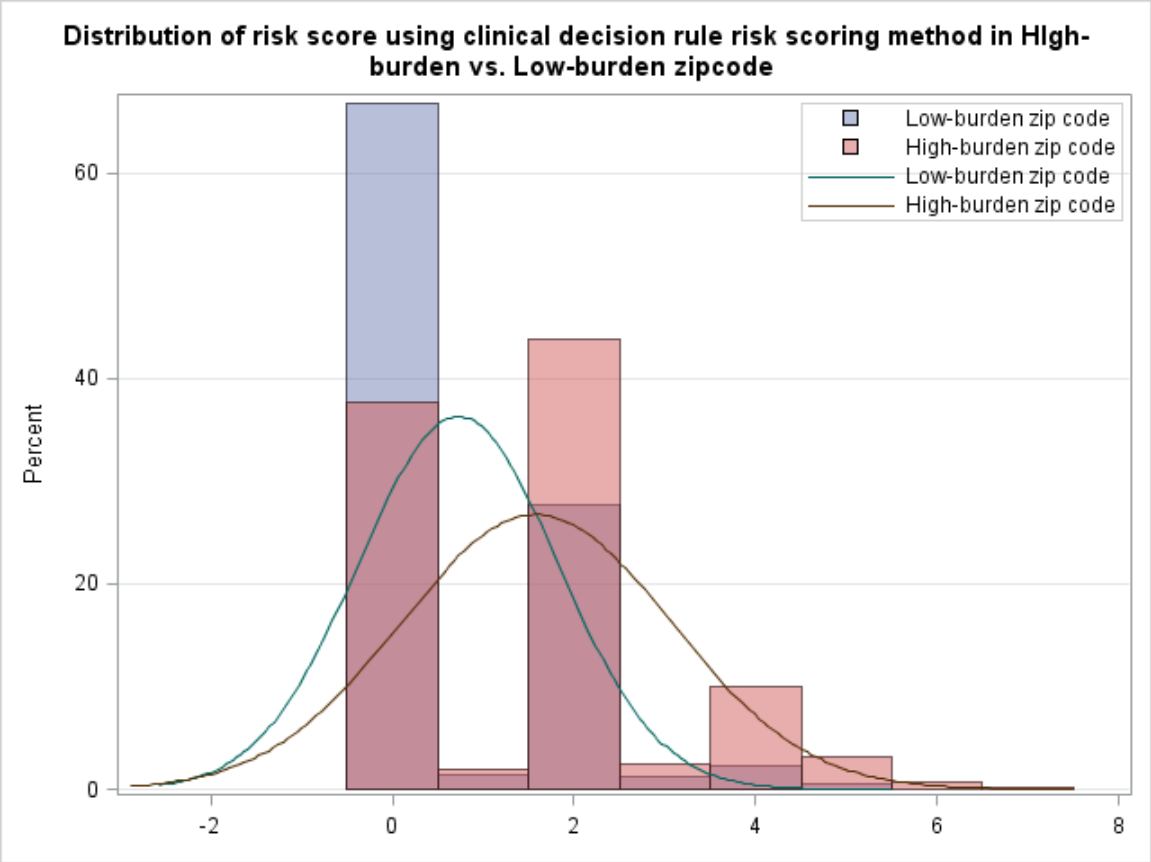

Fig. 6(b). Distribution of individual risk scores assigned using the Clinical Decision Rule method by low and high zip code groups.

	N	%
Lower Risk Area	70	16.60%
Higher Risk Area	194	41.50%
Areas combined	266	29.8%

z =8.142
p<0.001

Table 11. Comparison of composite risk using clinical decision rule to identify proportion of high-risk individuals living in high-burdened zip code groups.

4.3.4 Binary Risk Indicator

Distribution of risk scores among individuals showed most individuals with scores of three or less (Table 12). ROC and Youden-index show the optimal cutoff point for both

HIV-status and high-burden zip code group prediction to be 3, categorizing "High-HIV risk" to 3 and above for both outcomes. Area under the ROC curve for HIV-status prediction was 0.68 (Fig. 8).

Using this risk scoring method showed an increase in HIV positive participants as well as among high-burden zip code groups. Mean (SD) risk score of HIV positive participants is 4.9 (4.2) and median is 3 (Fig. 7(a)). The mean difference is 2.03. Mean risk score of the high-burden zip codes is significantly higher than that of low-burden zip code. High-burden population had a mean (SD) risk score of 6.8 (4.6) and median of 3.0. Low-burden had a mean of 3.8 (3.5) with a median of 2.0 (Fig. 7(b)). Mean difference between low and high burden zip groups was 2.3.

ROC Curve (Fig. 9) and Youden-Index show the best cutoff to be at 3 and an AUC for prediction of HIV status is 0.67 and for living in high burden zip codes is 0.68. Using this method, the proportion of participants who had a risk score >3 are significantly higher than the proportion of those living in low burden zip code (p-value <0.0001) (Table 13).

Score	N	%
0	18	1.9
1	6	0.6
2	352	38
3	267	28.8
4	1	0.1
6	11	1.2
7	53	5.7
8	3	0.3
9	1	0.1
10	41	4.4
11	97	10.5
14	7	0.8
15	70	7.6
Total	927	100

Table 12. Distribution of individuals with scores derived from Binary Risk Indicator method

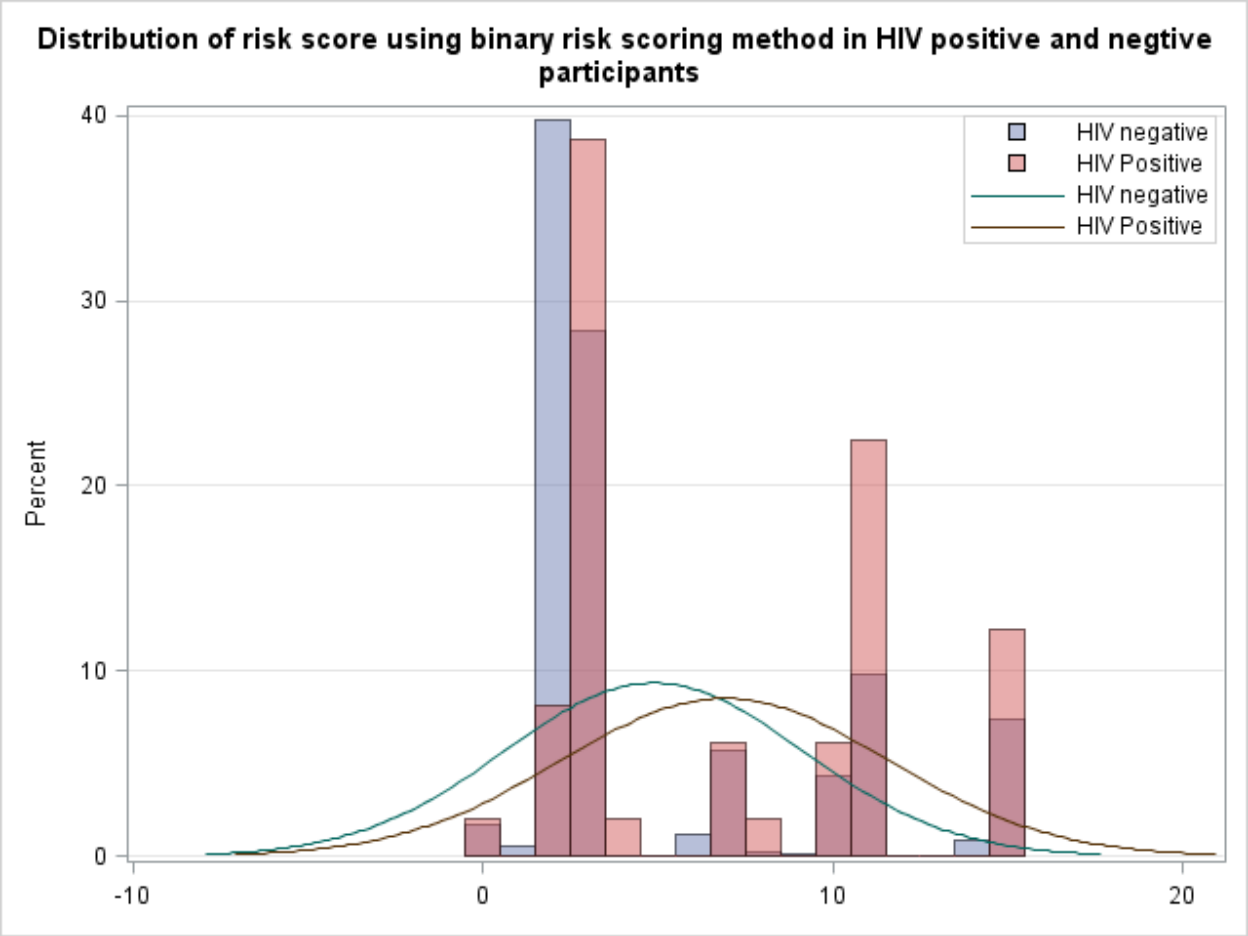

Fig. 7(a). Binary risk score distribution. Binary Risk Indicator score distribution among HIV positive and negative participants

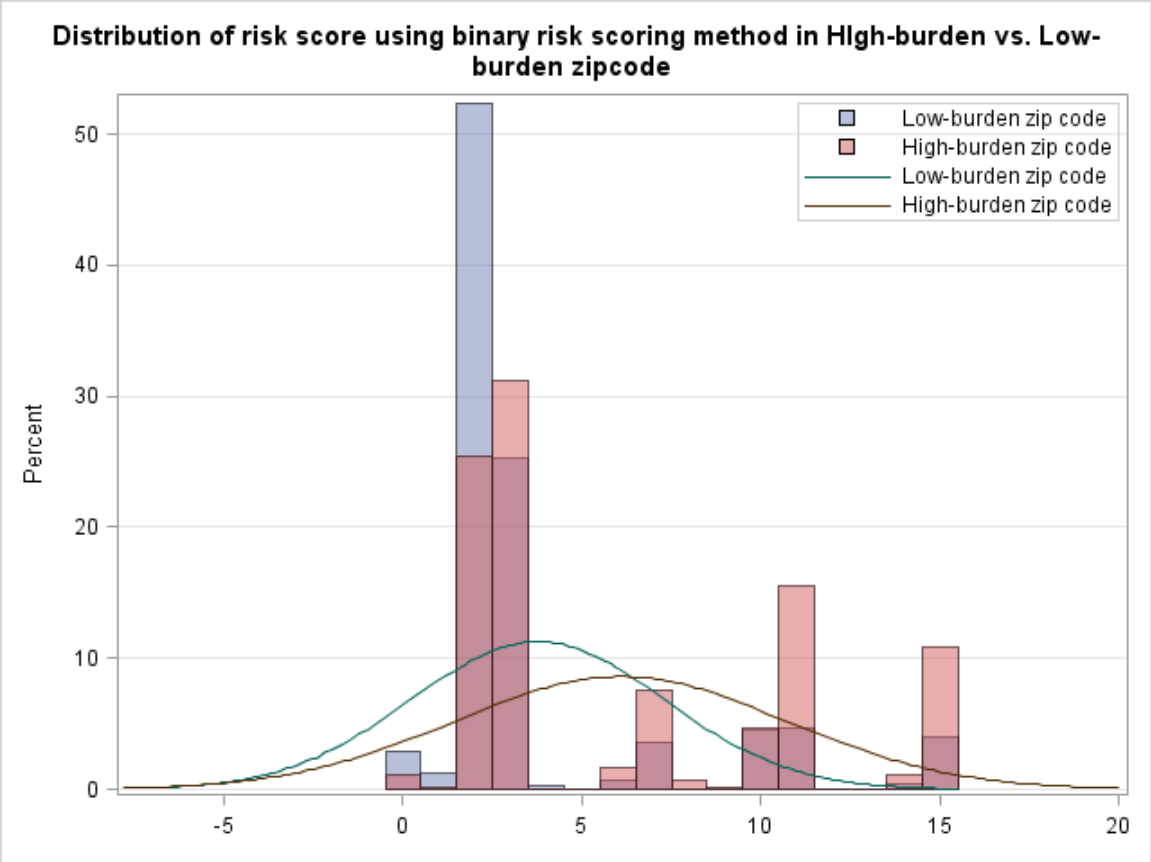

Fig. 7(b). Binary risk score distribution. Binary Risk Indicator score distribution for High-burden vs. Low burden zip code groups

	N	%
Lower Risk Area	184	43.60%
Higher Risk Area	346	73.30%
Areas combined	530	59.30%

z = -9.0242
p <0.0001

Table 13. Binary Risk Indicator scored high-risk proportion. Comparison of a binary risk score of 3 or more in High-burden vs. Low-burden zip code

4.4 Comparison and Predictive Validity of Risk Assessment Methods

4.4.1 Comparing Risk Assessment Tests by Sensitivity and Specificity.

Sensitivity and specificity of optimal cut off points of the four risk scoring methods

for HIV-status and residence of high HIV-burdened zip code group prediction are presented in Tables 14 and 15, respectively. When comparing the sensitivity of the best cutoff point in prediction of HIV status by McNemar's chi-square two-sided exact test, the sensitivity of the best cutoff point of the Binary Risk Indicator risk assessment method is higher than the three other methods but only significantly higher in CDR method ($p=<0.05$) (table 16(a) to 18(a)). In prediction of living in high burden zip codes, the sensitivity of the best cutoff point of the Binary Risk Indicator risk assessment method is significantly higher than the other three risk scoring models ($p=<0.0001$) (table 16(b) – 18(b))

Unit weight score				SME differential weight score				Clinical decision rule				Binary Risk Score			
Cutpoint	Sensitivity	Specificity	Youden-Index	Cutpoint	Sensitivity	Specificity	Yodane Index	Cutpoint	Sensitivity	Specificity	Youden Index	Cutpoint	Sensitivity	Specificity	Youden Index
1				1-5				0				0			
2	0.98	0.01	-0.01	6-10	0.98	0.02	0	2	0.9	0.36	0.26	1	0.98	0.02	0
3	0.9	0.3	0.2	11-15	0.9	0.31	0.21	3	0.83	0.53	0.36	2	0.98	0.02	0
4	0.8	0.52	0.32	16-20	0.78	0.55	0.33	4	0.71	0.72	0.43	3	0.9	0.42	0.32
5	0.59	0.71	0.3	21-25	0.59	0.73	0.32	5	0.65	0.74	0.39	4	0.51	0.7	0.21
6	0.42	0.84	0.26	26-30	0.41	0.86	0.27	6	0.2	0.94	0.14	6	0.49	0.7	0.19
7	0.18	0.93	0.11	31-35	0.2	0.94	0.14	7	0.2	0.95	0.15	7	0.5	0.71	0.21
8	0	0.97	-0.03	36-40	0	0.97	-0.03	8	0.1	0.98	0.08	8	0.43	0.77	0.2
9	0	0.98	-0.02	41-45	0	0.99	-0.01	9	0.04	0.99	0.03	9	0.4	0.78	0.18
10	0	0.99	-0.01	46-50	0	0.99	-0.01	10	0.02	1	0.02	10	0.4	0.78	0.18
												11	0.35	0.82	0.17
												14	0.12	0.92	0.04
												15	0.12	0.93	0.05

Table 14. Sensitivity and specificity of different cut off points of three risk scoring methods in prediction of HIV-status

Unit weight score				SME differential weight score				Clinical decision rule				Binary Risk Score			
Cutpoint	Sensitivity	Specificity	Youden-Index	Cutpoint	Sensitivity	Specificity	Youden-Index	Cutpoint	Sensitivity	Specificity	Youden-Index	Cutpoint	Sensitivity	Specificity	Youden-Index
1				1-5				0				0			
2	98.52%	2.37%	0.0089	6-10	98.52%	2.37%	0.0089	2	74.15%	45.50%	0.1965	1	98.94%	2.84%	0.01784
3	80.08%	41.94%	0.2202	11-15	79.45%	42.18%	0.2163	3	55.93%	60.43%	0.1636	2	98.73%	4.03%	0.02758
4	62.71%	66.82%	0.2953	16-20	59.32%	68.01%	0.2733	4	41.53%	83.41%	0.2494	3	73.31%	56.40%	0.2971
5	40.47%	81.04%	0.2151	21-25	37.29%	83.41%	0.207	5	37.50%	84.60%	0.221	4	42.16%	81.75%	0.2391
6	24.36%	92.42%	0.1678	26-30	22.67%	94.08%	0.1675	6	9.75%	98.10%	0.07846	6	42.16%	81.99%	0.2415
7	11.86%	97.87%	0.0973	31-35	10.17%	97.87%	0.0804	7	8.90%	98.34%	0.07238	7	40.47%	82.70%	0.2317
8	4.45%	99.05%	0.03499	36-40	4.87%	99.05%	0.03923	8	2.54%	99.76%	0.02302	8	32.84%	86.26%	0.191
9	1.91%	100%	0.01907	41-45	1.91%	99.76%	0.01667	9	0.42%	100%	0.004237	9	32.20%	86.26%	0.1846
10	0.42%	100%	0.004237	46-50	0.21%	100%	0.002119	10	0.21%	100%	0.002119	10	31.99%	86.26%	0.1825
												11	27.33%	90.76%	0.1809
												14	11.86%	95.50%	0.0736
												15	10.81%	95.97%	0.0678

Table 15. Sensitivity and specificity of different cut off points of three risk scoring methods in prediction of residence in high HIV-burdened zip codes

	Binary Risk Indicator		
Unit Weighted Risk score	High risk score	Low risk score	Total
High risk score	36	8	44
Low risk score	3	2	5
Total	39	10	49

Sensitivity of Binary Risk Indicator=44/49=90%
Sensitivity of Unit Weighted Risk score=39/49=80%

Table 16(a). Unit Weighted Risk Score specificity compared with Binary Risk Indicator among HIV positive and negative patients.

	Binary Risk Indicator		
CDR risk score	High risk score	Low risk score	Total
Low risk score	35	0	35
High risk score	9	5	14
Total	44	5	49

Sensitivity of Binary Risk Indicator=44/49=90%
Sensitivity of CDR risk score=35/49= 71%

Table 17(a). Clinical Decision Rule Risk Score specificity compared with Binary Risk Indicator among HIV positive and negative patients.

	Binary Risk Indicator		
SME risk score	High risk score	Low risk score	Total
High risk score	9	35	44
Low risk score	2	3	5
Total	11	38	49

Sensitivity of Binary Risk Indicator=44/49=90%
Sensitivity of SME risk score=38/49=78%

Table 18(a). Subject Matter Expert Weighted Risk Score specificity compared with Binary Risk Indicator among HIV positive and negative patients.

	Binary Risk Indicator		
Unit Weighted Risk score	High risk score	Low risk score	Total
High risk score	280	16	296
Low risk score	66	110	176
Total	346	126	472

Sensitivity of Binary Risk Indicator=346/472=73.30%
Sensitivity of Unit Weighted Risk score=296/472=62.70%

Table. 16(b) Unit Weighted Risk Score specificity compared with Binary Risk Indicator among patients who live in High-HIV burden zip code.

	Binary Risk Indicator		
CDR risk score	High risk score	Low risk score	Total
High risk score	196	0	196
Low risk score	150	126	276
Total	346	126	472

Sensitivity of Binary Risk Indicator=346/472=73.30%
Sensitivity of CDR risk score=196/472=41.50%

Table. 17(b). Clinical Decision Rule Risk assessment specificity compared to Binary Risk Indicator specificity among patients who live in High-HIV burdened zip code.

	Binary Risk Indicator		
SME risk score	High risk score	Low risk score	Total
High risk score	278	16	294
Low risk score	68	110	178
Total	346	126	472

Sensitivity of Binary Risk Indicator=346/472=73.30%
Sensitivity of SME risk score=294/472=62.30%

Table. 18 (b). Subject Matter Expert Weighted Risk Score specificity compared to Binary Risk Indicator specificity among patients who live in High-HIV burden zip code.

4.4.2. Comparing risk assessment methods by difference of means.

Risk scoring performance compared by mean difference after standardization between HIV positive and negative group show CDR risk scoring has the highest difference (table 19(a)). The different among low and high-burden zip code groups show Binary Risk Indicator to have the greatest difference between the mean at 1.5 units, followed by CDR at 1.2 units, and SME and Burgess both at 1.1 units (Table 19(b)).

Risk-Assessment Method	Mean difference between HIV positive and negative
Clinical Decision Rule Method	**2.2**
Binary Risk Indicator Method	1.3
SME-Weighted Method	1.2
Burgess Unit-weighted Method	1.2

Table. 19(a). Mean difference of risk score between HIV positive group and negative group

Risk-Assessment Method	Mean difference between zip code groups
Binary Risk Indicator Method	1.5
Clinical Decision Rule Method	1.2
SME-Weighted Method	1.1
Burgess Unit-weighted Method	1.1

Table. 19(b). Mean difference of risk score between high- and low-burden zip codes

4.4.3. Comparing risk assessment methods by AUC. Area under the curve was determined from the ROC Curve for each risk assessment model. Comparison of ROC curves using different risk assessment models for prediction of HIV status and living in high-burden zip codes were shown in Fig. 8 and 9. The area under the curve (AUC) for the probability of the risk score to correctly predict HIV status of the participant was 0.75 using clinical decision rule, 0.71 using SME-weighted method, 0.70 using Burgess unit method, and 0.67 by Binary Risk Indicator (Table 20). Area under ROC curve for zip group prediction showed similar areas, but CDR performing slightly less the other diagnostic tests (Table 21).

Risk-Assessment Method (HIV-Status)	AUC (95% CI)
Clinical Decision Rule	0.75 (0.68 – 0.82)
SME-Weighted Method	0.71 (0.64 – 0.79)
Burgess Unit-Weighted Method	0.70 (0.63 – 0.78)
Binary Risk Indicator Method	0.67 (0.61 – 0.74)

Table. 20. Comparison of Area Under the ROC curve for HIV-status outcome.

Risk-Assessment Method (Zip group prediction)	AUC (95% CI)
Binary Risk Indicator Method	0.68 (0.64 – 0.71)
Burgess Unit-Weighted Method	0.68 (0.64 – 0.71)
SME-Weighted Method	0.67 (0.64 – 0.71)
Clinical Decision Rule	0.64 (0.61 – 0.68)

Table 21. Comparing area under ROC curve among different risk models in prediction of zip code group

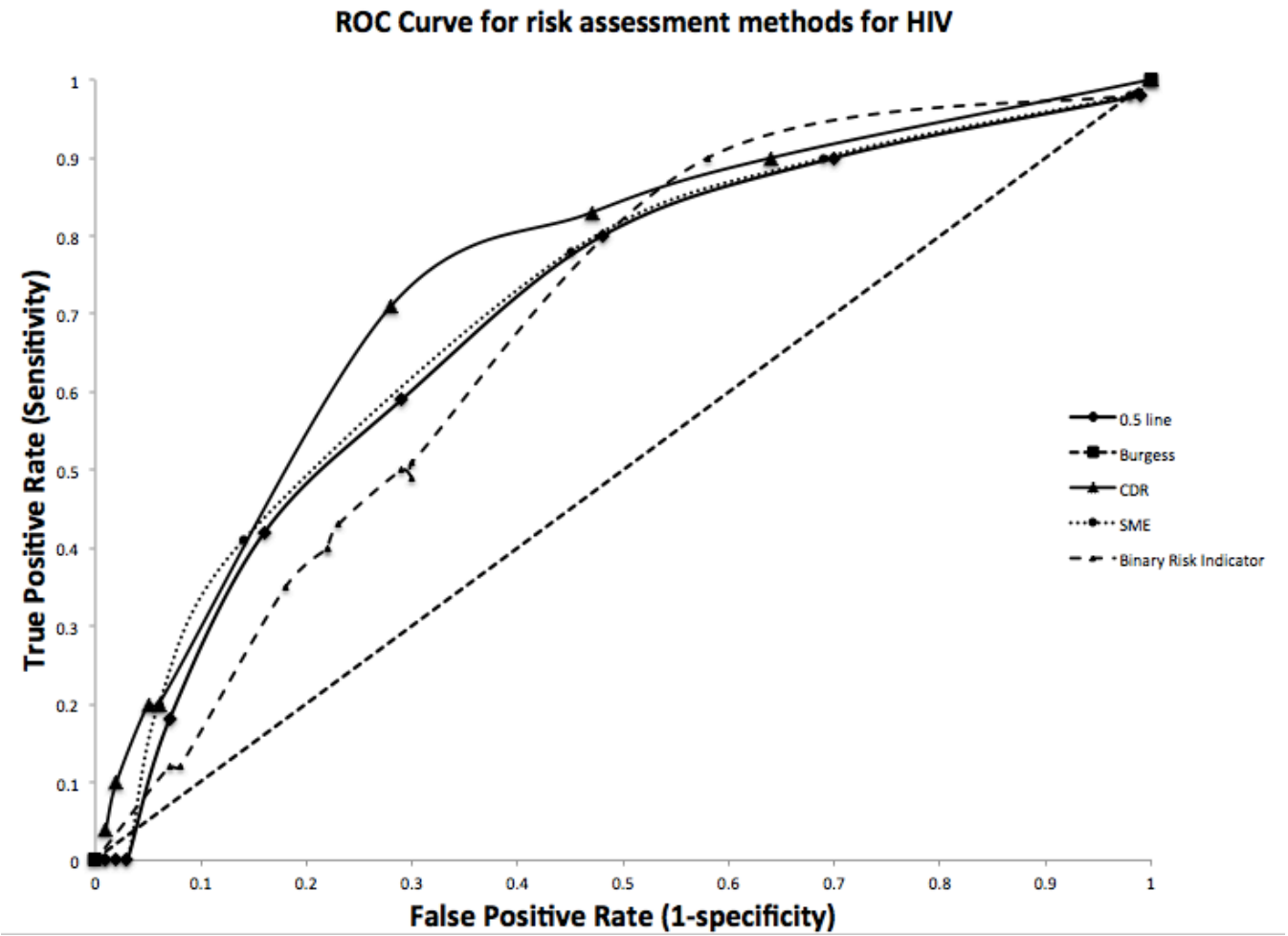

Fig. 8. ROC for different risk assessment methods applied to predict HIV status.

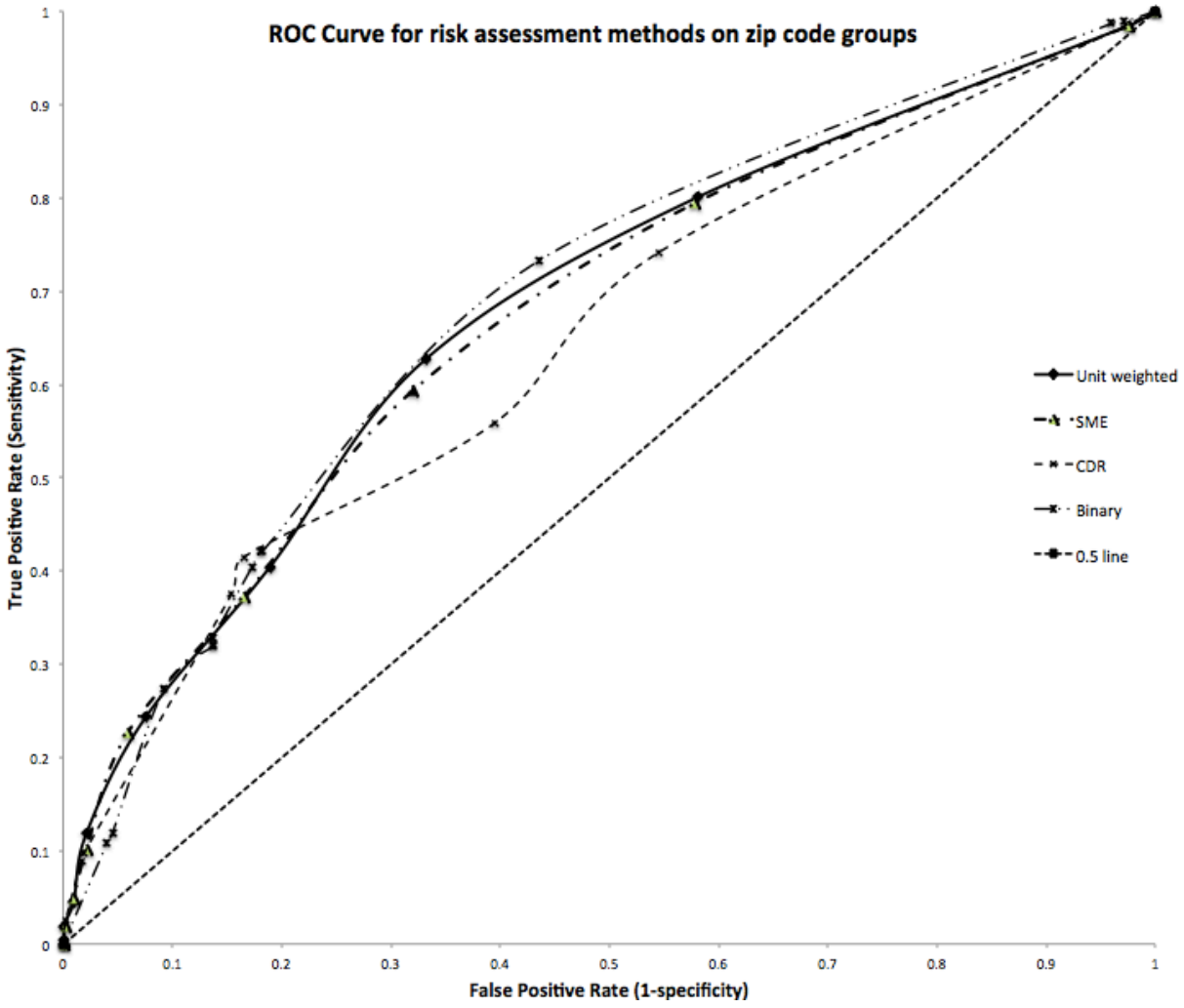

Fig. 9. ROC for different risk assessment methods applied to predict high-burden zip code grou

CHAPTER V.

DISCUSSION AND CONCLUSION

Three methods were used to evaluate risk assessment models to predict HIV status and predict high-risk participants in high-burdened zip codes: 1) comparing difference of the means of risk score, 2) comparing sensitivity for each method by McNemar's chi-square test, and 3) comparing area under the ROC curve.

5.1 HIV-status Prediction.

Area under the ROC curve. HIV risk was assessed using four risk assessment models. Performance was compared by area under the ROC curve. CDR was expected to outperform Binary Risk Indicator, SME, and Burgess methods because it was developed internally using the Geography Project dataset. Highest AUC was CDR, which was 0.75 (95% CI, 0.68 – 0.71), and this value is considered a fair diagnostic test. The other three tests: Burgess 0.71 (95% CI, 0.64 – 0.79), SME 0.67 (95% CI, 0.64 – 0.71), and Binary Risk Indicator 0.67 (95% CI, 0.61 – 0.74) would be considered a poor to fair diagnostic test. The shape of CDR (Fig. 8) shows acceptable sensitivity (0.71) and specificity (0.72) at the cutoff point of 4, suggesting some predictive ability. Further confirmation of this method would be external validation using a similar dataset or population.

Specificity comparison. When comparing sensitivity, Binary Risk indicator model showed the highest sensitivity compared to other three methods, however, only significantly higher than the CDR scoring methods.

Mean difference. Clinical decision rule risk scoring methods has the highest mean difference between HIV positive and negative group.

5.2 High-burden zip code group prediction. Diagnostic test evaluation of risk assessment methods for zip code group prediction was used for comparative purposes to test how well each method will predict those of high-risk in high-burden zip codes.

Area under the ROC curve. Area under the ROC curve for all four risk assessment methods showed little difference in performance, with AUCs between 0.6 and 0.7 and similar confidence intervals. High-risk participants living in high-burden zip code groups is not expected to be 100%, so there is not a perfect test to predict risk. A moderate shift in sensitivity and predictive value may indicate some usefulness in characterizing risk. All risk assessment methods show potential for predictability, suggesting some predictive validity and confirmation of high-risk characteristics among high-burdened zip codes. Analyses addressing predictability of risk assessment methods in low-burden zip code groups may confirm how risk assessment tools or sampling methodologies differ in low and high burden areas.

Specificity comparison. McNemar's chi-square test compared specificity of each diagnostic test to Binary Risk Indicator at optimal cutoff point for comparison. Binary

Risk Indicator remained the highest performance among all other risk assessment tests, with a specificity of 0.73. Burgess and SME were both 0.63 and 0.62 in relation to Binary Risk Indicator, respectively. CDR showed specificity at 0.42 in relation to Binary Risk Indicator. Binary Risk Indicator performed similar to other risk assessment methods in AUC-ROC, but more specific.

Mean difference. Distribution of risk scores in high and low-burdened zip codes, using different risk assessment models showed greatest difference in means to be Binary Risk Indicator (Table 16), suggesting this model is best among these risk assessment methods to be able to identify those who scored high in risk assessment in high-burdened zip code areas and categorize participants who scored lower in low-burdened zip codes. All risk assessment methods showed significant difference in risk score distribution between populations in low and high-burdened zip codes, suggesting these methods are congruent in characterizing risk.

5.3 Conclusion

This study may provide a template for assessing risk and evaluating performance of risk assessment tools through multiple methods. Studying multiple risk assessment methods and evaluation tools on multiple prediction outcomes may show relationships or inconsistencies between real risk and risk behaviors. How these converge is complex and is likely to vary in different environments and comparing assessment tools among risk and risk outcome may be an approach not frequently used. Another new approach this

study offers is the use of a simplified CDR risk assessment method for this type of study and population.

One of the limitations of the study is that the risk assessment models were not externally validated. Another limitation is the risk variable selection was not tested to include the highest predictive variable and the risk scoring models was not developed to get the highest predictive test score. Therefore, risk assessment methods can be further developed by computer modeling to mine combinations of risk factors most likely to give the highest predictive test scores. Modeling for a single risk assessment tool may help optimize diagnostic tools and programmers and statisticians are equipped to develop these types of analyses. In addition, further research is needed to evaluate the performance of these risk scoring methods in low HIV burden zip codes and high HIV burden zip codes separately.

VI. WORKS CITED

1. Centers for Disease Control and Prevention. HIV Incidence. (2013). Available at: http://www.cdc.gov/hiv/statistics/surveillance/incidence/index.html.
2. Long, E. F., Brandeau, M. L. & Owens, D. K. The cost-effectiveness and population outcomes of expanded HIV screening and antiretroviral treatment in the United States. *Ann. Intern. Med.* **153,** 778–789 (2010).
3. Fox, J. *et al.* Quantifying sexual exposure to HIV within an HIV-serodiscordant relationship: development of an algorithm. *Aids* **25,** 1065–82 (2011).
4. Wawer, M. J. *et al.* Rates of HIV-1 transmission per coital act, by stage of HIV-1 infection, in Rakai, Uganda. *J Infect Dis* **191,** 1403–9 (2005).
5. Seidlin, M., Vogler, M., Lee, E., Lee, Y. S. & Dubin, N. Heterosexual transmission of HIV in a cohort of couples in New York City. *Aids* **7,** 1247–54 (1993).
6. Wald, A. & Link, K. Risk of human immunodeficiency virus infection in herpes simplex virus type 2-seropositive persons: a meta-analysis. *J Infect Dis* **185,** 45–52 (2002).
7. Koelle, D. M., Abbo, H., Peck, A., Ziegweid, K. & Corey, L. Direct recovery of herpes simplex virus (HSV)-specific T lymphocyte clones from recurrent genital HSV-2 lesions. *J Infect Dis* **169,** 956–61 (1994).

8. Mayer, K. H. & Venkatesh, K. K. Interactions of HIV, other sexually transmitted diseases, and genital tract inflammation facilitating local pathogen transmission and acquisition. *Am. J. Reprod. Immunol. N. Y. N 1989* **65,** 308–316 (2011).
9. Fleming, D. T. *et al.* Herpes simplex virus type 2 in the United States, 1976 to 1994. *N Engl J Med* **337,** 1105–11 (1997).
10. Wald, A. & Ashley-Morrow, R. Serological Testing for Herpes Simplex Virus (HSV)–1 and HSV-2 Infection. *Clin. Infect. Dis.* **35,** S173–S182 (2002).
11. Koziel, M. J. & Peters, M. G. Viral hepatitis in HIV infection. *N. Engl. J. Med.* **356,** 1445–1454 (2007).
12. Sherman, K. E., Rouster, S. D., Chung, R. T. & Rajicic, N. Hepatitis C Virus prevalence among patients infected with Human Immunodeficiency Virus: a cross-sectional analysis of the US adult AIDS Clinical Trials Group. *Clin Infect Dis* **34,** 831–7 (2002).
13. Teri Shors. *Understanding Viruses*. (2013).
14. Johnson, L. F. & Lewis, D. A. The effect of genital tract infections on HIV-1 shedding in the genital tract: a systematic review and meta-analysis. *Sex Transm Dis* **35,** 946–59 (2008).
15. Fastring, D. R. *et al.* Co-occurrence of Trichomonas vaginalis and bacterial vaginosis and vaginal shedding of HIV-1 RNA. *Sex Transm Dis* **41,** 173–9 (2014).
16. Hall, H. I. *et al.* Estimation of HIV incidence in the United States. *Jama* **300,** 520–9 (2008).

17. AIDSVu (www.aidsvu.org). Emory University, R. S. of P. H. [May 28th, 2014.]. AIDSVu (www.aidsvu.org). Emory University, Rollins School of Public Health. [May 28th, 2014.]. (2014).
18. National Center for HIV/AIDS, Viral Hepatitis, STD, and TB Prevention & Division of HIV/AIDS Prevention. *HIV Among African Americans*. (2013).
19. Herbst, J. H. *et al.* Estimating HIV prevalence and risk behaviors of transgender persons in the United States: a systematic review. *AIDS Behav* **12,** 1–17 (2008).
20. Baggaley, R. F., White, R. G. & Boily, M. C. HIV transmission risk through anal intercourse: systematic review, meta-analysis and implications for HIV prevention. *Int J Epidemiol* **39,** 1048–63 (2010).
21. Boily, M. C. *et al.* Heterosexual risk of HIV-1 infection per sexual act: systematic review and meta-analysis of observational studies. *Lancet Infect Dis* **9,** 118–29 (2009).
22. Dyer, T. P. *et al.* Differences in substance use, psychosocial characteristics and HIV-related sexual risk behavior between Black men who have sex with men only (BMSMO) and Black men who have sex with men and women (BMSMW) in six US cities. *J. Urban Health Bull. N. Y. Acad. Med.* **90,** 1181–1193 (2013).
23. Stein, M. D. *et al.* Alcohol use patterns predict high-risk HIV behaviors among active injection drug users. *J. Subst. Abuse Treat.* **18,** 359–363 (2000).
24. Fisher, J. C., Bang, H. & Kapiga, S. H. The association between HIV infection and alcohol use: a systematic review and meta-analysis of African studies. *Sex Transm Dis* **34,** 856–63 (2007).

25. DeBeck, K. *et al.* Smoking of crack cocaine as a risk factor for HIV infection among people who use injection drugs. *Cmaj* **181,** 585–9 (2009).

26. Edlin, B. R. *et al.* Intersecting epidemics--crack cocaine use and HIV infection among inner-city young adults. Multicenter Crack Cocaine and HIV Infection Study Team. *N Engl J Med* **331,** 1422–7 (1994).

27. Kaplan, E. H. & Heimer, R. A model-based estimate of HIV infectivity via needle sharing. *J Acquir Immune Defic Syndr* **5,** 1116–8 (1992).

28. HIV Infection and HIV-Associated Behaviors Among Persons Who Inject Drugs — 20 Cities, United States, 2012. Available at: http://www.cdc.gov/mmwr/preview/mmwrhtml/mm6410a3.htm. (Accessed: 23rd April 2016)

29. Illinois., Bruce, A. A, Harno, A. J., Landesco, J. & Burgess, E. W. *The workings of the Indeterminate-sentence law and the parole system in Illinois. A report to the Honorable Hinton G. Clabaugh, chairman, Parole board of Illinois.* (1928).

30. Bobko, P., Roth, P. L. & Buster, M. A. The Usefulness of Unit Weights in Creating Composite Scores: A Literature Review, Application to Content Validity, and Meta-Analysis. *Organ. Res. Methods* **10,** 689–709 (2007).

31. Davis-Stober, C. P., Dana, J. & Budescu, D. V. A constrained linear estimator for multiple regression. *Psychometrika* **75,** 521–541 (2010).

32. Rothenberg, R., Baldwin, J., Trotter, R. & Muth, S. The risk environment for HIV transmission: results from the Atlanta and Flagstaff network studies. *J. Urban Health Bull. N. Y. Acad. Med.* **78,** 419–432 (2001).

33. McGinn, T. G. *et al.* Users' guides to the medical literature: XXII: how to use articles about clinical decision rules. Evidence-Based Medicine Working Group. *Jama* **284,** 79–84 (2000).

34. Menza, T. W., Hughes, J. P., Celum, C. L. & Golden, M. R. Prediction of HIV Acquisition Among Men Who Have Sex with Men. *Sex. Transm. Dis.* **36,** 547–555 (2009).

35. Kahle, E. M. *et al.* An empiric risk scoring tool for identifying high-risk heterosexual HIV-1-serodiscordant couples for targeted HIV-1 prevention. *J Acquir Immune Defic Syndr* **62,** 339–47 (2013).

36. Trajman, A. & Luiz, R. R. McNemar chi2 test revisited: comparing sensitivity and specificity of diagnostic examinations. *Scand. J. Clin. Lab. Invest.* **68,** 77–80 (2008).

37. Park, S. H., Goo, J. M. & Jo, C.-H. Receiver Operating Characteristic (ROC) Curve: Practical Review for Radiologists. *Korean J. Radiol.* **5,** 11–18 (2004).