EPIDEMIOLOGICAL ANALYSIS OF SOCIOSEXUAL HIV NETWORKS IN CENTRAL NORTH CAROLINA

Dana Kristine Pasquale

A dissertation submitted to the faculty at the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Epidemiology (Infectious Diseases) in the Gillings School of Global Public Health.

Chapel Hill
2018

Approved by:

William C. Miller

Ann M. Dennis

Irene A. Doherty

Peter A. Leone

Kimberly A. Powers

ABSTRACT

Dana Kristine Pasquale: Epidemiological Analysis of Sociosexual HIV Networks in Central North Carolina
(Under the direction of William C. Miller)

Disparities in HIV incidence are seen by race and sexual orientation,[3-5] although race and sexual orientation do not sufficiently explain differential risk within sexual networks. Race and sexual orientation, however, influence partner selection,[8] risk behavior,[9] and access to care.[10] Partner selection and risk behavior underlay differences in HIV acquisition and can be studied within the context of a sociosexual network.[11] Marginalized or stigmatized persons are more likely to be diagnosed with HIV later in the course of infection[12,13] and less likely to achieve viral suppression,[14,15] which both increase the amount of time that a person remains infectious.

Infectious persons who are active in a sexual network risk onward transmission of HIV, becoming superinfected, or acquiring another sexually-transmitted infection (STI) such as syphilis, which has a synergistic effect with HIV.[16-19] Knowing the HIV prevalence within high-risk sexual networks, HIV and STI history of network members, and partnership patterns may provide sufficient information to guide targeted interventions to reduce the amount of time that HIV-positive persons remain infectious.

This study uses newly diagnosed HIV cases reported 2012-2013 to create a “baseline” sexual network. We examined HIV transmission cluster involvement and followed new cases through 2016 for post-baseline partnerships investigated for public health HIV prevention efforts as a marker for transmission risk potential. Network structure and partner selection behaviors were modeled to predict which cases were likely to be involved in a transmission event and were a candidate for enhanced linkage to care support.

PUBLIC HEALTH SIGNIFICANCE

Too few diagnosed HIV cases in the United States are in care and virally suppressed,[10,20] thereby increasing the likelihood of onward transmission. Identifying factors associated with HIV acquisition and transmission by following persons in the sexual network or assessing transmission cluster growth may help us understand how partnerships form and whether any baseline network structures predict remaining active in high-risk sexual networks.

For Marc
thank you for the 10,000 hours

ACKNOWLEDGMENTS

The project described in Aim 1 was supported by the National Center for Advancing Translational Sciences (NCATS), NIH, through Grant Award Number UL1TR001111 (awarded to Ann M. Dennis, MD and Irene A. Doherty, PhD, MPH via the NC Translational and Clinical Sciences Institute at UNC). The project described in Aim 2 was supported by the National Institute of Allergy and Infectious Diseases (NIAID), NIH, through Grant Award Number 5T32AI070114-10 to UNC-Chapel Hill (awarded to Steven Meshnick, MD, PhD and Brian Pence, PhD). The content in Aims 1 and 2 is solely the responsibility of the investigators and does not necessarily represent the official views of the NIH.

I would like to thank Dr. Steven Meshnick and Dr. Brian Pence for their support of me through the T32 training grant, above and far beyond the financial support provided. Dr. Meshnick and Dr. Pence met with me and other Trainees regularly to discuss progress, answer questions, and support our professional development. I have gained much from their ability to quickly identify the most important question. I would also like to thank the other Trainees, notably Sara Levintow and Jane Chen; the thought-provoking discussions during TIDE meetings certainly strengthened these analyses. Steve was also my preceptor through grant-writing, when many of these ideas were conceived.

As a student and prior to the T32, I was supported by grants NIH/NIAID R01 AI083059 and NIH/NIAID K08AI112432. I would like to thank the PIs of those studies: Dr. Bill Miller and Dr. Audrey Pettifor for MP3 and Dr. Ann Dennis for several studies beyond the study which supported the Aim 1 analyses. Bill, Ann, and Irene went on to serve on my committee.

I cannot say enough about my committee. Each of the 5 members, Bill Miller, Ann Dennis, Irene Doherty, Peter Leone, and Kim Powers, serves a vital role which complements

the rest of the group. I returned to UNC as Irene's research assistant prior to enrolling in the doctoral program. Irene taught me everything that I know about network analysis. Through this work, I was introduced to Bill and Peter who provide excellent balance for me. Both have the ability to instantly see the implications of an analysis (Peter's "30,000 foot view"). The balance struck between Peter's big ideas and Bill's gentle focus guided me through this dissertation. Along the way, Bill introduced me to Ann, who took me on as a research assistant. I worked with Ann for years, during which time she taught me how to apply phylogenetic analysis in a variety of settings. Though I did not work with Kim outside the dissertation analyses, her thoughtful, thorough review of everything that I presented to her has made me a better epidemiologist. Beyond the professional aspects of working with my committee, each member has also been an incredible source of support. Bill, especially, has been my advisor for the entirety of the PhD program. His calm guidance has shaped the program for me.

My family deserves recognition for their support as well. I must express gratitude to my parents, Deb and Joe, and my brother, Joseph, for their encouragement of all of my schooling, their help especially toward the end of this program, and their perspective conveyed through humor.

I will never be able to adequately express my thanks to my husband, Marc. Marc's generous and unwavering support started with my decision to apply to the doctoral program and has been a constant in our life throughout. Marc inspires me to try to be better; I am grateful every day that he is my best friend, husband, and partner. Though this dissertation is dedicated to him, I would also like to acknowledge our daughter, Ellie, who has brought a joyful new dimension to my life. Through her, I finally, truly understand stochasticity.

Most of all, I would like to acknowledge the dedicated members of the North Carolina Division of Health and Human Services Communicable Disease Branch, the disease intervention specialists serving North Carolina, and the HIV patients of North Carolina. Each person whom I have met at DHHS works tirelessly and remains committed to improving the

health of all North Carolinians. The entire group has been welcoming and supportive of me and other doctoral students and UNC, but I would especially like to thank Lynne Sampson, John Barnhart, Jason Maxwell, and Erika Samoff for their support and time. Lynne, John, Jason, Erika, and Anna Cope, a CDC fellow who works with the group, were always more than willing to make time for me to ask questions. Although I did not work with them directly on this project, the disease intervention specialists also deserve my gratitude. I have been fortunate enough to get to know some of them during past studies and can attest to their dedication to the people of North Carolina.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

AA	Amino acid
AHI	Acute HIV infection
AIDS	Acquired immune deficiency syndrome
ART	Antiretroviral therapy
ARV	Antiretroviral
AUC	Area under the curve
AZT	Azidothymidine (Zidovudine)
CDC	Centers for Disease Control
CTRS	Counseling, testing, and referral services
DIS	Disease Intervention Specialist
DRM	Drug resistance mutation
ERGM	Exponential random graph model
EVC	Eigenvector centrality score
GEE	Generalized estimating equation
GI	Gastrointestinal
GIS	Geographic information system
GUD	Genital ulcer disease
HIV	Human immunodeficiency virus
HSV	Herpes simplex virus
ID	Infectious disease
IDU	Intravenous drug use(r)
INI	Integrase inhibitor
MCMC	Markov chain Monte Carlo
MSM	Men who have sex with men
MSM/W	Men who have sex with men and women

MSW	Men who have sex with women
NC	North Carolina
NC EDSS	North Carolina Electronic Disease Surveillance System
NIH	National Institute of Health
NNRTI	Non-nucleoside reverse transcriptase inhibitor
NRTI	Nucleoside reverse transcriptase inhibitor
PI	Protease inhibitor
PLWHA	People living with HIV or AIDS
PEP	Post-exposure prophylaxis
PrEP	Pre-exposure prophylaxis
R6	North Carolina HIV/STD Control Region 6
ROC	Receiver operator characteristics
RWB	Random walk betweenness
SAS	Statistical Analysis Software®
SNA	Social network analysis
STD	Sexually transmitted disease
STI	Sexually transmitted infection
TC	Transmission cluster
TDRM	Transmitted drug resistance mutation
UNC	University of North Carolina at Chapel Hill
US	United States
YBMSM	Young Black men who have sex with men

I. CHAPTER ONE: SPECIFIC AIMS

Despite intervention, HIV continues to spread in NC, particularly among young Black men.[21,22] HIV prevention requires timely identification of infected persons, intervening to stop onward transmission, identification of persons at high risk of acquiring the disease, and intervening to prevent acquisition. Identifying persons at highest risk requires consideration of the people and the environment since assessing HIV risk based solely upon individual behaviors and attributes fails to place persons in context.

HIV acquisition risk is a product of behavior and HIV prevalence in a person's sexual network.[23,24] HIV transmission in a population is a complex process, involving individual-level factors (personal risk behaviors, STI co-infection[16]); partnership-, or dyad-,level factors (sexual practices during the act, HIV serodiscordance, condom use), and network-level factors (HIV prevalence in the sexual network,[24] behavior regarding partnership length, and concurrent partnerships[25]). Thus, analysis of local sexual networks is crucial to reduce HIV transmission. However, this must include assessment of the data gaps and limitations.

In North Carolina (NC), persons newly diagnosed with HIV are interviewed to elicit partners, creating a robust source of individual, partner, and behavioral data. Due to the comprehensive interviews conducted by Disease Intervention Specialists (DIS) at the time of diagnosis in NC, individual-level and partnership-level data are available for each person interviewed. A sexual network can be constructed based upon the name-based partnership data elicited by DIS for partner tracing and testing. Network-level covariates can be calculated for each person in the sexual network once the network is constructed.

For this study we constructed sexual networks from named partner ties and phylogenetic linkages. All cases age 14 years or older at diagnosis in our area under study were used,

allowing us to make inter-group comparisons without excluding persons who may unexpectedly be involved in transmission. Combining sexual network and phylogenetic analyses provided a more complete picture of transmission. The aims are steps to support the stated purpose of intervening to disrupt transmission.

Knowing network structure, partner selection habits, and place-associated traits could provide a more accurate estimation of HIV acquisition risk to inform pre-exposure prophylaxis (PrEP) program targets. Combining partnership data with phylogenetic analysis from an epidemiological perspective is novel and has real-world application. The study will improve use of surveillance data, particularly through its exploration of limitations, with the intention of translating this research into both clinical and public health practice. Specifically, we:

Aim 1: Compare phylogenetic links and named partner ties in a small, defined geographic area.

Overview: We compared phylogenetic links and named partner ties in a small, defined geographic area. The purpose was to identify the limitations of contact tracing, so that the gaps in the sexual network could be identified. The sexual network is the basis for some public health HIV prevention efforts. We constructed a sexual network from HIV cases diagnosed 2012-2013 in Wake County, NC and their partners. Transmission clusters were constructed from HIV *pol* gene sequences. Available HIV sequences were analyzed to determine whether there was co-occurrence of named partner ties and phylogenetic ties. Generalized estimating equations (GEE) compared gene sequence linkages and sexual network ties to identify contact tracing gaps.

Hypothesis 1.1: Phylogenetic analysis will identify gaps in contact tracing, including characteristics of persons newly diagnosed whose risk is not well-represented by interview and partner elicitation.

Hypothesis 1.2: Behavioral or contextual factors are associated with not having well-enumerated partners in the contact tracing network.

Aim 2: Determine which individual- and network-level traits predict future HIV outcomes for persons who are part of sexual networks with circulating HIV in central NC.

Overview: The primary purpose of this study is the Aim 2 analysis, which is to analyze whether certain individual-level characteristics or behaviors, partnership-level attributes, or network-level structures were associated with being named as a case or a partner in future transmission in our high-risk sexual network components. We constructed a sexual network from HIV cases diagnosed 2012-2013 in central NC and their partners. We included partnership data from 2012-2013 syphilis investigations involving these persons in the same geographic area to increase network completeness. HIV index cases were followed through 2016 to determine continued involvement in the sexual network, defined as being identified as a partner on future cases or being diagnosed with a STI 6 or more months after HIV diagnosis.

Predictors included individual- and dyad-level characteristics routinely collected during DIS interviews so that the derived risk score can easily be applied during DIS interview. The predictive model was constructed so that DIS may note any factors elicited during the primary interview that may lead to the long-term (3-year) outcome. Highlighting opportunities for DIS to identify patients who would benefit from enhanced support to link to care may lead to fewer cases of onward transmission if patients with higher transmission risk potential are engaged in care and achieve viral suppression. This model has the potential to effectively disrupt transmission in NC if new cases are averted. Unlike other areas of the United States (US) where HIV transmission can be disproportionately attributed to acutely-infected persons[26] or those unaware of their infection,[27] HIV transmission in NC tends to involve persons who are chronically infected and aware of their own status.[28]

Hypothesis 2.1: Behavioral and contextual factors, collected for public health purposes, will successfully predict which HIV-positive persons have highest transmission risk potential.

II. CHAPTER TWO: BACKGROUND AND SIGNIFICANCE

A. Background

1. HIV Trends in the United States

HIV rates and acquisition risk differ by race, age, geographic region, and behavioral risk factor. In 2014, 44% of new HIV infections were diagnosed among Black persons, despite only comprising 12% of the population.[29] Though the rate of new diagnoses has declined in Black persons since 2010, the rate of new HIV diagnoses in 2015 was 2.7 times the rate of new diagnoses amongst Hispanics and Latinos, the racial/ethnic group with the next highest rate, and 8 times higher than the rate amongst Whites (per 100,000 population, rates were 44.3 for Black persons, 16.4 for Hispanic/Latino, and 5.5 for White).[30] From 2006-2009, HIV incidence increased by 48% among young African American men in the US.[5]

Sexual transmission is responsible for most new cases of HIV in the US.[30] In the US, men who have sex with men (MSM) are at a higher risk of acquiring HIV than men of the same race who have sex with women.[4,29] Among MSM, most diagnoses are made in the younger age groups (20-29 years);[5,31] women with heterosexually-acquired HIV tend to be diagnosed slightly later (25-34 years).[30]

Geographically, the rates of incident HIV diagnoses are much higher in the US Northeast and South than the West and the Midwest. Further comparison of the higher rates US Northeast and the US South yields racial differences; in both places, 75% of people living with AIDS were either Black or Hispanic, although Hispanics accounted for a smaller percentage in the US South than in the Northeast.[32]

On a smaller scale, geographic core areas have been identified for several sexually transmitted infections.[33] Risk of having primary or secondary syphilis was 4.6 times higher for

persons living in a certain area of San Francisco between 1985 and 2007. The spatial analysis was able to separate core and outbreak areas.[34] HIV-positive persons resided closer to their partners in Colorado Springs than at risk persons who are not HIV-positive and their partners.[35] During a syphilis outbreak in Baltimore, two areas were identified as core areas from which the outbreak spread and a new core area was created. Even after the outbreak ended, density of cases remained higher in all 3 core areas.[36]

2. HIV Transmission Dynamics in North Carolina

a. Epidemiology

The South has the highest rate of new HIV infections in the US[30] and the highest number of adult and adolescent persons living with HIV or AIDS (PLWHA).[32,37] North Carolina (NC) had the 10th highest rate of new HIV infections in the U.S. in 2015.[30] There were 1,345 new HIV diagnoses in NC in 2015, with the overwhelming majority due to sexual transmission.[22] In NC, generally more than one-quarter of persons diagnosed with HIV are concurrently diagnosed with AIDS.[38] NC has already met 90-90-90 goal for diagnosis; it is estimated that 10% of HIV-infected individuals in NC are unaware of their status.[39]

Approximately 70-80% of non-pediatric HIV cases diagnosed in NC each year are among men. Being MSM is the most significant HIV acquisition risk factor for men in NC. In 2015, 747 of the 1,078 (69%) non-pediatric HIV cases diagnosed among men had male sexual partners noted as a risk factor.[22] Among all young men in NC, the major risk factor is being MSM.[4,38]

As is typical of other areas in the Southeastern US, the largest proportion of HIV cases diagnosed are in Black persons (2015 adult/adolescent diagnoses: 62% men and 71% women were Black). Black persons also had the highest rates of new HIV diagnoses in NC in 2015, at 80.3 and 18.7 per 100,000 adult/adolescent population among men and women, respectively. The rate among Black persons of both genders was 47.0 per 100,000 population, 2.5 times

higher than the rate among Hispanic/Latino persons (18.7 per 100,000 population) and 8.5 times higher than the rate among White persons (5.5 per 100,000 population).[22]

The highest rates of AIDS diagnoses were also made among Black persons, including both concurrent diagnoses and progression from HIV.[38] The rate of non-pediatric AIDS diagnoses was 40.3 per 100,000 population for Black men and 17.9 per 100,000 Black women in NC in 2012. For comparison, the rates among White men and women were 4.2 and 2.7 per 100,000 adult/adolescent population, respectively.[22]

b. Social Factors

A previous study of HIV diagnoses in NC identified an epidemic among college students, particularly young Black men.[40] The investigators reviewed HIV records for newly diagnosed men age 18-30 from 69 of 100 counties in NC in years 2000-2003 and found 735 new diagnoses, of which 84 (11.4%) were college students. Only 3.6% of the college men reported having sex with women only in the 12 months prior to diagnosis, compared to 29.5% of the non-college men. 33.3% of the college men reported having sex with both men and women, as did 11.7% of the non-college men.[40] When compared to men who were not enrolled in a college or university, college men were significantly more likely to go to bars and clubs, meet partners in bars and clubs, use ecstasy, and use the internet to meet partners. However, the men in college were significantly less likely to have partners with known HIV infection, exchange sex for drugs or money, use crack-cocaine, or have a history of incarceration. The men in college were significantly more likely to be African American and significantly less likely to be Latino. The HIV incidence rate among African American college students increased dramatically during the study period, from 15 per 100,000 persons to 79 per 100,000 persons.[40]

Having college students as sexual partners, along with having anonymous or internet partners, is associated with being coinfected with HIV and syphilis in NC.[41] Despite the significance of these trends, neither race nor college enrollment impart any biological risk of HIV or STI acquisition. Instead, there are social and contextual factors at play and trends among

broad demographic categories do not adequately describe the epidemic. Methods that discern trends on a smaller scale are necessary to understand where and why transmission continues to occur.

A second study of young men in NC increased the study period of the first study through 2004 and added behavioral data collected by voluntary counseling and testing centers with the purpose of constructing the sexual network. The network was based on a total of 1013 available records of incident HIV diagnoses made among men age 18-30 that were deemed to have sufficient partner information for the 12 month period prior to diagnosis (279 records were unavailable and excluded). MSM/W (n=161) were compared with MSM (n=573); being MSM/W was significantly associated with having more than 10 partners in the 12 month period prior to diagnosis and being enrolled in college. MSM/W were also more likely to be central in the sexual network, although the investigators do not specify whether they are using betweenness or degree centrality, the latter of which is synonymous with number of partners. They do report, however, that MSM/W bridged several network components,[42] a finding that has been demonstrated elsewhere.[43]

Several outbreaks of HIV and sexually-transmitted infections (STIs) in NC have been investigated in different populations throughout the state, highlighting the importance of studying the sexual networks (see section II.5. for an explanation of sociosexual network analysis and structures). Counties with fluctuating syphilis incidence or incidence only rising toward the end of the study period had visibly different network structures.[44] All counties were found to be assortative with respect to race/ethnicity. Non-outbreak counties were found to be more assortative with respect to number of partners (degree), largely because persons in non-outbreak counties tended to form smaller monogamous dyads.[45]

Exchange sex and crack-cocaine use are important drivers of syphilis in NC; in one rural county, half of the people diagnosed with primary, secondary, or early latent syphilis during a 14-month period reported either using crack-cocaine or having a partner that did.[46] Syphilis is

endemic in NC with epidemic outbreaks that can be seen in the sexual networks. Sexual networks in counties in rural southern/southeastern NC with epidemic syphilis were more densely connected than in counties with fluctuating incidence. The counties with the highest incidence had distinct features of the sexual network, including more closed loops and larger, more connected components overall.

NC has one of the fastest-growing Latino populations in the US.[47] US- and foreign-born Latinos in NC have distinct HIV risk factors. US-born Latinos report more lifetime sexual partners and males are more likely to report engaging in sexual activity with another man. Foreign-born Latinos in NC are more likely to report exchange sex and are more likely to speak only Spanish or report low acculturation.[48] This likely contributes to why foreign-born Latinos in NC are more likely than US-born Latinos to present to care with advanced disease;[48] HIV-infected Latinos in NC are predominantly immigrants and are more likely to present to HIV care later than Black or White patients.[12] Males in each racial/ethnic group are more likely to present later to care than women in the same group, with a similar proportion of female Latinas presenting late as White men.[12] Phylogenetic analysis of sequences collected from a clinical cohort showed that Latinos were more likely to be in linked pairs but less likely to be in clusters when compared to White and Black men.[49] This may be due to foreign-born patients acquiring infection prior to arriving in NC; US-born Latinos were more likely to integrate in MSM clusters.

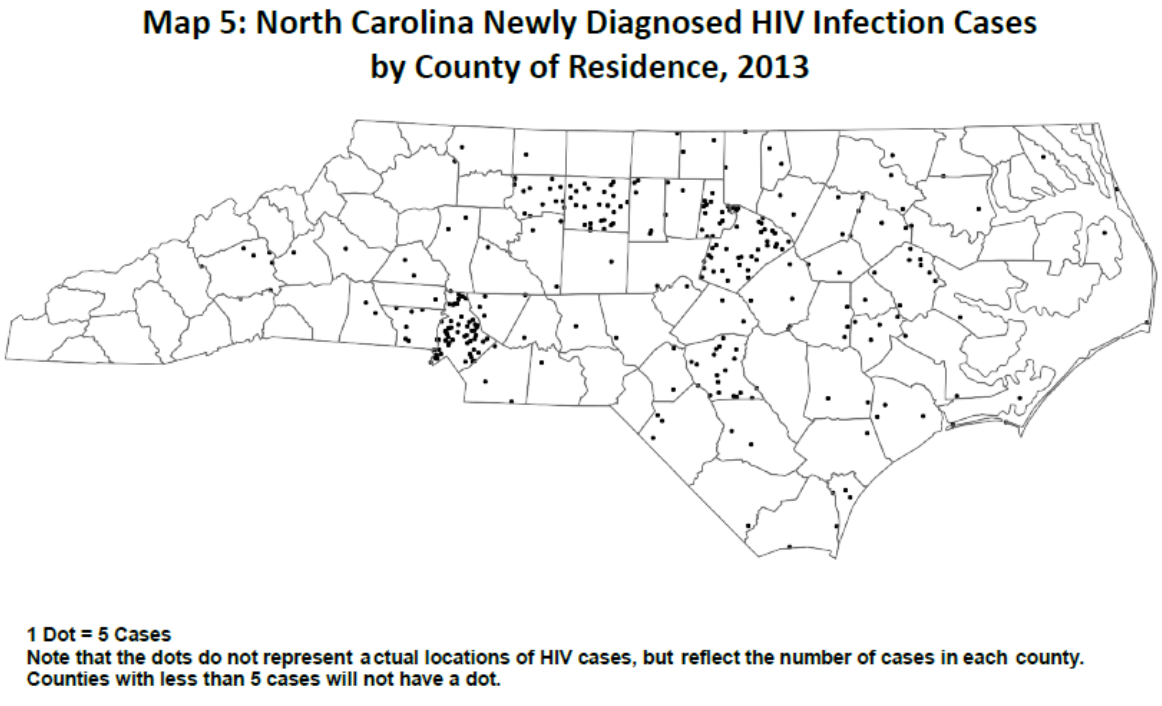
c. Geography

Distinct geographic regions and populations make HIV initiatives in NC challenging, as interventions are often more successful when they are tailored to a population. Syphilis was found to co-cluster with gonorrhea in NC; state-wide mapping of gonorrhea and syphilis over time identified 20 core areas for gonorrhea and 10 for syphilis. All of the syphilis core areas were found to have at least some overlap with at least one gonorrhea core area. All clusters, for gonorrhea and syphilis, were found to be associated with an urban area; some areas existed entirely in urban areas and some encompassed but urban and rural but none were entirely

rural.[50] The rural-urban divide has been noted for other STIs in NC. Drug susceptible HIV strains were significantly more genetically similar than drug resistant strains for rural-rural and rural-urban partnerships, but not for urban-urban partnerships. Urbanicity of residence was not associated with TDRM v. drug susceptible virus among persons with acute HIV infection.[51] Geographically-associated network cores have been found to significantly contribute to STI spread,[50,52] but it is unknown whether the same relationship is found with drug resistant HIV, particularly in rural areas.

Geographic and spatial investigation of HIV and STI outbreaks in NC has yielded important information. The CDC found that NC had the highest burden of HIV in non-urban areas in 2006.[4] Figure 1, from the 2013 HIV/STD NC Epidemiologic Profile report, shows incident and prevalent HIV cases across the state. The largest clusters are in the most urban areas, but the rural eastern part of the state has a high burden of HIV without having as many specialists and providers as the central part of the state.

Figure 1. Newly-diagnosed HIV cases by county of residence across North Carolina, 2013[7]



Spatiotemporal analysis was applied to identify core clusters of gonorrhea and syphilis in NC, with an additional assessment of rurality. All of the syphilis and gonorrhea core areas included at least one urban area, and all of the syphilis core areas (N=10) overlapped with gonorrhea core areas (N=20).[50] Similarly, in Wake County, a single urban county in central NC, chlamydia, gonorrhea, primary and secondary syphilis, and HIV were found to cluster with a single identifiable core area; all four core areas overlapped.[53] These studies show that in NC, there is geographic overlap of several STIs. Without added analysis of the sexual networks, it is unknown whether the STIs are circulating among different groups.

As HIV is more easily transmitted in the presence of certain STIs, future areas of research should include both network and geographic or spatial analysis. Sexual transmission of HIV is likely to occur along racial and geographic lines, as partnerships tend to be assortative with respect to race.[9,25,49,54] Examining partnership patterns and geographic differences, and variation between populations that tend to form discrete sexual network components with little overlap, may provide new avenues for intervention.

3. Individual Characteristics and Differential Risk

Despite many targeted interventions, incidence continues to rise among young Black MSM. There are many levels of reasons to account for this. Higher level sexual network structure variables include HIV prevalence and network density. A particularly high risk network comprised largely of Black MSM in NC was found to have a HIV prevalence of 29%,[21] so people entering that network assume the risk of such a high prevalence. Dyad-level variables are those surrounding each partnership, such as condom use. Examples of individual-level characteristics include demographics and infection with ulcerative STIs. In order to simplify, much of the research focuses on individual-level or dyad-level characteristics in aggregate, although often the dyad-level characteristics are aggregated at the individual-level: for instance, instead of looking at each sexual act, condom use is often attributed as a category value to the individual. One study found that Black MSM were less likely to know the HIV status of their

partners,[9] which is an example of a compelling dyad-level factor that was measured at an individual-level.

To continue with the use of Black MSM as an example, some network-level factors are important in aggregate, such as network prevalence or proportion of anonymous partners because both cascade to the more focused levels. However, it does not serve the individual people to look at consistency of condom use within a network if the individual consistently uses (or does not use) condoms, as his risk is mediated by his own behavior. Number of partners is similarly most often attributed to a person, but is clearly modified by his extended network: someone with one partner is at great risk if his partner has many other partners. The converse is true as well, where someone with many partners isn't substantially increasing his own risk if each of his partners has no others. In some way and barring an active infection with an ulcerative STI, number of partners should not be a separate variable from percent condom use if both variables are aggregated to an individual observation, as it is plausible that there would then be multiplicative modification within the individual-level observation.

HIV trends by group are invaluable to assess risk, but there is more nuance in risk as traits such as race or age do not perfectly correlate. Agent-based approaches have long been eschewed because they are time- and resource-intensive, although recent adoption of electronic medical records and more flexible databases has made it easier to obtain partner information and the HIV sequences performed at entry into care. Sociometric networks permit complete construction of a case, including his or her partners, and an iterative gathering of their partners.[11] With the complete network, demographic data, and behavioral data, we can start with complete cases in order to assess risk by network-level, dyad-level, and individual-level characteristics, particularly if we categorize by HIV infection status (chronically, recently, or not).

Risk factors for acquiring drug resistance with HIV infection may differ somewhat from the risk factors for acquiring drug susceptible HIV. Phylogenetic cluster analysis showed serial acute transmission among MSM.[27] As there is a probabilistic element to certain regions of the

transmitted/founder strain,[55] it stands to reason that someone who is acutely infected with drug resistant HIV would be more likely to then pass that strain along while the resistant strain is still predominant. In support of this, MSM in Europe have significantly higher TDRM prevalence than other groups.[56]

Although Black MSM have disproportionately high risk of acquiring both drug susceptible and drug resistant HIV, we believe there is merit in using all cases in this geographic area. First, other groups are at risk, even if that risk is smaller. Second, studying the similarities and differences between risk groups may provide insight to the cause of the higher risk, as there were comparison groups. Third, we were able to see if there is overlap in or bridging among the network components by race and risk group. Fourth and finally, networks are often incomplete due to the high number of anonymous partners and undiagnosed cases, both of which warrant as much investigation as possible. Interview of other racial, ethnic, and gender groups may be able to describe inter-racial partnerships which would be unknown due to the inability to interview persons missing from the network, particularly since Black MSM are less likely to get diagnosed than other groups.

4. Population Mixing Patterns and Examination of Person-to-Person Linkages

Several years ago, a rash of studies claimed that obesity and behaviors such as smoking are 'contagious'.[57-60] The idea of behavioral contagion is that people tend to cluster with others who practice similar behaviors and also influence the behaviors of their peers, which can be to the detriment of the entire group when the behavior is associated with greater risk of disease.[61] Assortativity in risk behaviors such as substance use combined with sexual activity,[62] engaging in group sex, or engaging in unprotected sex could all be related to influence or homophily in the social network: Black men who can identify an 'enabler', defined by the investigators as someone who engaged in risky behaviors, in their social (not sexual) network were more likely to engage in unprotected anal intercourse.[61]

Modeling studies have found that the largest epidemics result from an assortative core group with high rates of partner change that links disassortatively to peripheral groups.[63-65] The STI spreads rapidly within the core group that practices similar risky behaviors, such as rapid partner change. Sexual mixing between core and peripheral groups that do not tend to practice the same risky behaviors then leads to a wider-spreading infection that becomes an epidemic.[63,64,66] There is a tendency to select sexual partners of similar age, race, education level, and demographics,[45,54,57,62,67] which could also lead to an increase in circulating infection within a core high-risk group.[66,68] Such assortativity was observed in a syphilis epidemic in southeastern NC. Non-outbreak counties were found to have high rates of assortativity with respect to number of partners, while outbreak counties were more likely to be disassortative.[45] Substantial disassortative partnerships have been observed in some cases, and linked to bridging populations for infection transmission.[45,69,70] Dis-assortativity in number of partners was observed in central NC among women trading sex for crack-cocaine (unpublished, PI: Irene Doherty, PhD). This sustained the local syphilis epidemic among heterosexuals, where a few females had many male partners.

Networks of people are unique because ties between individuals can form on the basis of extra-actor features of social processes.[71] Actors can form ties with someone new because that person is already connected to someone with whom he or she is connected (transitivity).[71-73] Special classes of models are required to parse the effects of actor traits and extra-actor processes and study the linkages between persons in a social or sexual network.[72,74,75] If all of the actors share a trait then the new connection could incorrectly be attributed to homophily rather than the underlying influence of the social network itself.[76]

Inherent difficulties in the study of sexual behavior and HIV include a difference in timing of infection and diagnosis, which can lag for years; more heavily affected hard-to-find marginalized populations; and difficulty in measuring societal and external forces, which exert pressure upon partner selection and sexual behavior. In order to simplify, behavior following

diagnosis is often used as a measure of behavior at the time of infection[77,78] and an individual's behavior is often treated as fixed (until diagnosis or declining health).[79] However, these simplifying assumptions may obscure vital information. A prospective cohort study found that individual behavior does fluctuate, even within short periods of time.[80] Several modeling studies have shown that variable individual behavior affects outbreak size[81] and observed rates of transmission by HIV stage.[81-84] Zhang, *et al.* (2013)[85] modeled changing periods of high- and low-risk behavior. They found that not only were individuals more likely to be infected during high-risk periods but that they were also more likely to transmit as the period of high risk did not end abruptly with infection. Additionally, many studies simplify by using more broad demographic groups or geographic regions which may mask the specifics of the person-to-person interactions leading to transmission – examination at a finer scale has the potential to reveal mixing patterns important to transmission; for instance, local TDRM trends are distinct even from regional trends.[86]

We used reportable disease data combined with surveillance data to examine the network of individual partnerships. Sexually transmitted infections allow analysis of transmission and behaviors that is not possible with other types of diseases. As opposed to other infectious diseases, which can infect anyone, sexually transmitted infections only infect the sexually active population[87] and are not transmitted through casual contact. Thus, with enough information about partnerships, one can construct transmission chains of the virus through the population. Molecular techniques allow us to differentiate between acute/recent infections and chronic infections. Categorizing based upon recency of infection permits analysis of behaviors prior to infection. Comparing recently infected, uninfected, and chronically infected persons will demonstrate differences in behaviors and allow identification of practices associated with HIV acquisition. This analysis has the potential to avoid issues in assessment related to the timing of diagnosis rather than the timing of infection. Behavior in the time prior to diagnosis has not been well-studied, although there is some evidence of higher rates of risk-

taking behavior immediately preceding infection.[79] It is the time prior to infection which is related to risk of acquiring HIV, and it is unknown whether the behaviors and partnerships of individuals who are HIV-uninfected but at risk differ from those who are HIV-negative and not at risk or chronically infected persons.

Social network analysis (SNA) and phylogenetic analysis are two relatively new fields which permit *post hoc* study of person-to-person linkages. SNA can quantify likeness among partner selection[88] in terms of risk or demographic characteristics, which is an analytical approach to population mixing scaled down to the level of the individual. Phylogenetic analysis shows likeness of infection and can identify clusters of acute transmission. Both allow us to link individuals in the sexual transmission of HIV.

5. Social and Sexual Network Analysis

Social network analysis (SNA) is a method that permits examination of relationships (dyad-level) and individual-level characteristics.[23,89,90] A social network is comprised of all individuals participating in the social system under study.[91] A pair of linked persons is a dyad. The network is grouped into components, which are clusters of linked individuals.[89,90] The "network" is the entire set of components (Table 1 includes network terminology used in this document). A sexual network is a specific type of social network in which linked individuals are grouped into components based upon having self-reported sexual relationships during the time period under study.[11] Network analysis includes a visual representation of the individuals (nodes) and their linkages (edges) to each other. Network summary measures are often descriptive, and include various measures of a node's centrality in the network, the number of edges per node (degree), and the proportion of possible connections that are made (density).[89,90]

Table 1. Network terminology used in this paper.

Object Being Described	Network Analysis Term
Person in the network	Node, actor, individual
Partnership between persons	Edge, link, linkage, tie
Group of linked persons	Component, Network component

Once the network is drawn and components (a group of linked persons) are identified, the combination of characteristics of individuals in high-risk network components can be compared to individuals in other components to determine the self-sorting factors that influence risk in the local epidemic. Disparities in risk of sexual acquisition of HIV continue to exist in part because social and sexual networks tend to be positively assortative, which means that connections are more frequently made between individuals of like age, race, educational status, and behaviors, including number of sexual partners.[11,45,54,62,92] This encourages transmission within a like group of people. While it is not unusual to see groups of demographically-similar persons linked together,[93] racial mixing has increased in some areas.[94] Additionally, some groups which may appear to be a homogeneous group are not: a study of Latinos in NC found that foreign-born Latinos and US-born Latinos had different risk factors.[48] These studies demonstrate that it is a mistake to oversimplify and focus on risk by demographic group alone.

The combination of factors contributing to HIV risk is complex and distinctions such as race, age, and sex are often too coarse to gain an understanding of the epidemic.[49] Interventions based upon coarse or crude distinctions are limited and may be missing individual-level factors leading to transmission. Sexual network analysis permits examination of local and individual-level characteristics.[23] Examining network structure and the characteristics in a network component provides insight into the dynamics influencing risk at a much finer scale.[44,95]

Previous studies employing sexual networking methods in NC have found distinct sexual networks with limited interaction between the networks and different circulating STIs,[40,42,44,96] suggesting that HIV drug resistance mutations may also be distinct by network component. The most-studied networks in NC are young, Black MSM; older individuals trading sex for crack-cocaine; and rural, White individuals. Partner selection tends to be assortative,[11,45,92] and communities also tend to exhibit homophily over diversity.[97]

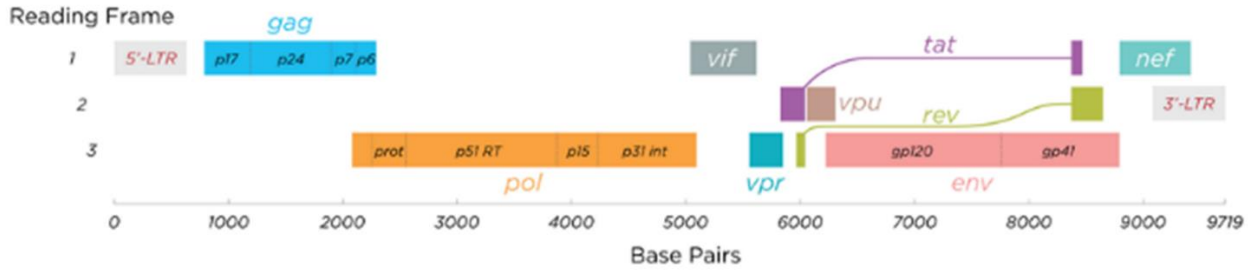
The structure of the sexual network may also provide clues to the processes including tie formation.[98] Closed triangles in a social network indicate that existing relationships influence

the formation of other relationships. This is an endogenous process, where the existing structure of the network supports the formation of another tie. Assortativity of race or another attribute is an exogenous process, where people tend to form relationships based upon shared characteristics. Closed triangles are often seen in MSM sexual networks, though not in heterosexual networks. As noted above, assortativity does play a role in partner selection. Certain network structures, such as *k*-cores (a closed loop where each person in the loop has at least *k* edges) and closed triangles have been shown to influence disease transmission in modeling studies, particularly if concurrency is present.

6. HIV Sequence Analysis

The plasma *pol* region of the HIV genome (Figure 2) is variable and encodes drug resistance mutations.[99] Knowing the infection phenotype is valuable for clinical care decisions, as it conveys which combination of ARV drugs would be most effective for the patient. Thus, obtaining this information is now standard care in most of the US.[100] In NC, blood samples are routinely collected at entry into care for drug resistance testing to determine whether the patient's virus has encoded drug resistance.

Figure 2. HIV genome showing position of genes.[6] The polymerase (*pol*) gene encodes drug resistance.



The *pol* gene sequences collected during drug resistance testing can also be used to identify transmission clusters of persons with genetically similar virus and assess transmission chains, which has public health utility. The potential to link infections through phylogenetics may help link to[101] or identify anonymous individuals in the network who aren't identified by during partner notification.

Acute/recent infections are more likely to be identified in clusters defined by short genetic distances because less time has elapsed between transmission and sampling, so there is less genetic divergence.[102] Coalescent models were used to demonstrate that some of the excess clustering of sequences obtained during acute infection is due to this and not solely excess transmission during acute infection. A study of publicly available *pol* sequences (n=84,527) representing 141 countries was undertaken with the intent to build a "global transmission network" by looking for similarities worldwide.[103] Interestingly, the investigators found an inverse relationship between the number of linked sequences and the amount of drug resistance mutations (DRM) encoded in the sequences, demonstrating that we do not have a clear picture yet of how drug resistance circulates in populations because it appears as though having a lot of transmission is associated with having less risk drug resistance. An alternate explanation, however, not presented by the investigators is that groups with more DRM are those who have more access to ARV which explains both the lower transmission and the higher circulating DRM.

a. HIV transmission cluster analysis applied to population-based research

A recent investigation of an increase in acute HIV diagnoses in the area around Charlotte, NC and in Western NC was unable to phylogenetically link all of the acute cases with the infections acquired from either chronically-infected individuals in a single cluster, or from anonymous partners who could not be located for testing.[104] However, sexual networks were constructed, with two distinct groups noted. Although both locations are nearby, they are geographically distinct and differ by rurality. In the metropolitan area around Charlotte, young Black MSM accounted for most of the diagnoses while older white MSM accounted for most of the diagnoses in rural Western NC. No significant overlap between the two groups was found using partner trace back or HIV sequence analysis, so the increase in acute diagnoses was likely due to better case finding and diagnosis rather than an outbreak of acute HIV. Importantly, the acute cases had on average fewer than one located HIV-positive partner, so

partner finding was not very successful for these cases as each case must have had at least 1 recent HIV-positive partner to be acutely infected.

A study of 1671 HIV-positive persons enrolled at two spatially-near, university-based HIV clinical cohorts in NC was able to link 557 of the patients, the largest cluster including 36 patients.[49] Clustering was largely seen along racial lines, although not by ethnicity as Latinos were significantly less likely to cluster than non-Latinos. There were MSM and heterosexual clusters, although there were mixed clusters as well. [49] Phylogenetic analysis is a powerful tool for examining transmission patterns and delineating trends, although the likelihood of finding clusters can be reduced if time has passed between samples[102] or one of the patients has been exposed to ART. The ability to construct a large network of individuals using partner data obtained via interview supplements the linkages identified using phylogenetic analysis.

b. Limitations

Due to the limitations of HIV sequence analysis, phylogenetic data alone is not as powerful as the combination of phylogenetic and partnership data. First, neither first-degree partnerships nor directionality can be inferred from HIV sequence analysis. Second, observed cluster size may not represent actual transmission if there is a high proportion of missing data, which can occur at any of the first steps along the HIV care cascade. Third, cluster size is affected by cluster definition; if percent difference is used then cluster size changes with the cut-off selected. Fourth, sequences obtained for clinical care are a consensus sequence, where the sequence returned represents the most frequent base pair observed at any position after sampling multiple viruses within the host. Therefore, minor variants are not captured and intrahost variability is unknown. A consensus sequence may also have ambiguous sites if a position has undergone mutations and is not clearly represented by any single base pair. Having many ambiguous sites may affect clustering.

7. Combining Sexual Network and HIV Sequence Analysis

Sexual network analysis and HIV sequence analysis examine the relationships between individuals which account for HIV acquisition risk. The network constructed from DIS interviews can be compared to the phylogenetic tree. In a previous study of sexual networks in NC, 50% of partners were anonymous, as defined by inability of the State to locate the individual for testing due to lack of identifying information,[96] and the addition of gene sequences to the network may help identify that persons are directly or indirectly linked in a transmission chain even without having disclosed partner information. The distribution of branching points for each terminal node in the phylogenetic tree can yield the underlying network structure, which may provide a clearer picture of transmission in NC, as the tree topology is not hindered by anonymous contacts or encumbered by contacts that do not result in virus transmission.[105] Additionally, adding phylogenetic data to network data may clarify temporal trends in transmission, which are not always clear with the network contact data alone. The network data complements the phylogenetic data because it is not always clear in a tree how the transmission events occurred from partner to partner.[106]

Previous studies have only found little to moderate overlap between contact data and phylogenetic trees constructed from sequences,[101,107] but the analysis described here is improved because it used all reported cases in a geographic area rather than a sample, interview questions about partner contact dates were targeted based upon the stage of infection, and sequences are now routinely collected at entry into care. Using all cases provides a complete picture of sexual transmission of HIV in this geographic area, allows assessment of homophily and bridging in network component, and allows comparisons of risk between groups without confounding by spatial or geographic parameters.

a. Sexual Network Analysis and Spatial or Geographic Analysis

Applying a combination of methods to infectious disease processes can often yield surprising or unexpected results. Even though shigellosis is often transmitted from person to

person, shigellosis (and cholera) in Matlab, Bangladesh from 1983-2003 was found to cluster in space more so than among kinship networks. This is particularly true for sexual networks, which form within a context that includes geography, place, and culture.[11] Geographic locations can include bars or clubs where people meet partners[11] and important partner-finding venues can be identified through network analysis.[108] In NC, universities are known to play a central role in STI transmission networks.

Spatial analysis permits assessment of distance between partners which can be influenced by many factors. People who meet partners on the internet may not have a strong association with any physical locations, in which case geographic place data would be less helpful. The distances people travel to meet partners may also be influenced by societal or contextual factors. NC is largely a rural state with a few urban centers. Persons living in more rural areas may travel different distances than people living in urban areas. People seeking anonymous encounters may travel greater distances to ensure anonymity.

Applying both SNA and geographic analysis to groups of people at-risk for HIV in Colorado Springs, Colorado demonstrated that spatial relationships varied by social relationship. Examinations of the types of relationships showed that HIV-positive persons and their sexual partners had a significantly smaller distance between residences than prostitutes and their paying partners. Amongst persons who were connected via sexual contact or injection drug use, injection drug users resided closer to their IDU partners (mean distance 4.0 km) than individuals who were only linked to each other through sexual activity (mean distance 6.0 km). However, persons who engaged in sexual activity and injection drug use together had an even smaller mean distance of 3.2 km. Overall, the at-risk persons were more closely related by residence than would be expected compared to the entire population of Colorado Springs.[35]

Analysis of partnerships and sexual networks can enhance geospatial findings. Geographic data alone do not necessarily reflect sexual risk or transmission trends if two discrete networks are present in one geographic area. Geographic information[109] and spatial

data have the power to enrich sexual network analysis. Spatial data may be a better fit with network data than geographic data, however. In addition to being more suited for models and being interpretable on continuous scales, spatial data are a good fit with network data as both require functions to account for clustering.

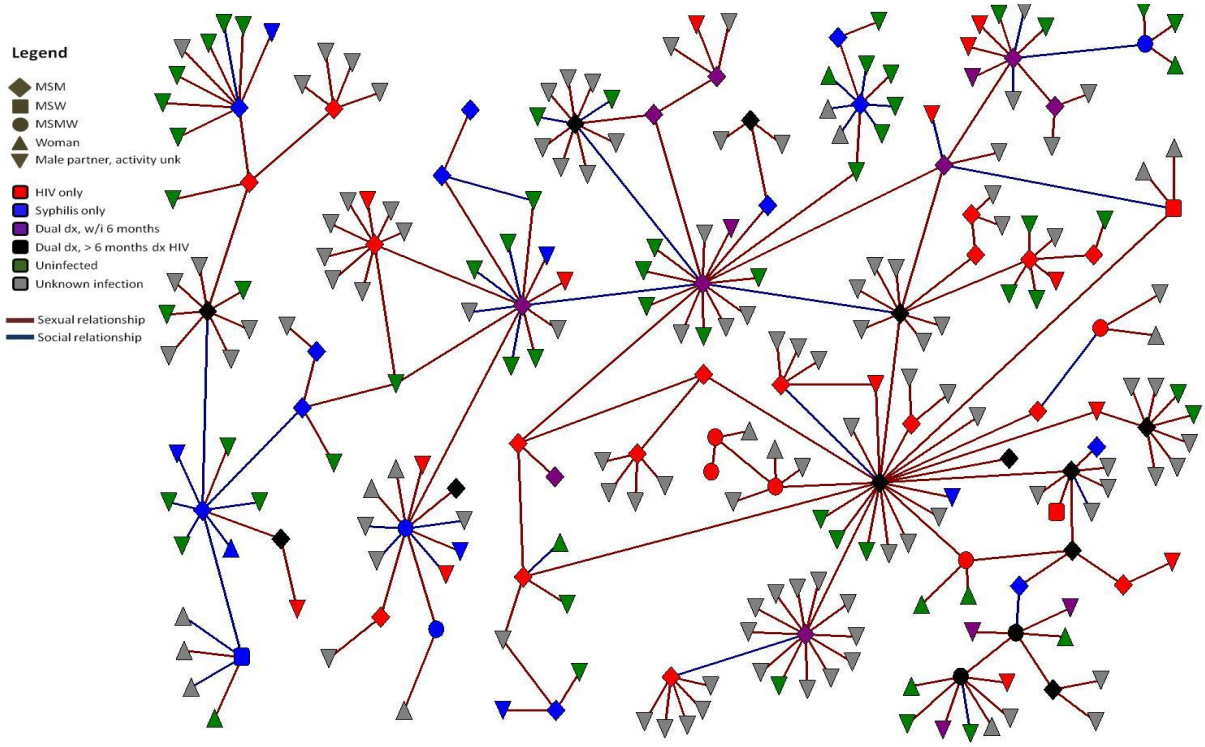
B. Preliminary Studies

A previous study by co-investigator Dr. Irene Doherty, *"The nexus of drugs, sex networks, HIV, and syphilis in young African American MSM"* (hereafter called '*Nexus*') constructed a sexual network using new HIV and syphilis diagnoses made among Black men age 15-30 in the areas around Winston-Salem and Raleigh, NC. Nexus abstracted both sexual partners and at-risk social contacts for the study. We abstracted 1,100 cases and 3,438 contacts (both social and sexual) from name-based reportable disease records and used to construct sexual networks. Dr. Doherty and Ms. Pasquale successfully linked the 4,538 cases and contacts by name to create 508 network components, the largest of which comprises 1,403 individuals (30.9% of the total population). Overall, the network was primarily composed of disjoint components; 90.4% of the nodes could not reach each other.[96]

All 3,438 partners were used to construct the network; of the 3,438 partners, 2080 (60.5%) were unknown because they could not be located for the interview or refused testing.[96] Despite the high proportion of anonymous partners, enough data were present to construct the network, which included several large components. The largest component in Region 3 centered on young Black MSM was comprised of 261 cases and contacts (Figure 3), many of whom were in college. Positive assortativity was present; distinct and discrete network components were seen: 1) one network involved young, Black MSM infected with HIV and syphilis and 2) another network comprised older individuals engaging in transactional sex and mainly infected with syphilis.[96] Nexus demonstrated that in a limited geographic area during a limited time period, Black men age 15-30 separated themselves by age, college status, sexual orientation, and drug use. This present study added to the knowledge gained by Nexus by

abstracting all HIV diagnoses during the region under study from 2012-2013 to compare self-sorted network components by factors such as age, race, education, and risk behavior.

Figure 3. Largest sociosexual network component from the Nexus study (N=261). Graph shows gender/sexual orientation, HIV and syphilis status at the time of network construction, and the types of relationships between persons (sexual or social).



A newer study (*"Integrating HIV Phylogenetics and Sexual Networks to Inform HIV Prevention"*: co-PIs Dr. Ann Dennis and Dr. Irene Doherty; hereafter called '*Phylogenetics/Networks*'), combined data for HIV cases diagnosed in Wake County, NC in 2012 and 2013 with clinical cohort and phylogenetic data to identify HIV clusters and the study the overlap between reported partners and transmission clusters (TC). Sexual network components included all persons linked together based upon named partner ties. TC were defined as being phylogenetically linked based upon short branch lengths and high bootstrap values. The sexual network was constructed using 280 index cases and 422 locatable sexual and social contacts; the final network included 663 unique persons as some of the partners were named by multiple index cases or were index cases themselves. HIV-positive persons

(N=411) were probabilistically matched using date of birth and gender to a dataset of 15,247 background *pol* gene sequences obtained from persons in care in NC; 230 (56.0%) of 411 matched to a sequence. The sexual network components only partially overlapped the phylogenetic clusters, demonstrating the utility of adding phylogenetic information to the network (see Chapter 5). One hundred and seventy-one persons were in a TC with at least one other sequence in the background dataset. Eighty-seven persons were in a TC with at least one other person in the study. Overall, only 42 of 87 (48%) persons in the study were in a TC with at least one other person in the cluster. This could possibly be due to poor recall, transmission prior to the DIS interview period of interest, or having anonymous partnerships with a person that could not be identified and located. However, persons in named-tie heterosexual partnerships were always in the same study TC when both persons had a genetic sequence available; this was not the case for MSM partnerships. Phylogenetics/Networks is the parent study for Aim 1 of this study and demonstrated the feasibility of the larger present study. The co-PIs of the Phylogenetics/Networks project were collaborators for this study.

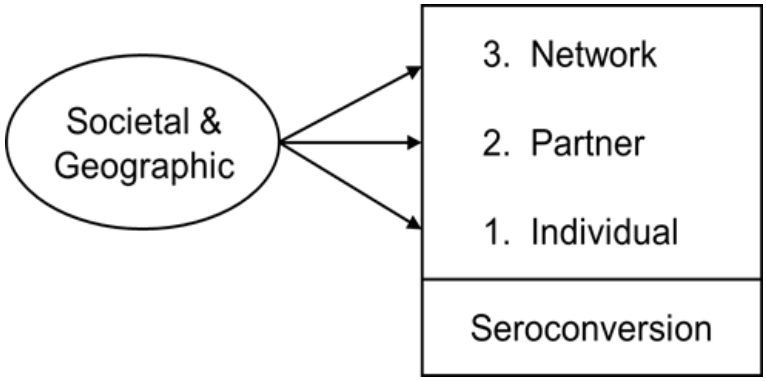
C. Conceptual Framework

This is one of the first studies to link HIV behavioral and phylogenetic data to examine HIV trends by creating a comprehensive sexual network using all cases in a defined and diverse geographic area over a span of several years. Using all cases reported in a geographic area during a defined time period permitted analysis of the complete observed network, thereby allowing a comparison of distinct network components and independent analysis of people as they sort themselves by demographic and risk category.

Sexual network studies tend to simplify by focusing on a subset of patients and treating the connections between people in the network as if they have arisen from spontaneous or wholly exogenous processes, although in reality, partnership formation is not random.[69,110] We used the structure of the sexual network to determine whether any factors related to the network itself were associated with continued involvement in the sexual network following HIV diagnosis.

Figure 4 depicts the levels of factors which we believe affect a person's risk of HIV acquisition or transmission. Network structure indicators, such as density of ties or the clustering coefficient, affect how likely someone who enters the network is to become a part of a cluster. Network-level characteristics also include HIV prevalence. This is the top level of effects which we will study; network-level characteristics such as assortativity or prevalence influence partner-level characteristics via partner selection.[45,54,62,69] Partner- (or dyad-) level characteristics include frequency and types of contact along with likeness of demographics and risk behaviors between nodes. Partner-level characteristics influence individual-level characteristics,[61] which include everything unique to the person, including his demographic traits and STI status. The amalgamation of these levels affect each person's HIV status.

Figure 4. Theoretical framework for factors related to HIV acquisition.



We tested a new method to better identify persons at risk in this study. First, we used all cases, allowing us to compare persons in the network. Second, combining networks and the phylogenetic tree may have more utility in identifying clusters than the sexual network due alone to the high proportion of unknown individuals in the network. Finally, using surveillance data to follow cases and partners forward in time allowed us to look specifically at our population to determine which factors increase potential for being involved in future disease transmission as an HIV-positive person continuing to engage in high-risk partnerships. We identified a set of traits at diagnosis which are predictive of onward transmission at a later time point, providing valuable information in shaping how we follow persons who are first introduced into the network.

D. Synopsis

Incident HIV ultimately results from interactions between discordant individuals. This study combined sexual network analysis and phylogenetic analysis to examine the person-to-person interactions that result in HIV transmission. Variation in human behavior and its effect on HIV requires more complex models and a deeper understanding of human interaction. HIV transmission risk, either acquired or onward, is associated with partner selection,[8,9] partnerships,[111] the existing structure of the network.[76,98]

III. CHAPTER THREE: DATA

A. Overview of Data Sources

HIV testing in NC is confidential and name-based, and HIV is a reportable disease by mandate in North Carolina. The state has a well-defined network of disease intervention specialists (DIS), employed by the State or county, who conduct interviews, perform counseling, and attempt to trace the partners of all located cases when a positive diagnosis of HIV or syphilis disease is received by the state. DIS interviews with new cases elicit testing history, prior history of STIs, risk factors, demographic information, employment and incarceration history, and information about partners in the 12 months prior to diagnosis. As testing for HIV in North Carolina is name-based, sexual networks of cases and partners can be constructed from the DIS case reports.

Aim 1 used a network of Wake County residents with new diagnoses made in 2012-2013 and their partners to compare the overlap of phylogenetic transmission clusters and sexual network ties based upon DIS interviews.

The network constructed for Aim 2 used all HIV cases, syphilis cases, and located partners of each diagnosed during 2012-2013 in NC HIV/STD Control Region 6 (R6) [Figure 7], which includes 11 contiguous counties. A sociosexual network is a depiction of partnerships where people are connected to their disclosed sexual partners and high-risk social contacts. As such, the network includes singletons (persons not connected to anyone else) and components (a set of persons who are connected through partnership ties).

This study used reportable disease data to create a sexual network that includes demographic data, partnership data, HIV viral gene sequences, and geospatial data. The network constructed was a static representation of a dynamic network that changed over a two

year period (2012-2013); individuals linked in the network diagram only needed to be linked once during the period under study to appear together. The network was constructed based upon DIS interview, so a person with a new HIV diagnosis will only be able to identify the partners that s/he had up until that point. Any other linkages made required being named by a future partner who was diagnosed during the study period. As such, the network was undirected, which means that a linkage was made if any actor identified any other actor as a contact during the study period. Where available, temporal data including dates of contact, date of diagnosis, and date of infection were incorporated, but the data were still considered cross-sectional for analysis. All persons in the network were followed for 3 years (1,095 days) after the date of diagnosis for future linkages and changes in STI status.

The network structure itself was analyzed for density and clustering. Actors in the network were described along with their network components so that key traits related to risk of ongoing HIV acquisition could be identified and assessed. Sexual network components were categorized according to the prevalent demographic and risk characteristics of the actors in the component. Actor involvement in network structures of interest, including triangles and *k*-cores were collected and used as model predictors.

Samples are collected for drug resistance testing during clinical care to evaluate for transmitted or acquired drug resistance. Resistance mutations are reported back to the medical provider to guide clinical decisions, although the sequence can be used to construct putative transmission clusters based upon gene sequence similarity. This study linked cases in the network with those phenotypic and phylogenetic clinical data.

B. Data Collection and Management

There are two main data sources for this study. First, the majority of data were abstracted from the electronic HIV database maintained by the State of North Carolina Department of Health and Human Services. Second, HIV nucleotide sequences were obtained from LabCorp®, which is the largest reference laboratory in NC conducting genotyping.

The State of NC keeps electronic records with patient interviews for all new diagnoses in the North Carolina Electronic Diseases Surveillance System (NC EDSS). DIS employed by the State of NC interview all persons with HIV new diagnoses to collect risk information, elicit partners for tracing and testing, and initiate a relationship with a clinic where the person can receive care. Partner information is name-based, and records for the index case and all locatable partners are entered in an electronic database. Each record is unique to an individual, and all contacts are linked to his or her record.

A line listing generated by the State was used to identify cases newly diagnosed in R6 during 2012-2013. Data for identified cases and their partners were electronically abstracted for the present study. Abstracted data included demographic information, diagnosis information including acute status, any lab results including syphilis titers or CD4 count and viral load, HIV testing history, STI history, incarceration history, college status, immigration history, how sexual partners are met, injection drug use, number and gender of partners in the last 12 months, and sexual risk factors including anonymous partners, such as types of sex, partners who were known to be HIV+ at the time of sexual contact, and engagement in transactional sex. Partner information was abstracted using SAS v9.4[112] from a copy of the surveillance system data.

All data used for network construction were abstracted from the DIS case reports, which were collected for the purpose of public health action. No contact were made with patients for the purposes of this study. An executed Data Use Agreement with the NC Communicable Disease Branch of the Division of Public Health indicating agreement to provide access to the study population was obtained.

HIV *pol* gene sequences were obtained from LabCorp for Aim 1 as collaborative research with Dr. Ann Dennis at the UNC School of Medicine supported by a NIH K08. Most of these sequences were generated by GenoSureMG through mid-2014 from clinics throughout NC. Some larger clinics started moving to Monogram Biosciences in 2012 because LabCorp

uses the GenoSure MG assay, which doesn't sequence the portion of *pol* encoding for integrase, but Monogram Bioscience's GenoSure PRIme assay does.[113]

Sequences in Aim 1 were probabilistically matched to patients by gender, date of birth, and specimen collection date. Sequences that cannot be matched or sequences which aren't uniquely identified by gender and date of birth may be able to be matched by more identifiable information, such as initials or name. Through these methods, we matched available sequences against all HIV-positive cases and partners abstracted for this study.

This study required linking identified data collected for State surveillance with HIV sequences obtained once patients are linked into care. Upon completion of abstraction, datasets were merged in a secured facility at the State prior to de-identification. A unique person identifier assigned in NC EDSS was used to create the named-partner sexual network. All subjects, both cases and partners, were also issued a randomly-generated unique study ID number which was used in lieu of identifying information once the network linkages were established and all datasets were merged. Datasets were merged programmatically using SAS software v9.4.[112]

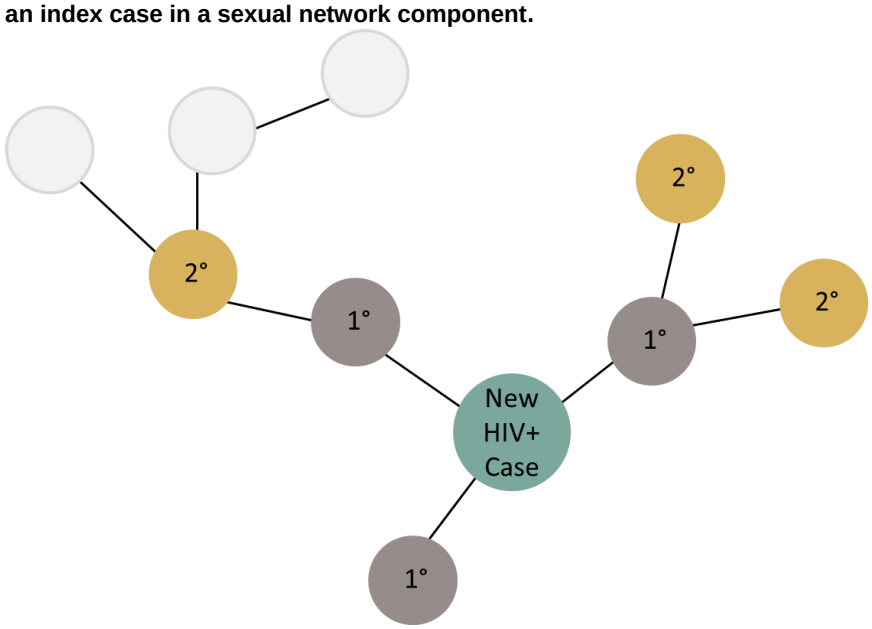
All data were completely de-identified prior to leaving the secure facilities housing the State data. The Principal Investigator had already developed a system for manually tracking linkages using de-identified data, while also ensuring that linkages are de-duplicated, which was successfully applied in the Nexus and Phylogenetics/Networks studies (section II.B). Network visualization and analyses were performed using the igraph[114] package in *R*.[115] Unless otherwise specified, all other analyses were performed using Stata software v15.[116]

IV. CHAPTER FOUR: METHODS

A. Study Design

A cross-sectional study design was employed for Aims 1 and 2 where all index cases newly diagnosed during the study period in the geographic area of interest and their primary degree partners (Figure 5) were abstracted. Demographic, behavioral, and county of residence of cases and partners were analyzed for the purpose of understanding which local social processes were associated with network structure and HIV prevalence. A sociosexual network was constructed from named first-degree partner ties. A phylogenetic tree was constructed from available *pol* sequences using a background dataset of over 15,246 sequences collected from individual persons receiving care in NC.

Figure 5. Illustration of primary (1°) and secondary (2°) degree partners of an index case in a sexual network component.

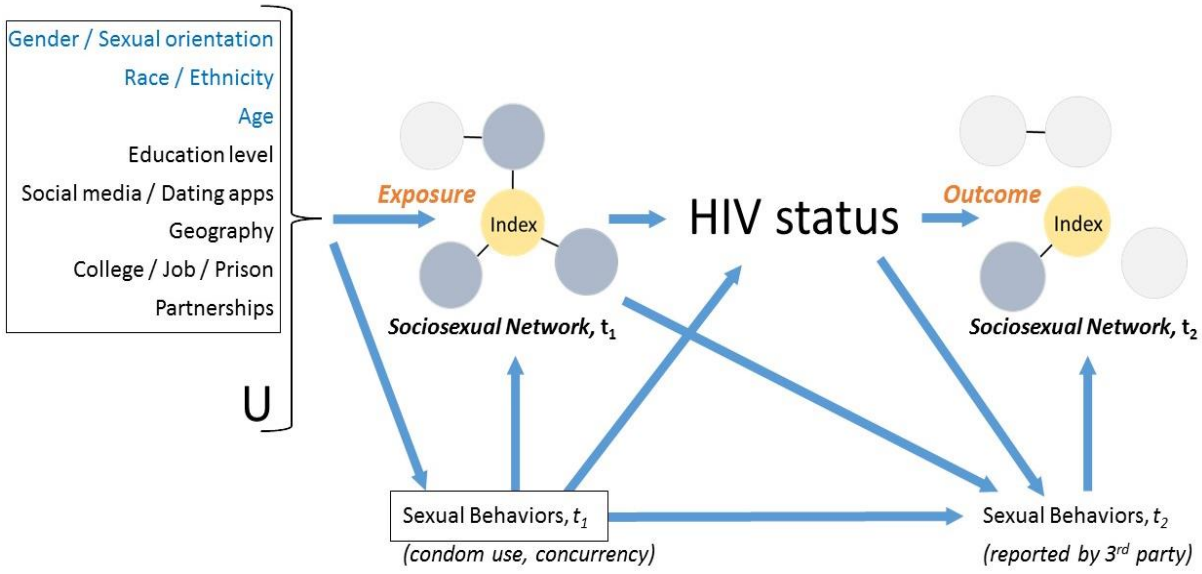


For Aim 2, the 2012-2013 network served as the "baseline" period before the outcome accrual period (2014-2016). There is significant overlap of syphilis and HIV sexual networks in NC, particularly among MSM,[21,96] so the partnerships elicited during contemporaneous syphilis investigations in the same geographic region were used to increase the size of the network. Some of the syphilis cases or partners were already in the network, in which case only new syphilis partnership information was abstracted and added to the HIV network. Components

that did not include at least one HIV index case were removed from the sexual network, and singletons could only be HIV index cases. All cases and partners abstracted for the 2012-2013 HIV network and all 2012-2013 syphilis cases and partners added to create the baseline network were "followed" through 2016 for post-diagnosis partnerships, although persons were still retrospectively followed as all partnerships occurred by the time that data abstraction began. Outcomes included whether HIV index cases were diagnosed with a new STI or were named as partners on HIV or STI cases following their own their own HIV diagnoses.

Simulations along a sexual network showed that post diagnosis behavior change does occur,[117,118] although analysis of acutely- and recently-infected diagnosed persons in NC showed that persons who are aware of their own seropositive status appear to be responsible for a higher proportion of onward infections[28] than is typically understood to occur in the United States.[119] Thus, the purpose of the predictive model was used to identify newly diagnosed with the highest risk of onward transmission in the years following diagnosis.[10] Figure 6 shows the hypothesis. Predictors inside boxes were abstracted. "U" denotes unmeasured factors influencing sociosexual network contacts.

Figure 6. Hypothesis for predictive model. Sexual behavior and partner-seeking characteristics at the time of diagnosis are hypothesized to be predictive of sexual behaviors and partner-seeking behaviors post-diagnosis.



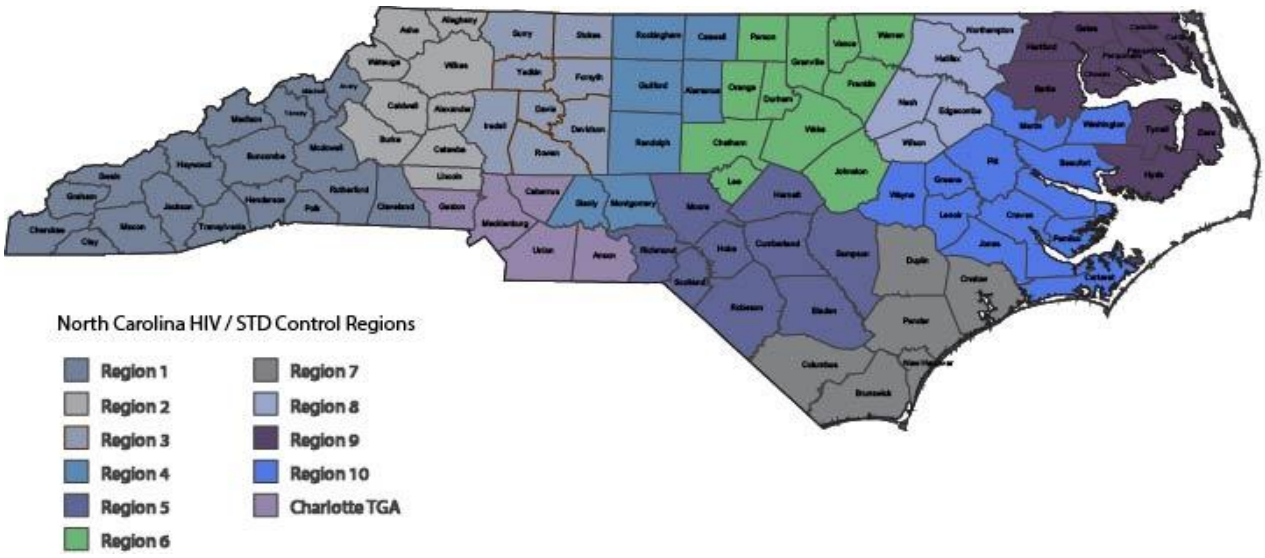
For Hypothesis 2.1, index cases were traced to determine whether they were reported as partners for any HIV or STI case investigated by the State of NC through the end of 2016. A binary outcome indicating presence of either outcome was calculated for all index cases. Data were treated as cross-sectional for this analysis.

B. Subject Identification / Sampling

1. Source Population, Identification of Cases, and Identification of Controls

A line listing including State case number, patient's name, and date of diagnosis were generated by State personnel for chart abstraction. The Aim 1 listing included all year 2012-2013 HIV diagnoses made among residents of Wake County, NC along with their partners. The Aim 2 listing included all year 2012-2013 HIV diagnoses among residents of HIV/STD Control Region 6 (Figure 7), which includes Wake County and 10 other contiguous counties. HIV cases determined to be perinatally acquired or diagnosed in a minor less than 14 years of age were not included in the network analyses as the infection was likely perinatally acquired rather than due to activity in the sexual network. All persons diagnosed with syphilis, regardless of stage, during 2012-2013 in the Region under study were also abstracted to increase the completeness of the sexual network.

Figure 7. Eleven NC Communicable Disease Branch HIV/STD Planning and Care Regions. Region 6 (green) comprises 11 contiguous counties and is the area under study.



For Aim 2, the population of interest was newly-diagnosed HIV cases age 14 years or older at the time of diagnosis who were included on the line listing provided by the State of persons first diagnosed in Region 6 during 2012-2013. Partners and contacts who were abstracted included those who were linked in the NC EDSS to the case's electronic record AND who meet one of the following criteria:

- Those who were in contact with the case during the critical period, defined below, OR
    - HIV: 12 months prior to diagnosis for chronically infected, or 3-6 months prior to diagnosis for acute and recently infected cases
    - Syphilis: 3 months prior to diagnosis for primary syphilis, 6 months prior to diagnosis for secondary syphilis, 12 months prior to diagnosis for latent syphilis
- Those who were believed by the investigating DIS to be the source of infection for the case even if it was outside of the critical period.

It is expected that some cases were contacts of other cases, in which case they were abstracted as such and de-duplicated in the network during analysis based upon study ID.

2. Selection Criteria

Aim 1 included all HIV diagnoses made among Wake County, NC residents during 2012-2013 and their named sexual partners and social contacts. HIV *pol* gene sequences sampled from HIV-positive residents across the State of NC were used as background sequences for the construction of the phylogenetic trees.

The NC Communicable Disease Branch divided the State into eleven regions for HIV and STD prevention and care (Figure 7).[4,120] Each region has its own set of DIS. Aim 2 of this study included all persons diagnosed with HIV in 2012-2013 in Region 6 and their contacts. The entire population of HIV diagnoses made in Region 6 from 2012-2013 were analyzed, which allowed between-group comparisons to be made. There were 569 newly HIV diagnoses made among residents age 14 years or older during 2012-2013 in the 11 counties comprising Region 6 (Table 2).[38] Twelve additional cases determined to be perinatally acquired or acquired

in a minor who was not part of the sexual network were discarded (representing 2% of 581 total diagnoses made in R6 during 2012-2013).

3. Sample Size

Aims 1 and 2 used the entire population of diagnoses made in the areas under study, so no sample size calculation is necessary. Figure 8 shows the number of persons per aim.

Table 2. HIV cases first diagnosed among residents of Region 6 age 14 years or older, by county, 2012-2013.

	HIV Cases*		HIV Rate†	
County	2012	2013	2012	2013
Chatham	<5	<5	3.0	6.0
Durham	67	72	23.7	24.3
Franklin	5	7	8.1	11.2
Granville	15	7	26.0	12.1
Johnston	11	15	6.3	9.0
Lee	<5	<5	6.7	6.7
Orange	13	14	10.9	10.1
Person	6	7	15.3	17.8
Vance	11	7	24.4	15.7
Wake	137	161	14.4	16.9
Warren	0	<5	0.0	4.9
	271	298		

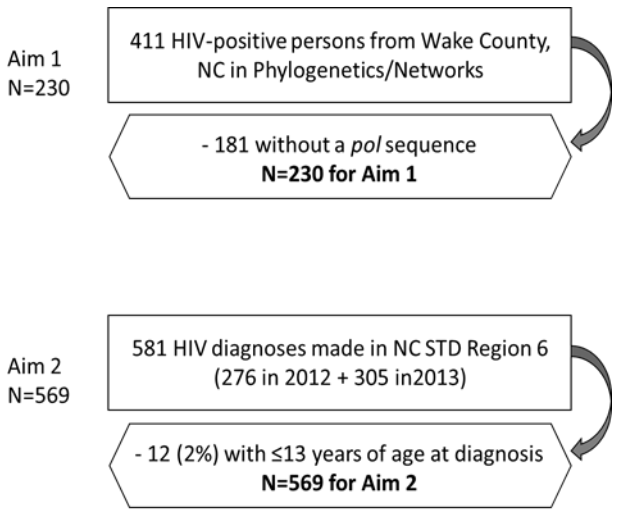
* number of cases age ≥14 years per data provided by NC Department of Health and Human Services

† per 100,000 total population, any age (to protect confidentiality)

a. Aim 1

The starting population for this analysis was the 230 HIV-positive persons (cases and partners) with available sequences abstracted for the Phylogenetics/Networks study (section II.B.) as the starting population (Figure 8). Eight-seven of those persons are in one of 34 transmission clusters.

Figure 8. Estimated population size by aim.



b. Aim 2

Among residents of Region 6 during 2012-2013, 569 new HIV diagnoses were first made among persons age 14 years or older at the time of diagnosis (Table 2 shows the county-by-county diagnoses reported to the CDC). An additional 12 diagnoses were among pediatric patients age 13 years or

younger at the time of diagnosis. These pediatric cases were excluded from analyses, as their infections were most likely acquired perinatally and not through the sexual network.

C. Measurements and Assessments

1. Aims

a. Aim 1

In the first aim, I compared phylogenetic links and named partner ties in a small, defined geographic area by constructing the phylogenies and sexual networks from the Wake County, NC data. We identified transmission clusters, defined as short branch lengths and high bootstrap support (≥90%), from the phylogenetic trees. I identified sociosexual network components, defined as groups of persons directly or indirectly linked through name-based interview and partner elicitation. Comparisons for this aim were made between the network components, instead of direct person-to-person links, and the phylogenetic trees since direct person-to-person transmission cannot be inferred from the phylogeny.[121] I used the component as the unit instead of the dyad because it is possible for two persons to be in a phylogenetic tree with an intervening person who may or may not be identified. Therefore, components are a more representative comparator for the clusters.

Sociosexual networks can be constructed and observed from contact tracing data, which is the public health "gold standard" for identifying partnerships for the purpose of intervening to halt some transmission. However, the Phylogenetics/Networks study found that there was only partial overlap between the transmission clusters identifying infection passage and the networks identifying partnerships with the potential to transmit infection. As not all partnerships lead to transmission, one would expect there to be partnerships in the network that are not represented in the phylogenies. However, any partnerships or transmission clusters identified in the trees that was not seen amongst person in the same network component suggests that contact tracing is not capturing transmission pathways.

The outcome for the model comparing the trees and the components was, therefore, whether or not any two persons in the same transmission cluster were in different network components. Covariates included several dyadic characteristics, including homophily of demographics (i.e., race assortativity or men with male partners), homophily of risk (amount of time between partner diagnosis dates), and characteristics that affect DIS case-finding ability, such as interview refusal, having anonymous partners, or either partner being previously diagnosed and therefore not interviewed for the recent investigation.

b. Aim 2

In Aim 2, I determined which individual- and network-level traits can predict future HIV outcomes for individuals who are part of sexual networks with circulating HIV and syphilis in NC. The post-diagnosis interview conducted by the DIS elicits partner information, which was used to construct a name-based sexual network. A semi-dynamic network was constructed to test the hypothesis; new ties were allowed after 2013 through 2016, but 2012-2013 ties did not dissolve. I built a predictive model with model terms that included individual-level exogenous characteristics collected by DIS and all network-level structures involving the persons being analyzed (endogenous effects) that were found when the sociosexual network was constructed.

Network configurations for assessing endogenous processes were collected, including involvement in ties (edges), *k*-cores, and closed triangles. Network centrality scores were calculated for each non-singleton node. Centrality scores were continuous, calculated from the observed network, and normalized for the model. CD4 count and having more than 5 sexual partners in the 3 months prior to diagnosis has been found to be associated with being in a phylogenetic transmission cluster,[101] so both variables were calculated and tested for association with the outcome.

Three different centrality scores, degree, adjusted degree, and betweenness, were calculated for each index case, as each centrality score represents infectious disease in a network differently.[122-124] Degree is simply the number of connections (partners) that a node

possesses in the network.[11,90] Although it is not representative of placement in an overall sexual network and therefore not well-representative of risk, degree is commonly used to assess sexual risk[69,77,101,125] so we used it because it is easy to calculate and provides a comparator to other studies. Degree is also considered very robust to missing data compared to other centrality measures.[89] However, previous sexual network studies in NC have shown that degree varies significantly depending upon whether the network is constructed with all partners indicated or only partners located, and the difference is a form of missing data which is not random[126] and it also impacts network structure.

I calculated a new centrality measure, which I called adjusted degree, based upon the principle of eigenvector centrality (EVC). EVC, also called Bonacich centrality,[127] is an adjusted form of degree centrality. EVC adjusts a node's degree based upon the degree or his or her partners.[90,128,129] EVC uses eigenvectors to find popularity or exploitation of a valued relation within a social network by calculating the largest vector within an adjacency matrix;[128,129] the eigenvector is the solution to the matrix.[130] Since the eigenvector is the solution to the network adjacency matrix, it is very sensitive to missing data[131] and cannot be calculated in a disconnected network, which is a network in which not all nodes are reachable by walks – that is, a network with multiple discrete components as any sexual network is.

Instead of calculating a matrix score, I used the same principles to calculate an egocentric EVC for each index case where his number of partners (degree) was modified by the average number of partners had by his partners (his partners' mean degrees). My rationale was that risk is not well-captured by degree if a case only has one partner, but that partner happens to have many partners. Weighting degree by partners' degrees creates a second order measure,[129] which I incorporated into our centrality measure equations. I stopped at the second order values instead of calculating a single vector for the entire matrix since partner elicitation interviews do not interview all partners and we wanted a cohesive measure that could be

calculated for all index cases in our network; additionally, temporality was ignored in the network so including higher-order partners may not have truly represented the index case's risk.

Traditional EVC score is calculated as

$$\lambda x_i = \sum_{j=1}^{n} a_{ij} x_j, \quad i = 1, \ldots, n$$

Equation 1. Eigenvector centrality score calculation.

where n is the total number of nodes or vertices; a_{ij} is 1 if i and j are connected and 0 if they are not; and x_j is the centrality score of node j.[128]

Instead, I used an adjusted degree formula where the index case's degree increased if the average degree of all first degree partners exceeds the index's degree. I multiplied degree by the log of the difference in degrees between the index and his or her nearest neighbors, and added that number to the index case's degree. The adjustment formula is mine. As far as I know, an adjustment of EVC is a novel approach and has not been previously used. The benefit of using the log difference is that it allowed me to increase an index case's degree if his or her partners had more partners on average, but not adjust the index case's degree lower if the index case's degree was higher than the mean degree of his partners.

The final centrality measure, betweenness centrality, is stochastic and more complex. It is the count of all paths between all other pairs of nodes that must pass through a given node, so it is a good model for disease transmission in the true network. However, it is difficult to calculate and cannot be applied without having much of the network constructed.

The majority of partnerships in Nexus and Phylogenetics/Networks (section II.B.) were dyads,[96,132] so I set a score of 0 for all singletons and dyads so as not to lose them in this calculation. As I had a largely disconnected network with smaller sized components, there were many betweenness values equal to 0.

I did not expect betweenness to approximate disease transmission well in the observed network; the manner of the DIS interviews is a severe limitation here. DIS are instructed to ask about all partnerships that occurred during the period of interest for new diagnoses, but do not contact partners who are known to be previously positive. This results in many isolated persons or artificially small components in the network. If a new case has a single partner who was previously diagnosed then that partnership will appear to be a dyad that is disconnected from the rest of the network, when in fact the previously-diagnosed person could be an active part of it.

For Hypothesis 2.1, Region 6 residents newly diagnosed with HIV during 2012-2013 Aim 2 network were included in the analysis, with one observation per new index case. The outcome was a binary variable indicating whether the index case was 1) named on a HIV case investigated in NC during 2014-2016, where the partnership continued for at least 2 months after the index case's diagnosis or began at any point after the index case's diagnosis or 2) diagnosed with a new STI at least 6 months after the index case's diagnosis. Limited additional information about 2014-2016 partnership(s) was collected and persons not meeting either of the outcome criteria were assumed to not have been in a high-risk partnership.

2. Constructing the Sexual Network

The name-based partnership data collecting during DIS interviews can be used to construct the sexual network. Interview partnership data were used to construct a static undirected sexual network of all partners named during the 2-year study period. A sociometric network[11] was constructed for this study. Cases and partners who met the entry criteria (section IV.B.1.) and were included on the line listing provided by the State were abstracted along with their sexual partners and social contacts for Aim 2. Partners and contacts who were abstracted included those who were linked in NC EDSS to the case's electronic record AND who met one of the following criteria:

- Those who were in contact with the case during the critical period (12 months prior to diagnosis for chronically infected and 3-6 months prior to diagnosis for acute and recently infected cases) *or*
- Those who are believed by the investigating DIS to be the source of infection for the case even if it is outside of the critical period.

Some cases were contacts of other cases, in which case they were abstracted as such and de-duplicated in the network during analysis based upon State-assigned unique ID.

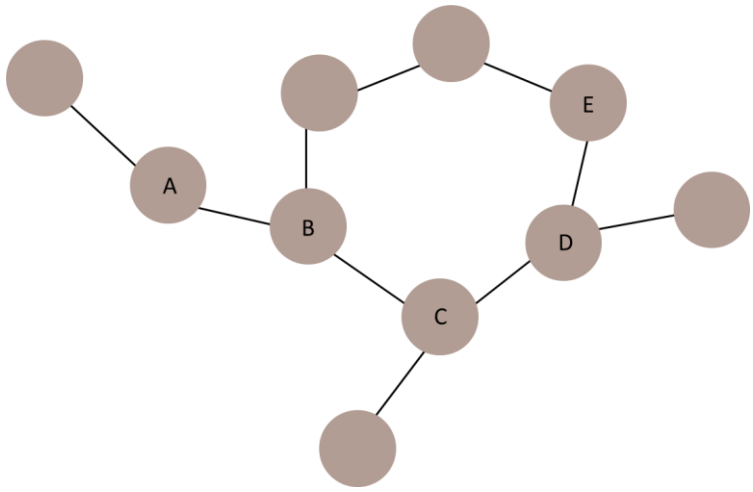
The agents' individual-level characteristics including age, race, education level, location of residence, and sexual risk behaviors were assigned to the agents. Between-agent characteristics (dyadic covariates), including homophily of race, age, and HIV status (i.e., same race/different race) and concurrency of partnerships were attributed to the ties between agents. All unique persons were represented once in the network as an agent with individual- and aggregate dyad-level attributes, whereas other attributes were characteristics of the partnership itself.

Accurate construction of the network relies upon the ability of the DIS to locate the case for interview, truthful interview with the DIS, accurate recording of the interview, and sufficient information to locate partners. Less than 1% of the cases in Nexus could not be located or refused DIS interview, although 20% of cases in Phylogenetics/Networks (Aim 1 parent study) refused to be interviewed (section II.B.). However, 40% of the partners in the Nexus study were anonymous and could not be located, resulting in "dead ends" in the network.[96] While this does impact the accuracy of the network, all cases are accurately placed in the network with respect to their partnerships.

The final network included multiple "components". A component is a group of at least 2 agents (persons) linked by name in the DIS interviews (a person without any network ties is a singleton and not a component).[11,90] For example, Figure 9 contains a single network

component that comprises several persons who were interviewed and are linked together by their named partner ties.

Figure 9. Illustration of a network component, which is a group of at least 2 persons linked together through defined ties.



3. Constructing the Phylogenetic Trees

a. Aim 1

More than half of HIV-positive patients in NC attend a clinic that uses LabCorp for genotyping. Sequences were available for approximately half of newly diagnosed cases (56%) in the parent study (Phylogenetics/Networks; section II.B.).[132] Sequence data was most likely to be unavailable for diagnoses made by smaller private physicians who do not use LabCorp. The sequences in the statewide dataset were collected starting in 1997 for clinical care from patients across NC, although not routinely collected prior to starting ART until 2006.[133] Indexes and HIV-positive partners were probabilistically matched to the sequence dataset using birth date, gender, and laboratory test dates. The earliest sequence per person in the statewide dataset, including those who were not in the sexual network, was retained and used for cluster construction. The final dataset included 15,246 sequences.

Sequences were aligned and phylogenetic trees were constructed from available *pol* gene sequences.[134,135] Maximum likelihood methods, which are less prone to overconfidence than Bayesian methods if the model is mis-specified,[136] were used to construct the initial phylogenetic tree, refine the tree, and identify transmission clusters.[49] Bootstrap percentages were assigned to each tree node. For these analyses, a phylogenetic transmission cluster was defined as a group of at least 2 individuals with sequences less than 3.5% genetic distance

apart across the entire cluster and a high branch support value defined as a bootstrap value greater than or equal to 95%.[49] Pairwise genetic distance between sequences was calculated.

Adding phylogenetic data to the network may lead to identification of linkages between individuals thought to be unlinked due to an unknown individual between them or identification of anonymous individuals themselves if two nodes are phylogenetically linked and each have anonymous partners. A study which inferred sexual network based upon phylogenetic linkages of ≤1.5% genetic distance between *pol* sequences found that transmission networks could be identified, although direct transmissions could not always be since directionality cannot be known and there may always be an unsampled third party.[137]

There are several limitations inherent in this approach. Not all HIV-positive persons will have a sequence available – 47% of index cases and 37% of partners in the Phylogenetics/Networks study could not be matched to a sequence.[132] The majority of sequences on patients in care in Wake County and Region 6 overall are run at LabCorp and there is an agreement to obtain those sequences and agreements with other labs are in process. A limitation of working with Sanger consensus sequences, however, is that, unlike deep pyrosequencing, the entire virus population in an individual is not sequenced and it must be assumed that the sequences associated with any given sample represent the dominant strain. Finally, risk factor and demographic data are abstracted from DIS interview records, which are not standardized. Risk factors must be explicitly stated in the interview, so anything implied or omitted from the written interview record will bias the result toward the null if the value in question appears to be an important predictor in the model but is not explicitly recorded for some.

b. Aim 2

Similar to Aim 1, indexes and HIV-positive partners were matched to a statewide dataset of HIV-1 *pol* gene sequences, though identifiers such as name, medical record number, date of birth, gender, and residence location were used for direct matching. These sequences (full

length protease, partial reverse transcriptase, and integrase) were sequenced by LabCorp or Monogram Biosciences and collected 2010-2016 statewide. Rachael Billock identified putative transmission clusters of these sequences using HIV-TRACE[138] with the TN93 model.[139] Clusters were defined as clades with ≤1.5% pairwise genetic distance between sequences, one per person. I matched the cases in these analyses to the defined clusters.

D. Statistical Analyses

1. Aim 1

The first aim of this study was to compare phylogenetic links and named partner ties in a small, defined geographic area. Two identical, empty *N*x*N* matrices were constructed, where N was the total number of persons in the network. One matrix represented the sexual network, where the value were filled in as 1 if two agents were in the same sexual network component and 0 otherwise. The other matrix represented the phylogenetic trees; agents not matched to a sequence, and therefore not a part of the phylogeny, were removed from the matrix as they could not inform this analysis. For all remaining agents, the value was filled in as 1 if the two agents were in the same transmission cluster and 0 otherwise. All agents removed from the phylogeny matrix were removed from the network matrix. Both matrix diagonals were set to 0, as self-loops were not allowed in this analysis.

The next step in the Hypotheses 1.2 analysis was to subtract the network matrix from the phylogenetic matrix. There were 4 possible outcomes:

Table 3. Comparison of phylogeny and network connection *N*x*N* matrices showing outcomes values for presence or absence of the type of tie represented in each matrix.

Initial Matrix Value			
Phylogeny	Network	Result	Meaning
1	1	0	Both phylogenetically- and network- linked
1	0	1	In a transmission cluster without being in the same network component
0	1	-1	Not phylogenetically linked, but in the same network component
0	0	0	Neither a phylogenetic nor a network link

All values of -1 in the resultant matrix were reset to 0. The rationale for this was that we were not interested in persons who were in partnerships that did not result in transmission, as these people were already identified in the DIS investigation. However, persons who were in the same transmission cluster, and were therefore highly likely to be involved in the same transmission chain, but who were not in the same set of name-linked partnerships indicated gaps in contact tracing and case finding. This matrix was the outcome for a GEE model using an exchangeable correlation matrix, as we can infer transmission clusters but not direct person-person transmission.

Future directions for this dataset include building a valued exponential random graph model (ERGM). In a valued model, the outcome can vary by type of tie between persons, instead of a non-valued model where the outcome is simply presence or absence of a link between persons. Here, our ties included being sexually linked, being phylogenetically linked, both, or neither.[140]

2. Aim 2

The second aim of this study is to determine which individual- and network-level traits can predict future HIV outcomes for individuals who are part of sexual networks with circulating HIV and syphilis in central NC. I constructed a predictive model based upon characteristics collected at HIV diagnosis (number of partners, proportion of anonymous partners, being a student) of persons who are later named as a partner on a new HIV case in NC, as the person may have been involved in virus transmission. I selected the predictor variables for the Hypothesis 3.1 model using backwards elimination and summed the β coefficients of the final bootstrapped model, from 1,000 trials with replacement and robust standard errors, to obtain a risk score. I calculated sensitivity and specificity scores at various levels of summed β.

The outcome for Hypothesis 2.1 was being named as a partner on an HIV case in NC 2014-2016 or being phylogenetically linked to a case diagnosed with HIV in NC 2014-2016. I abstracted a dataset that included the 2012-2013 HIV index cases diagnosed in R6, with one

observation per case. Any subsequent NC EDSS partnership records or STI diagnoses within 3 years (1,095 days) after HIV diagnosis were abstracted per case. I constructed a logistic regression model with robust standard errors to account for lack of independence of the outcome among network members. The model was internally validated with 1,000 bootstrapped samples with replacement from the population of cases.

Predictors included individual-, dyad-, and network-level variables. Concurrent diagnosis with HIV and AIDS (defined as diagnosis of AIDS within 6 months of diagnosis with HIV) was included in this model as a predictor, as behavior changes at this point and persons with AIDS may not represent the behavior of persons with HIV overall.[79] Network structures identified in this analysis, such as involvement in closed triangles or *k*-cores (section II.A.5.), were the network-level predictors. Degree was adjusted as previously described (section IV.C.1.b.) based upon the *k*-nearest neighbor algorithm in igraph.[114]

Missing data were an issue in this analysis and affected calculation of network measures. The proportion of unknown partners is a limitation which had an effect on the model since missing nodes are “dead ends” in the network diagram and cannot be used to link cases in the network. Nor could I use risk behavior data, which has been demonstrated to be predictive of future behavior, since it was not available for the 25% of index cases who refused interview.[78,79]

V. CHAPTER FIVE: LEVERAGING PHYLOGENETICS TO UNDERSTAND HIV TRANSMISSION AND PARTNER NOTIFICATION NETWORKS[1]

A. Overview

1. Background

Partner notification is an important component of public health test and treat interventions. To enhance this essential function, we assessed the potential for molecular methods to supplement routine partner notification and corroborate HIV networks.

2. Methods

All persons diagnosed with HIV infection in Wake County, NC during 2012-2013 and their disclosed sexual partners were included in a sexual network. A dataset containing HIV-1 *pol* sequences collected in NC during 1997-2014 from 15,246 persons was matched to HIV-positive persons in the network and used to identify putative transmission clusters. Both networks were compared.

3. Results

The partner notification network comprised 280 index cases and 383 sexual partners and high-risk social contacts (n=131 HIV-positive). Of the 411 HIV-positive persons in the partner notification network, 181 (44%) did not match to a HIV sequence, 59 (14%) had sequences but were not identified in a transmission cluster, and 171 (42%) were identified in a transmission cluster. More than half (59%) of transmission clusters bridged sexual network partnerships that were not recognized in the partner notification; most of these clusters were dominated by men who have sex with men.

---

[1]This chapter has been accepted in *JAIDS*: Pasquale, DK, Doherty, IA, Sampson, LA, Hué, S, Leone, PA, Sebastian, J, Ledford, SL, Eron, JJ, Miller, WC, Dennis, AM. Leveraging Phylogenetics to Understand HIV Transmission and Partner Notification Networks. *J Acquir Immune Defic Syndr*, Accepted 2018.

4. Conclusions

Partner notification and HIV sequence analysis provide complementary representations of the existent partnerships underlying the HIV transmission network. The partner notification network components were bridged by transmission clusters, particularly among components dominated by men who have sex with men. Supplementing the partner notification network with phylogenetic data highlighted avenues for intervention.

B. Introduction

Across the Southern United States (US), including North Carolina (NC), the HIV epidemic has persisted in large connected sexual networks, particularly among men who have sex with men (MSM).[21,49,108,141,142] The South is the epicenter of the US epidemic, accounting for a disproportionate number of HIV infections.[141] HIV incidence continues to rise among Black and Hispanic/Latino MSM,[3] despite widespread prevention efforts. Entry into a sexual network composed largely of Black MSM increases the likelihood of contracting HIV,[21] highlighting the importance of enumerating sexual networks. An improved understanding of sexual networks will aid in the development of enhanced interventions to reach Black and Hispanic/Latino MSM. Time-intensive efforts to reach members of densely-connected sexual networks often result in analysis of incomplete networks, due in part to anonymous partners, persons who cannot be located, and interview refusal.[143]

Phylogenetic analysis of HIV sequences is an excellent adjunct to enumerating networks and allows tracking of local transmission patterns. HIV phylogenies based on sequence similarity and inference of common ancestors can identify putative transmission clusters.[121,137] While these methods are increasingly used to understand HIV transmission dynamics within sub-populations,[93,144,145] use of sequence data to complement sexual networks as understood by contacts elicited during partner notification services (PNS) is understudied.[146] Sequence data has potential to add structure to the sexual network through genetic linkage of network components that erroneously appear disjointed due to inability to locate network members.[147-149]

In San Diego, for example, HIV genetic clusters combined with PNS data from recently-infected MSM increased membership in putative transmission networks.[148] In an investigation of spatiotemporally-clustered acute HIV infections in NC, phylogenetics revealed multiple transmission chains rather than a single outbreak.[104] Such analyses demonstrate that sequence data can enhance our knowledge of sexual networks. Analysis of phylogenetic transmission cluster growth can also point to groups in which HIV transmission continues to occur,[106] signaling the need for immediate intervention.[26,150]

We investigated the sexual network constructed from PNS data in Wake County, NC, and compared this with HIV transmission clusters using *pol* sequences routinely collected statewide. Our objective was to assess the overlap between networks derived through PNS and sequence analysis to identify areas where interventions could be intensified.

C. Methods

1. Study Setting and Design

Wake County is a metropolitan county in central NC that accounts for approximately 10% of statewide annual new HIV diagnoses.[38] In 2012, Wake County had a population of approximately 963,000 persons, including >2,800 persons living with HIV and an incidence of 16.3 cases per 100,000 person-years.[38]

We conducted a cross-sectional analysis of Wake County residents ≥18 years of age who were newly diagnosed with HIV-1 during 2012-2013 and their social and sexual contacts reported during routine PNS. These data were compared with HIV genetic 15,246 sequences collected among HIV cases in NC 1997-2014. The University of North Carolina Biomedical Institutional Review Board approved the study.

2. Study Population

Disease intervention specialists (DIS), employed by NC Department of Health and Human Services (DHHS) or Wake County DHHS, attempt to interview all newly diagnosed persons (referred to as index cases) and collect information about their partners for tracing and

testing. In NC, high risk social contacts are elicited at the discretion of each DIS when perceived to increase case finding without overly burdening investigations.[151,152] Using standardized data abstraction, we collected demographics, HIV testing history, and HIV-related laboratory results for index cases, and sexual and social contact data.

Acute HIV infection (AHI) was identified through the NC Screening and Tracing Active Transmission (STAT) Program,[153] and defined by a positive HIV RNA test and negative or indeterminate HIV antibody, or a positive HIV antibody within 30 days of confirmed negative testing. Cases who did not meet the AHI definition but were reported to STAT with a positive antibody test with seronegative documentation and/or symptoms compatible with AHI within 3 months of first positive HIV test were classified as recent HIV infection (RHI). For persons diagnosed with AHI or RHI, DIS interviews focus on partnerships within 2 or 6 months prior to diagnosis, respectively.

3. Sexual Network Construction

We constructed the sexual network using name-based partnership data collected during PNS interviews with index cases. All network members were de-identified after network construction to preserve patient confidentiality. A socio-sexual network comprises discrete components (at least two people directly or indirectly connected) and singletons (isolated persons if no partners are disclosed or located). The network was created using the igraph[114] package in R.[115]

4. HIV-1 Sequences and Transmission Cluster Identification

HIV-1 *pol* sequences (full length protease and partial reverse transcriptase) were extracted from genotypes performed by LabCorp®, the largest reference laboratory in NC, and sampled between 1997 and mid-2014 from patients accessing clinical care. Demographic variables available included birth date, gender, and sampling site. Geographic location of sampling site was categorized by NC-DHHS HIV Field Service Region (Figure 10).

Index and HIV-positive partners were probabilistically matched to the statewide sequence dataset by birth date, gender, and laboratory test dates. We considered nonmatching sequences as background references for cluster construction. All analyses used the earliest sequence per individual. The final dataset included 15,246 sequences. A random subset of 100 sequences is available in GenBank, accession numbers KY579388-KY579812.

Sequences were aligned using MUSCLE[154] and edited manually in BioEdit,[155] with a final sequence alignment length of 1,497 bases. Maximum-likelihood (ML) phylogenies were constructed in FastTree[156] with the generalized time-reversible model.[157] Statistical support of clades was assessed with local support values using the Shimodaira-Hasegawa-like test (SH-test).[158] Putative transmission clusters were identified using ClusterPicker v1.3[159] and defined as clades with 1) high branch support (≥0.90 SH-test), 2) maximum pairwise genetic distance <3.5% between all sequences, and 3) inclusion of a sequence from at least one index or partner case.

Putative clusters were confirmed with the Bayesian Markov Chain Monte Carlo (MCMC) approach in BEAST v1.8.2.[160] Analyses were conducted using the SRD06 nucleotide substitution model, a lognormal relaxed molecular clock model, and the Bayesian Skyline model as coalescent tree prior. The MCMC chain was run for 50-100 million generations, sampling every 10,000 generations. Convergence of the estimates was considered satisfactory when the effective sample size calculated in Tracer v1.6.0[161] was >200 in all parameters; 10% of generations were discarded as burn-in. The maximum clade credibility tree was summarized using TreeAnnotator v1.8.2,[160] keeping the median height over the posterior distribution of trees. Clades with posterior probability ≥0.95 were considered highly supported and analyzed further.

5. Statistical Analyses

We compared membership in transmission clusters and sexual network components. Clusters involving ≥2 cases (index or partners) were characterized by demographic features and compared to case location within and across network components. Time of most recent

common ancestor (MRCA) and cluster age were estimated based upon timing of branching in the phylogeny.

D. Results

1. Study Population

In total, 280 persons newly diagnosed with HIV were reported in Wake County from 2012-2013; 83% (n=232) were male, 65% (n=183) were Black, and 40% (n=112) were younger than 30 years. Many (27%, n=75) were concurrently diagnosed with AIDS and 4% (n=11) were diagnosed during AHI. Among 235 index cases with CD4 count data, the median first CD4 count was 338 cells/mm^3 (IQR 130-525 cells/mm^3); 31% had CD4 count <200 cells/mm^3. Among 147 cases with viral load results within 3 months of diagnosis, the median was 4.9 log copies/mL (IQR 4.3-5.3 log copies/mL) (Table 4).

2. Partner Notification Network

DIS interviewed 225/280 index cases (80%), who reported 854 sex partners and 34 social contacts (average 4 contacts per person; number of sex partners ranged 0-50). Approximately half (50%; 446/888) of contacts (414 sexual and 32 social contacts) had enough locating information for DIS to begin investigation. The 446 partnerships investigated (Table 2) resulted in 383 unique non-index case partners (Table 4): 36 were index cases themselves, 19 were named by ≥2 index cases, and 3 were index cases who were also named as partners more than once. Although 48/383 (13%) partners were not located during investigation, we included them in the network. Of 383 partners, 39% were HIV-negative, 34% (n=131) were HIV-positive, and 27% HIV status was unknown. Most HIV-positive non-index partners (81%; 106/131) were diagnosed before 2012. Thirty-six percent (138/383) of partners resided outside of Wake County, including 22 (6%) residing out of state and 6 (2%) with unknown location of residence.

The PNS network included 663 persons (Table 4), with 280 index cases and 383 partners. Most network members were Black (63% vs. 29% White and 5% Hispanic/Latino),

MSM or men who have sex with transgender women (MST) (61%), and young (median age 30 years, IQR 24-42). Persons of color were more likely to be HIV-positive (74% Latino and 66% Black) compared to White persons (53%). MSM index cases were more likely to have partners who could not be located than men only reporting female partners (37% vs. 29%).

Overall, 176/280 index cases were connected to at least one other person in the network. The remaining 104 singletons represented 37% of index cases; 55 (53%) reported zero partners and 49 provided information for 1-50 partners, though none could be located. The sexual network was sparsely connected, comprising 104 singletons and 137 network components (≥2 persons). Component sizes ranged from 2-65 persons; the three largest included 20, 26, and 65 people (Figure 11a). Most (62%, n=85) components only included MSM and MST.

We assessed characteristics of the 446 partnerships (93% sexual and 7% social), which included 559 persons across 137 network components (excluding 104 singletons) [Table 5]. Most partnerships involved either MSM or MST (81%), were among people of the same race (82%), and included at least one Black person (71%). Nearly 25% (n=106) of partnerships were between an index case and a person with unknown HIV status. Among 340 partnerships where HIV status was documented for both people, 53% involved two HIV-infected persons (n=181). Most (80%) of the 131 HIV-infected partners received their diagnoses before the index cases (median 2.5 years, IQR 1 month-5.5 years).

3. Transmission Clusters

Over half of HIV-positive cases (56%; 230/411) matched to a *pol* sequence, including 53% (148/280) index cases and 63% (82/131) HIV-positive partners. Cases who had sequences were similar to those without sequences with respect to gender and age. Among index cases, Whites were more likely than persons of color to have sequences (64% vs. 49%, p=0.04), as well as those diagnosed in 2012 compared to 2013 (63% vs. 44%, p=0.002).

We identified 116 clusters involving ≥1 person from the network, with a total of 800 persons including 103 index cases (70% those with sequences), 66 partners (80% those with sequences), and 631 background sequences (Figure 12). In the initial ML analysis, 117 clusters were identified but two sequences failed to cluster in the confirmatory BEAST analyses. The 116 confirmed clusters had median size two members (range 2-36 persons); only three clusters were non-B subtypes (A1, CRF02_AG, CRF06_cpx).

Among 230 index cases (n=148) and partners (n=82) with sequences, we evaluated associations with cluster membership. Cluster members were more likely to be male (77% vs. 52% female, p=0.006), men reporting male contacts (83% vs. 67% heterosexual and 57% no partners reported, p<0.001), Black (80% vs. 69% White and 33% Latino, p=0.001), and younger (mean age 35 vs. 38 years, p=0.04), compared to cases with sequences who were not in a cluster. Cluster members had more connections in the network than did cases with sequences who did not cluster (2 vs. 1 mean partners, p=0.001).

Most clusters included only one index case or partner from the network; 34 (29%) including ≥2 index cases were denoted "Wake" clusters for further analysis (Table 6 shows Wake clusters with ≥5 total cluster members). Wake clusters included 287 persons (56 index cases, 31 partners, and 200 background sequences) [Figure 2]; two (6%) comprised only two partners with no index cases. All Wake clusters were subtype B and most were male-dominated; seven (21%) included ≥50% women. More than half (59%; n=20) of Wake clusters only included persons sampled from the same eleven-county geographic region (Figure 2). Most (74%; 61/82) clusters with only one person from PNS were clusters with ≥50% members sampled in the same region, including 22 clusters with 100% members sampled in the same region.

Wake cluster maximum genetic distance was 1.67% (IQR: 1.04-2.93%) statewide and 0.95% (IQR: 0.32-1.28%) when restricted to network members (Table 6). Median estimated

cluster age prior to the index case diagnosis was 8.5 years (IQR: 5.1-12.9 years) with median MRCA estimated to occur in 2005 (range 2000-2007).

4. Partner Notification Network and Transmission Cluster Overlap

The PNS network included 663 persons: 280 index cases and 383 contacts who formed 104 singletons plus 559 persons in 446 partnerships (Figure 11a). Among 230 network members with sequences, including 45 singletons, 169/230 (74%) were in one of 116 statewide transmission clusters that included at least one network member. The 169 persons spanned 82 network components and 23 singletons; the remaining 61 persons who were not in a cluster spanned 36 network components and 22 singletons. Among the 23 singletons in a cluster (51% singletons with sequences), 8 (35%) did not name any partners and the remainder disclosed at least one partner, though none could be located. The median cluster size among singletons was 4 persons (range 2-23).

Among 446 partnerships, 70 (16%) included two HIV-positive persons with sequences; of these, 83% (58/70) were sexual connections. All male-female pairs were in the same cluster, whereas only 34% of male-male pairs were in the same transmission cluster (χ^2 $p<0.001$). Of the 383 contacts, 27 (7%) were only identified as social contacts of an index case; 11 had a sequence, of which 9 were in a statewide cluster with no one else from the PNS network and 2 were in a Wake cluster; one clustered with another PNS social contact (statewide cluster size 2) and the other clustered with the index case who disclosed the contact as a social connection (pairwise genetic distance 1.3%, statewide cluster size 14).

Eighty-seven persons were in 34 Wake clusters (defined as ≥2 persons from PNS network), which included 2-6 network members and spanned 56 PNS network components plus 12 singletons. Overall, 41% (14/34) Wake clusters covered only one network component; 1 included three network members and the rest included two. The Wake clusters that covered only one component were more likely to include ≥50% women (36% [5/14] vs. 10% [2/20] spanning multiple components).

Among 19 Wake clusters with ≥5 persons statewide (Table 6), 6 (32%) covered only one component, where all network persons in the cluster were also linked by named partner ties. The remaining 13 spanned multiple components, where the phylogenetic relationships bridged located partnerships: 7 (37%) spanned two components, 5 (26%) spanned three, and 1 (5%) spanned four components. For example, the three network members in Cluster J spanned two components and one singleton (Figure 11a, quadrants a, c, and d), although there were 12 people in the cluster statewide (Figure 11b). The maximum genetic distance between any pair of network members in Cluster J was 1.24%, despite each of the 3 network members being in different components (Table 6, Cluster J). Of the 13 clusters with ≥5 members statewide (Table 6) that spanned multiple components, 9 (69%) included only men.

There was no significant difference by sampling year, cluster age, or statewide genetic distance between Wake clusters that covered single or spanned multiple components. However, the mean genetic distance among persons in the Wake cluster was significantly smaller when the cluster covered only one component (0.66% vs. 1.23%, p=0.03).

E. Discussion

This study sought to explore the benefits of combining molecular data with sociosexual network data obtained during routine partner notification services from persons newly diagnosed with HIV in a single large county in NC. The study drew on a statewide dataset of over 15,000 HIV-1 sequences from persons sampled between 1997 and mid-2014. We overlaid the genetic data and sociosexual network constructed from partner notification records to obtain a more comprehensive picture of the epidemic and identify gaps in PNS, particularly among male-dominated sexual network components.

More than half of local transmission clusters bridged sexual network components that appeared disconnected, demonstrating that molecular data can detect unobserved links in the sexual network. Furthermore, despite not having any partners identified in the network, over half of singletons with sequences were in a statewide cluster. For each set of disconnected

network components or singletons in the same transmission cluster, at least one connection is not represented in the PNS network. Some of the disagreement may be explained by differing collection periods, as sequence sampling time for the clusters was not limited by time period. Many index cases were likely infected for years, so partners reported at diagnosis may not reflect the network at the time of infection. Additionally, some persons in the network were only social contacts, so their inclusion increased PNS network component size and may have increased the effect of bridging by the transmission clusters if they were in a different cluster than the index case. However, they represented only 2 of 87 network members in a Wake cluster.

Partner notification is limited by missing data due to persons not being diagnosed or located and partnerships not being disclosed or not occurring during the DIS interview time period. Stigma and discrimination faced by MSM contribute to interview bias and may reduce willingness to disclose partners to health authorities. Previous HIV sexual network studies in NC found that a high proportion of partners cannot be located[21,28] and MSM tend to have more undisclosed partners,[162] causing components to appear disjointed and impacting PNS network completeness. However, this completeness is precisely what we wanted to investigate and adding sequences offered some correction to the observed network.

Accordingly, local transmission clusters, particularly those that spanned multiple components, were more likely to be male-dominated. This reflects the current epidemic in NC, where the overall rate of new diagnoses remains elevated with ongoing transmission among young men[22] and demonstrates the value of supplementing partner notification with another method that portrays transmission networks differently. By overlaying phylogenetic data onto the sexual network, we were able to identify components with ongoing transmission. Persons in either network may benefit from interventions such as offering pre-exposure prophylaxis to HIV-negative partners or linkage to care support to HIV-positive persons who are not virally suppressed.[163] A substantial proportion of incident HIV cases in NC are attributed to persons

who are diagnosed and aware of their status at the time of transmission;[28] determining which network components have unidentified partners and which clusters have unsuppressed members may help guide intervention targets. Additionally, the smaller genetic distance amongst persons in the sexual network compared to other cases in NC indicates that applying these interventions locally could have an immediate local benefit.

We combined methodologies previously used to describe HIV transmission networks. While several studies have used sequence data to construct transmission networks,[49,103,164-170] few have compared these to PNS networks.[149,150,171] To our knowledge, none compared PNS networks constructed from surveillance data using all known incident HIV diagnoses made in a large, defined administrative area. We used all incident diagnoses in our area of interest and matched to all available sequences from one laboratory that serves most patients in this area. We included partnership and demographic data, allowing us to compare groups. We found that male-male PNS pairs were less likely to be in the same transmission cluster, and that male-dominated clusters are more likely to bridge PNS components. The percentage of named partners with genetically similar virus in this largely-MSM population was similar to what was found among MSM in New York City (NYC). Similar to NYC, heterosexual pairs in this population were more likely to cluster than MSM pairs.[149]

Combining PNS and molecular data can lead to an improved representation above what is possible with either alone,[144,145] as both methods have limitations. Sequence analysis is limited by inability to infer directionality and missing data for persons who have not been diagnosed or who do not have sequences available.[172] In NC, genotyping is routinely performed at entry to clinical care, so failure to receive a diagnosis or link to care will impact phylogenetic network completeness. Black persons with HIV infection are less likely to link to care,[173-175] which is reflected in the lower proportion of Black persons in our study with sequences. Additionally, sequences stemmed from only one laboratory and some of the cases without sequences may seek care from providers who use other labs, affecting cluster

comprehensiveness. Still, characteristics associated with cluster membership in our study, including younger age,[49,176,177] Black race,[176] being male,[176] and being MSM,[162,169,176] agree with previous studies in the US.

While there is no accepted genetic distance criteria to define transmission clusters,[121] traditional cut-offs of <1.5% genetic distance difference allow a focus on only on recent transmissions. We used a higher cluster threshold within the range of multiple other studies[121] to permit the characterization of transmission dynamics over longer time periods in the region. Our focus is not on source attribution or using the sequences to confirm transmissions between known partners, but to identify ongoing, local transmission networks using available sequence and routinely collected PNS data. Additionally, most sequences were from chronically-infected persons, so genetic distances between connected persons are expected to be larger due to greater time since infection and we did not want to restrict our analysis to recent partnerships.

Both HIV phylogenetic and PNS data portray networks differently and care must be taken not to misinterpret results. Although the combination of these data provide new insights into network structure, potential ethical and privacy concerns must be considered. HIV genetic clustering does not imply direct person-to-person transmission or direction of transmission;[172] thus these data should not be used for identification of first-degree partnerships or confirming transmission from one person to another.

The HIV sequence analysis recognized ongoing transmission chains among high-risk persons, notably MSM, which was not detected through routine partner notification. Persons who experience the most stigma and those at highest risk, MSM or not, such as those who engage in transactional sex or have anonymous partnerships, are more difficult to reach and may therefore be absent from the PNS network. Molecular approaches provide clues to gaps in PNS and direction for case finding and partner elicitation efforts.[178]

By adding HIV sequences to the PNS network, we were able to successfully identify localized areas where infected persons were missing from the network, demonstrating the value of integrating molecular data into routine partner tracing and testing.

Table 4. Index cases diagnosed 2012-2013 in Wake County, NC and their partners in the sociosexual network (N=663).

	Index (n=280)		Partner (n=383)	
	n	(%)	n	(%)
Gender				
Male	232	(83)	327	(85)
Female	44	(16)	53	(14)
Transgender (M to F)	4	(1)	3	(1)
Race/ethnicity				
non-Hispanic White	69	(25)	120	(31)
non-Hispanic Black	183	(65)	238	(62)
Hispanic or Latino	23	(8)	12	(3)
Other	5	(2)	8	(2)
unknown	0		5	(1)
Age at index case's HIV diagnosis (years)*				
≤ 19	5	(2)	28	(7)
20-29	107	(38)	178	(46)
30-39	54	(19)	87	(23)
40-49	67	(24)	49	(13)
≥ 50	47	(17)	26	(7)
unknown	0		15	(4)
median (IQR)	34	(25-45)	28	(23-37)
HIV status				
Positive, with HIV sequence	148	(53)	82	(21)
Positive, no HIV sequence	132	(47)	49	(13)
Negative	---		148	(39)
unknown	---		104	(27)
Year of HIV diagnosis			n=131	
< 2006	---		31	(24)
2006-2010	---		59	(45)
2011	---		16	(12)
2012	131	(47)	11	(8)
2013	149	(53)	9	(7)
2014	---		5	(4)
HIV stage at diagnosis				
Acute / Recent	23	(8)	---	
Chronic, non-AIDS	182	(65)	---	
Chronic, AIDS	75	(27)	---	
CD4 count closest to diagnosis (cells/mm^3)	n=235			
< 200	74	(31)	---	
≥ 200	161	(69)	---	
Viral load (log copies/mL)†	n=147		n=60	
≤ 3	8	(5)	29	(48)
> 3-5	78	(53)	10	(17)
> 5-5.7	44	(30)	1	(2)
> 5.7	17	(12)	20	(33)
median (IQR)	4.9	(4.3-5.3)	3.3	(2.9-7.7)
Number of sexual and social partners reported‡	n=225			
0	15	(7)	---	
1	78	(35)	---	
2	42	(19)	---	
3-5	61	(27)	---	
≥ 6	29	(13)	---	

* Among partners, for earliest record associated with an index case

† Within 3 months of diagnosis for index patients and within 12 months prior to index case diagnosis for partners

‡ Among those reached for interview; includes located and anonymous partners

Table 5. Partnerships reported by index cases with located members of the sociosexual network (N=446).

	Sociosexual Network Partnerships (N=446)*	
	n	(%)
Partnership type		
Sexual	414	(93)
Social only	32	(7)
Pair gender		
Male - Male	355	(80)
Male - Transgender	5	(1)
Male - Female	85	(19)
Female - Female	1	(0.2)
Index case		
Index - Index	42	(9)
Index - Partner	404	(91)
HIV serostatus		
Positive - Positive (concordant)	181	(41)
Positive - Negative (discordant)	159	(36)
Positive - unknown	106	(24)
Pair race		
Black - Black	261	(59)
White - White	98	(22)
Hispanic - Hispanic	8	(2)
Black - White	40	(9)
White - Hispanic	17	(4)
Black - Hispanic	7	(2)
Other	15	(3)

* 104 singletons in the network are not represented in this table

Table 6. Transmission clusters that included 5 or more persons statewide and at least two members of the Wake County-based sexual network of adults diagnosed with HIV during 2012-2013 and their contacts (n=235).

	Statewide						Wake Network			
Cluster ID	Cluster Size	Max Genetic Distance (%)	Sampling Year (median (IQR))	Estimated Cluster Age (years)	Most Recent Common Ancestor	# Male: # Female	# of Network Persons	# Index: # Partner	Max Genetic Distance (%)	# Components Spanned*
A	5	0.95	2013 (2013-2013)	4.7	2009	4:1	2	1:1	0.00	1
B	5	1.65	2012 (2012-2013)	7.6	2005	5:0	2	2:0	1.02	2
C	5	2.58	2011 (2009-2012)	11.7	2002	4:0†	3	1:2	2.58	2
D	6	2.05	2012 (2008-2014)	8.1	2006	6:0	2	1:1	0.96	2
E	6	0.95	2012 (2010-2013)	6.5	2007	6:0	3	2:1	0.68	1
F	6	1.56	2012 (2012-2012)	5.5	2007	6:0	6	3:3	1.56	3
G	7	2.92	2010 (2005-2012)	18.5	1993	1:6	2	2:0	1.40	1
H	8	2.94	2007 (2003-2012)	18.6	1995	8:0	2	2:0	0.61	1
I	8	1.56	2013 (2012-2014)	7.6	2006	8:0	4	3:1	1.15	2
J	12	3.42	2011 (2009-2012)	12.2	2001	11:1	3	2:1	1.24	3
K	14	3.33	2007 (2007-2010)	18.3	1995	6:7†	2	1:1	1.27	1
L	14	3.22	2011 (2008-2013)	18.3	1995	14:0	2	1:1	0.07	1
M	15	3.59	2010 (2008-2011)	14.7	1999	15:0	2	1:1	0.94	2
N	16	2.33	2009 (2008-2013)	12.5	2001	16:0	2	2:0	0.47	2
O	16	2.11	2010 (2009-2012)	8.8	2004	15:1	3	1:2	1.24	3
P	20	3.26	2010 (2008-2012)	13.8	2000	20:0	3	3:0	3.22	3
Q	23	3.24	2008 (2007-2011)	17.3	1997	10:13	2	2:0	0.07	2
R	23	2.95	2012 (2012-2013)	12.0	2002	23:0	5	4:1	1.83	4
S	36	3.26	2012 (2011-2013)	12.4	2002	34:2	4	4:0	2.54	3

* Includes number of network singletons and components that included at least one person from the Wake County sexual network

† Gender unknown for one person in this cluster

Figure 10. NC HIV Field Service Regions Prior to Office Redistribution.

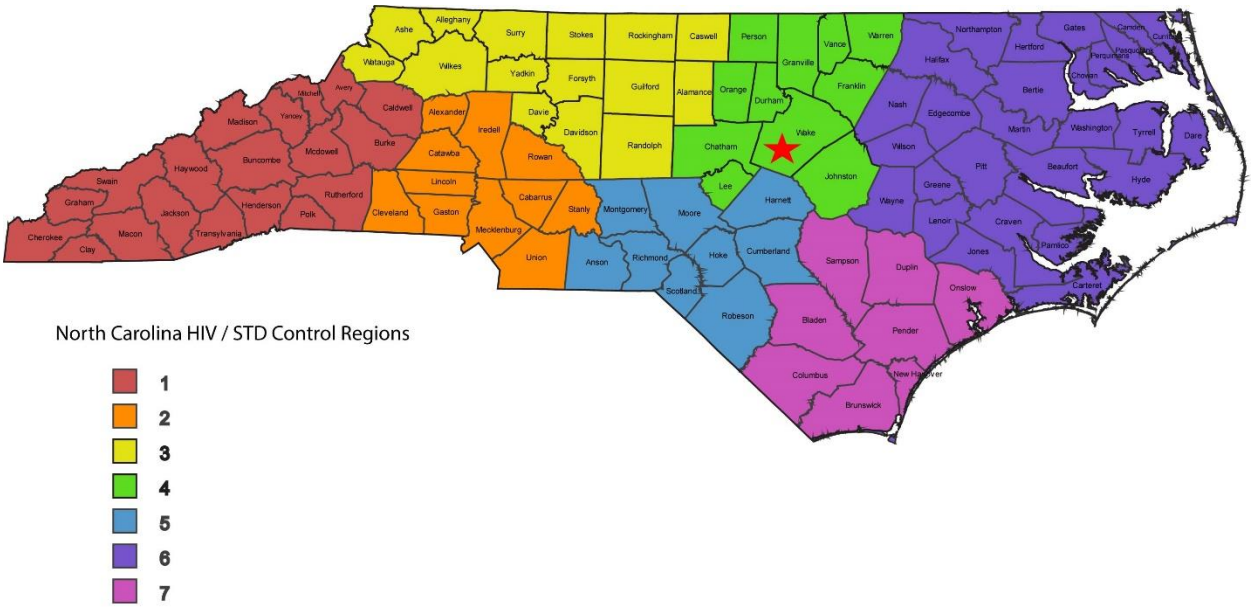
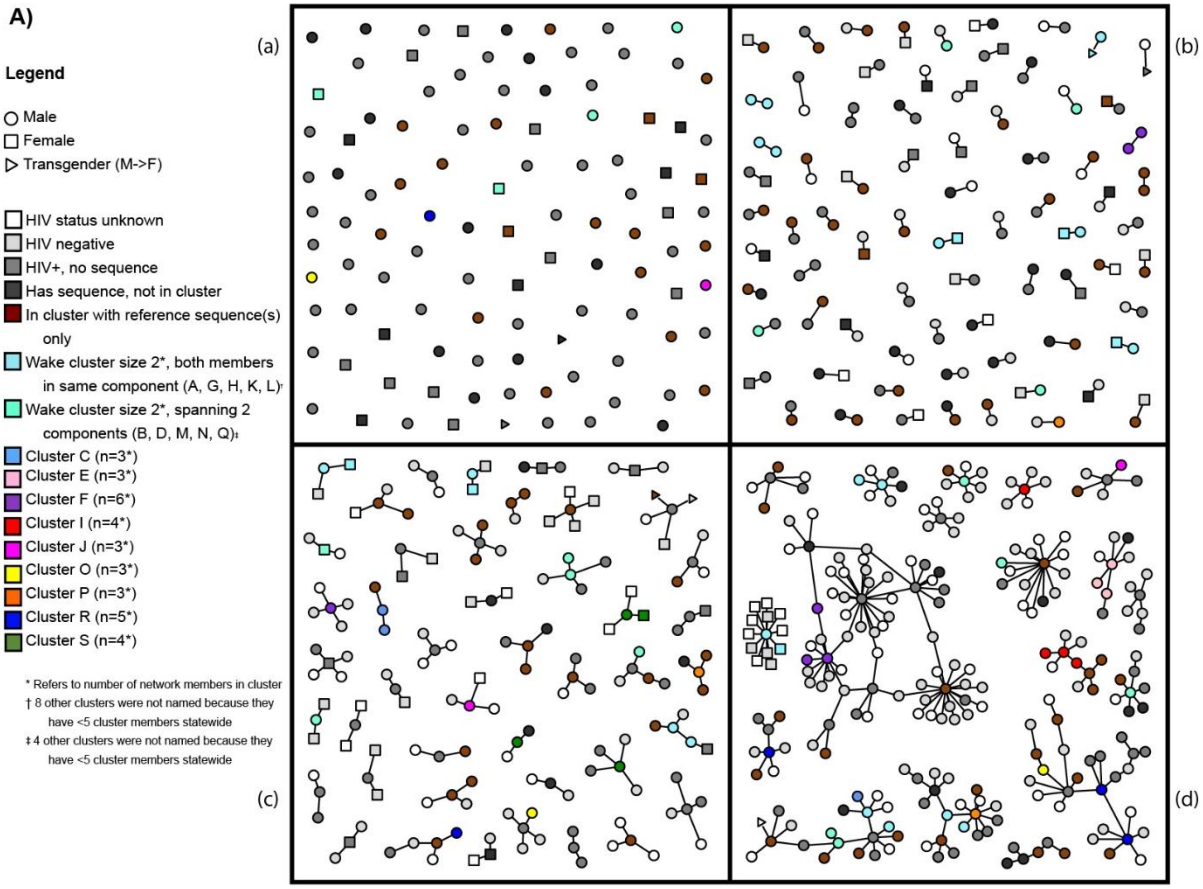


Figure legend:

Map showing the previous North Carolina field service regions for HIV and STD control, as they were during the conception of this study. Wake County, marked with a star, is one of eleven counties in Region 6.

Figure 11. Sexual network showing phylogenetic cluster membership and gender (A), and selected sexual network components showing cluster members and genetic distance statewide (B)



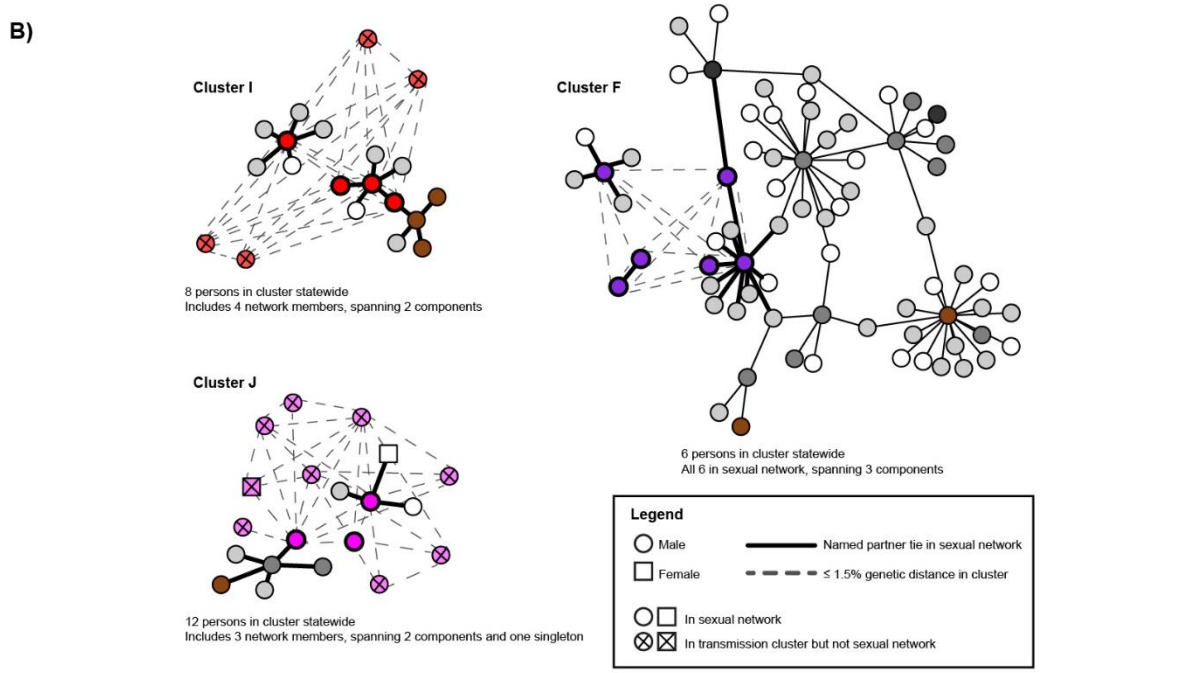


Figure legend:

1A) Sexual and social network compiled from contract tracing depicting HIV status and phylogenetic transmission cluster, Wake County, NC during 2012-2013. Graph shows gender (node shape), cluster membership with respect to gene sequence availability and cluster membership of other persons represented in this sexual network (node color), and partnerships disclosed by index cases (lines connecting nodes). The graph is split into quadrants by number of persons in each component: (a) singletons (n=104 persons), (b) dyads (n=75 components), (c) components size 3 (n=22), 4 (n=10), or 5 (n=12), and (d) components size 6 or larger (n=18 components comprising 243 persons).

1B) Selected phylogenetic transmission clusters (F, I, and J) show sexual network components spanned and additional cluster members statewide who were not part of the Wake County-based sexual network. Graph shows gender (node shape), appearance in sexual network or only transmission cluster (diagonal cross in node shape), transmission cluster status (node color), and connections between nodes. Having a named partner tie (i.e., connection in the sociosexual network) is represented by a solid line and being ≤1.5% pairwise genetic distance in the transmission cluster is represented by a dashed line.

Figure 12. Phylogenetic tree of HIV pol gene sequences showing transmission clusters.

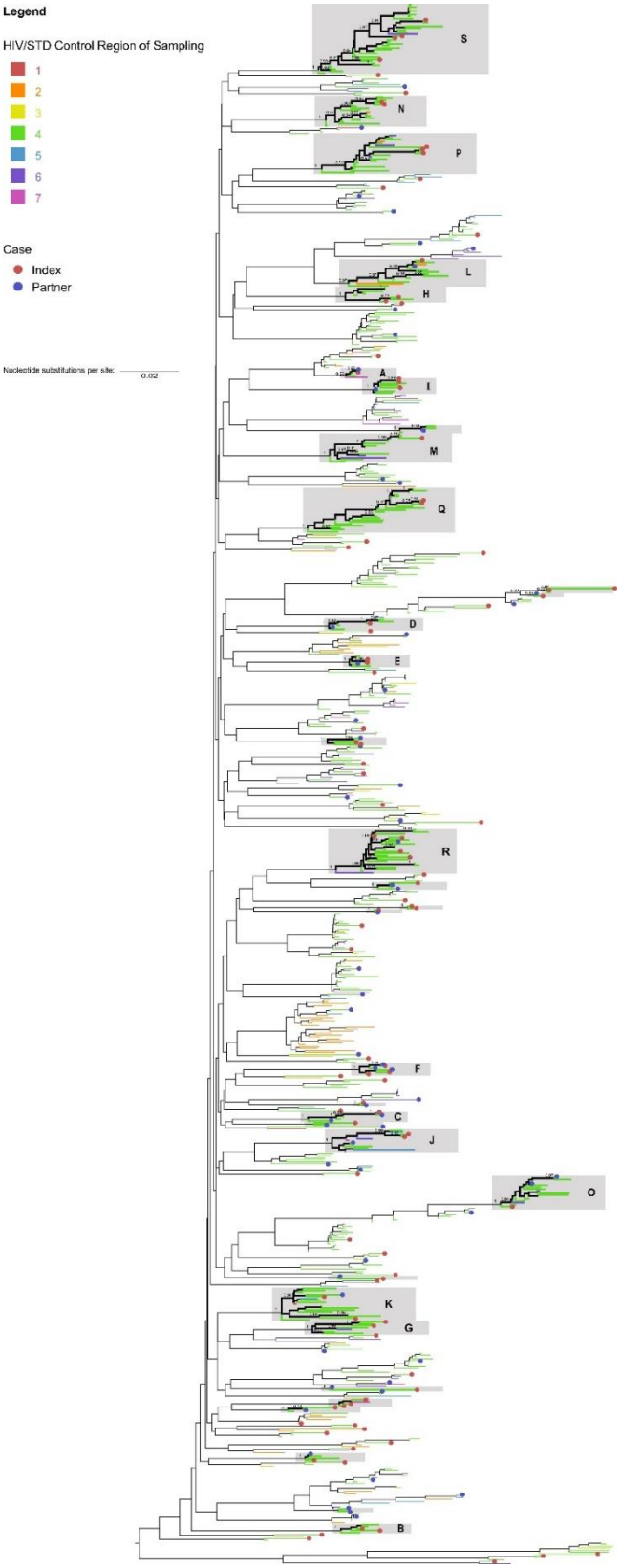


Figure legend:

Maximum-likelihood tree constructed for display purposes using sequences (n=800) identified in confirmed phylogenetic transmission clusters among 15,246 HIV-1 positive persons sampled in North Carolina 1997-2014. Confirmed clusters had posterior probability >0.98 in the Bayesian analysis and include at least one index or partner case identified during partner notification of new HIV diagnoses in Wake County, 2012-2013. Index cases (new diagnoses in 2012-2013) are indicated by red circles and partner cases are indicated with blue circles at the tips of the tree. Clusters in grey boxes involve ≥2 cases from the partner notification network. Clusters with letters (A-S) are the Wake clusters that meet these criteria and also include ≥5 persons statewide. Branch support, using the Shimodaira-Hasegawa-like test values, is included for the Wake clusters.

VI. CHAPTER SIX: MINING THE GAPS: LEVERAGING GENERALIZED ESTIMATING EQUATIONS TO UNDERSTAND HOW PHYLOGENETICS CAN COMPLEMENT CONTACT TRACING[2]

A. Overview

1. Background

An estimated 13-15% of HIV-infected persons are unaware of their status. Public health efforts directed at diagnosis improve community outcomes by increasing the proportion of persons who can be linked to HIV care and eventually virally suppressed.

2. Methods

We combined contact tracing and gene sequence data for all persons diagnosed with HIV in urban Wake County, NC in 2012-2013. We created a dataset of pairs of persons in the same phylogenetic transmission clusters and analyzed whether these persons in the same transmission cluster were in the same sexual network component. We applied a set of generalized estimating equations to the differences amongst phylogenetic cluster and sexual network component membership.

3. Results

Age homophily, race homophily, biological sex, partner number assortativity, the number of months between HIV diagnoses of the persons in the pair, and not having an interview for the person diagnosed second were all significantly associated in the multivariable model with phylogenetically clustered persons not being linked in the contact tracing.

[2]Pasquale, DK, Doherty, IA, Miller, WC, Powers, KA, Sampson, LA, Leone, PA, Sebastian J, Ledford, SL, Eron, JJ, Dennis, AM. Mining the Gaps: Leveraging Generalized Estimating Equations to Understand How Phylogenetics Can Complement Contact Tracing.

4. Conclusions

Interview refusal has a significant impact on disaggregating the network. Encouraging partner elicitation during contact tracing interviews, especially of male cases, would lead to better partner finding and subsequent diagnoses in this network. This novel approach can systematically identify where to direct efforts for improved case finding.

B. Introduction

The Seek, Test, Treat, and Retain strategy put forth by the NIH and WHO aims to identify persons living with HIV infection who do not know their status and remove barriers to retaining them in care and on therapy.[179,180] "Seeking" and "testing" rely on several actions by public health personnel, including contact tracing. Once an HIV-infected person is diagnosed, contact tracing frequently leads to finding more HIV-infected persons.[28,143] However, contact tracing is hindered by inability to locate some partners or interview refusal,[143] which results in persons who cannot be tested and thereby diagnosed. These persons then contribute to losses along the continuum of HIV care.[180,181]

Analysis of partnership dynamics in North Carolina reveals high rates of anonymous exchange sex among heterosexuals during a syphilis outbreak[44] and a substantial proportion of partners who could not be located.[125,182] Consequently, the observable sociosexual network based upon the disclosed and located partners has many "missing" partnerships. As contact tracing is name-based, elicited partnerships can be used to construct a sexual network of index cases and their located partners. Identifying where and why partnerships are "missing" can provide clues to find persons who are active in the network but not identified during contact tracing. To assess the comprehensiveness of contact tracing efforts, another method of representing HIV transmission at a similar scale is needed for comparison.[146] Like sexual network analysis, HIV phylogenetic analysis depicts groups of persons thought to be close together in the transmission chain.

In a qualitative comparison of a contact tracing-based sexual network with transmission clusters inferred from HIV *pol* gene sequences, we found that many index cases had no located partners in the sexual network, despite being recently diagnosed, and more than half of the isolated persons who had an HIV gene sequence were still in a phylogenetic transmission cluster.[183] The combination of these outcomes – recent diagnosis with an HIV gene sequence similar enough to cluster with at least one other person, yet having no locatable partners – indicates gaps in the contact tracing network. Contact tracing could be improved by identifying factors contributing to the network gaps. Here, we quantitatively compared the contact tracing sexual network and the gene sequence analysis network and tested associations where they showed different relationships amongst groups of persons. Using transmission clusters inferred from HIV gene sequence analysis as the "gold standard", we compared the clusters of like infections based upon HIV gene sequence analysis with a sexual network constructed from partnerships elicited during contact tracing.

C. Methods

1. Parent Study

The parent study for the subset of persons included in this multivariable analysis qualitatively compared a sexual network constructed from HIV surveillance data to phylogenetic trees constructed from HIV *pol* gene sequences collected for clinical care from persons in the same geographical area. Methods for sexual network and phylogenetic tree construction for the parent study are previously described.[183] Briefly, all 2012-2013 incident HIV cases residing in urban Wake County, North Carolina (NC) and their partners were abstracted from public health contact tracing records. Disclosed partnerships elicited during contact tracing were used to construct a name-based sexual network. The resultant sexual network, based upon named, located partners, most closely resembles exponential non-discriminative snowball sampling, as any number of partners can be named; an element of discrimination is introduced as only partners who are newly diagnosed with HIV are then asked to name partners.

HIV-positive persons in the parent study sexual network were probabilistically matched to available HIV *pol* gene sequences. The *pol* gene sequences were used to construct phylogenetic trees, calculate pairwise genetic distance between persons, and infer transmission clusters where sequences in the cluster differed by no more than 3.5%. Genetic distance calculations were performed in FastTree[156] using the generalized time-reversible model. Clusters were confirmed with a BEAST analysis.[184]

2. Measures

a. Study Population

We restricted to transmission clusters that had at least 2 persons in the parent study sociosexual network. For the parent study, we created a phylogenetic tree using 15,246 HIV *pol* gene sequences collected (and reported) 1997-2014 in NC and confirmed the transmission clusters with a BEAST analysis.[183] We also created a sociosexual network from Wake County HIV diagnoses made 2012-2013 and their located partners and high risk social contacts (N=663 total persons, n=411 (62%) HIV-positive). We probabilistically matched 56% (230/411) HIV-positive persons in the sociosexual network to the set of 15,246 sequences. Overall, 73% (169/230) of the parent study participants (index or HIV-positive partner with a linked *pol* gene sequence) who matched to a sequence were in a transmission cluster. When using all 15,246 persons with HIV *pol* sequences in the state of North Carolina to create a phylogenetic tree, we found 116 putative transmission clusters which included at least 1 person from the sociosexual network and 34 transmission clusters which included at least 2 persons from the parent study sociosexual network. We noted the 34 “local” transmission clusters, which included 800 total persons, 87 of whom were from the Wake County sociosexual network. Using the 87 persons, we created dyads with each possible pair of persons from the sociosexual network who were in the same transmission cluster (Figure 13).

This quantitative multivariable analysis was restricted to the 87 persons who were in one of 34 local clusters. The purpose was to evaluate if HIV-infected network indexes and partners

identified in the same cluster were directly or indirectly connected within the same sexual network component. The outcome was being in the same sociosexual network component, based upon being in a transmission cluster 'dyad'.

Collected person-level traits were retained from the parent study. Dyad-level traits including age difference, difference in race/ethnicity, gender and sexual preference, and time between HIV diagnoses were calculated for all possible pairs of persons within each local cluster. Pairwise percent genetic distance between *pol* gene sequences was retained from the parent study.

b. Study Design

We set cluster membership as the "gold standard" for relationship linkages and contact tracing network component membership as the comparator. We set component membership as the comparator, rather than restricting only to first-degree partners, because there may be unsampled third persons in the cluster who are unknown, but the known first (or second) degree partners of the unsampled persons would still cluster; being in a cluster doesn't necessarily indicate being first-degree partners. For that reason, we instead used the contact tracing network component as the sexual network cluster since it captures k^{th}-degree partners who are all linked together.

Additionally, we would expect many more links in the contact tracing network than what is represented in the cluster because not every sexual contact results in transmission. Therefore, it is expected that contact tracing network components could contain persons from many different clusters. However, if all partnerships are elicited during contact tracing (in an ideal world) then a cluster should not contain more than one sexual network component since everyone with like infection should have all of their partnerships represented in the contact tracing, thereby linking those persons together.

The contact tracing sexual network components were considered "clusters" of linked persons for the purpose of comparison to the phylogenetic cluster. Among persons in the same

phylogenetic cluster, membership in different contact tracing components suggests that persons or relationships between persons were missing thereby disaggregating the full network into multiple components (Figure 14). This commonly occurs when not all of the links between persons in the transmission chain are discovered during contact tracing and partner elicitation.

3. Statistical Analyses

Each possible pair of persons in the same local cluster (a "dyadic pair") was set up as an observation in the dataset. Each person in the pair retained his or her information (i.e., gender, race) and dyad-level variables (i.e., age difference) were calculated. Each contact tracing-based sexual network component was assigned a number and component membership was assigned as a person-level attribute. Whether the persons in each dyadic pair were or were not in the same sexual network component, knowing that they were in the same cluster due to the way that the pairs were created, was our outcome variable.

a. Generalized Estimating Equations to Compare Linkages

The analysis examined characteristics of dyadic pairs of persons in the same cluster that were associated with being in *different* network components using a set of generalized estimating equations (GEE) with a binomial distribution, logit link function, and robust variance. Robust standard errors accounted for the clustering of the outcome. We selected an exchangeable correlation matrix to treat each person in the cluster as equally likely to have been the transmitting partner. *A priori*-selected model covariates included whether persons in the dyadic pair shared the same race and whether persons in the pair were within 5 years of age, as these factors influence partner selection and network formation.[45,49,54,88] Other covariates tested were related to risk factor assortativity, as both sexual network[62] and transmission cluster[162] analysis have found that partners tend to group by risk behavior. The remaining covariates were selected on the basis of the bivariate relationship between the covariate and the outcome using the odds ratio and confidence interval, at an alpha level of 0.20. The quasilikelihood independence model criterion (QIC)[185] was used to refine the GEE[186]

using the remaining covariates. The final model was the one with the lowest QIC. Covariates were considered significant in the final multivariable model if the *p*-value was ≤0.05. Stata 12[187] was used for all modeling.

D. Results

1. Study Population

The 34 local clusters included 287 persons overall: 87 were part of the parent study and 200 were part of the dataset containing HIV *pol* gene sequences collected in NC and only included to construct the clusters. The 87 persons formed 83 dyadic pairs of persons who were in the same local cluster. These 83 dyadic pairs comprise the observations included in this analysis.

These 87 persons were significantly more likely to be Black when compared to persons included in the parent study but who did not meet the criteria for this nested study (76% v. 62%, p=0.01). Persons in the nested study were younger (mean 32 v. 36 years, p<0.01) when compared to HIV-positive persons from the parent study who were not included in this analysis (Table 7). HIV-positive persons in the nested study were significantly less likely to be a sexual network isolate (8% v 30%, p=<0.01) than HIV-positive persons excluded from this analysis because they were not part of a local cluster.

2. Transmission Cluster and Network Overlap

About half of persons in a local cluster (45/87, 52%) were linkable in the same contact tracing network component as at least one other person in his or her local cluster. The other 48% (n=42) were phylogenetically linked to at least one other person in the study but not identified as being in the same linked group as those persons via the network constructed from contact tracing interviews.

Two-thirds of the dyadic pairs (56/83, 67%) were not in the same network component. The mean genetic distance for the dyadic pairs was 0.9% (95%CI: 0.8-1.1%). Amongst dyadic pairs in the same network component, genetic distance was skewed positively and was

significantly smaller in that group (0.5% v. 1.1%, *P*=0.0001). Genetic distance amongst dyadic pairs in different network components appears to have a bimodal distribution (Figure 15).

3. Generalized Estimating Equations

Age and race were selected *a priori*. When both persons were age 30 years or younger at the index case's diagnosis, they were twice as likely to be in the same transmission cluster but different sexual network components than pairs where at least one person was older than 30 years, although the difference was not significant in the bivariable analysis (odds ratio (OR)=2.2, *P*=0.12; Table 8). Pair members being of different race was not associated with sociosexual component membership.

Both persons in the dyad being male was statistically significantly correlated with being in different network components (OR=4.7, *P*=0.03). The continuous number of months between HIV diagnoses (OR=1.0, *P*=0.12), the person diagnosed later not being interviewed (OR=2.8, *P*=0.12), and having less than 10 sex partners difference between both persons in the dyad (OR=2.5, *P*=0.18) were included in the first multivariable GEE iteration as they met the *a priori* cutoff in the bivariate comparisons. Having anonymous partners was not associated with being in the same cluster without being linked in the contact tracing.

In addition to age, race, and the factors that met the criteria for being included in the multivariable GEE, neither person being acutely infected was also included in the model as it improved the QIC. The number of months between HIV diagnoses was not significant in the multivariable GEE and its removal improved the QIC, so it was dropped from the equation.

The final model included age, race, pair gender, whether the person diagnosed later was interviewed, difference in number of located sex partners, and acute infection at diagnosis. Neither age (adjusted OR (AOR)=1.9, *P*=0.27) nor race (AOR=3.6, *P*=0.06) was significantly associated with being in the same cluster but not the same network component in the multivariate analysis (Table 8). However, the remaining covariates were significantly associated with the outcome of being in a different sexual network component while being in the same

transmission cluster in the multivariable analysis. Both partners being male was highly significant (AOR=25.7, *P*<0.01), as was the person diagnosed later not being interviewed (AOR=8.1, *P*<0.01). The difference in the number of sex partners being less than 10 was an indicator (AOR=4.9, *P*=0.03) as was neither person being acutely infected at diagnosis (AOR=7.9, *P*=0.02).

E. Discussion

Nearly half of the pairs in our analysis were part of the same transmission cluster without being represented as being connected in the contact tracing network. We used a set of generalized estimating equations to identify factors associated with clusters of linked persons in the contact tracing network failing to approximate clusters estimated in the phylogenetic analysis, which is an indication that partnerships were not elicited during contact tracing. Specifically, we observed that contact tracing missed pairs where both partners were male, neither was diagnosed with acute HIV, and the person diagnosed later in time was not interviewed.

For each male-female pair, both persons were in the same network component. Therefore, being in a male-male pair was predictive of not being in the same network component. This has potential public health implications, as it appears that men with male partners are not captured as well in the contact tracing. This could be due to failure to disclose male partners to the disease intervention specialists (DIS) performing the interviews or to the inability of the DIS to locate partners.

There are several reasons why pairs which include someone acutely diagnosed are less likely to be missing. First, DIS pursue persons diagnosed with acute HIV more thoroughly. Second, elicited partnerships would have occurred just prior to diagnosis, so recall bias may be less of an issue. Third, persons with acute infection are more likely to cluster in the phylogeny,[102] and partner testing tends to yield more new diagnoses than when the partners of a chronically-infected person are tested.[188] Newly-diagnosed persons identified during contact

tracing are in turn interviewed to elicit partnerships, which can increase the size, density, and completeness of the sexual network component. This leads to more opportunities for the transmitting partner of the acutely infected person to be included in the network and have an available sequence for the phylogeny.

Missing an interview for the person diagnosed later in an index case-index case pair could mean that we do not have an accounting of partnerships that occurred between the earlier index case diagnosis and the later index case diagnosis. Since we do not have interview data on the later person, we are unable to verify this theory, but the larger genetic distance amongst persons in different network components supports several possibilities. First, the earlier mode may represent partnerships which occurred between the earlier person's diagnosis in the dyad and the later person's diagnosis, so that the earlier person didn't have the opportunity to disclose the partnership and the later person was then not interviewed. Second, the larger genetic distance may represent a partnership that occurred longer ago and was not captured due to recall bias or having occurred prior to the partnership period of interest. The third possibility is that it indicates that there are unsampled intervening persons in the transmission cluster due to the high number of untraced partners and we are seeing second-degree partners clustered together instead of first-degree partners. If the last is true then having anonymous partnerships would be substantively significant even if not statistically significant in the GEE. Inferences must be made with caution from the phylogeny, however, as we used a consensus gene sequence instead of performing deep sequencing and we do not know ART history, which affects viral evolution and therefore clustering.

This analysis provides an initial examination into possible gaps in contact tracing. The magnitude, direction, and statistical significance of the odds ratios suggest that this is worth exploring even if the estimates from the multivariable GEE presented are imprecise due to the small sample size. Future studies should aim to increase the proportion of cases with available sequences or decrease the proportion of missing interviews and untraced partners. A larger

sample size may also reveal more information about partnerships as they relate to HIV transmission. We used GEE because there is certainly some amount of clustering in the data since we are looking at person-to-person transmission, but the exact nature of it is unknown. Despite these limitations, the value of this novel analysis is its quantitative approach to examining contact tracing comprehensiveness related to public health efforts aiming to curb the HIV epidemic in North Carolina.

Diagnosis is critical to ending the HIV epidemic, in the context of currently available interventions. Diagnosis is the first step in the process of engaging patients in care, access to treatment, and viral suppression, which improves quality of life, increases duration of survival, and reduces the likelihood of onward transmission.[180,181,189,190] However, an estimated 13-15% of HIV-infected persons in the United States are unaware of their status,[174,189,191,192] which prevents these people from entering care, initiating antiretroviral therapy, and becoming virally suppressed.[181] The culmination of losses to care along these of the HIV care continuum steps results in only approximately 30% of HIV-infected persons in the United States being virally suppressed.[174,192] Despite the limitations of the depiction of the HIV care continuum,[193-195] diagnosis is the necessary first step to engagement and retention in care. NC has already met the National HIV/AIDS Strategy 2020 program target of diagnosing at least 90% of HIV-infected persons residing in the state,[39] which is excellent. Efforts to identify persons who are missed during contact tracing can keep this proportion steady or help increase it among groups that tend to have lower rates of diagnosis, including persons of color[173] and men who have sex with men (MSM). In support of this, we found that MSM are more likely to be part of sexual network components that are more disjointed than the phylogenetic analysis suggests. The GEE that we created underscores the importance of surveillance and diagnosis by examining which factors are most associated with two persons being in the same local cluster without being in the same contact tracing sexual network component, which is an indication of missing links or persons in the elicited sexual network.

The incomplete contact tracing interviews substantively and significantly affect the structure of the sexual network. They were a significant predictor of not seeing that persons who were in the same transmission cluster were in the same sexual network component, meaning that persons with phylogenetically similar HIV infections were not known to be partners or have partners in common. Partners who are not traced due to incomplete interviews means that there are potentially infectious persons not traced for diagnosis or linkage to care. Public health efforts directed at completing interviews among men who refuse will elicit more partners for testing, thereby filling in some of the gaps in the contact tracing network and keeping NC on track to reduce the burden of HIV.

This analysis allowed us to quantitatively identify traits associated with being part of a phylogenetic transmission cluster that does not approximate one's contact tracing connections. Using GEE, we were able to identify factors that were significantly associated with losses in the sexual network, as measured by having transmission clusters that spanned several contact tracing-based sexual network components. To our knowledge, this is the first time that a sexual network and a phylogenetic tree based upon reported HIV cases have been systematically, quantitatively compared. This analysis provides a novel way to identify characteristics associated with HIV-positive persons whose partners would benefit from testing but are not being located during routine contact tracing.

Figure 13. Flow chart of inclusion into the analysis dataset (N=87 in 83 dyadic pairs), starting from the parent study Wake County network (N=663, 62% HIV-positive (n=411)).

"Local clusters" are transmission clusters which include at least 2 persons from the local sociosexual network.

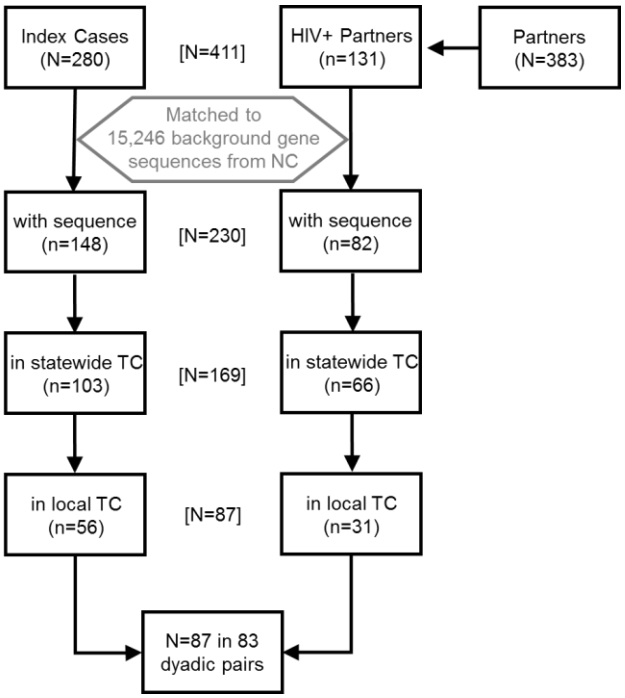


Figure 14. Illustration of possible missing relationships (dotted lines) if the "gold standard" transmission cluster shows persons who are not linked in the sexual network. The person, C, who is disconnected in the contact tracing-based sexual network could have (1) been a partner to A, (2) been a partner to B, or (3) been connected to an unsampled person who is in turn connected to A or B.

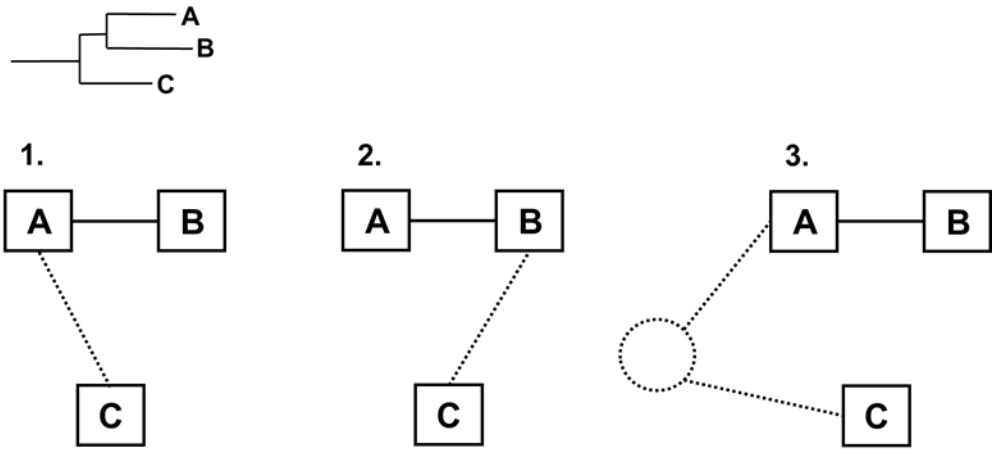


Figure 15. Percent genetic distance by outcome.

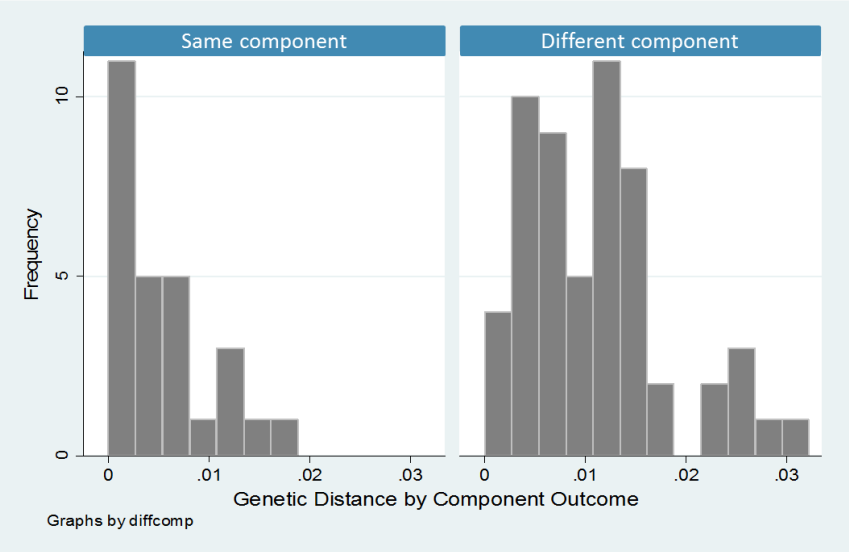


Table 7. Description of index cases diagnosed 2012-2013 in the Wake County, NC parent study and their partners in the sexual network (N=663) and a description of the substudy analysis set restricted on the basis of number of study participants within each transmission cluster (N=87).

	Sexual Network (N=663)		Persons in Analysis Pairs (N=87)	
	Index (n=280)	**Partner (n=383)**	**Index (n=56)**	**Partner (n=31)**
	n (%)	n (%)	n (%)	n (%)
Gender				
Male	232 (83)	327 (85)	49 (88)	27 (87)
Female	44 (16)	53 (14)	6 (11)	4 (13)
Transgender (M to F)	4 (1)	3 (1)	1 (2)	0
Race/ethnicity				
non-Hispanic White	69 (25)	120 (31)	13 (23)	7 (23)
non-Hispanic Black	183 (65)	238 (62)	42 (75)	24 (77)
Hispanic	23 (8)	12 (3)	1 (2)	0
Other	5 (2)	8 (2)	0	0
unknown	0	5 (1)	0	0
Age				
mean (SD)	36 (12)	31 (11)	32 (12)	31 (11)
HIV status				
Positive, with HIV sequence	148 (53)	82 (21)		
Positive, no HIV sequence	132 (47)	49 (13)		
Negative	0	148 (39)		
unknown	0	104 (27)		

Table 8. Bivariable and multivariable relationships in the GEE between explanatory variables and the outcome of being in the same transmission cluster but not in the same sexual network component, by odds ratio (OR) and confidence intervals (CI) (N=83 pairs unless otherwise indicated).

		Different component?							
	Total (N=83)	Yes (N=56)	No (N=27)	Bivariable			Multivariable		
	N (col %)	N (row %)	N (row%)	OR	95%CI	P-value	OR	95%CI	P-value
Different race persons in pair	20 (24)	15 (75)	5 (25)	1.6	0.5-5.0	0.40	3.6	1.0-13.1	0.06
both persons in the pair identify as the same race	63 (76)	41 (65)	22 (35)	1.0			1.0		
Neither person African-American race	17 (20)	13 (76)	4 (24)	1.8	0.5-6.8	0.37			
at least one person in the pair identifies as Black	66 (80)	43 (65)	23 (35)	1.0					
Age difference > 5 years	37 (45)	26 (70)	11 (30)	1.2	0.5-3.0	0.71			
age difference <= 5 years of age	46 (55)	30 (65)	16 (35)	1.0					
Both age 30 or younger at diagnosis	50 (60)	37 (74)	13 (26)	2.2	0.8-5.8	0.12*	1.9	0.6-5.7	0.27
at least one person more than 30 years of age at diagnosis	33 (40)	19 (58)	14 (42)	1.0			1.0		
Both reside in Wake County at diagnosis	55 (66)	38 (69)	17 (31)	1.5	0.5-4.1	0.46			
at least one person doesn't reside in Wake Co at dx	28 (34)	18 (64)	10 (36)	1.0					
Less than 5 years between HIV diagnoses	68 (82)	48 (71)	20 (29)	2.0	0.6-6.3	0.26			
>= 5 years between HIV diagnoses	15 (18)	8 (53)	7 (47)	1.0					
Both persons are index cases	34 (41)	28 (82)	6 (18)	3.4**	1.2-9.5	0.02*	16.2**	6.7-39.2	<0.01
index/partner or partner/partner pair	49 (59)	28 (57)	21 (43)	1.0			1.0		
Person diagnosed second not interviewed	18 (22)	15 (83)	3 (17)	2.8	0.8-10.0	0.12*	8.1**	2.0-33.2	<0.01
Person in pair diagnosed second is interviewed	65 (78)	41 (63)	24 (37)	1.0			1.0		
Total number of sex partners	6.07 +/- 4.86	5.74 +/- 5.50	6.23 +/- 4.56	1.0	0.9-1.1	0.83			
Less than 10 partners difference	73 (88)	51 (70)	22 (30)	2.5	0.7-9.9	0.18*	4.9**	1.1-21.3	0.03
more than 10 sex partners difference among pair	10 (12)	5 (50)	5 (50)	1.0			1.0		
Number of anonymous partners [N=34]	1.74 +/- 2.92	3.00 +/- 6.00	1.46 +/- 1.82	0.9	0.7-1.1	0.32			
Neither person has anonymous partner(s) [N=53]	18 (34)	14 (78)	4 (22)	1.1	0.3-3.9	0.88			
at least one person in pair has at least one anonymous ptn	35 (66)	27 (77)	8 (23)	1.0					
Neither person acutely infected at diagnosis	71 (86)	49 (69)	22 (31)	1.6	0.5-5.8	0.46	7.9**	1.5-43.0	0.02
at least one person in pair acutely infected at diagnosis	12 (14)	7 (58)	5 (42)	1.0			1.0		
Male-male pair	72 (87)	52 (72)	20 (28)	4.7**	1.2-18.4	0.03*	25.7**	7.1-92.9	<0.01
male/female or female/female pair	11 (13)	4 (36)	7 (64)	1.0			1.0		

* Tested in the multivariable model based upon bivariate relationship with outcome

** Significant at alpha=0.05

VII. CHAPTER SEVEN: PREDICTING INDICATORS OF ONGOING HIV TRANSMISSION RISK AFTER HIV DIAGNOSIS IN NORTH CAROLINA[3]

A. Overview

1. Background

Transmission potential of HIV can be reduced by providing antiretroviral drugs as therapy to infected persons or as pre-exposure prophylaxis to seronegative persons. Engaging previously-diagnosed persons with indicators of onward transmission potential in care is an efficient way to reduce transmission potential.

2. Methods

We used HIV surveillance data to create a retrospective cohort of all persons newly HIV-diagnosed over a two-year period in the area around Raleigh, North Carolina. We assessed two surveillance-based indicators of ongoing network involvement among these cohort members over the subsequent three years: incident sexually transmitted infection (STI) or being named as a sexual contact of a newly diagnosed case of HIV or syphilis. We used logistic regression to construct two predictive models, one using only simple information collected through routine surveillance and one that also used sociosexual network statistics, to identify cases with these indicators.

3. Results

Of 569 newly HIV-diagnosed cases included in the cohort, one quarter (N=x) had one of the two outcomes indicating continued involvement in the network within the first three years of diagnosis. Combining demographic characteristics, HIV/STI testing history, and sociosexual

[3]Pasquale, DK, Powers, KA, Doherty, IA, Dennis, AM, Samoff, E, Maxwell, J, Barnhart, J, Leone, PA, Miller, WC. Predicting Indicators of Ongoing HIV Transmission Risk After HIV Diagnosis in North Carolina.

network data was predictive of future indicators of risk behavior. A network model including demographics, HIV/STI testing history, and sociosexual network measures correctly classified 80% of cases, and a simple model without network measures correctly classified 74% of cases.

4. Conclusions

Our predictive models indicate that data collected at the time of HIV diagnosis may help identify persons who are likely to have future outcomes that are consistent with ongoing HIV transmission risk behavior. Information gathered from the sociosexual network reduced the number of cases followed while increasing the percent of cases correctly classified.

B. Introduction

Most HIV infections in the United States (US) are acquired sexually.[3] Sexual transmission of HIV is driven by uncontrolled HIV prevalence among active members of the sexual network,[24,190] which can be reduced with antiretroviral therapy (ART).[196,197] Immediate ART is the current recommendation, but only 60% newly diagnosed cases in the US link to and remain in care[20] and fewer than half of diagnosed adults achieve sustained viral suppression.[10] For newly diagnosed persons who fail to link to care within one year of diagnosis,[20,39] contact with medical or public health professionals occurs only at the time of diagnosis.

Across the US, too few people receive or are retained in care following HIV diagnosis,[20] which is a barrier to achieving viral suppression and thereby increases the possibility of onward transmission. In North Carolina (NC), the proportion of new sexually-transmitted HIV infections due to partners who were aware of their HIV-positive status at the time of transmission appears higher than in the US as a whole.[28] NC has achieved significant increases since 2009 across the HIV care continuum,[181] including meeting the 90-90-90 target goal[180] for diagnosis in 2015.[39] But too few cases are engaged and retained in care to meet the viral suppression goal.[39] While losses at any step of the HIV care continuum increase the likelihood of onward transmission,[190] engaging previously diagnosed persons is likely to be more efficient than identifying people who are undiagnosed, since diagnosed persons are already known to medical and public health

professionals. Therefore, directing limited public health resources towards intensive linkage and retention services for diagnosed persons with a high potential for onward transmission may be more effective for reducing HIV incidence.

Targeting high-risk persons might be further refined with epidemiological analysis of the socisexual network.[11,23,198] Partnership patterns are associated with sexual risk and disease transmission in both observed[8,9,45,199] and simulated networks.[25,70,111,126] The structure of the network itself provides clues to the underlying processes driving relationship formation[76,98] and certain features are typical across networks where infections are sexually transmitted, even with differences in partner selection behaviors.[200] Reduction in risk behavior following diagnosis reduces HIV prevalence in the network,[117] although risk behavior reduction is not constant.[78,79] Therefore, assessment of network position at diagnosis may provide clues to behavior and partnerships following diagnosis.

We sought to develop a risk score algorithm to be used at the time of diagnosis with the goal of identifying persons with the highest transmission risk potential, taking into account activity in the sexual network after HIV diagnosis. Using demographic, sexually transmitted infection (STI) and HIV testing history, and sexual network characteristics collected at the time of HIV diagnosis, our model predicts who would benefit from enhanced efforts to link to and remain engaged in care throughout the course of this chronic infection.

C. Methods

1. Study Population, Setting, and Data

NC is divided into 10 regions for HIV and sexually transmitted disease (STD) control activities. Region 6 (R6) in north central NC (Figure 7) comprises 4 urban/suburban and 7 rural counties based upon 2012 USDA rural-urban continuum codes,[201] with a total population in 2012 of ~1.9 million persons[202] including ~5,700 persons living with HIV or AIDS (PLWHA).[38] The rate of new diagnoses in R6 was 14.4 per 100,000 population in 2012, with ~300 new diagnoses each year.[203]

Disease intervention specialists (DIS) attempt to interview all persons newly diagnosed with HIV and syphilis to assist with care linkage and elicit sexual and needle-sharing partners for testing. DIS elicit partners in the previous 12 months for established HIV cases; 6 months for persons thought to be recently infected based upon acute viral illness or recent negative HIV test; and 2 months for persons diagnosed during acute HIV infection per 4th generation antibody or RNA test results. DIS also elicit and trace high risk social contacts at their discretion due to the overlap between social networks and sexual partners,[150] particularly among Black men who have sex with men (MSM).[8,61]

We identified R6 residents age ≥14 years and first diagnosed with HIV during 2012-2013 from State of NC surveillance records. For each of these index cases ("indexes"), we abstracted from the surveillance system demographic characteristics, risk factors (transactional sex, drug use, anonymous partners) immediately prior to diagnosis or lifetime, HIV-related laboratory results, STI and HIV testing history, and recent partnerships. For each partner identified, we abstracted demographic characteristics, risk factors, and HIV and STI history if known. The University of North Carolina Biomedical Institutional Review Board approved the study.

2. Study Design

Continued involvement in the sociosexual network within 3 years (1,095 days) after HIV diagnosis date was inferred for each index based on surveillance system records (more information below). We also abstracted HIV viral load results, which are reportable in NC, during this period to assess viral suppression overall and at each date where the index had an outcome. Indexes were determined to be virally suppressed if viral load was <200 cells/mm^3 or undetectable, with the index designated as durably suppressed during the outcome if the outcome was reported between two clinical visits occurring ≤200 days apart and the index was virally suppressed at each.

3. Sociosexual Network Construction

A sociosexual network is composed of discrete components, groups of persons joined together by social or sexual relationships, and singletons, solitary persons who are not joined to anyone else. Partner notification (PN) data collected by DIS during HIV and syphilis investigations for indexes age 14 or older at diagnosis were used to create an undirected sociosexual network using the igraph[114] package in R.[115] The network immediately surrounding the index was treated as a local ego network and network characteristics and structures were collected based upon index network position. Partnerships elicited during syphilis investigations in R6 during 2012-2013 were included to better understand network position among high-risk persons. As such, syphilis network components that did not include at least one person from the HIV investigation were discarded from the analysis network since they would not affect index network position.

Network structures ascertained included size of index's component, density of index's component, and inclusion in triangles or *k*-cores (closed loops where each person in the loop has *k* partners). Dyad (partnership)-level covariates, including race homophily (agreement of race between partners), age difference, and seroconcordance, were calculated for each partnership. Multiple node-level measures of centrality were calculated after network construction, including simple degree (number of directly linked people in the network), degree weighted by partners' average degree ("adjusted degree"), and betweenness. Betweenness measures the number of times that a path between any two persons (nodes) in the network passes through the index.[89] All network members were de-identified after network construction to preserve confidentiality.

4. Statistical Analyses

We created a composite outcome that included the first time an index was reported in the surveillance system to be: 1) identified as a partner of a new HIV or syphilis case where the relationship began before index HIV diagnosis and continued for ≥2 months after, or started at

any point after index diagnosis, or 2) diagnosed with a new STI ≥6 months after HIV diagnosis; STIs included syphilis, gonorrhea, chlamydia, and non-gonococcal urethritis (Figure 16). These composite measures were chosen to reflect continued sexual risk behavior with potential consequences after diagnosis. A binary outcome measure was created to indicate the presence of either outcome within 3 years (1,095 days) of HIV diagnosis date.

Model predictors were based on variables collected at the time of the index's HIV diagnosis and partner notification interview. To reflect the time at which the model would be applied in future public health practice, we used information as it was understood at the time of diagnosis, even if that information later changed. Predictors abstracted from the surveillance system included demographics (gender, estimated transmission risk category, age, race/ethnicity, marital status), social environment (college student, job, prison), sexual risk factors, STI and HIV testing history. Other predictors were calculated based upon the constructed network (dyadic factors such as race homophily, network structures).

Predictors were categorized as demographic, STI testing history, and network structure; we were not able to use risk/behavior in the multivariable model due to the amount of missing data. We calculated unadjusted odds ratio for the association between each predictor and the outcome. Predictors in each category were tested against each other for collinearity; if meaningful collinearity was observed, one variable was selected for multivariable model candidacy based on predictive ability against the outcome and ease of calculation.

Due to the high number of predictors associated with the outcome, we performed backwards elimination in two stages for model construction. First, we created a multivariable logistic model containing all predictors with unadjusted odds ratio $p<0.20$ for each predictor category using backwards elimination within each predictor category model at $\alpha=0.05$. The variables selected for each category model were then added as a set sequentially in each possible order to the sets of variables in the other category models, with a second round of

backwards elimination at α=0.05 and refined by of Akaike's information criteria[204] to arrive at a final, combined model.

We repeated this process twice to create two final models. One included all information categories, including network terms, as predictors and the other only used standardly-collected demographic information and STI/HIV history as predictors. For each model, 1,000 bootstrap samples of indexes were randomly selected with replacement to internally validate the model. We used robust standard errors to account for non-independence of the indexes with respect to the network predictors and outcome. *β*-coefficients were evaluated in each bootstrap for consistency in relation to the outcome. We report the variables retained in the final bootstrap multivariable models, with bias-corrected conference intervals (Table 2). *β*-coefficients from the final "network" and "simple" bootstrap models were summed to estimate risk for each index. Modeling was performed using Stata 15.[116]

We compared sensitivity and specificity of each model at various model score cut points. *A priori*, we decided to maximize sensitivity over specificity because the current protocol is to link all newly diagnosed persons to care, so the risk is greater for failing to identify cases at high risk of onward transmission than it is for identifying low-risk persons who would benefit from enhanced linkage to care support. Per a discussion with public health officials regarding this population, we weighted false negatives by a factor of 2.5 and calculated total error rate (false positive + weighted false negative) based upon observed outcome prevalence. We compared predictive capability and total weighted errors at various levels of sensitivity and specificity for the network and simple models to determine whether only routinely-collected predictors could be applied in the field with acceptable loss of capability.

D. Results

1. Study Population

A total of 569 new HIV diagnoses were reported among R6 residents age ≥14 years during 2012–2013. Most newly-diagnosed persons were male (79%) and Black (66%). Median

age at diagnosis was 33 years (IQR: 24–45) [Table 9]. Based upon laboratory results, 32 (6%) were acutely or recently infected with HIV and 144 (25%) were concurrently diagnosed with AIDS; there was imperfect agreement of cases when comparing HIV stage at diagnosis per laboratory results with stage estimated at interview.

In the 3 years following diagnosis, 483 (85%) indexes had 2,746 viral load results reported (median=6 (IQR: 3–8)). Nearly 2/3 (62%, n=1,695) of all viral load tests indicated viral suppression, with 51% (290/569) indexes having durable viral suppression at least once in the 3 year period following diagnosis (length of time ranging from 7 months through censoring at 3 years). However, nearly half (43%) of indexes either had no viral loads reported (n=86), had tests reported but none indicated viral suppression (n=94), or had periods of viral rebound interspersed with periods of suppression (n=65). Black persons were less likely to ever achieve durable viral suppression, while Whites, Hispanic/Latinos and people of other races were more likely (p=0.009).

2. Elicited Contacts and Baseline Sociosexual Network

Nearly all indexes (97%) were interviewed, although 26% indexes (n=146) declined to discuss contacts. Those who disclosed contacts (n=423, 74%) reported a total of 1,850 sexual partners (median=2 (IQR: 1–4), range 0–60), 130 high-risk social contacts (range 0–19), and 5 needle-sharing partners (range 0–3) in the 2, 6, or 12 months prior to diagnosis (depending on infection stage; see Methods). Of the sexual partners reported, 521/1,850 (28%) did not have enough locating information to initiate partner notification. DIS attempted to locate and notify the remaining 1,329 sexual partners, along with all social and needle-sharing contacts.

Indexes and their located first-degree contacts formed 845 relationships: 749 sexual partners, 92 social contacts, and 4 needle-sharing partners, representing 40%, 71%, and 80% of total claimed, respectively. These contacts formed the sociosexual network used for analyses. Most sexual partnerships were among people of the same race (78%), included at least one person of Black race (77%), and were among two men (72%). Half (51%) of sexual

partners were ≤5 years of age apart. Among the 749 sexual and 4 needle-sharing partners, 42% (319/753) were HIV-positive (diagnosis year range 1995–2017, 46% diagnosed ≥6 months prior to index), 30% were documented to be HIV-negative during the investigation, and 27% had unknown serostatus. Among 425 (75%) indexes who were not concurrently diagnosed with AIDS, <1 HIV-positive sexual or needle-sharing partner on average was identified during the investigation (mean=0.64, SD=1.05; total=270).

After de-duplication (106/845 contacts were indexes themselves and the other 39 were contacts of >1 index), 700 unique first-degree contacts were added to the network along with 201 additional people from concurrent syphilis investigations who were linked to HIV indexes or contacts. The total network size was 1,470 persons (Figure 17). Excluding the 569 newly-diagnosed indexes, 283 (31%) network members were HIV-positive (median diagnosis year 2009 (IQR: 2006–2012) excluding 80 (28%) unknown diagnosis years), 272 (30%) were HIV negative based upon a test during the investigation period, and serostatus was unknown for the remaining 346 (38%).

Nearly half of indexes (44%, n=248) were isolated in the network, with no located partners: 54% (n=134) isolates did not disclose partners, 17% (n=42) claimed zero sexual partnerships, and 29% (n=72) claimed 1-50 partners (median=2 (IQR: 1-4). The remaining 321 indexes formed 220 discrete components of ≥2 people (Figure 1), most of which contained two people (n=238 people across 119 dyads) or three people (n=144 people across 48 components) [Figure 17]. The largest component included 320 persons (22% total network, 8% (46/569) indexes). Component size, dominant demographic characteristics, and member HIV status are shown in Table 10a and Table 10b. Ten (4%) of 220 components contained second-degree partnerships from syphilis investigations. Seven (1%) indexes were in at least one triangle and 22 were in a *k*-core: 21 (4%) indexes were in a 2-core and 1 (0.2%) was in a 3-core. Due to the high number of isolates and smaller-sized components in the network, transitivity and betweenness scores were uninformative.

3. Outcomes

One-quarter of indexes (141/569) met the outcome definition within 3 years after HIV diagnosis. Over three-fourths of those with the outcome (78%, n=110) were diagnosed with a new STI ≥6 months after HIV diagnosis. The remaining 22% (n=31) were identified as a partner by someone who was diagnosed with HIV or syphilis after the index's HIV diagnosis, with the dates of the relationship as described by the new case continuing for at least 2 months or starting any time after the index's diagnosis. Time to outcome following HIV diagnosis ranged from 9 days–2.9 years, where the smallest time was the start of a new relationship with someone who was later diagnosed with HIV. The median time was 1.4 years (IQR: 8 months–2.1 years).

Among 141 outcomes, 33 (23%) occurred during a period of durable viral suppression: 17/110 (15%) STI outcomes and 16/31 (52%) partnership outcomes. Median length of durable viral suppression periods with an outcome was 1.7 years (IQR: 11 months–2.5 years), with the outcome occurring 8.5 months after the start of the period (median; IQR: 3 months–1.4 years). Cases who met the outcome definition were more likely to have been durably suppressed at least once during the follow-up period ($p<0.05$).

4. Bivariable Analysis

Demographic predictors were highly associated ($p\leq0.0001$) with having the outcome, including being male (unadjusted odds ratio (uOR)=3.5, 95% confidence interval (CI): 1.8–6.7), younger than 30 years of age at HIV diagnosis (uOR=3.9, 95%CI: 2.6–5.9), and unmarried (uOR=9.9, 95%CI: 4.3–23.1); race/ethnicity was not associated (Table 2).

Categorized degree (number of partners), categorized degree modified by mean number of partners' partners, and component size were each highly predictive (uOR $p<0.0001$) [Table 11]. Other network measures are sensitive to number of partnerships, so the network terms were not associated with the outcome due to the high proportion of cases who were singletons or in smaller sized components. Being in a *k*-core was associated with the outcome, although

the confidence interval was wide (uOR=9.0, 95%CI: 3.4–23.5). Indexes who had more sexual partners in the network than they disclosed were twice as likely to have the outcome (uOR=2.3, 95%CI: 1.3–4.1).

5. Development of Multivariable Models

Not being married was the strongest predictor of the outcome in both models. Younger age, non-rural county of residence, not being concurrently diagnosed with AIDS, and choosing to disclose partners were predictive in both models, although rurality was not significant in the bootstrapped simple model (Table 11). The network model additionally included component size, categorized as isolated, 2–4 persons, and ≥5 persons in the component, where each categorical increase in component size increased likelihood of being an outcome. However, being in a component size 2–4 persons was not significant compared to being a singleton in the bootstrapped model.

6. Predictive Capabilities of Multivariable Model

The simple model ROC area was 0.83 (Figure 18). Based upon the observed outcome prevalence of 25%, among a hypothetical population of 1,000 newly-diagnosed persons, the lowest number of unweighted errors across all cut-points for the simple model was 207 (at sensitivity=47%, specificity=90%). The lowest total weighted errors with the simple model was 339 (sensitivity=80%, specificity=72%) [Table 12a]. At the weighted sensitivity and specificity levels, bridge counselors would intervene on 415 people, of whom 200 (48%) would actually be likely without intervention to later arise as STI cases or partners of newly diagnosed HIV/STI cases, while 50 such people would not be identified for enhanced public health support.

The network model ROC area was 0.84, which was not significantly different than the simple model area under the curve. The lowest number of unweighted (n=201) and weighted (n=321) errors occurred when sensitivity=68% and specificity=84% (Table 12b). Bridge counselors would intervene on 292 people, of whom 170 (58%) would actually to have the outcome in the absence of intervention, while missing 80 higher-risk persons. At the sensitivity

and specificity levels where each model had the lowest number of weighted errors, the simple and network models correctly classified 74% and 80% cases, respectively.

E. Discussion

We sought to identify who is likely to remain active in the sociosexual network following HIV diagnosis, with the goal of assisting these patients to remain in care and ultimately achieve viral suppression. We developed two models with sufficient predictive ability to identify most new HIV cases with higher transmission risk potential. Both models used information collected at the time of diagnosis, including demographics and STI and HIV testing history. One model also included the number of people directly or indirectly linked to the case in the sociosexual network as a predictor, which increased the likelihood of correctly classifying cases by post-diagnosis network involvement.

Ours is not the first study to leverage information from the time of diagnosis to reduce onward transmission. Similar to what we found by using network component size, number of partners expressed as a percentile was predictive of onward transmission risk among cases in a network constructed from gene sequences instead of partnerships.[171] Several studies have recognized the importance of the time of diagnosis to encourage reduction in risk behaviors,[205,206] though results are mixed.[207,208]

Care engagement, however, appears to reduce risk behaviors,[190] which highlights the importance of linkage to care support. A reduction in risk behavior may prevent HIV and syphilis co-infection, which is a concern given the frequency of STI infections that occur after diagnosis, both in our study and elsewhere,[21] and the overlap of HIV and syphilis networks,[209,210] as shown by the components which were bridged by second-degree partnerships captured during syphilis investigations.

Even with the addition of syphilis cases, it is clear that partnerships are missing when fewer than one HIV-infected partner was found per new case who was not diagnosed with late-stage infection. Multiple network measures that we calculated are sensitive to missing data,[211]

notably degree and the centrality measures in sparse networks.[212] It is possible that the network structures would have been associated with transmission potential had we observed more of the network, which may partially explain why our results did not agree with simulation studies showing the importance of network structures.[98,213,214]

We may have had fewer prediction errors had we been able to use pre-diagnosis risk behavior, given its associated with post-diagnosis behavior.[77,205] Some degree of outcome misclassification is possible, too. Outcomes required reporting to the State within the 3 year follow-up period; partnerships or STIs not reported at all or reported after the follow-up period, even if the risk behavior occurred during the follow-up period, would not be known. However, STI diagnosis was the more frequent outcome basis, which does not rely on partner disclosure. Additionally, the STI outcome may be the marker for higher-risk sex[215] and indicates that condoms may not have been used.

While this model was predictive in our population, it has not been validated in another population. It may be challenging to validate elsewhere since the predictors may not be collected the same way. However, a potential future analysis includes validating the model in another geographic area in NC or among cases diagnosed later than those in this sample. Even so, the method of using demographic, STI and HIV testing history, and network information to predict post-diagnosis network involvement could be applied in another population.

This study is timely in that it coincides with changes to NC law decriminalizing sexual activity provided the PLWHA is in care and has been virally suppressed for at least 6 months.[216] Viral loads are reportable in NC, so bridge counselors could access current suppression status for any persons who meet the model criteria before offering additional support to establish or re-establish care. The racial disparities in durable viral suppression agree with recent findings,[10,15,217] so although race and ethnicity were not predictive in our models, it is possible that enhanced support offered would still address racial disparities in HIV incidence since the

majority of partnerships included at least one Black person and/or were among people of the same race.

Several of our findings support our belief that newly diagnosed cases would be valuable intervention targets to stop the spread of HIV. We found fewer than one known HIV-positive sexual or needle-sharing partner per new case who was not diagnosed with late-stage disease, implying that many HIV-positive sexual network members cannot be located. Therefore, regular contact with newly diagnosed cases may prevent them from becoming "lost". Among these newly diagnosed persons, one-fifth were diagnosed with a new STI 6 months–3 years after HIV diagnosis. Inflammatory and ulcerative STIs can increase the likelihood of transmitting HIV,[218] and the time period during which this occurred in our population is beyond what is considered the most infectious period since all new cases would have passed out of the acute stage by that point; it is possible that this contributes to the high proportion of new infections from previously-diagnosed persons in NC and is worth considering for future study.

We used PNS data collected at the time of HIV diagnosis to predict continued sexual network involvement after diagnosis. Using demographic and STI data, we developed a model to correctly categorize three-quarters of newly diagnosed HIV cases by risk of onward transmission. The addition of sociosexual network data to a second model, while complicated, increased classification to 80%. Sexual network position is associated with risk behavior, and changes in risk behavior following diagnosis[78,219] are likely from the patient's own baseline, so collecting these characteristics at the time of HIV diagnosis may inform patient needs post-diagnosis. This analysis has the potential to refine targets for bridge counseling and direct efforts towards patients who would benefit most from support to establish or re-establish HIV care.

Table 9. Index cases age 14 and older diagnosed 2012-2013 in NC HIV/STD Control Region 6 and their first-degree contacts in the sociosexual network (N=1,269).

	Index (n=569)		Contact (n=700)	
	n	(%)	n	(%)
Gender				
Male	451	(79)	581	(83)
Female	114	(20)	98	(14)
Transgender (M to F)	4	(0.7)	0	---
not indicated	0	---	21	(3)
Race / Ethnicity				
non-Hispanic White	114	(20)	164	(23)
non-Hispanic Black	378	(66)	459	(66)
Hispanic	58	(10)	33	(5)
Other	19	(3)	31	(4)
not indicated	0	---	13	(2)
Region of birth				
USA-50 states	530	(93)	344	(49)
Latin / South America, Caribbean (incl. US Territories)	24	(4)	6	(0.9)
Europe, Asia, Oceania	3	(0.5)	2	(0.3)
Africa	12	(2)	0	---
not indicated	0	---	348	(50)
Marital status				
Currently married	39	(7)	44	(6)
Divorced / separated / widowed	20	(4)	9	(1)
Never married	413	(73)	413	(59)
not indicated	97	(17)	234	(33)
County of residence				
Urban	437	(77)	---	
Suburban	27	(5)	---	
Rural	105	(18)	---	
Age at index case's HIV diagnosis (years)*				
≤ 19	29	(5)	46	(7)
20-29	214	(38)	316	(45)
30-39	101	(18)	169	(24)
40-49	136	(24)	88	(13)
≥ 50	89	(16)	52	(7)
not indicated	0	---	29	(4)
median (IQR)	33	(24-45)	28	(23-37)
HIV status				
Positive	569	(100)	221	(35)
Negative	---		243	(32)
unknown	---		236	(34
Year of HIV diagnosis			n=221	
< 2006	---		50	(23)
2006-2010	---		71	(32)
2011	---		21	(10)
2012	271	(48)	13	(6)
2013	298	(52)	7	(3)
≥ 2014	---		35	(16)
not indicated	---		24	(11)
HIV stage at diagnosis†				
Acute / Recent	32	(6)	---	
Chronic, non-AIDS	393	(69)	---	
Chronic, AIDS	144	(25)	---	

* Among partners, for earliest record associated with any index case

† Based upon laboratory results

Table 10a. Dominant characteristics of sociosexual network components size 7 and smaller (n=248 isolates and n=201 components size 2–7), comprising 794 persons (54% total network).

Component size	Components (N)	People (N)	Number of indexes (mean)	Components at least half Black (n)	Components at least half female (n)	Components with transwomen (n)	Number of residents of Region 6 (mean)	Number known HIV+ (mean)
1	248	248	1.0	149	52	1	1.0	1.0
2	119	238	1.1	88	64	1	1.8	1.3
3	48	144	1.2	39	14	0	2.6	1.8
4	20	80	1.1	15	6	0	3.1	1.8
5	5	25	1.2	3	0	0	2.8	2.6
6	4	24	1.3	3	0	0	3.8	2.8
7	5	35	1.4	4	1	0	4.8	3.0

Table 10b. Demographic characteristics of 19 largest network components, comprising 676 persons (46% total network).

ID	People (N)	Total network (%)	Indexes (n)	White* (%)	Black (%)	Hispanic (%)	Other (%)	Male* (%)	Female (%)	Trans-women (%)	R6 residents (n)	Known HIV+ (n)
A	8	0.5	1	0	100	0	0	63	38	0	8	2
B	8	0.5	2	0	100	0	0	100	0	0	6	5
C	8	0.5	2	38	0	63	0	88	0	13	8	3
D	9	0.6	1	22	56	0	0	100	0	0	8	4
E	9	0.6	1	0	100	0	0	100	0	0	5	4
F	9	0.6	1	11	89	0	22	78	22	0	9	2
G	9	0.6	1	22	78	0	0	11	89	0	2	1
H	9	0.6	2	0	100	0	0	100	0	0	4	7
I	10	0.7	2	30	50	10	0	60	40	0	9	2
J	10	0.7	3	50	40	10	10	100	0	0	10	6
K	11	0.7	1	18	73	0	0	9	73	0	10	1
L	13	0.9	4	0	100	0	0	77	15	0	8	9
M	13	0.9	2	0	92	8	0	100	0	0	7	8
N	15	1.0	2	100	0	0	0	100	0	0	15	5
O	16	1.1	2	0	100	0	0	100	0	0	5	10
P	26	1.8	1	62	31	4	0	100	0	0	19	5
Q	81	5.5	9	1	90	0	2	94	5	0	39	28
R	92	6.3	18	8	85	1	7	93	4	1	47	45
S	320	21.8	46	38	51	5	5	97	0.1	0	164	136

* Ethnicity / race and gender do not sum to 100% for some components due to missing information

Table 11. Bivariable and multivariable relationships between predictors and post-diagnosis continued involvement in the sexual network among 569 persons first diagnosed with HIV during 2012-2013 in NC HIV/STD Control Region 6. Confidence intervals for adjusted odds ratios (aOR) from the multivariable models are bias-corrected from 1,000 bootstrapped samples made with replacement.

Predictor	Total		Remain Involved		Bivariable Analysis		Network Multivariable		Simple Multivariable	
	N	(col %)	n	(row %)	OR	(95% CI)	aOR	(95% CI)	aOR	(95% CI)
	569	(100)	141	(25)						
Demographic Characteristics										
Gender										
Woman or Transwoman	118	(21)	12	(10)	1.0					
Male	451	(79)	129	(29)	3.5	(1.8-6.7)‡				
Ethnicity / Race										
non-Hispanic White	114	(20)	28	(25)	1.0					
Person of color	455	(80)	113	(25)	1.0	(0.63-1.6)				
Country of birth										
USA	530	(93)	137	(26)	1.0					
Other	39	(7)	4	(10)	0.33	(0.11-0.94)‡				
Marital status*										
Married	137	(24)	6	(4)	1.0		1.0		1.0	
Unmarried	432	(76)	135	(31)	9.9	(4.3-23.1)‡	6.5	(2.7-15.9)‡	7.2	(3.0-17.6)‡
County of residence at diagnosis										
Rural	105	(18)	14	(13)	1.0		1.0		1.0	
Urban or suburban	464	(82)	127	(27)	2.4	(1.3-4.5)‡	2.2	(1.1-5.1)‡	2.0	(1.1-4.3)
Age at diagnosis										
≥ 30 years	326	(57)	46	(14)	1.0		1.0		1.0	
< 30 years	243	(43)	95	(39)	3.9	(2.6-5.9)‡	2.0	(1.2-3.3)‡	2.5	(1.6-4.1)‡
HIV and STD History and Care										
HIV interview										
No interview or no partner disclosure	146	(26)	13	(9)	1.0		1.0		1.0	
Interviewed and disclosed partners	423	(74)	128	(30)	4.4	(2.4-8.1)‡	2.5	(1.2-5.7)‡	3.6	(2.0-8.5)‡
Estimated HIV stage at diagnosis†										
Concurrently diagnosed with AIDS	136	(24)	13	(10)	1.0		1.0		1.0	
Acute, recent, or chronic HIV	433	(76)	128	(30)	4.0	(2.2-7.3)‡	2.1	(1.1-4.3)‡	2.4	(1.3-5.1)‡

Predictor		Total		Remain Involved		Bivariable Analysis		Network Multivariable		Simple Multivariable	
		N	(col %)	n	(row %)	OR	(95% CI)	aOR	(95% CI)	aOR	(95% CI)
History of any STI											
	No	385	(68)	67	(17)	1.0		1.0		1.0	
	Yes	184	(32)	74	(40)	3.2	(2.1-4.7)‡	1.5	(0.89-2.5)	1.9	(1.1-3.0)‡
History of syphilis, specifically											
	No	510	(90)	112	(22)	1.0					
	Yes	59	(10)	29	(49)	3.4	(2.0-6.0)‡				
Any STI infection at diagnosis											
	No	432	(76)	75	(17)	1.0				1.0	
	Yes	137	(24)	66	(48)	4.4	(2.9-6.7)‡			2.2	(1.2-3.5)‡
STI other than syphilis at diagnosis											
	No	474	(83)	87	(18)	1.0		1.0			
	Yes	95	(17)	54	(57)	5.9	(3.7-9.4)‡	2.8	(1.4-4.8)‡		
Prior negative HIV test result											
	No	415	(73)	80	(19)	1.0					
	Yes	154	(27)	61	(40)	2.7	(1.8-4.1)‡				
Linked to care within 3 months											
	No	357	(63)	76	(21)	1.0					
	Yes	212	(37)	65	(31)	1.6	(1.1-2.4)‡				
HIV gene sequence available											
	No	231	(41)	57	(25)	1.0					
	Yes	338	(59)	84	(25)	1.0	(0.68-1.5)				
Network Structures											
Singleton											
	No	321	(56)	111	(35)	1.0					
	Yes	248	(44)	30	(12)	0.26	(0.17-0.40)‡				
Degree‖											
	0 (singleton)	248	(44)	30	(12)	1.0					
	1	159	(28)	33	(21)	1.9	(1.1-3.3)‡				
	2	62	(11)	22	(35)	4.0	(2.1-7.6)‡				
	3-5	67	(12)	35	(52)	7.9	(4.3-14.7)‡				
	≥ 6	33	(6)	21	(64)	12.7	(5.7-28.5)‡				

Predictor	Total		Remain Involved		Bivariable Analysis		Network Multivariable		Simple Multivariable	
	N	(col %)	n	(row %)	OR	(95% CI)	aOR	(95% CI)	aOR	(95% CI)
Adjusted degree¶										
0	248	(44)	30	(12)	1.0					
> 0-1	140	(25)	25	(18)	1.6	(0.89-2.8)§				
> 1 to < 2.5	62	(11)	23	(37)	4.3	(2.3-8.1)‡				
2.5 to 5	79	(14)	38	(48)	6.7	(3.8-12.1)‡				
> 5	40	(7)	25	(63)	12.1	(5.7-25.5)‡				
Network sex partners > disclosed										
No	515	(91)	119	(23)	1.0					
Yes	54	(9)	22	(41)	2.3	(1.3-4.1)‡				
Component size										
1 (singleton)	248	(44)	30	(12)	1.0		1.0			
2-4	202	(36)	46	(23)	2.1	(1.3-3.5)‡	1.4	(0.74-2.6)		
≥ 5	119	(21)	65	(55)	8.7	(5.2-14.8)‡	3.0	(1.5-5.7)‡		
Closed triangle involvement										
No	562	(99)	137	(24)	1.0					
Yes	7	(1)	4	(57)	4.1	(0.91-18.7)§				
k-core involvement										
No	547	(96)	125	(23)	1.0					
Yes	22	(4)	16	(73)	9.0	(3.4-23.5)‡				

* Designated as ‘married’ only if noted as such in case record, otherwise classified as ‘unmarried’

† Based upon estimation of investigating DIS at diagnosis, not confirmed by labs

‡ Significant at p<0.05 (bivariable analysis and multivariable analyses)

§ Significant at p<0.20 (bivariable analysis only)

‖ Degree is the number of connections in the sociosexual network

¶ Adjusted degree is number of sociosexual network connections adjusted by partners’ average number of network connections

Table 12a. Sensitivity and specificity of simple model score to predict post-diagnosis involvement in the sexual network at selected risk score cut-offs, including false negative and false positive rates based upon observed 25% prevalence of continued involvement in the sexual network among a hypothetical population of 1,000 persons newly diagnosed with HIV.

Risk score	Sensitivity (%)	Specificity (%)	Patients correctly identified as at risk (True Positive) (n)	Patients missed (False Negative) (n)	Patients correctly identifed as not at risk (True Negative) (n)	Patients unnecessarily followed (False Positive) (n)	2.5 x False negative (n)	Total correctly classified (n)*	Total unweighted errors (n)	Total weighted errors (n) †	Patients followed (n) ‡
7.14	16	97	41	209	725	25	523	766	234	548	66
5.85	47	90	117	133	676	74	333	793	**207**	407	191
5.17	70	76	176	74	571	179	185	747	253	364	355
4.88	80	72	200	50	536	214	125	736	264	**339**	415
4.76	91	62	227	23	478	282	58	695	305	340	510
4.14	95	50	238	12	375	375	30	613	387	405	613
2.16	100	12	250	0	89	661	0	339	661	661	661

* True positive + true negative

† False positive + weighted false negative

‡ True positive + false positive

Table 12b. Sensitivity and specificity of network model score to predict post-diagnosis involvement in the sexual network at selected risk score cut-offs, including false negative and false positive rates based upon observed 25% prevalence of continued involvement in the sexual network among a hypothetical population of 1,000 persons newly diagnosed with HIV.

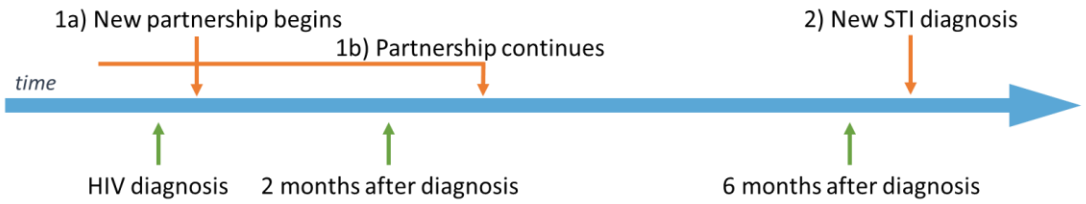
Risk score	Sensitivity (%)	Specificity (%)	Patients correctly identified as at risk (True Positive) (n)	Patients missed (False Negative) (n)	Patients correctly identifed as not at risk (True Negative) (n)	Patients unnecessarily followed (False Positive) (n)	2.5 x False negative (n)	Total correctly classified (n)*	Total unweighted errors (n)	Total weighted errors (n) †	Patients followed (n) ‡
7.51	11	100	27	223	746	4	558	773	227	562	32
6.09	36	94	90	160	708	42	400	798	202	442	133
5.31	68	84	170	80	629	121	200	799	**201**	**321**	292
4.81	72	77	181	69	578	172	173	759	241	345	353
4.13	94	59	234	16	442	308	40	676	324	348	543
3.20	99	33	246	4	249	501	10	495	505	511	748
2.00	100	11	250	0	86	664	0	336	664	664	915

* True positive + true negative

† False positive + weighted false negative

‡ True positive + false positive

Figure 16. Depiction of indicators of remaining active in the sociosexual network following HIV diagnosis used to calculate outcomes.



1) (a) Engagement in a partnership which began after index's HIV diagnosis or (b) began before index's HIV diagnosis date and continued ≥2 months, or
2) Diagnosis with new STI > 6 months after index HIV diagnosis

Figure 17. Sociosexual network showing 569 index cases newly diagnosed with HIV in the area around Raleigh, NC during 2012-2013. Total graph includes 1,470 persons distributed in 468 network components.

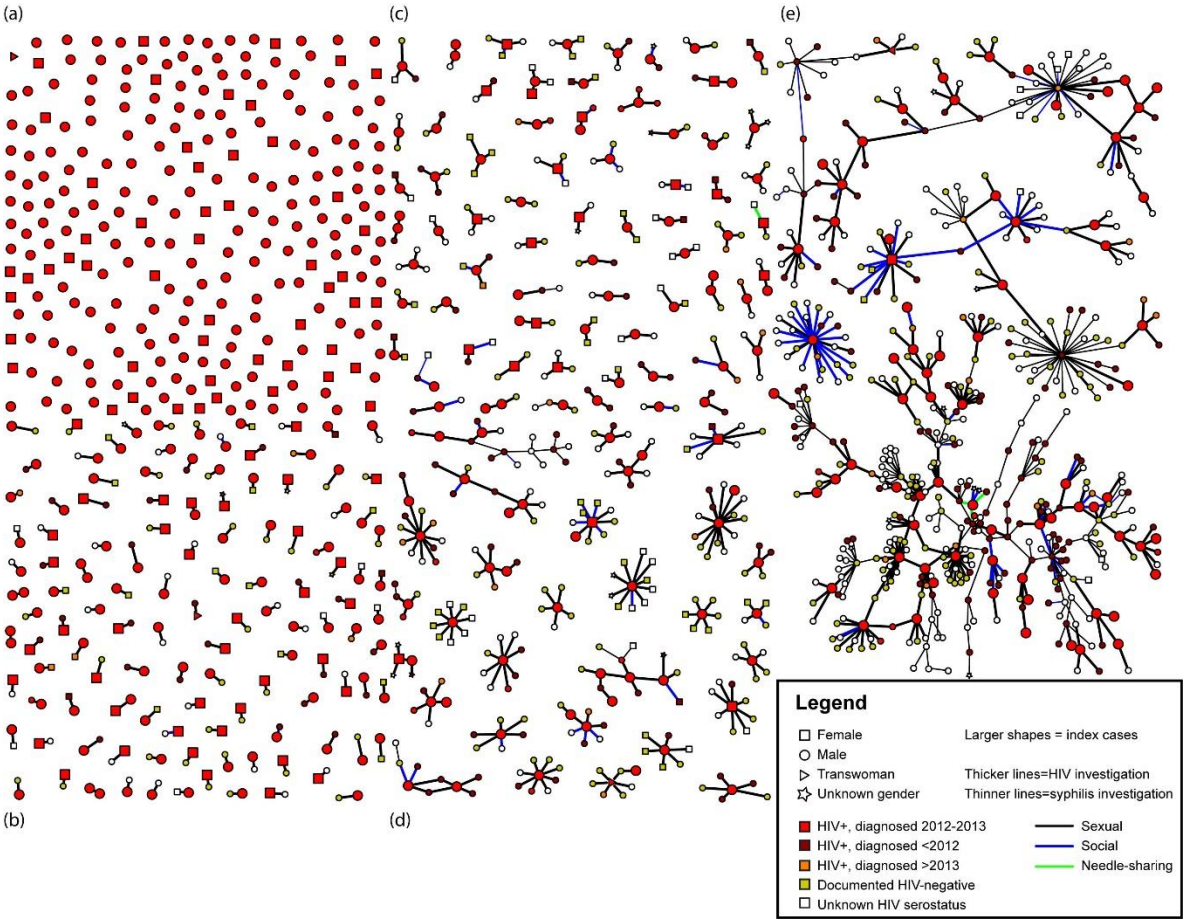
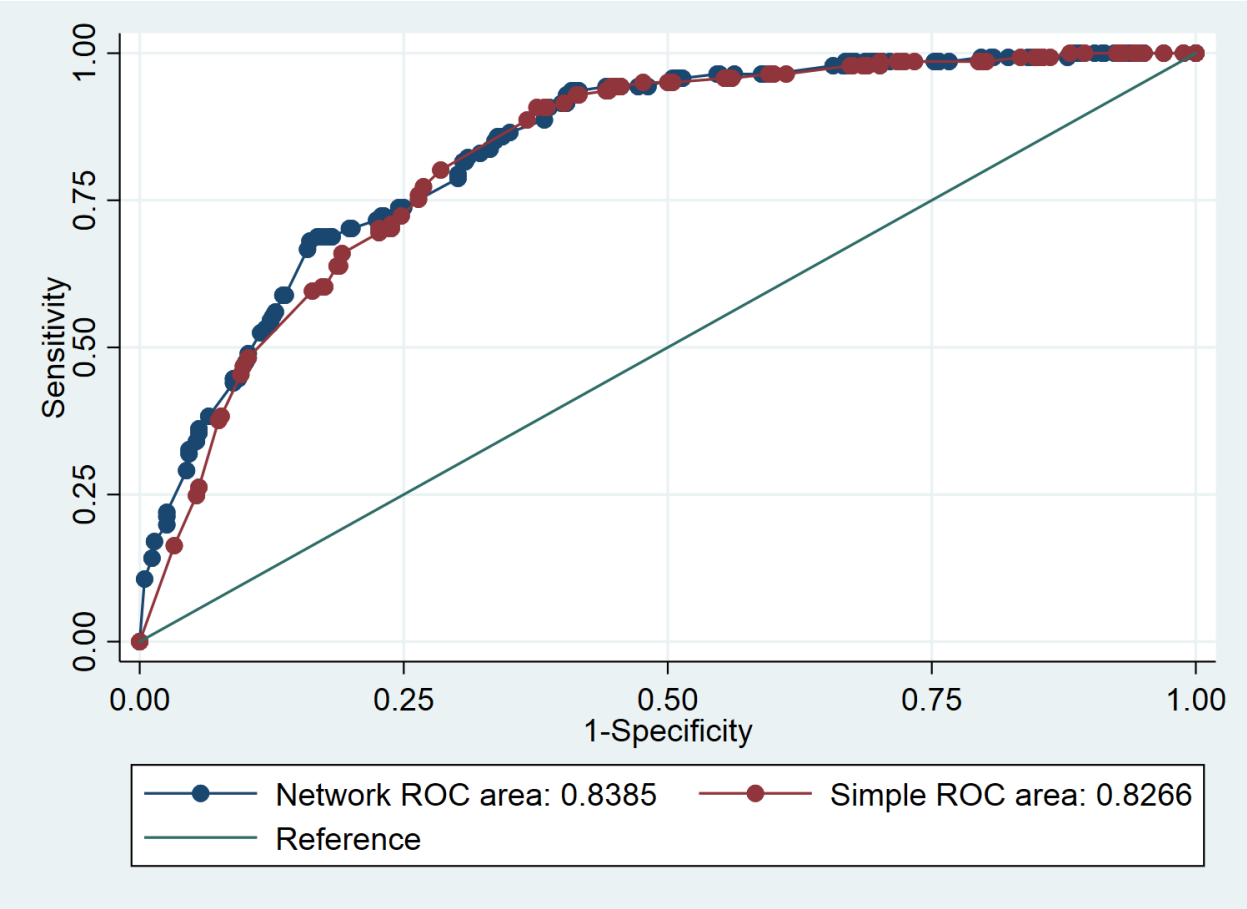


Figure legend:

Graph shows gender (node shape), HIV status (node color), index case status (node size), type of contact (color of line connecting the nodes), and whether the contact was part of an HIV or syphilis investigation (thickness of line connecting the nodes). Graph is loosely grouped by size of sociosexual network component: a) isolates (n=248 persons), b) dyads (n=238 persons across 119 components), c) components size 3-4 (n=224 persons across 68 components), d) components size 5-16 (n=241 persons across 29 components), and e) components size 26, 81, 92, and 320 (n=519 persons across 4 components).

Figure 18. Receiver operator characteristics (ROC) curves showing area under the curve for the network and simple predictive multivariable models among a population of 569 persons newly-diagnosed with HIV in the area around Raleigh, NC. Cases prospectively followed for 3 years to determine continued activity in the sexual network.



VIII. CHAPTER EIGHT: CONCLUSIONS

A. Summary of Aims and Findings

Both aims constructed a sociosexual network based upon partners elicited during contact tracing, and in both aims, fewer than one HIV-positive partner was found per case in the sociosexual network. Among 280 cases in the Aim 1 network based in Wake County, 131 HIV-positive contacts were found (mean 0.47 HIV-positive contacts/case). Among 569 cases in the Aim 2 network based across the Region that contains Wake County, contact tracing identified 221 HIV-positive first-degree contacts (mean 0.39 HIV-positive contacts/case). Since approximately one quarter of cases in both aims were concurrently diagnosed with AIDS, and late diagnosis is an issue across the South, we do not expect that a recent HIV-positive partner ought to be identified. However, the low number of HIV-positive partners and high proportion of partners with unknown HIV status (27% and 38% in Aims 1 and 2, respectively) is alarming and is an indicator that case finding may not be capturing infection sources or possibly even cases of onward transmission since so few partners are located and tested.

1. Aim 1

In Aim 1, I sought to identify some of the gaps in the observed contact tracing network. The descriptive and quantitative analyses in this aim both showed that contact tracing provides a better representation of heterosexual network components than components dominated by MSM. Among all pairs of sexual partners in the Wake County-based sociosexual network where each partner had a sequence available, all male-female pairs were in the same transmission cluster, although only 34% of male-male pairs were in the same cluster. Transmission clusters dominated by MSM were more likely to span multiple components in the descriptive analysis and the quantitative analysis; being in a male-male transmission cluster-

based dyad was significantly associated with spanning components in the multivariable model (Table 8 in Chapter VI). The quantitative analysis also identified missing interviews as associated with having network components that do not approximate the transmission clusters.

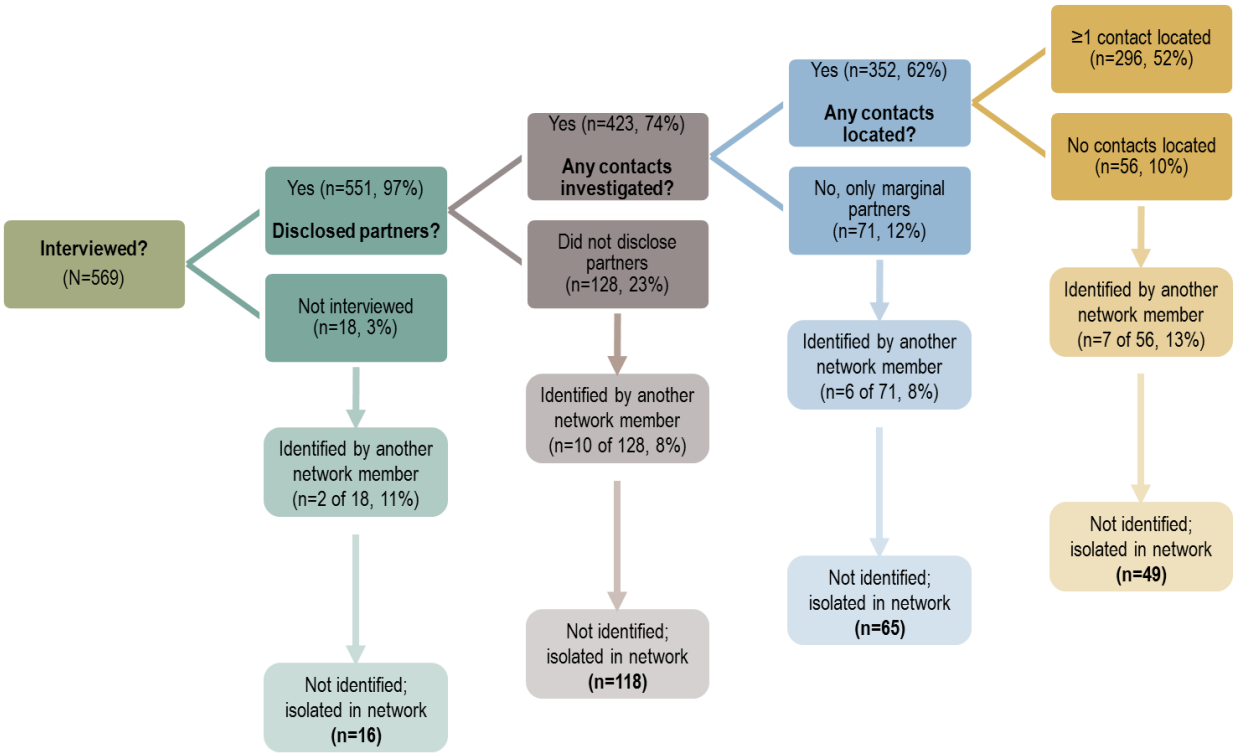
2. Aim 2

In the second aim, I used the network as it was observed, including missing information, and created a predictive model to identify persons who have the highest transmission risk potential, based upon continued activity in the sociosexual network that is documented in the HIV and STI surveillance system. To assess the utility of the sociosexual network in predicting post-diagnosis behavior, I compared a logistic regression model which used network data as a predictor with a simpler model that used only routinely-collected demographic and HIV/STI testing history information.

I expected to use risk behavior information and multiple network predictors in the multivariable predictive model, but the amount of missing data made this impossible. The high number of isolated cases in this sociosexual network (248/569; 44%) was unexpected and likely impacted the predictive ability of the model. During case interview, DIS elicit number of needle-sharing partners, social contacts, and sexual partners before asking for identifying information so that partner tracing can be initiated. If there is enough identifying information then a partner record is created in the surveillance system and investigation begins. Not all contacts with records are found in the course of the investigation, but the record exists in the system and those partners were included in the sexual network. If a DIS judges that there is not enough information to begin investigation then the partner is deemed “marginal” and is present in partner count, but a record is not created. Therefore, cases in this sociosexual network have partners if 1) he or she discloses at least one social contact, needle-sharing partner, or sexual partner who is not deemed marginal, or 2) he or she is successfully identified by another HIV or syphilis case investigation during 2012-2013 in Region 6.

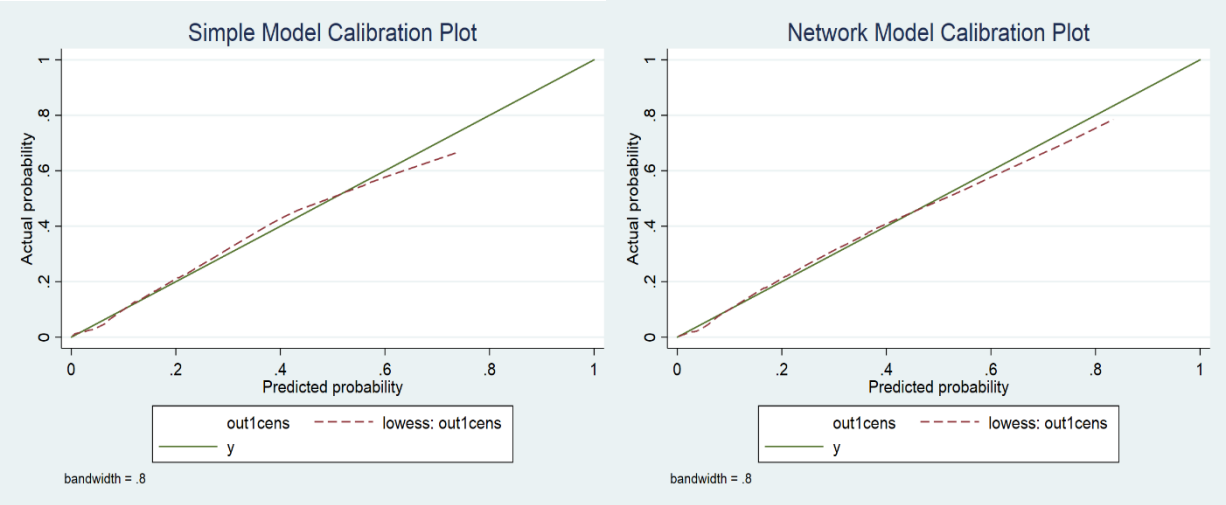
Of 569 cases, 18 (3%) were not interviewed, 128 (23%) were interviewed but did not disclose partners, and 423 (74%) were interviewed and chose to disclose partners. Among the 423 people who chose to disclose partners, 71 only had marginal partners and 56 had partners that DIS attempted to locate but could not. However, some of the cases without any located partners were not isolated in the network because they were identified by another network member (Figure 19). It is important to note, too, that most of these cases were not concurrently diagnosed with AIDS so the likelihood is higher that there was a recent HIV-positive partner who was not identified in the network. In fact, among cases who were not identified by another network member, 2/118 who did not disclose partners, 2/65 who only had marginal partners, and 3/49 with no located contacts after investigation were diagnosed with acute or recent HIV infection.

Figure 19. Mechanisms for appearing as one of the 248 isolated cases in the sociosexual network.



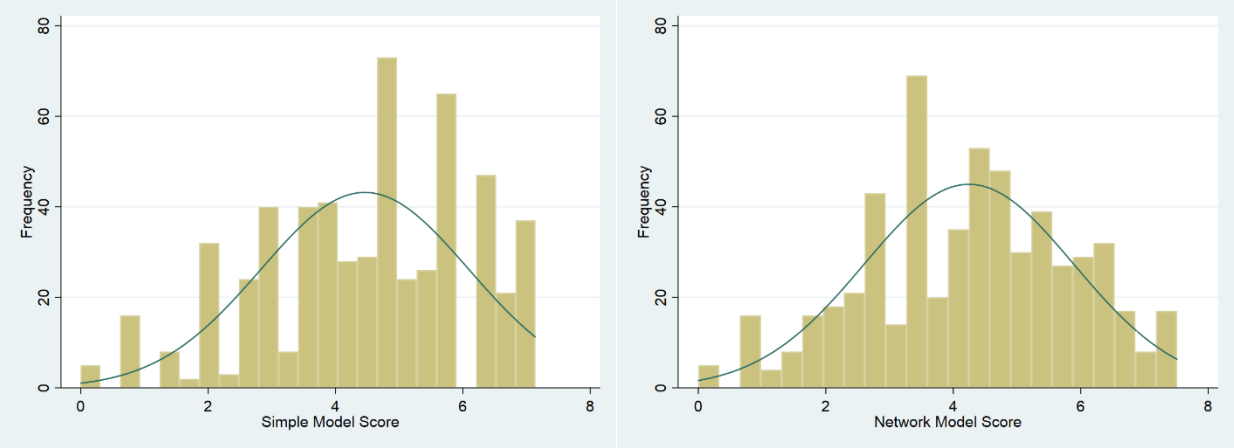
Though we could not use all of the information that we planned, the models still had good predictive ability (Chapter VII, Figure 18). The AUC was >0.83 in both models. The difference among the AUC for the simple and network predictive models was not statistically significant, but the shape of the curves led to fewer weighted errors overall with the network model (Chapter VII, Table 12). Additionally, the model calibration curves (Figure 20) show that the network model has better performance when risk score indicates highest risk.

Figure 20. Calibration plots for Simple (top) and Network (bottom) models.



Histograms of the risk scores reveal that some covariate patterns are common among the cases, which is worth further study. Of 569 cases, there were 77 different risk scores in the simple model (range 1-47 cases; 47/569 = 8%) and 115 different risk scores in the network model (range 1-29 cases; 29/569 = 5%). The histogram for the network model is more normally distributed with fewer peaks than the simple model (Figure 21); the addition of the network predictor appears to normalize and smooth the covariate pattern frequencies which may indicate that the information gathered from the network adds another dimension to transmission risk potential that is not understood solely by demographic and STI information.

Figure 21. Frequency of model scores for Simple (top) and Network (bottom) models.



In agreement with other studies, I found that Black MSM are less likely to be durably virally suppressed.[10,15,217] This supports why I did not ignore outcomes where the case was virally suppressed. Black MSM are at highest risk for acquiring HIV in NC.[22] Black MSM are more likely to have partners who are of the same race, as demonstrated in this study and in others.[9,25,54] A substantial proportion of new HIV cases in NC are attributed to people who were already diagnosed and aware of their status at the time of transmission.[28] If Black MSM are also less likely to remain suppressed then it stands to reason that cases who are identified as having a high potential for onward transmission should be followed by bridge counselors or patient navigators to ensure that they remain engaged in care.

B. Strengths

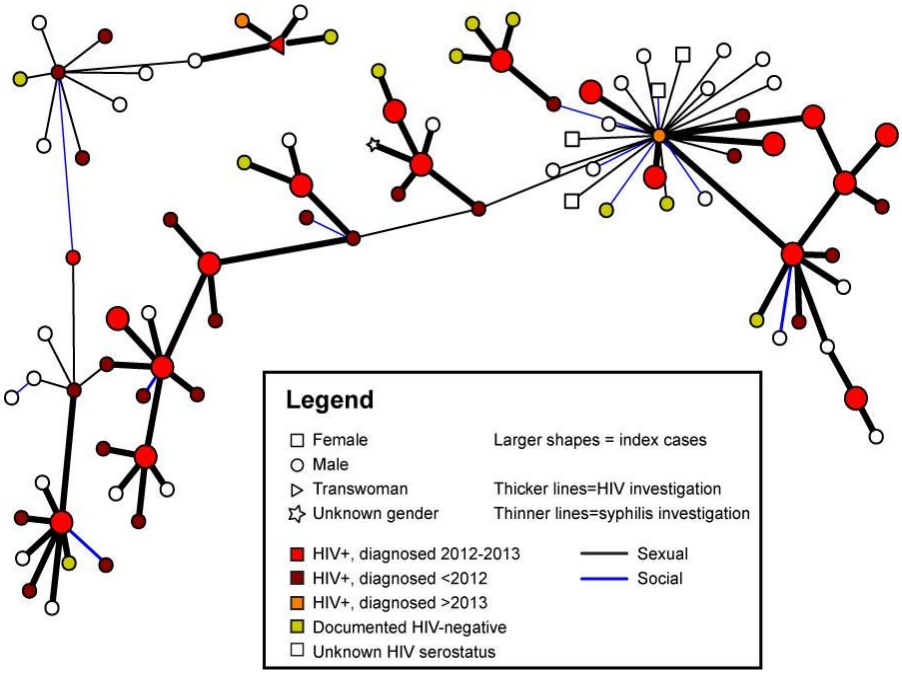
The study population for Aim 2 included all HIV and syphilis cases diagnosed in the entire control region and not a sample or a subset of the population, which allowed me to examine partnerships across the entire sociosexual network where there is risk of HIV or STI transmission. Using all cases in the region also has the potential for direct application in local prevention efforts. NC already has an infrastructure of DIS who interview cases and assist with linkage to care. Cases are assigned bridge counselors and patient records containing both DIS and bridge counselor contact with the case is kept in NC EDSS. Bridge counselors can access

this record. The predictive model constructed in Aim 2 of this project identified predictors associated with HIV-positive persons being more likely to continue to engage in sexual partnerships after HIV diagnosis. I was then able to calculate a risk score and assign sensitivity and specificity to assist bridge counselors with ensuring identified persons are engaged in care, with the goal of viral suppression.

The Aim 2 model used information as it is understood at the time of diagnosis, giving the model real-world applicability. The simple model, which did not use any predictors from the network, still had good predictive capability and it may be judged that the increase in errors, largely false positives, is worth the ease of calculation. Both models could be piloted relatively easily since there are already protocols to collect the information used when new cases are identified.

The addition of the syphilis investigation partnerships may have served to correct some of the missingness in the sociosexual network, since some of the cases had partnerships only as a result of the syphilis investigation. The overlap between the HIV and syphilis networks, particularly among MSM, is in my opinion one of the more interesting findings in this aim. One of the components included 92 people; we would have observed 6 smaller components had only the HIV investigations been used (Figure 22). This

Figure 22. Sociosexual network component comprising 92 people, showing bridging of investigated HIV partnerships by investigated syphilis partnerships.

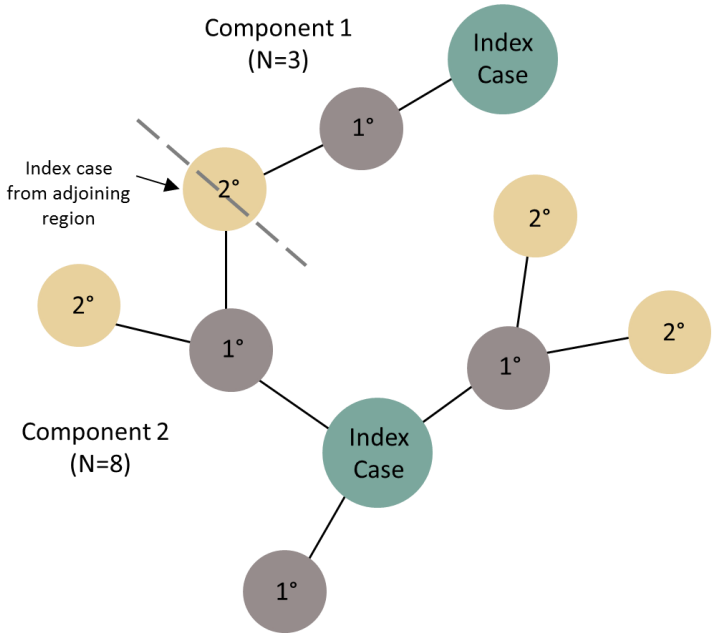


demonstrates that MSM who are being investigated for syphilis in NC would be good candidates for PrEP.

C. Limitations

There are inherent limitations in any phylogenetic analysis. Missing data is certainly an issue. Specific to this study, approximately half of HIV-positive persons did not have a sequence available since not everyone who is diagnosed enters care and some patients enter care with providers who use labs that do not provide samples to UNC for research purposes. While NC does have DIS case reports on all new diagnoses, sequences are only available for individuals who had genotypes performed by LabCorp or Monogram Biosciences. However, the majority of persons who are linked into care in NC, and would thus have a sequence available prior to beginning ART, have samples run at one of these labs. New agreements to obtain sequences from other labs have been made and sequences are now reportable to the State, but the current study did not benefit from those changes.

Figure 23. Depiction of artificial shrinking of components by administrative boundaries. Index cases in other areas that are not investigated do not contribute partnership information to the network, even if they are network members.



The method itself has limitations: 1) acute and recent infections are more likely to cluster since there are shorter branch lengths between them which may affect the trees used to calculate genetic distance;[102] 2) directionality of transmission cannot be inferred; 3) unsampled third parties may be involved in transmission chains; and 4) the genetic distance measurements will miss substitutions if multiple occurred and resulted in a substitution back to the original amino acid.

Incomplete networks and anonymous partners are an issue for studies of sexual networks,[220] which pertains to both Aims. We defined cases as residents of Wake County (Aim 1) or Region 6 (Aim 2) and only abstracted first degree partners, potentially causing the components to appear artificially small (Figure 23 depicts this problem). For each of the two index cases in the example component in Figure 23, we would abstract the index and his primary degree partners, resulting in two distinct observed components of sizes 2 and 4. However, a slight change of partnerships to also include recent second degree partners or index cases from an adjoining region would result in 10 people in a single component.

Partners who cannot be located would bias both the observed network and outcome collection for Aim 2. Index cases who engage in anonymous partnerships after diagnosis may not be captured for Aim 2 outcomes, creating a source of bias, since cases are only counted as Aim 2 outcomes if they are identified as partners in the State HIV database 2014-2016. This requires the DIS having enough information about the new partner to initiate partner trace back to the index case. Therefore, index cases engaging in anonymous partnerships could appear to not be an Aim 2 outcome, despite engaging in potentially high risk partnerships.

The high proportion of cases who did not have risk behavior information was a limitation for the Aim 2 model, as noted in the manuscript discussion (section VII.E). Several studies have shown that risk behavior appears to decrease after diagnosis, although the durability of that decrease is questionable.[78,79,117] This decrease is meaningful as it is relative to baseline behavior, which we could not collect for one-quarter of the cases. This information may then have been predictive of subsequent activity. The high number of singletons in the Aim 2 network (44% cases) also resulted in the ability to use only one network term per full model. To keep the cases who were not isolated in the model, a category indicating singleton status needed to be used, and that category was then collinear across multiple predictors.

Finally, the results of this findings from either Aim may not be generalizable to other geographic areas. I bootstrapped the predictive models in Aim 2 to validate the model

internally, but neither model is externally validated. However, some of the findings in Aim 1 and Aim 2 agreed with previous findings. Attempting to apply these methods in other areas would be an interesting and important next step.

D. Future Directions

1. Testable Interventions

Aim 2 can be piloted in NC. Upon identification, DIS and bridge counselors can be alerted to prioritize getting and keeping in care the persons who are diagnosed and at risk of onward transmission. Any additional time that bridge counselors spend tracking persons who are identified by the predictive model to be most likely to be involved in onward transmission can be tracked. The expense of resources can be offset by cases successfully re-engaged in care to calculate cost of cases averted.

2. Proposed Future Analyses

a. Expansion of Network to Increase Accuracy

The methods applied in these analyses can be applied to a larger area with the goal of correcting some of the bias in the observed network due to missing nodes. Regions 3-6 and Charlotte TGA (Figure 7) comprise the central area of NC, which is geographically and contextually distinct from the western mountain and eastern coastal areas. The Nexus study (section II.B.) found overlap in the sexual networks between adjoining regions; adding cases from Regions 3 and 4 to the Region 6 cases resulted in large sexual network components that appeared to be smaller, discreet components prior to the addition of the cases and their partners from other regions, without artificial reduction of the sexual network components due to the constraints of administrative boundaries (Figure 23).

I demonstrated this in our study by including syphilis cases and partners, which increased the size of the network, while resulting in a smaller number of components. The component size 92 in this study (Figure 22) would have been 6 smaller components without the inclusion of the syphilis investigations. Adding cases and partners, including both HIV and

syphilis, from a neighboring region would likely result in a better representation of the true underlying sexual network than what was observed. This may have implications for the model, since I used only the observed Region 6 network to calculate network predictors that are sensitive to network size.

b. Exponential Random Graph Models to Analyze Network Growth Drivers

Markov chain Monte Carlo (MCMC) estimation can be used to create a distribution of randomly-generated networks based upon selected attributes of the observed network, such as number of nodes or edges (see Appendix C section A). Upon comparing the observed network to the distribution of random networks, important properties of the observed network, such as clustering or transitivity above or below what is expected, can be identified based upon divergence from the distribution of randomly-generated networks. Inferences can then be drawn estimating which processes influence tie formation in NC, which may reveal avenues for intervention.

c. Comparison of Centrality Scores and Network Structure Involvement by Infection Recency

Behavior changes following diagnosis,[77-79] so chronically-infected persons aren't well-suited for an assessment of behavior and risk though they are frequently used in publications to describe risk behavior in local populations. An analysis of HIV-positives nodes comparing the centrality scores and risk behaviors of persons estimated to be infected recently (less than 12 months between infection and diagnosis) to persons estimated to be chronically-infected (12 months or more) may provide clues to behaviors and partnership patterns associated with HIV acquisition, and help us determine whether network position differs during the period in which infected was acquired. Comparing recently-infected, chronically-infected, and uninfected individuals may be less biased than networks that are egocentric with respect only to the infected individuals.[11] Centrality score mean, across multiple centrality measures, can be calculated and compared to determine whether there is a significant difference between groups at alpha = 0.05.

d. Geographic or Spatial Analysis Combined with Social Network and Gene Sequence Analyses

The majority of location-based analyses in NC thus far have been geographic rather than spatial. A spatial analysis based on residence would allow location data to be incorporated into models. In addition to being more suited for models and be interpretable on continuous scales, spatial data are more suitable to combine with both network and gene sequence relatedness data; phylogenetic relatedness, spatial closeness, and network distance can all be defined as continuous measures. A previous study of spatial distance and genetic distance used acutely infected individuals with *pol* gene sequences who were consented into a clinical cohort in NC. The patients were assessed for transmitted drug resistant or drug susceptible virus. Sequenced virus was found to differentiate less within rural areas than within urban areas.[51] The application of the combination of these three methods to an entire population of geographically-defined incident diagnoses is novel and has the potential to significantly improve the understanding of the spread of HIV.

e. Transmitted Drug Resistance

Prevalence of TDRM by sexual network component could be ascertained from a study, which could then be used to guide both clinical and public health practice at a more precise sub-population level. The ability to link drug resistance data extracted from sequences to defined sexual network components would allow assessment not only HIV risk, but also TDRM risk, for individuals within the component.[49] Finding a similar mutation profile may result in network component linkages through an individual believed to be HIV-negative or thought to be anonymous.

Once the assortative factors of the network components were determined, HIV and TDRM risk by type of mutation could be assessed by those factors. Linking the resistance profile to cases who are newly diagnosed or who are failing treatment would allow calculation of a crude prevalence of DRM to be established by drug class within each component – the

prevalence by type of TDRM for the unique set of demographic and risk characteristics which define each component. In another state in the US South, Mississippi, ART-naïve young Black MSM clustered phylogenetically and had TDRM strains, while phylogenetic transmission clusters (TC) containing only older Black MSM did not have TDRM,[93] indicating two discrete sexual networks separated by age and supporting the idea that discrete sexual network components may each have their own circulating drug resistant variants.

One limitation is that a consensus sequence is returned by LabCorp. Minor variants have been shown to incorporate DRM, contributing to the risk of virologic failure and a resurgence in resistant strains,[221-223] which may not be reflected using this sequencing method. If minority variants coding for DRM are significant in this population then prevalence estimates were underestimated. However, one study that employed ultra-deep pyrosequencing found that half of patients with TDRM only had the mutations in <20% of their virus population.[224]

f. Predictive Model to Identify HIV-Negative Persons Who Would Benefit Most from PrEP

The converse of the model presented in Chapter VII is a model to identify HIV-negative persons who would benefit most from PrEP. A sexual network study that began with 2 acutely-infected patients in NC resulted in a sexual network of 398 persons, nearly half of whom (47%) were of unknown serostatus due to inability to locate or testing refusal. Ninety-two persons in the network were confirmed to be HIV-negative, but 24 of those persons (26%) seroconverted within 3 years.[21] HIV-negative partners of new HIV cases and HIV-negative persons involved in syphilis investigations can be followed through the same 3-year period as the HIV-positive index cases to determine who becomes infected within 3 years. There is significant overlap of syphilis and HIV sexual networks in NC, particularly among MSM,[21,96] so the syphilis cases will serve as an additional HIV-negative population for the baseline period. A predictive model can then be constructed, using the same types of information as the predictive model already developed.

E. Public Health Implications

HIV is an epidemic. Prevention efforts must focus on disrupting sexual transmission. Previous studies have shown that identifying central persons in networks and "cutting" the network at those points has the potential to break it up into smaller components and disrupt disease transmission.[11,90] In the Aim 1 quantitative analysis, we identified factors associated with being in an observed sexual network component that was not well-represented by inferred transmission chains (the phylogenetic clusters). In Aim 2 of this study, we identified HIV-positive persons who are most likely to engage in post-diagnosis partnerships and remain part of transmission chains. Identification of these persons, with the purpose of engaging them in care and helping them to achieve viral suppression, leads to a sharp reduction of infectiousness and prevents onward transmission. While these cases may not be removed from the network, nullifying the likelihood of infection is essentially disrupting the network at these nodes. This is of particular importance in NC, where persons who are aware of their infection are responsible for a higher proportion of new infections than in other places in the US.[28]

In both aims, we were able to show that partner tracing data can be leveraged to help guide interventions. Understanding that the observed sexual network may not represent transmission well, as for Black MSM in Aim 1, demonstrates that interventions based upon partner tracing may need to be tailored differently to this high-risk group. In Aim 2, I take that further by showing that despite the limitations in the data, those data can still help predict where transmission is likely to occur.

F. Conclusions

Both aims of this analysis demonstrate that sociosexual network analysis can be harnessed to help guide interventions as long as the limitations of the observed sociosexual network are understood. In conducting this dissertation research in North Carolina, it is my hope that these findings can be directly applied to improve the health of people here. Black and Latino MSM in NC fail to see the reductions in number of cases seen by other races.[22] This is

likely due to social factors affecting partner selection and risk behavior, including stigma, which appears to influence behavior.[225]

The higher rate of HIV incidence observed among Black MSM may also be exacerbated by biological factors such as higher prevalence of STIs circulating in network components dominated by Black MSM.[226] Again, the sexual network can be studied as proposed in Future Analyses (VIII.D.2.). A census of the components by demographic factor, STI co-infection, sexual behaviors, and drug resistant variants of HIV might provide valuable information to assess risk. Further study of component density, size, and structure may reveal the generative behaviors driving partner selection among component members. Knowing demographic information, risk, and partner selection traits might be the key needed to be able better identify candidates for PrEP and intervene early to prevent HIV acquisition.

I have not discussed needle-sharing as a mechanism for HIV transmission in this dissertation. There were 4 needle-sharing partnerships among the 845 first-degree partnerships in the sociosexual network; 3 occurred in the largest component (n=320). I understand that there is an increase in intravenous drug use in North Carolina, and it is concerning to see the overlap of partnership types in a single network component. Intervening in the large component which comprised both sexual and needle-sharing to offer HIV testing and PrEP or linkage to care might have great benefit.

Sexual partner selection is not random. Studying the sexual network may provide insight into forces guiding partner selection. Understanding what drives partner selection may provide opportunities to identify people at risk earlier, with the goal of intervening to prevent HIV acquisition or transmission.

APPENDIX A: ADDITIONAL RESULTS RELATED TO AIM 1

A. Results

The PNS network included 663 persons: 104 singletons plus 559 persons in 446 partnerships (Figure 11a). Among 446 partnerships, 70 (16%) included two HIV-positive persons with sequences; of these, 26 (37%) were in the same cluster. A subset of the 70 includes 12 network dyads with no other persons linked to the pair; 7 (58%) were in the same cluster, 1 (8%) pair was in different clusters, and 4 (33%) included one person in a cluster and one not in a cluster. Among the 7 dyads where both persons were in the same cluster, 2 were in a cluster size two, where both the component and cluster contained only those persons, and the rest were larger (median 7 members, range 3-14).

APPENDIX B: ADDITIONAL METHODS, RESULTS, AND DISCUSSION RELATED TO AIM 2

A. Methods

1. Statistical Analyses

One-quarter (n=146, 26%) of indexes chose not to be interviewed, which resulted in a loss of risk behavior and partnership information. For the final models, we ultimately chose to exclude risk behavior variables because there was a high degree of data missing not at random. The data were missing across the entire set of risk behaviors, so add a "missing" category to the predictor would have overwhelmed the information in each predictor. Performing a complete case analysis, and dropping the patients who were missing risk behavior, may have resulted in a sample that was not representative of newly-diagnosed cases locally.

Before deciding to proceed without using risk behavior data, we split the population of indexes into two sets by interview/disclosure status and constructed predictive models for each group. The group who chose not to disclose had many candidate predictors with perfect separation in at least one category, so we used Firth logistic regression to construct the model for that group. The models did not predict the outcome well for cases in the non-interview group (Figure 24), so we decided to test the predictive ability of the models on the entire group of cases using only demographics, HIV/STD testing history, and information gleaned from the sociosexual network.

B. Results

1. Sociosexual Network

Figure 25 is the *k*-nearest neighbor plot. It shows that people with low degree in the network are typically connected to people with high degrees, which demonstrates degree disassortativity. It is likely that this disassortativity is an artefact of the data, explained by DIS interviewing new cases and not interviewing previously-diagnosed cases, who would then only have a degree representing the number of new cases who identified the previously-diagnosed case, and not actually a true representation of their number of partners.

2. Transmission Cluster Involvement

Over half (59%, n=338) indexes and 83/221 (38%) HIV-positive first-degree contacts had an available HIV gene sequence, of which 197 indexes and 61 partners (N=258) with sequences were in one of 137 putative statewide transmission clusters with ≥1 network member. These clusters ranged in size from 2–53 people, including 1–17 network members, for a total of 870 people statewide. Median year of diagnosis for contacts in a cluster was 2011 (IQR: 2005–2012).

Among the 845 first-degree contacts in the network, 72 (9%) included two people where each had a sequence. Of these, 40/72 (56%) were in the same transmission cluster where 21 contacts were diagnosed ≥6 months prior to the index, 18 were diagnosed within 6 months before or after the index, and 1 was diagnosed ≥6 months after the index.

When assessing outcomes, we found 19 partnerships between indexes and future partners occurred where both had gene sequences available. Both persons were identified in a transmission cluster 7 (37%) times; once in the same cluster (cluster size=3) and 6 times in two different transmission clusters.

3. Bivariable Analyses

We tested many forms of the variables before selecting predictors for inclusion into the final multivariable models. Some of these are presented in Table 13. Both estimated and laboratory-confirmed AIDS at diagnosis were associated with the outcome (24% (p<0.0001) and 25% (p=0.001), respectively, with imperfect agreement of cases across both classifications), where DIS-estimated concurrent AIDS diagnosis was protective (OR for established, acute, or recent infection=4.0, 95%CI: 2.2–7.3). Being acutely- or recently-infected based upon laboratory results was also associated with the outcome (p=0.0003). Log viral load within 6 months of diagnosis was not associated with the outcome. Linkage to care within 3 months of diagnosis was associated with the outcome (p=0.01), although having an available HIV gene sequence, which is essentially a proxy for being in care, was not. Having any prior negative HIV

test, any prior STI history, prior history of syphilis specifically, any STI at diagnosis, and co-infection with syphilis at diagnosis were associated with the outcome (each $p<0.0001$).

Linkage to care within 3 months of diagnosis was associated with the outcome (OR=1.6, 95%CI: 1.1–2.4), although having an available HIV gene sequence, which is essentially a proxy for being in care, was not. Several behavioral predictors were highly associated with the outcome ($p\leq0.01$), including not always using a condom, meeting partners online, having a male sex partner known to be HIV-positive, number of recent sex partners (total of located and not), proportion of sex partners who could not be located, recent MSM activity, and multiple recent new sexual partners, although these predictors were missing for the 74% indexes who were not interviewed.

4. Multivariable Analyses

There were over 30 multivariable models with all predictors significant at $p<0.05$ prior to bootstrapping. We selected our final two models based upon AIC and AUC. Selected candidate models with significant predictors using different network terms are presented in Table 14. Demographic characteristics were often retained in the multivariable models, particularly marital status, age at diagnosis, and urbanicity of residence. Having any STI at diagnosis was associated with continued involvement in the sexual network. Most HIV and STI history variables were retained in some of the candidate models, with general STI history and STI status at HIV diagnosis most often retained. Partner disclosure was the only risk-related predictor that was tested in the multivariable models, and was retained in both.

We used estimated HIV stage based upon DIS investigation, rather than lab-confirmed AIDS, in the multivariable models since we wanted the model to reflect the available information at diagnosis, and the predictor was retained in both. Acute infection status was not retained in either model. Having history of any STI was retained in both models, while the simple model also included having any concurrent STI and the network model included having an STI other than syphilis at diagnosis.

Most of the models that we tested before selecting the final network model retained at least one network predictor. The final network model included component size, categorized as isolated, 2–4 persons, and ≥5 persons in the component.

C. Discussion

These findings support previous studies despite the data limitations noted in the manuscript. Indexes concurrently diagnosed with AIDS were less likely to remain active in the sexual network.[79] Similarly, in a network was constructed from gene sequences instead of partnerships, an increase in log viral load at baseline increased the likelihood of onward transmission. That study also found that number of partners expressed as a percentile was predictive of onward transmission risk, while race was ultimately not in their multivariable model.[171]

Future exploration of the reasons why cases do not remain engaged in care would be valuable. Being engaged in care appears to reduce sexual risk behaviors regardless of suppression status,[190] so this model has utility for high-risk patients who are new to care or choose not to take ART.

However, this analysis raises ethical concerns. If misinterpreted, it could be used to advocate for criminalizing behavior. However, the purpose of this analysis is to recognize that most HIV patients in the US are still, unfortunately, diagnosed quite young;[30] that HIV is a chronic, lifelong infection; and that committing to care at least twice per year (as is the recommendation for viral suppression) for the duration of life is particularly difficult, especially for younger patients who are less likely to access regular care.[227] In attempting to support this population, bridge counselors need to assess the possibility of late reporting of viral loads,[228] then balance the privacy of the patient against risks and benefits of follow-up contact.[229]

Table 13. Bivariable analyses for predictors not presented in the manuscript. The outcome is whether any cases were identified as an active member of the sexual network in the 3 years following HIV diagnosis.

Predictor	Total		Remain Involved		Bivariable Analysis	
	N	(col %)	n	(row %)	OR	(95% CI)
	569	(100)	141	(25)		
Demographics and History						
Gender						
Male	451	(79)	129	(29)	1.0	
Female	114	(20)	11	(10)	0.27	(0.14-0.51)‡
Transgender (M->F)	4	(0.7)	1	(25)	0.83	(0.09-8.1)
Student*						
No	543	(95)	131	(24)	1.0	
Yes	26	(5)	10	(38)	2.0	(0.87-4.4)§
Ever incarcerated*						
No	533	(94)	133	(25)	1.0	
Yes	36	(6)	8	(22)	0.86	(0.38-1.9)
HIV and STD History and Care						
Laboratory-confirmed HIV stage at diagnosis						
Established / chronic	393	(69)	103	(26)	1.0	
Acute or recent	32	(6)	17	(53)	3.2	(1.5-6.6)‡
Concurrent AIDS diagnosis	144	(25)	21	(15)	0.48	(0.29-0.80)‡
Log viral load within 6 months of diagnosis	N=306		n=83	(27)		
≤ 2.3	42	(14)	8	(19)	1.0	
> 2.3 to < 2.7	8	(3)	2	(25)	1.4	(0.24-8.4)
2.7 - 4.0	41	(13)	11	(27)	1.6	(0.55-4.4)
> 4.0 - 6.0	200	(65)	56	(28)	1.7	(0.72-3.8)
> 6.0	15	(5)	6	(40)	2.8	(0.78-10.3)§
Sexual Risk Behaviors†						
Condom use	N=402		n=122	(30)		
Always	38	(9)	6	(16)	1.0	
Sometimes	263	(65)	97	(37)	3.1	(1.3-7.7)‡
Never	101	(25)	19	(19)	1.2	(0.45-3.4)
Met partners online	N=423		n=128	(30)		
No	301	(71)	69	(23)	1.0	
Yes	122	(29)	59	(48)	3.1	(2.0-4.9)‡
Met partners at bar	N=423		n=128	(30)		
No	388	(92)	110	(28)	1.0	
Yes	35	(8)	18	(51)	2.7	(1.3-5.4)‡
Ever had anonymous partner	N=423		n=128	(30)		
No	311	(74)	79	(25)	1.0	
Yes	112	(26)	49	(44)	2.3	(1.5-3.6)‡
Transactional sex, ever	N=423		n=128	(30)		
No	402	(95)	121	(30)	1.0	
Yes	21	(5)	7	(33)	1.2	(0.46-3.0)

Predictor	Total		Remain Involved		Bivariable Analysis	
	N	(col %)	n	(row %)	OR	(95% CI)
Ever went to a sex party	N=423		n=128	(30)		
No	405	(96)	116	(29)	1.0	
Yes	18	(4)	12	(67)	5.0	(1.8-13.6)‡
Ever used intravenous drugs	N=423		n=128	(30)		
No	412	(97)	125	(30)	1.0	
Yes	11	(3)	3	(27)	0.86	(0.22-3.3)
Ever had sex with an intravenous drug user	N=423		n=128	(30)		
No	420	(99)	127	(30)	1.0	
Yes	3	(0.7)	1	(33)	1.2	(0.10-12.9)
Ever had sex with known HIV+	N=423		n=128	(30)		
No	260	(61)	68	(26)	1.0	
Yes	163	(39)	60	(37)	1.6	(1.1-2.5)‡
Transgender woman or male who has ever had sex with male	N=423		n=128	(30)		
No	168	(40)	24	(14)	1.0	
Yes	255	(60)	104	(41)	4.1	(2.5-6.8)‡
Ever had sex with male	N=423		n=128	(30)		
No	96	(23)	15	(16)	1.0	
Yes	327	(77)	113	(35)	2.9	(1.6-5.2)‡
Ever had sex with female	N=423		n=128	(30)		
No	292	(69)	99	(34)	1.0	
Yes	131	(31)	29	(22)	0.55	(0.34-0.89)‡
Sex partners during period of interest	N=423		n=128	(30)		
0	56	(13)	8	(14)	1.0	
1-2	192	(45)	49	(26)	2.1	(0.91-4.7)§
3-4	71	(17)	28	(39)	3.9	(1.6-9.5)‡
5-60	104	(25)	43	(41)	4.2	(1.8-9.8)‡
> mean recent partner number	N=423		n=128	(30)		
No	316	(74)	85	(27)	1.0	
Yes	107	(26)	43	(40)	1.8	(1.2-2.9)‡
Proportion recent partners anonymous	N=423		n=128	(30)		
0	313	(74)	81	(26)	1.0	
1-25	11	(3)	7	(64)	5.0	(1.4-17.6)‡
26-50	43	(10)	19	(44)	2.3	(1.2-4.4)‡
51-75	13	(3)	4	(31)	1.3	(0.38-4.3)
76-100	43	(10)	17	(40)	1.9	(0.97-3.6)§
Any new partners during period of interest	N=423		n=128	(30)		
No	246	(58)	57	(23)	1.0	
Yes	177	(42)	71	(40)	2.2	(1.5-3.4)‡
Sex with male during period of interest	N=423		n=128	(30)		
No	245	(58)	60	(24)	1.0	
Yes	178	(42)	68	(38)	1.9	(1.3-2.9)‡

Predictor	Total		Remain Involved		Bivariable Analysis	
	N	(col %)	n	(row %)	OR	(95% CI)
Sex with female during period of interest	N=423		n=128	(30)		
No	350	(83)	114	(33)	1.0	
Yes	73	(17)	14	(19)	0.49	(0.27-0.92)‡
	N=423		n=128	(30)		
Network Information						
Number of network sex partners						
0 (singleton)	275	(48)	38	(14)	1.0	
1	156	(27)	35	(22)	1.8	(1.1-3.0)‡
2	55	(10)	20	(36)	3.6	(1.9-6.8)‡
3-5	60	(11)	33	(55)	7.6	(4.1-14.1)‡
≥ 6	23	(4)	15	(65)	11.7	(4.6-29.5)‡
Didn't disclose partners, but identified by others in network						
No	557	(98)	139	(25)	1.0	
Yes	12	(2)	2	(17)	0.60	(0.13-2.8)
Disclosed contacts, none of whom could be located						
Did not disclose, or ≥1 contact located	351	(62)	113	(32)	1.0	
Yes	218	(38)	28	(13)	0.31	(0.20-0.49)‡
Network social contacts greater than number disclosed						
No	560	(98)	137	(24)	1.0	
Yes	9	(2)	4	(44)	2.5	(0.65-9.3)§
Network needling sharing partners greater than number disclosed						
No	569	(100)	141	(25)	---	
Yes	0	---				
Bonacich power score						
0	248	(44)	30	(12)	1.0	
> 0 to < 0.3	125	(22)	19	(15)	1.3	(0.70-2.4)
0.3 - < 0.4	20	(4)	7	(35)	3.9	(1.4-10.6)‡
0.4-0.66	57	(10)	23	(40)	4.9	(2.6-9.4)‡
> 0.66	119	(21)	62	(52)	7.9	(4.7-13.4)‡
Betweenness score						
0	409	(72)	64	(16)	1.0	
> 0	160	(28)	77	(48)	5.0	(3.3-7.5)‡

* Yes only if documented as such in the interview record, else recorded as No
† Per the number of cases who agreed to be interviewed and discuss the topic
‡ Significant at p<0.05
§ Significant at p<0.20

Table 14. Multivariable models not presented in the manuscript. Relationship between predictors tested but not retained in the final models and the outcome of remaining active in the sociosexual network, showing odds ratio (OR), 95% confidence interval (CI) using robust standard errors, Akaike's information criterion (AIC), and receiver operator characteristics area under the curve (AUC) calculated from the collected population.

Predictor	Total		Remain Involved		Bivariable Analysis		Network Multivariable 1		Network Multivariable 2		Network Multivariable 3	
	N	(col %)	n	(row %)	OR	(95% CI)	aOR	(95% CI)	aOR	(95% CI)	aOR	(95% CI)
	569	(100)	141	(25)								
Model AIC							493.70		494.36		493.38	
Model AUC							0.8377		0.8393		0.8379	
Demographic Characteristics												
Gender												
Woman or Transwoman	118	(21)	12	(10)	1.0		1.0				1.0	
Male	451	(79)	129	(29)	3.5	(1.8-6.7)‡	2.3	(1.1-5.1)‡			2.3	(1.1-4.9)‡
Marital status*												
Married	137	(24)	6	(4)	1.0		1.0		1.0		1.0	
Unmarried	432	(76)	135	(31)	9.9	(4.3-23.1)‡	6.8	(2.8-16.3)‡	6.9	(3.0-16.0)‡	6.7	(2.8-16.0)‡
County of residence at diagnosis												
Rural	105	(18)	14	(13)	1.0							
Urban or suburban	464	(82)	127	(27)	2.4	(1.3-4.5)‡						
Age at diagnosis												
≥ 30 years	326	(57)	46	(14)	1.0		1.0		1.0		1.0	
< 30 years	243	(43)	95	(39)	3.9	(2.6-5.9)‡	1.8	(1.1-3.0)‡	2.0	(1.2-3.3)‡	1.7	(1.1-2.8)‡
HIV and STD History and Care												
HIV interview												
No interview or no partner disclosure	146	(26)	13	(9)	1.0		1.0		1.0		1.0	
Interviewed and disclosed partners	423	(74)	128	(30)	4.4	(2.4-8.1)‡	2.1	(1.0-4.3)‡	2.2	(1.1-4.5)‡	2.2	(1.1-4.5)‡
Estimated HIV stage at diagnosis†												
Concurrently diagnosed with AIDS	136	(24)	13	(10)	1.0		1.0		1.0		1.0	
Acute, recent, or chronic HIV	433	(76)	128	(30)	4.0	(2.2-7.3)‡	2.3	(1.2-4.4)‡	2.1	(1.1-4.0)‡	2.4	(1.2-4.4)‡
History of any STI												
No	385	(68)	67	(17)	1.0		1.0		1.0		1.0	
Yes	184	(32)	74	(40)	3.2	(2.1-4.7)‡	1.7	(1.0-2.7)‡	1.7	(1.0-2.7)‡	1.7	(1.1-2.8)‡

Predictor	Total		Remain Involved		Bivariable Analysis		Network Multivariable 1		Network Multivariable 2		Network Multivariable 3	
	N	(col %)	n	(row %)	OR	(95% CI)	aOR	(95% CI)	aOR	(95% CI)	aOR	(95% CI)
Any STI infection at diagnosis												
No	432	(76)	75	(17)	1.0		1.0		1.0		1.0	
Yes	137	(24)	66	(48)	4.4	(2.9-6.7)‡	2.1	(1.2-3.5)‡	2.1	(1.2-3.5)‡	2.0	(1.2-3.4)‡
Network Structures												
Degree												
0 (singleton)	248	(44)	30	(12)	1.0		1.0					
1	159	(28)	33	(21)	1.9	(1.1-3.3)‡	1.5	(0.83-2.8)				
2	62	(11)	22	(35)	4.0	(2.1-7.6)‡	1.8	(0.81-4.1)				
3-5	67	(12)	35	(52)	7.9	(4.3-14.7)‡	2.7	(1.3-5.7)‡				
≥ 6	33	(6)	21	(64)	12.7	(5.7-28.5)‡	5.0	(1.9-12.9)‡				
Adjusted degree												
0	248	(44)	30	(12)	1.0							
> 0-1	140	(25)	25	(18)	1.6	(0.89-2.8)§			1.4	(0.74-2.8)		
> 1 to < 2.5	62	(11)	23	(37)	4.3	(2.3-8.1)‡			1.8	(0.82-4.0)		
2.5-5	79	(14)	38	(48)	6.7	(3.8-12.1)‡			2.5	(1.3-5.1)‡		
> 5	40	(7)	25	(63)	12.1	(5.7-25.5)‡			4.8	(1.9-11.8)‡		
Difference between index degree and partner average degree												
0	248	(44)	30	(12)	1.0						1.0	
> 0 to < 1	141	(25)	69	(49)	7.0	(4.2-11.5)‡					2.8	(1.5-5.3)‡
1 to < 2	138	(24)	24	(17)	1.5	(0.85-2.7)§					1.3	(0.65-2.5)
2-32	42	(7)	18	(43)	5.5	(2.6-11.2)‡					2.2	(0.98-4.9)

* Designated as ‘married’ only if noted as such in case record, otherwise classified as ‘unmarried’
† Based upon estimation of investigating DIS at diagnosis, not confirmed by labs
‡ Significant at p<0.05 (bivariable analysis and multivariable analyses)
§ Significant at p<0.20 (bivariable analysis only)

Figure 24. ROC curves for predictive models constructed after splitting the group of 569 indexes by whether or not the index chose to be interviewed and disclose partners (n=423, 74%) or not (n=141, 26%).

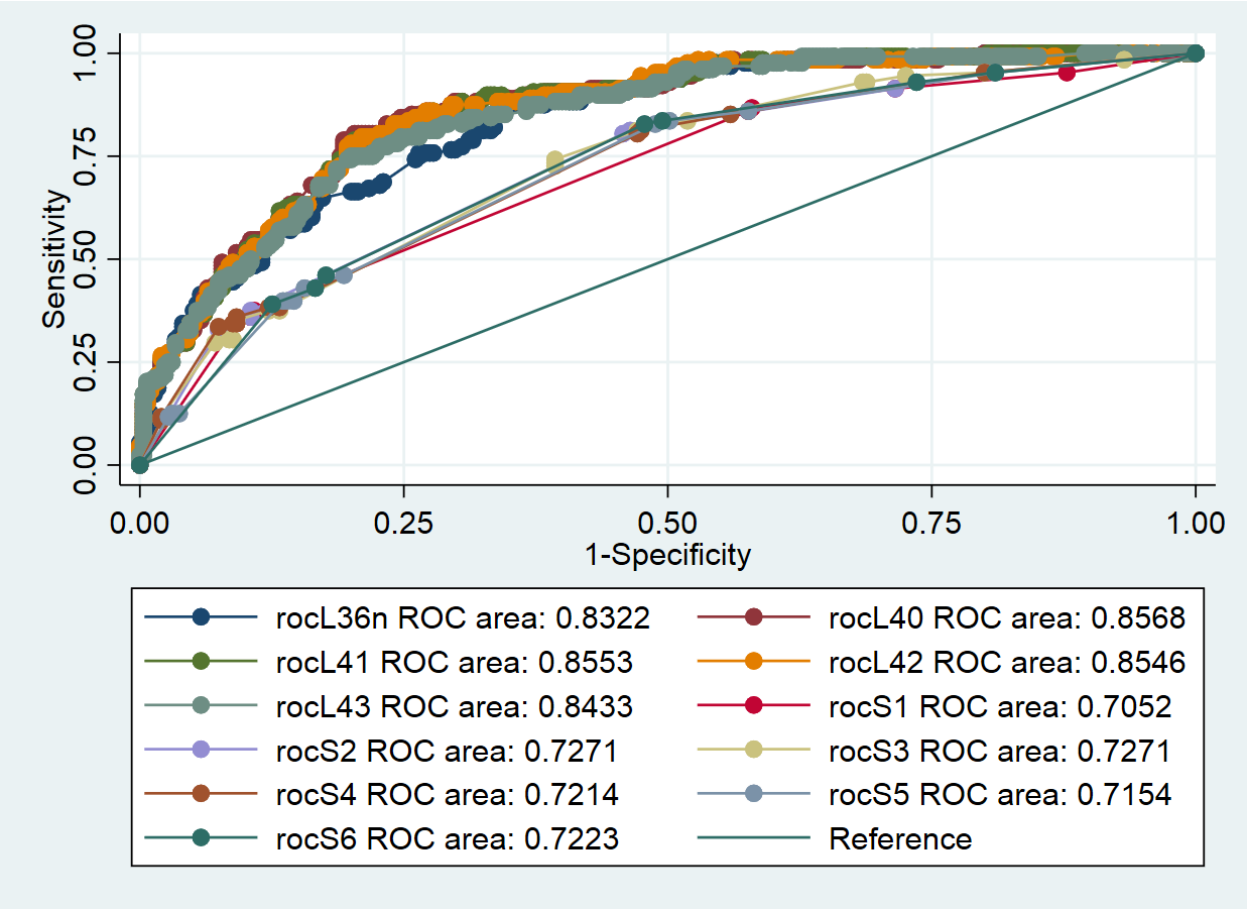
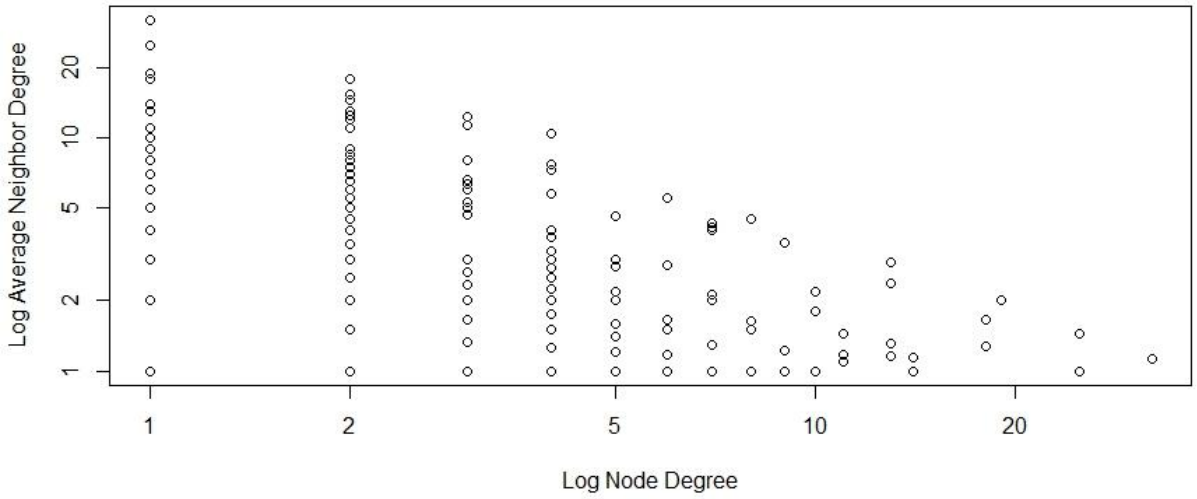


Figure legend:

The rocL models represent the larger group who chose to disclose partners and the rocS models represent the smaller group who chose not to disclose partners. All models depicted met the model evaluation criteria, including all predictors chosen through backwards elimination and significant at $p<0.05$.

Figure 25. Log node degree of persons in network compared to the log average node degree of their first-degree contacts.



APPENDIX C: BACKROUND AND METHODS RELATED TO FUTURE ANALYSES

A. Exponential Random Graph Models

Regression models applied to network data must adequately account for interdependence between actors. Exponential random graph models (ERGMs) are "tie-based" models of social and sexual networks, which indicates that the presence or absence of a tie between two nodes is the most basic unit of data. Data for the ERGM is adjacency matrix-based, where the vertices are the nodes and each cell has a 1 if two nodes are connected and 0 if not; this value is also the value of the regressand for each observation.[72] ERGMs use the summary and more complex measures of a network as the regressors, which allows calculations of parameters which convey the importance of those measures in the form of the local network. More complex measures of network structure include closed loops (called *k*-cores) and attribute-based homophily, where it is noted whether two nodes share an attribute of interest or are discordant for that attribute. Categorical and continuous regressors can be used, as in a standard regression equation. The outcome is the probability of the observed network given the values of the network statistics in the model. ERGMs use both endogenous and exogenous factors as regressors. Endogenous effects arise from social processes, including the influence that the presence (or absence) of ties in the network have on the formation or dissolution of other ties. Exogenous effects are actor attributes, such as age or gender, which affect tie formation.[72,74]

There are several advantages to using ERGMs for social network analysis. ERGMs model the structure of social networks and also allow for randomness through stochasticity,[72,74] which is valuable because the complexities of social processes can't (yet) be well-measured. A new class of valued models can be applied to the analysis in the first aim, where ties between partners may include being sexually linked, being phylogenetically linked, both, or neither.[140] ERGMs can be applied to large datasets[98] and allow for partial dependence between actors who aren't directly connected, which is suitable for social networks. Recent advances in model

specification better fit observed data[75,98] and permit inference when applied to social networks regarding social processes that influence network ties.[98]

B. Spatial Analysis of Sexually Transmitted Infections

Many studies have examined geographic or spatial heterogeneity of HIV or STI diagnoses.[36,44,51,52,230] On a large scale, the rates of incident HIV diagnoses are much higher in the US Northeast and South than the West and the Midwest. Further comparison of the higher rates US Northeast and the US South yields racial differences; in both places, 75% of people living with AIDS were either Black or Hispanic, although Hispanics accounted for a smaller percentage in the US South than in the Northeast.[32]

On a smaller scale, geographic core areas have been identified for several sexually transmitted infections.[33] Risk of having primary or secondary syphilis was 4.6 times higher for persons living in a certain area of San Francisco between 1985 and 2007. The spatial analysis was able to separate core and outbreak areas.[34] HIV-positive persons resided closer to their partners in Colorado Springs than at risk persons who are not HIV-positive and their partners.[35] During a syphilis outbreak in Baltimore, two areas were identified as core areas from which the outbreak spread and a new core area was created. Even after the outbreak ended, density of cases remained higher in all 3 core areas.[36] Syphilis was found to co-cluster with gonorrhea in NC; state-wide mapping of gonorrhea and syphilis over time identified 20 core areas for gonorrhea and 10 for syphilis. All of the syphilis core areas were found to have at least some overlap with at least one gonorrhea core area. All clusters, for gonorrhea and syphilis, were found to be associated with an urban area; some areas existed entirely in urban areas and some encompassed but urban and rural but none were entirely rural.[50] The rural-urban divide has been noted for other STIs in NC. Drug susceptible HIV strains were significantly more genetically similar than drug resistant strains for rural-rural and rural-urban partnerships, but not for urban-urban partnerships. Urbanicity of residence was not associated with TDRM v. drug susceptible virus among persons with acute HIV infection.[51] Geographically-associated network

cores have been found to significantly contribute to STI spread,[50,52] but it is unknown whether the same relationship is found with TDRM, particularly in rural areas.

Geographic and spatial investigation of HIV and STI outbreaks in NC has yielded important information. The CDC found that NC had the highest burden of HIV in non-urban areas in 2006.[4] Figure 1 (section II.A.2.c.), from the 2013 HIV/STD NC Epidemiologic Profile report, shows incident and prevalent HIV cases across the state. The largest clusters are in the most urban areas, but the rural eastern part of the state has a high burden of HIV without having as many specialists and providers as the central part of the state.

Spatiotemporal analysis was applied to identify core clusters of gonorrhea and syphilis in NC, with an additional assessment of rurality. All of the syphilis and gonorrhea core areas included at least one urban area, and all of the syphilis core areas (N=10) overlapped with gonorrhea core areas (N=20).[50] Similarly, in Wake County, a single urban county in central NC, chlamydia, gonorrhea, primary and secondary syphilis, and HIV were found to cluster with a single identifiable core area; all four core areas overlapped.[53] These studies show that in NC, there is geographic overlap of several STIs. Without added analysis of the sexual networks, it is unknown whether the STIs are circulating among different groups. As HIV is more easily transmitted in the presence of certain STIs, future areas of research should include both network and geographic or spatial analysis.

1. Phylogenetic Analysis Combined with Spatial or Geographic Analysis

Gene sequence analysis combined with spatial and geographic analyses have helped trace the spread of HIV around the world through history; looking at virus evolution in different places gives a sense of divergence.[230-232] The F1 subtype is the second most common HIV-1 type in Italy, although two distinct clades are seen and its origin was not known until recently. HIV-1 subtype F1 is believed to have arisen in West Africa before arriving in South America in the 1950s. From South America, the F1 subtype spread to Angola.[232] Subtype C is most common in Angola,[233] but the F1 variant spread from Angola to Romania.[232] The F1 subtype

arrived in Italy in the 1970s directly from South America and indirectly through Angola and Romania before being introduced, which account for the distinct clades seen today.[232]

Angola is a unique environment. A war for independence from the Portuguese took place from 1961-1975, leading to mass migrations and mixing of groups, just after the time when the F1 variant was introduced from South America. In 1975, the Portuguese conceded and in the same year, a war for independence began between the two most dominant ethnic groups. The civil war lasted until 2002, with movement of troops and civilians throughout the country. Due to lengthy deployments, it was not uncommon for men in the military to have wives in different parts of the country. The civil war factions were split along ethnic lines and traditionally lived in different parts of the country; the creation of the colony of Angola by the Portuguese in the 1500s encompassed groups speaking 7 different Bantu languages. The ethnic group originally granted power in 1975 and currently in power retained the northern and most of the central territory and the "rebel" group retained the southern part of the country.

Subtype F1 would have been introduced just prior to the 1961 start of the war for independence[232] and its mass migration. Subtype C appears to have been introduced multiple times from several different African countries, leading to distinct lineages.[233] *Pol* gene subtype analysis somewhat follows ethnic lines as they were influenced by the war. Subtype F1 is the predominant strain in the north (20% of sequences analyzed) and subtype C is the predominant strain in the south (46% of sequences analyzed). The central part of the country is dominated by recombinants (42% of sequences analyzed).[234] Combining geographic and phylogenetic data can indicate the presence of historical or contextual reasons why groups with similarities in a contained area don't cluster.

US-born and foreign-born Latinos with acute HIV in NC did not cluster with each other and appeared to have distinct sexual networks; US-born Latinos were more likely to cluster with Black persons.[49] Therefore, adding the genetic data to the geographic data showed that despite a common ethnicity, Latinos in NC should be treated as distinct groups for intervention. This is

important to recognize for the present study, as 12.8% of persons living in Wake County from 2009-2013 were foreign-born.[235]

C. Sexual Transmission of HIV

In the US, HIV is primarily transmitted through sexual intercourse,[30] which requires the virus to cross a mucous membrane.[236-238] Intact vaginal mucosal epithelium has many defenses which reduce likelihood of HIV transmission.[236,238,239] The size of the gaps in the deeper layers of the epithelium is smaller than the virus particles, neither the lining cells of the genital stratified squamous epithelium (male and female) nor the endocervical columnar epithelium (female) easily transcytose free HIV virions,[240] and the surface stratified cells lack CD4 target receptors.[236] However, conditions which increase the permeability of the mucosa, including lesion-causing STIs or thinning of the lining due to progesterone treatment, in turn increase the likelihood of HIV transmission.[236-238] Rectal epithelium does not provide the same protection because it lacks many of the features of genital epithelium. Rectal epithelial cells transcytose virions[241,242] and express the CXCR4 coreceptor, the rectal epithelium is a single layer,[243] and trauma which breaches the epithelium is more likely.[236,242] Both genital and rectal epithelium are susceptible to physical breaches and both allow transmigration of the virus via HIV-susceptible cells in the epithelium.[236,239,244] Vaginal epithelium will transcytose infected cells bound to its surface[239] and dendritic and T cells in the deeper layers of the epithelium can become infected.[245] Innate immune protection such as secretory leukocyte protease inhibitors secreted by endocervical cells may also be protective in receptive vaginal sex, but there is not a similar secretion from rectal cells to protect during receptive anal sex.[236]

Certain conditions increase the likelihood of transmission in both the genital and rectal tracts by diminishing the physical defenses. Exposure to HIV-1 may increase inflammatory processes which increase gap size in both genital and rectal epithelium in vitro.[246] Lesion-causing STIs in the anogenital tract facilitate HIV transmission by providing direct contact with target cells and bypassing the physical defenses of the epithelium.[236,247,248] Genital ulcer

disease (GUD) has been associated with HIV-1 infection, and many studies have advocated for treating GUD to reduce HIV transmission.[248-252] HSV-2 infection has been associated with HIV-1 infection in several populations,[253-256] although HSV-2 has been shown to have no effect on HIV viral load or CD4 count[257,258] so HSV-2 may facilitate HIV transmission without affecting the course of HIV infection. Syphilis, however, has been shown to have a synergistic effect with HIV. Being infected with syphilis increases the likelihood of both acquiring and transmitting HIV.[16,17] Being infected with HIV was associated with larger chancres which took longer to heal among patients in a French cohort,[18] which could facilitate transmission of both infections. This is further compounded by the fact that monocytes in HSV-1, HSV-2, and syphilis lesions in the female genital tract were found to express higher levels of CCR5, as do monocytes elsewhere in the female genital tract during primary and secondary syphilis infection, which increases the likelihood of viral entry into the cell.[19]

The increase in CCR5-expressing white cells is important. R5-tropic strains preferentially attach to monocytes, which are more likely to transmigrate across the endocervical epithelium than lymphocytes.[259] This could indicate that without a significant breach in the epithelium, a single R5 strain is likely to establish infection, which is observed. The majority of heterosexually-acquired HIV cases are single-variant infections established by R5 strains,[260] which appear to be preferentially transcytosed as macrophage-associated virus across the epithelium in the female reproductive tract, when compared to X4-tropic strains.[259] Intestinal epithelium also preferentially transcytoses R5 strains in both the upper[261] and lower GI tract.[262] As expected since R5 strains are able to navigate the epithelium more easily and therefore establish infection, the majority of virus in early infection is R5-tropic, whereas X4-tropic strains are dominant in later infection.[263] The higher proportion of R5 virus in early infection could compound the increased infectiousness during the acute stage of infection, as the R5 strains are more likely to establish a new infection.

Acute transmission is thought to account for a disproportionately high number of new infections,[26,83,264-267] although questions have been raised as to whether some of the infections attributed are due to incorrect measurement.[102,268] However, most research supports the impact of the acute phase of infection on onward transmission. People with acute HIV infection (AHI) are often co-infected with STIs,[101] which in turn increases the likelihood of transmission.[250] Acutely infected persons have higher viral loads,[269] are not ill enough yet to reduce sexual activity,[79] and are more likely to be in a high risk group for HIV transmission[101] or a high risk sexual network.[21]

D. Intra-Host Viral Dynamics

HIV's short generation time and high mutation rate[270,271] leads to more intra-host variability in resistance and susceptibility to ART than is present at the population level.[270,272-274] Due to its high mutation rate, HIV population variability changes within a host over time,[274] so sequences obtained at the beginning of an infection will change without the presence of selective pressure from ARTs, although it will occur more quickly under those selective pressures.[271] Recently-infected persons have less diversity in the HIV virus population than chronically-infected persons, although diversity increases much faster in acutely-infected persons.[275] As such, algorithms can be applied to phylogenetic data to help determine whether infection was recently acquired or if it is chronic.[276]

The high mutation rate results in minority quasispecies which can remain in the latent reservoir. The minority variants are generally less fit else they would become dominant variants, and some encode drug resistance.[277] With selective pressure of ART, the minority variants can resurge and lead to early treatment failure.[223] These minority variants can be transmitted to ART-naïve persons, potentially limiting future treatment options.[278] Drug susceptible strains have higher fitness for establishing infection in an uninfected recipient. However, despite the lower fitness, resistance profiles of newly-infected ART-naïve persons show that drug resistant strains are transmitted.[279-283]

The high mutation rate also plays a role in the long-term persistence of HIV as it relates to host co-receptors targeted. Nearly all HIV infections are caused by a CCR5-tropic virus strain,[284] which is more efficient at crossing the epithelial barrier.[285] However, the population usually switches in some proportion to CXCR4-tropic strains over the course of the infection[286-288] as CD4+ cells are depleted. X4 strains likely arise via mutation and sustain the virus population once the R5 strain target cells decline in numbers.[263] However, X4 strains aren't well-suited to establishing new infections.[289]

1. Founder Strains

Evidence from strain diversity analysis supports a transmission bottleneck, where recently infected individuals have low strain diversity even when infected by chronically-infected persons who have high strain diversity.[290,291] This is due to higher fitness in establishing infection, as there are many variants in the inoculum. A single transmitted founder strain causes 76% of HIV infections.[260] Heterosexually-acquired HIV is slightly higher at 81%.[292] However, this proportion is significantly reduced for MSM and IDU,[293] where the inoculum can breach the mucosal barrier. Approximately 62% of infections acquired by MSM are thought to be the result of a single variant,[292] and it drops to 40% for IDU.[294] Engaging in the high risk behaviors which put one at risk of acquiring multiple infecting variants is associated with the person's sexual network component.[11] IDU,[294] MSM,[292] and heterosexuals co-infected with ulcerative STIs[295] are more likely to have infection resulting from multiple variants. It is plausible that conditions which circumvent the mucosal barrier also increase the likelihood of transmission of drug resistant variants which have less replicative capacity; the mechanisms which circumvent the transmission bottleneck may also facilitate infection with less fit drug resistant strains. ARV-naïve MSM are more likely than ARV-naïve heterosexuals to have drug resistant strains at diagnosis.[56,296]

Most new infections result from a single founder strain,[243,290,291,297,298] although even in a mixed population, some infections are established by less fit[299-301] strains which demonstrate

drug resistance.[19,20,24-27,101] Drug resistant strains can be transmitted, although not as frequently or as easily as wild type.[55] HIV clinic survey data showed that drug resistant virus strains have approximately 20% of the capacity of wild type strains to establish new infections.[55] Drug resistance mutations (DRM) have different fitness costs; some conferred almost no fitness costs[302] and as such persist after ARV is stopped.[300] Certain mutations do confer fitness loss that would lead to being out-competed within a host, but not so much of a fitness loss that the resistant strains cannot establish new infections.[55] Additionally, with fewer strains in early infection, a donor can only pass one of those strains on to a recipient.[55] Therefore, transmission of drug resistant variants may be a compounding problem at the population level. Viral diversity is lower after initial infection than it is in later infection, due to first a small number of infecting variants and then increasing in diversity due to escape mutations.[293] If someone who is infected with a resistant strain transmits the infection to a new recipient during the acute phase of infection,[55] as happens often, then onward transmission of the drug resistant variant might be more likely if there is somewhat of a probabilistic element to the successful transmission of the found strain, thus contributing to the epidemic of transmitted DRM (TDRM).

2. Drug Resistance Mutations

Drug resistance occurs as a result of mutation during virus transcription then becomes encoded in the viral population due to selective pressure in the presence of ART.[303] Mutations classified as drug resistance mutations (DRM) confer at least partial resistance to at least one first line ARV.[272,304]

HIV infection tends to revert to wild type virus once the selective pressure of ARVs is removed because some replicative capacity is lost with certain resistance mutations.[272,305-307] However, this only occurs as a result of a mutation that increases both drug sensitivity and reproductive fitness.[272,304] Additionally, primary infection with a drug resistant strain does not revert to a completely susceptible virus population because resistant strains persist in the cellular reservoir. [272,305-307] Standard therapies may then be perpetually less effective for

individuals with acquired drug resistance due to a latent resistant population.[272] Additionally, at least one reverse transcriptase mutation seems to revert only to an intermediate strain that quickly mutates into resistant virus if AZT is started.[282] Time to viral suppression is slower with an acquired or transmitted DRM,[308-311] there is less sensitivity to first-line drugs,[312] and treatment options are limited in individuals who have a transmitted drug resistance mutation.[309,313,314]

Polymorphic drug resistance mutations occur spontaneously and do not require selective pressure from ART. Non-polymorphic drug resistance mutations typically do not occur without the presence of ART.[314] Non-polymorophic drug resistance mutations in a person who is ARV-naïve indicate likely infection with a drug resistance strain rather than spontaneous (acquired) mutation; limiting study to nonpolymorphic mutations increases the specificity as a marker for transmitted DRM (TDRM).[315]

Table 15 shows the individual mutations found with high prevalence among nonpolymorphic sites in HIV-1 diagnoses of ART-naïve persons made in 2006 in selected areas of the US,[2] though the prevalence has changed with the standard use of integrase inhibitors. The table lists the drug class to which the mutation confers resistance (PI= protease inhibitors, NRTI=nucleoside reverse transcriptase inhibitors, NNRTI=non-nucleoside reverse transcriptase inhibitors, INI=integrase inhibitors), the amino acid position of the mutation, the consensus B amino acid (Cons-B AA), the substituted amino acid (Mut-AA), and the percent of sequences found to have the mutation at

Table 15. Most common HIV genetic mutations in untreated persons.[1,2]

Class	Pos	Cons B-AA	Mut-AA	Prev (%)	Hi Res
PI	46	Methionine	Isoleucine	0.6	No
PI	90	Leucine	Methionine	1.2	No
NRTI	41	Methionine	Leucine	1.7	No
NRTI	70	Lysine	Arginine	0.5	No
NRTI	184	Methionine	Valine	1.0	Yes
NRTI	215	Threonine	Cysteine	0.6	No
NRTI	215	Threonine	Glutamic Acid	0.2	No
NRTI	215	Threonine	Serine	0.5	No
NRTI	219	Lysine	Arginine	0.1	Yes
NNRTI	101	Lysine	Glutamic Acid	0.3	No
NNRTI	103	Lysine	Asparagine	5.1	Yes

* adapted from Wheeler et al, 2010 and Stanford HIV Drug Resistance Database 2012. Class=class of ART; Pos=genome position; Cons B-AA=consensus B amino acid; Mut-AA=amino acid after mutation; Prev (%)=prevalence of mutation; Hi Res=considered by Stanford Drug Resistance Database to contribute to a moderate or high level of ART resistance in most of the ARTs commonly prescribed in the US.

that position. The study from which Table 15 was adapted found an overall TDRM prevalence of 14.6% in 2030 ART-naïve individuals. 7.8% had at least one mutation conferring resistance to NNRTIs, 5.6% to NRTIs, and 4.5% to PIs.[2] Certain NNRTI and PI mutations persist even after stopping ART, increasing the likelihood of onward transmission.[300] Newer sequencing methods include INI mutations. Although the genetic barrier to resistance is lower for INIs than NRTIs or PIs, it is believed that INI resistance is transmitted less often than NNRTI, NRTI, or PI resistance mutations.[316]

a. Transmitted Drug Resistance and Associated Outcomes

DRM can be transmitted with infection.[164] Persons infected with resistant virus may never achieve suppression, as they will always harbor resistant strains.[317] Even minority variants detected prior to ART exposure have been associated with higher rates of treatment failure.[221,309] Resurgence of minor drug resistant variants happens in less than half the time, on average, as reversion to wild type strains after stopping ART.[318] Individuals who are ever infected with or who acquire drug resistance mutations are at lifetime risk of lower treatment efficacy because there may always be a resistant population of minor variants within the host.[309,310,318,319] At the population level, efficacy of pre-exposure prophylaxis (PrEP) and post-exposure prophylaxis (PEP) is reduced in the presence of certain drug resistance mutations.[304,320] The potential for transmission of DRM therefore also has implications for at-risk persons: dissemination of PrEP to high risk persons has been demonstrated to reduce the number of new cases, although the resistance profile of the local DRM should be known to achieve optimal results with PrEP (or PEP[321]) as some mutations decrease the efficacy the treatments.

The prevalence of transmissible DRM (TDRM) is largely unknown, but estimated to be 10-20% for at least one major resistance mutation in most regions in the U.S.[182,322-324] TDRM are more common in areas in which antiretroviral therapy is available;[319] resistance is expected to increase as ARVs become more available worldwide.[304] Risk of drug resistance mutations is

higher when adherence is suboptimal.[325] Nonadherent persons are likely to engage in unprotected sex, which provides opportunity for onward transmission of drug resistant infections.[326-329]

Previous studies of TDRM have focused on broad groups, such as race, or used previously-defined cohorts, such as HIV clinic patients, to assess TDRM prevalence, which doesn't reflect actual transmission patterns. There are several benefits to assessing the local TDRM profile by sexual network component using all incident diagnoses. A better estimation of the population at risk for HIV can be made, particularly when demographically compared to the local population. Knowing the TDRM profile by component has the potential to inform interventions, including PrEP, that are tailored to individuals who have newly entered high-risk sexual networks, thereby decreasing the number of new HIV infections in NC. It is estimated that approximately 10% of HIV-positive individuals in NC are unaware of their status.[39] Obtaining a better profile of HIV and TDRM acquisition risk by demographic and risk group would allow estimation of individuals with unknown status. Resistance mutation prevalence has been associated with risk behavior,[296] which further supports the utility of phylogenetic analysis by sexual network component. The higher number of DRM in drug-naive MSM with resistance mutations could be due to either increased transmission during the acute infection phase or possibly a higher likelihood of multiclonal infection.[292] This could be due to sexual practices which increase trauma and reduce the effectiveness of the mucosal barrier; intravenous drug users also tended to have multiplicity of infection.[294] Analysis of timing of infection, risk behaviors, and number of partners may yield some information in support of one of those two hypotheses.

A Swiss study of 197 persons with acute or recent HIV infection found that the prevalence of DRM among persons infected through homosexual contact or intravenous drug use was approximately twice that of HIV acquired through heterosexual contact (11%, 13%, and 6%, respectively).[265] As drug resistant strains have lower replicative capacity than wild type

strains, there may be a biological mechanism allowing the drug resistant strains to bypass the infection bottleneck among MSM and IDU. Receptive anal intercourse and injection into the circulatory system both breach the innate defenses of the epithelium. This could plausibly allow the drug resistant strains to compete in establishing infection, leading to transmission of multiple variants or single drug resistant variants.

The local resistance profile in NC has not been studied in any amount of detail. A previous study of clusters in NC found that individuals in clusters were more likely to have at least one DRM,[276] which is similar to what was found among patients in the Swiss study.[267] Two university hospitals in NC analyzed a cohort of acutely- and recently-infected patients and found that 17.8% had at least one TDRM,[182] which has implications for public health given the infectiousness due to high viral loads early in infection. In a different but geographically related clinical cohort that includes both acute and chronic infections, 9.3% of patients had transmitted drug resistance, with prevalence among acute infections being 2.4 times that of persons chronically infected.[323] Without distinguishing between transmitted and acquired mutations, more than half of the patients in that cohort had at least one DRM, and many of the patients reported inconsistent condom use and suboptimal ART adherence.[326] The spread of TDRM in NC has implications for disease mitigation, first-line treatment failure in new infections, pre-exposure prophylaxis (PrEP), and post-exposure prophylaxis (PEP).[304,330,331]

REFERENCES

1. Stanford HIV Drug Resistance Database, *Major HIV Drug Resistance Mutations*. 2012, Stanford University.

2. Wheeler, W.H., R.A. Ziebell, H. Zabina, D. Pieniazek, J. Prejean, U.R. Bodnar, . . . H.I.V.S.G. Resistant, *Prevalence of transmitted drug resistance associated mutations and HIV-1 subtypes in new HIV-1 diagnoses, U.S.-2006.* AIDS, 2010. **24**(8): p. 1203-1212.

3. Centers for Disease Control and Prevention, *HIV in the United States: At a glance*. 2017: Atlanta, GA.

4. NC Department of Health and Human Services, *Epidemiologic Profile for HIV/STD Prevention & Care Planning*. 2011, Division of Public Health: Raleigh, NC.

5. Prejean, J., R. Song, A. Hernandez, R. Ziebell, T. Green, F. Walker, . . . HIV Incidence Surveillance Group, *Estimated HIV incidence in the United States, 2006-2009.* PLoS One, 2011. **6**(8): p. e17502.

6. Splettstoesser, T., *Structure of the RNA genome of HIV-1*, HIV-genome.png, Editor. 2014: Wikimedia Commons. p. HIV genome.

7. NC Department of Health and Human Services, *2013 North Carolina HIV/STD Epidemiologic Profile*. 2015, Division of Public Health: Raleigh, NC. p. 242.

8. Tieu, H.V., T.Y. Liu, S. Hussen, M. Connor, L. Wang, S. Buchbinder, . . . Hptn, *Sexual Networks and HIV Risk among Black Men Who Have Sex with Men in 6 U.S. Cities.* PLoS One, 2015. **10**(8): p. e0134085.

9. Fujimoto, K. and M.L. Williams, *Racial/Ethnic Differences in Sexual Network Mixing: A Log-Linear Analysis of HIV Status by Partnership and Sexual Behavior Among Most at-Risk MSM.* AIDS Behav, 2014.

10. Crepaz, N., X. Dong, X. Wang, A.L. Hernandez, and I. Hall, *Racial and Ethnic Disparities in Sustained Viral Suppression and Transmission Risk Potential Among Persons Receiving HIV Care -- United States, 2014.* MMWR, 2018. **67**(4): p. 113-118.

11. Doherty, I.A., N.S. Padian, C. Marlow, and S.O. Aral, *Determinants and Consequences of Sexual Networks as They Affect the Spread of Sexually Transmitted Infections.* J Infect Dis, 2005. **191**(Suppl 1): p. S42-54.

12. Dennis, A.M., S. Napravnik, A.C. Sena, and J.J. Eron, *Late entry to HIV care among Latinos compared with non-Latinos in a southeastern US cohort.* Clin Infect Dis, 2011. **53**(5): p. 480-487.

13. Torrone, E.A., J.C. Thomas, P.A. Leone, and L.B. Hightow-Weidman, *Late diagnosis of HIV in young men in North Carolina.* Sex Transm Dis, 2007. **34**(11): p. 846-848.

14. Arnold, E.A., G.M. Rebchook, and S.M. Kegeles, *'Triply cursed': racism, homophobia and HIV-related stigma are barriers to regular HIV testing, treatment adherence and disclosure among young Black gay men.* Cult Health Sex, 2014. **16**(6): p. 710-722.

15. Beer, L., A.M. Oster, C.L. Mattson, and J. Skarbinski, *Disparities in HIV transmission risk among HIV-infected black and white men who have sex with men, United States, 2009.* AIDS, 2014. **28**(1): p. 105-114.

16. Karp, G., F. Schlaeffer, A. Jotkowitz, and K. Riesenberg, *Syphilis and HIV co-infection.* Eur J Intern Med, 2009. **20**(1): p. 9-13.

17. Lynn, W.A. and S. Lightman, *Syphilis and HIV: a dangerous combination.* The Lancet Infectious Diseases, 2004. **4**(7): p. 456-466.

18. Pialoux, G., S. Vimont, A. Moulignier, M. Buteux, B. Abraham, and P. Bonnard, *Effect of HIV on the Course of Syphilis.* AIDS Rev, 2008. **10**: p. 85-92.

19. Sheffield, J.S., G.D. Wendel, Jr., D.D. McIntire, and M.V. Norgard, *Effect of genital ulcer disease on HIV-1 coreceptor expression in the female genital tract.* J Infect Dis, 2007. **196**(10): p. 1509-1516.

20. Centers for Disease Control and Prevention, *Understanding the HIV Care Continuum*. 2017, U.S. Department of Health and Human Services,. p. 4.

21. Hurt, C.B., S. Beagle, P.A. Leone, A. Sugarbaker, E. Pike, J. Kuruc, . . . L.B. Hightow-Weidman, *Investigating a sexual network of black men who have sex with men: implications for transmission and prevention of HIV infection in the United States.* J Acquir Immune Defic Syndr, 2012. **61**(4): p. 515-521.

22. NC HIV/STD Surveillance Unit, *2015 North Carolina HIV/STD Surveillance Report*. 2016, NC Division of Public Health: Raleigh, NC. p. 96.

23. Friedman, S.R. and S.O. Aral, *Social Networks, Risk-Potential Networks, Health, and Disease.* J Urban Health, 2001. **78**(3): p. 411-418.

24. Skaathun, B., A.S. Khanna, E. Morgan, S.R. Friedman, and J.A. Schneider, *Network Viral Load: A Critical Metric for HIV Elimination.* J Acquir Immune Defic Syndr, 2018. **77**(2): p. 167-174.

25. Morris, M., A.E. Kurth, D.T. Hamilton, J. Moody, and S. Wakefield, *Concurrent partnerships and HIV prevalence disparities by race: linking science and public health practice.* Am J Public Health, 2009. **99**(6): p. 1023-1031.

26. Brenner, B.G., M. Roger, J.P. Routy, D. Moisi, M. Ntemgwa, C. Matte, . . . H.I.V.I.S.G. Quebec Primary, *High rates of forward transmission events after acute/early HIV-1 infection.* J Infect Dis, 2007. **195**(7): p. 951-959.

27. Fisher, M., D. Pao, A.E. Brown, D. Sudarshi, O.N. Gill, P. Cane, . . . D. Pillay, *Determinants of HIV-1 transmission in men who have sex with men: a combined clinical, epidemiological and phylogenetic approach.* AIDS, 2010. **24**(11): p. 1739-1747.

28. Cope, A.B., K.A. Powers, J.D. Kuruc, P.A. Leone, J.A. Anderson, L.H. Ping, . . . W.C. Miller, *Ongoing HIV Transmission and the HIV Care Continuum in North Carolina.* PLoS One, 2015. **10**(6): p. e0127950.

29. Centers for Disease Control and Prevention, *HIV Surveillance Report, 2014: Diagnoses of HIV Infection in the United States and Dependent Areas*, in *HIV Surveillance Report*. 2015, US Department of Health and Human Services,: Atlanta, GA. p. 123.

30. Centers for Disease Control and Prevention, *HIV Surveillance Report: Diagnoses of HIV Infection in the United States and Dependent Areas, 2015*, in *HIV Surveillance Report*. 2016, US Department of Health and Human Services,: Atlanta, GA. p. 114.

31. Centers for Disease Control and Prevention, *HIV Among Men in the United States*. 2013, National Center for HIV/AIDS, Viral Hepatitis, STD, and TB Prevention. p. 2.

32. Centers for Disease Control and Prevention, *HIV and AIDS in the United States by Geographic Distribution*. 2012: Atlanta, GA.

33. Rothenberg, R.B., *The geography of gonorrhea. Empirical demonstration of core group transmission.* Am J Epidemiol, 1983. **117**(6): p. 688-694.

34. Gesink, D.C., A.B. Sullivan, W.C. Miller, and K.T. Bernstein, *Sexually transmitted disease core theory: roles of person, place, and time.* Am J Epidemiol, 2011. **174**(1): p. 81-89.

35. Rothenberg, R., S.Q. Muth, S. Malone, J.J. Potterat, and D.E. Woodhouse, *Social and Geographic Distance in HIV Risk.* Sex Transm Dis, 2005. **32**(8): p. 506-512.

36. Gesink Law, D.C., K.T. Bernstein, M.L. Serre, C.M. Schumacher, P.A. Leone, J.M. Zenilman, . . . A.M. Rompalo, *Modeling a syphilis outbreak through space and time using the Bayesian maximum entropy approach.* Ann Epidemiol, 2006. **16**(11): p. 797-804.

37. National Center for HIV/AIDS, V.H., STD, and TB Prevention;, *HIV and AIDS in the United States by Geographic Distribution*. 2012, Centers for Disease Control and Prevention,.

38. NC Department of Health and Human Services, *North Carolina 2012 HIV/STD Surveillance Report*. 2013, NC Division of Public Health: Raleigh, NC. p. 63.

39. NC Department of Health and Human Services, *HIV Continuum of Care in North Carolina Reported HIV Case Data, 2015*. 2017, NC DHHS Communicable Disease Branch,: Raleigh, NC. p. 2.

40. Hightow, L.B., P.D. Macdonald, C.D. Pilcher, A.H. Kaplan, E. Foust, T.Q. Nguyen, and P.A. Leone, *The Unexpected Movement of the HIV Epidemic in the Southeastern United States.* J Acquir Immune Defic Syndr, 2005. **38**(5): p. 531-537.

41. Sena, A.C., E.A. Torrone, P.A. Leone, E. Foust, and L. Hightow-Weidman, *Endemic early syphilis among young newly diagnosed HIV-positive men in a southeastern U.S. state.* AIDS Patient Care STDS, 2008. **22**(12): p. 955-963.

42. Hightow, L.B., P.A. Leone, P.D. Macdonald, S.I. McCoy, L.A. Sampson, and A.H. Kaplan, *Men who have sex with men and women: a unique risk group for HIV transmission on North Carolina College campuses.* Sex Transm Dis, 2006. **33**(10): p. 585-593.

43. Gorbach, P.M., R. Murphy, R.E. Weiss, C. Hucks-Ortiz, and S. Shoptaw, *Bridging sexual boundaries: men who have sex with men and women in a street-based sample in Los Angeles.* J Urban Health, 2009. **86 Suppl 1**: p. 63-76.

44. Doherty, I.A., M.L. Serre, D. Gesink, A.A. Adimora, S.Q. Muth, P.A. Leone, and W.C. Miller, *Sexual networks, surveillance, and geographical space during syphilis outbreaks in rural North Carolina.* Epidemiology, 2012. **23**(6): p. 845-851.

45. Doherty, I.A., A.A. Adimora, S.Q. Muth, M.L. Serre, P.A. Leone, and W.C. Miller, *Comparison of sexual mixing patterns for syphilis in endemic and outbreak settings.* Sex Transm Dis, 2011. **38**(5): p. 378-384.

46. Sena, A.C., S.Q. Muth, J.D. Heffelfinger, J.O. O'Dowd, E. Foust, and P.A. Leone, *Factors and the sociosexual network associated with a syphilis outbreak in rural North Carolina.* Sex Transm Dis, 2007. **34**(5): p. 280-287.

47. Brown, A. and M.H. Lopez, *Mapping the Latino Population, By State, County and City*. 2013, Pew Research Center: Washington, D.C. p. 24.

48. Dennis, A.M., J.B. Wheeler, E. Valera, L. Hightow-Weidman, S. Napravnik, H. Swygard, . . . J.J. Eron, *HIV risk behaviors and sociodemographic features of HIV-infected Latinos residing in a new Latino settlement area in the Southeastern United States.* AIDS Care, 2013.

49. Dennis, A.M., S. Hue, C.B. Hurt, S. Napravnik, J. Sebastian, D. Pillay, and J.J. Eron, *Phylogenetic insights into regional HIV transmission.* AIDS, 2012. **26**(14): p. 1813-1822.

50. Gesink, D.C., A.B. Sullivan, T.A. Norwood, M.L. Serre, and W.C. Miller, *Does core area theory apply to sexually transmitted diseases in rural environments?* Sex Transm Dis, 2013. **40**(1): p. 32-40.

51. Carrel, M., J.J. Eron, Jr., M. Emch, and C.B. Hurt, *Spatial epidemiology of recently acquired HIV infections across rural and urban areas of North Carolina.* PLoS One, 2014. **9**(2): p. e88512.

52. Bernstein, K.T., F.C. Curriero, J.M. Jennings, G. Olthoff, E.J. Erbelding, and J. Zenilman, *Defining core gonorrhea transmission utilizing spatial data.* Am J Epidemiol, 2004. **160**(1): p. 51-58.

53. Law, D.C., M.L. Serre, G. Christakos, P.A. Leone, and W.C. Miller, *Spatial analysis and mapping of sexually transmitted diseases to optimise intervention and prevention strategies.* Sex Transm Infect, 2004. **80**(4): p. 294-299.

54. Bohl, D.D., W. McFarland, and H.F. Raymond, *Improved measures of racial mixing among men who have sex with men using Newman's assortativity coefficient.* Sex Transm Infect, 2011. **87**(7): p. 616-620.

55. Leigh Brown, A.J., S.D. Frost, W.C. Mathews, K. Dawson, N.S. Hellmann, E.S. Daar, . . . S.J. Little, *Transmission fitness of drug-resistant human immunodeficiency virus and the prevalence of resistance in the antiretroviral-treated population.* J Infect Dis, 2003. **187**(4): p. 683-686.

56. Frentz, D., D. van de Vijver, A. Abecasis, J. Albert, O. Hamouda, L. Jorgensen, . . . Spread Programme, *Patterns of transmitted HIV drug resistance in Europe vary by risk group.* PLoS One, 2014. **9**(4): p. e94495.

57. McPherson, M., L. Smith-Lovin, and J.M. Cook, *Birds of a Feather: Homophily in Social Networks.* Annu Rev Sociol, 2011. **27**: p. 415-444.

58. O'Malley, A.J., F. Elwert, J.N. Rosenquist, A.M. Zaslavsky, and N.A. Christakis, *Estimating peer effects in longitudinal dyadic data using instrumental variables.* Biometrics, 2014.

59. Shalizi, C.R. and A.C. Thomas, *Homophily and Contagion Are Generically Confounded in Observational Social Network Studies.* Sociol Methods Res, 2011. **40**(2): p. 211-239.

60. Shoham, D.A., L. Tong, P.J. Lamberson, A.H. Auchincloss, J. Zhang, L. Dugas, . . . A. Luke, *An actor-based model of social network influence on adolescent body size, screen time, and playing sports.* PLoS One, 2012. **7**(6): p. e39795.

61. Schneider, J.A., B. Cornwell, D. Ostrow, S. Michaels, P. Schumm, E.O. Laumann, and S. Friedman, *Network mixing and network influences most linked to HIV infection and risk behavior in the HIV epidemic among black men who have sex with men.* Am J Public Health, 2013. **103**(1): p. e28-36.

62. Doherty, I.A., V.J. Schoenbach, and A.A. Adimora, *Sexual mixing patterns and heterosexual HIV transmission among African Americans in the southeastern United States.* J Acquir Immune Defic Syndr, 2009. **52**(1): p. 114-120.

63. Anderson, R.M., S. Gupta, and W. Ng, *The significance of sexual partner contact networks for the transmission dynamics of HIV.* J Acquir Immune Defic Syndr, 1990. **3**(4): p. 417-429.

64. Hertog, S., *Heterosexual behavior patterns and the spread of HIV/AIDS: the interacting effects of rate of partner change and sexual mixing.* Sex Transm Dis, 2007. **34**(10): p. 820-828.

65. Kault, D.A., *The impact of sexual mixing patterns on the spread of AIDS.* Math Biosci, 1995. **128**(1-2): p. 211-241.

66. Aral, S.O., *Patterns of sexual mixing: mechanisms for or limits to the spread of STIs?* Sex Transm Infect, 2000. **76**(6): p. 415-416.

67. Prah, P., A.J. Copas, C.H. Mercer, A. Nardone, and A.M. Johnson, *Patterns of sexual mixing with respect to social, health and sexual characteristics among heterosexual couples in England: analyses of probability sample survey data.* Epidemiol Infect, 2014: p. 1-11.

68. Garnett, G.P. and R.M. Anderson, *Contact tracing and the estimation of sexual mixing patterns: the epidemiology of gonococcal infections.* Sex Transm Dis, 1993. **20**(4): p. 181-191.

69. Aral, S.O., J.P. Hughes, B. Stoner, W. Whittington, H.H. Handsfield, R.M. Anderson, and K.K. Holmes, *Sexual mixing patterns in the spread of gonococcal and chlamydial infections.* Am J Public Health, 1999. **89**(6): p. 825-833.

70. Doherty, I.A., S. Shiboski, J.M. Ellen, A.A. Adimora, and N.S. Padian, *Sexual bridging socially and over time: a simulation model exploring the relative effects of mixing and concurrency on viral sexually transmitted infection transmission.* Sex Transm Dis, 2006. **33**(6): p. 368-373.

71. Newman, M.E.J. and J. Park, *Why social networks are different from other types of networks.* Phys Rev E, 2003. **68**(3): p. 036122.

72. *Exponential Random Graph Models for Social Networks: Theory, Methods, and Applications*. Structural Analysis in the Social Sciences, ed. M. Granovetter. 2013, New York, NY: Cambridge University Press.

73. Flynn, F.J., R.E. Reagans, and L. Guillory, *Do you two know each other? Transitivity, homophily, and the need for (network) closure.* J Pers Soc Psychol, 2010. **99**(5): p. 855-869.

74. Robins, G., P. Pattison, Y. Kalish, and D. Lusher, *An introduction to exponential random graph (p*) models for social networks.* Social Networks, 2007. **29**(2): p. 173-191.

75. Robins, G., T. Snijders, P. Wang, M. Handcock, and P. Pattison, *Recent developments in exponential random graph (p*) models for social networks.* Social Networks, 2007. **29**(2): p. 192-215.

76. Foster, D.V., J.G. Foster, P. Grassberger, and M. Paczuski, *Clustering drives assortativity and community structure in ensembles of networks.* Phys Rev E Stat Nonlin Soft Matter Phys, 2011. **84**(6 Pt 2): p. 066117.

77. Gorbach, P.M., L.N. Drumright, E.S. Daar, and S.J. Little, *Transmission behaviors of recently HIV-infected men who have sex with men.* J Acquir Immune Defic Syndr, 2006. **42**(1): p. 80-85.

78. Gorbach, P.M., R.E. Weiss, R. Jeffries, M. Javanbakht, L.N. Drumright, E.S. Daar, and S.J. Little, *Behaviors of recently HIV-infected men who have sex with men in the year postdiagnosis: effects of drug use and partner types.* J Acquir Immune Defic Syndr, 2011. **56**(2): p. 176-182.

79. Eaton, L.A. and S.C. Kalichman, *Changes in transmission risk behaviors across stages of HIV disease among people living with HIV.* J Assoc Nurses AIDS Care, 2009. **20**(1): p. 39-49.

80. Romero-Severson, E.O., S.J. Alam, E.M. Volz, and J.S. Koopman, *Heterogeneity in Number and Type of Sexual Contacts in a Gay Urban Cohort.* Stat Commun Infect Dis, 2012. **4**(1).

81. Alam, S.J., X. Zhang, E.O. Romero-Severson, C. Henry, L. Zhong, E.M. Volz, . . . J.S. Koopman, *Detectable signals of episodic risk effects on acute HIV transmission: strategies for analyzing transmission systems using genetic data.* Epidemics, 2013. **5**(1): p. 44-55.

82. Romero-Severson, E.O., S.J. Alam, E. Volz, and J. Koopman, *Acute-stage transmission of HIV: effect of volatile contact rates.* Epidemiology, 2013. **24**(4): p. 516-521.

83. Alam, S.J., E. Romero-Severson, J.H. Kim, G. Emond, and J.S. Koopman, *Dynamic sex roles among men who have sex with men and transmissions from primary HIV infection.* Epidemiology, 2010. **21**(5): p. 669-675.

84. Kim, J.H., R.L. Riolo, and J.S. Koopman, *HIV transmission by stage of infection and pattern of sexual partnerships.* Epidemiology, 2010. **21**(5): p. 676-684.

85. Zhang, X., L. Zhong, E. Romero-Severson, S.J. Alam, C.J. Henry, E.M. Volz, and J.S. Koopman, *Episodic HIV Risk Behavior Can Greatly Amplify HIV Prevalence and the Fraction of Transmissions from Acute HIV Infection.* Stat Commun Infect Dis, 2012. **4**(1).

86. Brooks, J.I., H. Niznick, M. Ofner, H. Merks, and J.B. Angel, *Local phylogenetic analysis identifies distinct trends in transmitted HIV drug resistance: implications for public health interventions.* BMC Infect Dis, 2013. **13**: p. 509.

87. Vynnycky, E. and R.G. White, *An Introduction to Infectious Disease Modelling*. 2010, New York, NY: Oxford University Press. 370.

88. Newman, M.E.J., *Assortative mixing in networks.* Phys Rev Lett, 2002. **89**(20): p. 208701.

89. Valente, T.W., *Social Networks and Health: Models, Methods, and Applications*. 2010, New York, NY: Oxford University Press, Inc. 277.

90. Wasserman, S. and K. Faust, *Social Network Analysis: Methods and Applications*. Structural analysis in the social sciences. 1994, New York, NY: Cambridge University Press. 825.

91. Sharma, S. and R.K. Gupta, *Improved BSP Clustering Algorithm for Social Network Analysis.* International Journal of Grid and Distributed Computing, 2010. **3**(3): p. 67-76.

92. Handcock, M.S., A.E. Raftery, and J.M. Tantrum, *Model-based clustering for social networks.* J. R. Statist. Soc. A, 2007. **170**(2): p. 301-354.

93. Oster, A.M., D. Pieniazek, X. Zhang, W.M. Switzer, R.A. Ziebell, L.A. Mena, . . . J.D. Heffelfinger, *Demographic but not geographic insularity in HIV transmission among young black MSM.* AIDS, 2011. **25**(17): p. 2157-2165.

94. Sudhinaraset, M., H.F. Raymond, and W. McFarland, *Convergence of HIV prevalence and inter-racial sexual mixing among men who have sex with men, San Francisco, 2004-2011.* AIDS Behav, 2013. **17**(4): p. 1550-1556.

95. Potterat, J.J., R.B. Rothenberg, and S.Q. Muth, *Network structural dynamics and infectious disease propagation.* International Journal of STD & AIDS, 1999. **10**: p. 182-185.

96. Doherty, I.A., *The nexus of drugs, sex networks, HIV, and syphilis in young African American MSM: Final progress report*. 2011: Chapel Hill, NC. p. 1-6.

97. Neal, Z.P. and J.W. Neal, *The (in)compatibility of diversity and sense of community.* Am J Community Psychol, 2014. **53**(1-2): p. 1-12.

98. Goodreau, S.M., *Advances in Exponential Random Graph (p*) Models Applied to a Large Social Network.* Soc Networks, 2007. **29**(2): p. 231-248.

99. Pasquier, C., N. Millot, R. Njouom, K. Sandres, M. Cazabat, J. Puel, and J. Izopet, *HIV-1 subtyping using phylogenetic analysis of pol gene sequences.* Journal of Virological Methods, 2001. **94**: p. 45-54.

100. Panel on Antiretroviral Guidelines for Adults and Adolescents. *Guidelines for the Use of Antiretroviral Agents in Adults and Adolescents Living with HIV*. 2016 July 14, 2016 17-March-2018]; Available from: https://aidsinfo.nih.gov/guidelines/html/1/adult-and-adolescent-arv/6/drug-resistance-testing.

101. Pao, D., M. Fisher, S. Hue, G. Dean, G. Murphy, P.A. Cane, . . . D. Pillay, *Transmission of HIV-1 during primary infection: relationship to sexual risk and sexually transmitted infections.* AIDS, 2005. **19**(1): p. 85-90.

102. Volz, E.M., J.S. Koopman, M.J. Ward, A.L. Brown, and S.D. Frost, *Simple epidemiological dynamics explain phylogenetic clustering of HIV from patients with recent infection.* PLoS Comput Biol, 2012. **8**(6): p. e1002552.

103. Wertheim, J.O., A.J. Leigh Brown, N.L. Hepler, S.R. Mehta, D.D. Richman, D.M. Smith, and S.L. Kosakovsky Pond, *The global transmission network of HIV-1.* J Infect Dis, 2014. **209**(2): p. 304-313.

104. Dennis, A.M., D.K. Pasquale, R. Billock, S. Beagle, V.L. Mobley, A.B. Cope, . . . P.A. Leone, *Integration of Contact Tracing and Phylogenetics in an Investigation of Acute HIV Infection.* Sex Transm Dis, 2017.

105. Leventhal, G.E., R. Kouyos, T. Stadler, V. Wyl, S. Yerly, J. Boni, . . . S. Bonhoeffer, *Inferring epidemic contact structure from phylogenetic trees.* PLoS Comput Biol, 2012. **8**(3): p. e1002413.

106. Lewis, F., G.J. Hughes, A. Rambaut, A. Pozniak, and A.J. Leigh Brown, *Episodic sexual transmission of HIV revealed by molecular phylodynamics.* PLoS Med, 2008. **5**(3): p. e50.

107. Leigh Brown, A.J., S.J. Lycett, L. Weinert, G.J. Hughes, E. Fearnhill, and D.T. Dunn, *Transmission network parameters estimated from HIV sequences for a nationwide epidemic.* J Infect Dis, 2011. **204**(9): p. 1463-1469.

108. Oster, A.M., C. Wejnert, L.A. Mena, K. Elmore, H. Fisher, and J.D. Heffelfinger, *Network analysis among HIV-infected young black men who have sex with men demonstrates high connectedness around few venues.* Sex Transm Dis, 2013. **40**(3): p. 206-212.

109. Chen, Y., C. Tseng, C. King, T.J. Wu, and H. Chen, *Incorporating Geographical Contacts into Social Network Analysis for Contact Tracing in Epidemiology: A Study on Taiwan SARS Data.* Advances in Disease Surveillance, 2007. **4**(4).

110. Handcock, M.S. and J.H. Jones, *Likelihood-based inference for stochastic models of sexual network formation.* Theor Popul Biol, 2004. **65**(4): p. 413-422.

111. Moody, J. and R.A. Benton, *Interdependent effects of cohesion and concurrency for epidemic potential.* Ann Epidemiol, 2016. **26**(4): p. 241-248.

112. SAS Institute Inc. 2013. *SAS System for Windows, Version 9.4*. Cary, NC:

113. Monogram Biosciences, *HIV specimen collection guide*, LabCorp Specialty Testing Group, Editor. 2014, Laboratory Corporation of America Holdings. p. 2.

114. Csardi, G. and T. Nepusz, *The igraph software package for complex network research.* InterJournal, 2006. **Complex Systems**: p. 1695.

115. R Core Team. 2015. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.

116. StataCorp. 2017. *Stata Statistical Software: Release 15*. College Station, TX: StataCorp LLC.

117. Khanna, A.S., S.M. Goodreau, P.M. Gorbach, E. Daar, and S.J. Little, *Modeling the Impact of Post-Diagnosis Behavior Change on HIV Prevalence in Southern California Men Who Have Sex with Men (MSM).* AIDS Behav, 2013.

118. Marks, G., N. Crepaz, J.W. Senterfitt, and R.S. Janssen, *Meta-Analysis of High-Risk Sexual Behaviors in Persons Aware and Unaware They are Infected With HIV in the United States: Implications for HIV Prevention Programs.* J Acquir Immune Defic Syndr, 2005. **39**(4): p. 446-453.

119. Marks, G., N. Crepaz, and R.S. Janssen, *Estimating sexual transmission of HIV from persons aware and unaware that they are infected with the virus in the USA.* AIDS, 2006. **20**(10): p. 1447-1450.

120. NC Department of Health and Human Services, *North Carolina 2010 HIV/STD Surveillance Report*. 2010, Division of Public Health: Raleigh, NC.

121. Hassan, A.S., O.G. Pybus, E.J. Sanders, J. Albert, and J. Esbjornsson, *Defining HIV-1 transmission clusters based on sequence data.* AIDS, 2017. **31**(9): p. 1211-1222.

122. Rothenberg, R.B., J.J. Potterat, D.E. Woodhouse, W.W. Darrow, S.Q. Muth, and A.S. Klovdahl, *Choosing a centrality measure: Epidemiologic correlates in the Colorado Springs study of social networks.* Social Networks, 1995. **17**: p. 273-297.

123. Ghosh, R. and K. Lerman, *Parameterized centrality metric for network analysis.* Physical Review E, 2011. **83**(6).

124. Kermarrec, A.-M., E. Le Merrer, B. Sericola, and G. Trédan, *Second order centrality: Distributed assessment of nodes criticity in complex networks.* Computer Communications, 2011. **34**(5): p. 619-628.

125. Hoots, B.E., P.D. MacDonald, L.B. Hightow-Weidman, P.A. Leone, and W.C. Miller, *Developing a predictive model to prioritize human immunodeficiency virus partner notification in North Carolina.* Sex Transm Dis, 2012. **39**(1): p. 65-71.

126. Goodreau, S.M., *Assessing the effects of human mixing patterns on human immunodeficiency virus-1 interhost phylogenetics through social network simulation.* Genetics, 2006. **172**(4): p. 2033-2045.

127. Bonacich, P. and P. Lloyd, *Eigenvector-like measures of centrality for asymmetric relations.* Soc Networks, 2001. **23**: p. 191-201.

128. Bonacich, P., *Some unique properties of eigenvector centrality.* Social Networks, 2007. **29**(4): p. 555-564.

129. Bonacich, P., *Factoring and weighting approaches to status scores and clique identification.* The Journal of Mathematical Sociology, 1972. **2**(1): p. 113-120.

130. Bonacich, P. and P. Lloyd, *Calculating status with negative relations.* Social Networks, 2004. **26**(4): p. 331-338.

131. Juher, D., J. Saldana, R. Kohn, K. Bernstein, and C. Scoglio, *Network-Centric Interventions to Contain the Syphilis Epidemic in San Francisco.* Sci Rep, 2017. **7**(1): p. 6464.

132. Dennis, A.M., I.A. Doherty, and D.K. Pasquale, *TraCS pilot grant project summary*. 2015, UNC-Chapel Hill School of Medicine. p. 7.

133. DHHS Panel on Antiretroviral Guidelines for Adults and Adolescents - A Working Group of the Office of AIDS Research Advisory Council (OARAC), *Guidelines for the Use of Antiretroviral Agents in HIV-1-Infected Adults and Adolescents* 2006, U.S. DHHS.

134. Posada, D., K.A. Crandall, and D.M. Hillis, *PHYLOGENETICS OF HIV*, in *Computational and Evolutionary Analysis of HIV Molecular Sequences*, A.G. Rodrigo and G.H. Learn, Jr, Editors. 2000, Kluwer Academic Publishers: Dordrecht, The Netherlands. p. 121-160.

135. Baldauf, S.L., *Phylogeny for the faint of heart: a tutorial.* Trends in Genetics, 2003. **19**(6): p. 345-351.

136. Douady, C.J., *Comparison of Bayesian and Maximum Likelihood Bootstrap Measures of Phylogenetic Reliability.* Mol Biol Evol, 2003. **20**(2): p. 248-254.

137. Aldous, J.L., S.K. Pond, A. Poon, S. Jain, H. Qin, J.S. Kahn, . . . D.M. Smith, *Characterizing HIV transmission networks across the United States.* Clin Infect Dis, 2012. **55**(8): p. 1135-1143.

138. Kosakovsky Pond, S.L., S. Weaver, A.J. Leigh Brown, and J.O. Wertheim, *HIV-TRACE (Transmission Cluster Engine): a tool for large scale molecular epidemiology of HIV-1 and other rapidly evolving pathogens.* Mol Biol Evol, 2018: p. msy016-msy016.

139. Tamura, K. and M. Nei, *Estimation of the Number of Nucleotide Substitutions in the Control Region of Mitochondrial DNA in Humans and Chimpanzees.* Mol Biol Evol, 1993. **10**(3): p. 512-526.

140. Krivitsky, P.N., *Exponential-family random graph models for valued networks.* Electron J Stat, 2012. **6**: p. 1100-1128.

141. National Center for HIV/AIDS, V.H., STD, and TB Preventio, *HIV in the Southern United States*, in *CDC Issue Brief*. 2016, Centers for Disease Control and Prevention: Atlanta, GA.

142. McKellar, M.S., A.B. Cope, C.L. Gay, K.S. McGee, J.D. Kuruc, M.G. Kerkau, . . . H.I.V.I.C. Duke-Unc Acute, *Acute HIV-1 infection in the Southeastern United States: a cohort study.* AIDS Res Hum Retroviruses, 2013. **29**(1): p. 121-128.

143. Hogben, M., T. McNally, M. McPheeters, and A.B. Hutchinson, *The effectiveness of HIV partner counseling and referral services in increasing identification of HIV-positive individuals a systematic review.* Am J Prev Med, 2007. **33**(2 Suppl): p. S89-100.

144. Grabowski, M.K. and A.D. Redd, *Molecular tools for studying HIV transmission in sexual networks.* Curr Opin HIV AIDS, 2014. **9**(2): p. 126-133.

145. Vasylyeva, T.I., S.R. Friedman, D. Paraskevis, and G. Magiorkinis, *Integrating molecular epidemiology and social network analysis to study infectious diseases: Towards a socio-molecular era for public health.* Infect Genet Evol, 2016.

146. Delva, W., G.E. Leventhal, and S. Helleringer, *Connecting the dots: network data and models in HIV epidemiology.* AIDS, 2016. **30**(13): p. 2009-2020.

147. Peters, P.J., C.L. Gay, S. Beagle, A. Shankar, W.M. Switzer, and L.B. Hightow-Weidman, *HIV Infection Among Partners of HIV-Infected Black Men Who Have Sex with Men — North Carolina, 2011–2013.* MMWR, 2014. **63**(5): p. 90-94.

148. Smith, D.M., S.J. May, S. Tweeten, L. Drumright, M.E. Pacold, S.L. Kosakovsky Pond, . . . S.J. Little, *A public health model for the molecular surveillance of HIV transmission in San Diego, California.* AIDS, 2009. **23**(2): p. 225-232.

149. Wertheim, J.O., S.L. Kosakovsky Pond, L.A. Forgione, S.R. Mehta, B. Murrell, S. Shah, . . . L.V. Torian, *Social and Genetic Networks of HIV-1 Transmission in New York City.* PLoS Pathog, 2017. **13**(1): p. e1006000.

150. Brenner, B.G., M. Roger, D. Stephens, D. Moisi, I. Hardy, J. Weinberg, . . . M.A. Wainberg, *Transmission clustering drives the onward spread of the HIV epidemic among men who have sex with men in Quebec.* J Infect Dis, 2011. **204**(7): p. 1115-1119.

151. Centers for Disease Control and Prevention, *Recommendations for Partner Services Programs for HIV Infection, Syphilis, Gonorrhea, and Chlamydial infection*, in *MMWR*. 2008: Atlanta, GA.

152. Dailey Garnes, N.J., Z.S. Moore, B.L. Cadwell, A.T. Fleischauer, and P. Leone, *Previously undiagnosed HIV infections identified through cluster investigation, North Carolina, 2002-2007.* AIDS Behav, 2015. **19**(4): p. 723-731.

153. Kuruc, J.D., A.B. Cope, L.A. Sampson, C.L. Gay, R.M. Ashby, E.M. Foust, . . . J.J. Eron, *Ten Years of Screening and Testing for Acute HIV Infection in North Carolina.* J Acquir Immune Defic Syndr, 2016. **71**(1): p. 111-119.

154. Edgar, R.C., *MUSCLE: multiple sequence alignment with high accuracy and high throughput.* Nucleic Acids Res, 2004. **32**(5): p. 1792-1797.

155. Hall, T., *BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 98/98/NT.* Nucl Acids Symp Ser, 1999. **41**: p. 95-98.

156. Price, M.N., P.S. Dehal, and A.P. Arkin, *FastTree: Computing Large Minimum Evolution Trees with Profiles instead of a Distance Matrix.* Mol Biol Evol, 2009. **26**(7): p. 1641-1650.

157. Tavaré, S., *Some Probabilistic and Statistical Problems in the Analysis of DNA Sequences.* American Mathematical Society: Lectures on Mathematics in the Life Sciences, 1986. **17**: p. 57-86.

158. Shimodaira, H. and M. Hasegawa, *Multiple Comparisons of Log-Likelihoods with Applications to Phylogenetic Inference.* Mol Biol Evol, 1999. **16**(8): p. 1114-1116.

159. Struck, D., G. Lawyer, A.M. Ternes, J.C. Schmidt, and D.P. Bercoff, *COMET: adaptive context-based modeling for ultrafast HIV-1 subtype identification.* Nucleic Acids Res, 2014. **42**(18): p. e144.

160. Drummond, A.J. and A. Rambaut, *BEAST: Bayesian Evolutionary Analysis by Sampling Trees.* BMC Evolutionary Biology, 2007. **7**: p. 214.

161. Rambaut, A., M. Suchard, D. Xie, and A.J. Drummond. 2014. *Tracer v1.6*.

162. Chan, P.A., J.W. Hogan, A. Huang, A. DeLong, M. Salemi, K.H. Mayer, and R. Kantor, *Phylogenetic Investigation of a Statewide HIV-1 Epidemic Reveals Ongoing and Active Transmission Networks Among Men Who Have Sex With Men.* J Acquir Immune Defic Syndr, 2015. **70**(4): p. 428-435.

163. Amirkhanian, Y.A., *Social networks, sexual networks and HIV risk in men who have sex with men.* Curr HIV/AIDS Rep, 2014. **11**(1): p. 81-92.

164. Callegaro, A., V. Svicher, C. Alteri, A. Lo Presti, D. Valenti, A. Goglio, . . . F. Maggiolo, *Epidemiological network analysis in HIV-1 B infected patients diagnosed in Italy between 2000 and 2008.* Infect Genet Evol, 2011. **11**(3): p. 624-632.

165. Skoura, L., S. Metallidis, A.J. Buckton, J.L. Mbisa, D. Pilalas, E. Papadimitriou, . . . N. Malisiovas, *Molecular and epidemiological characterization of HIV-1 infection networks involving transmitted drug resistance mutations in Northern Greece.* J Antimicrob Chemother, 2011. **66**(12): p. 2831-2837.

166. Ross, L.L., J. Horton, S. Hasan, J.R. Brown, D. Murphy, E. Dejesus, . . . M.S. Shaefer, *HIV-1 Transmission Patterns in Antiretroviral Therapy-Naive, HIV-Infected North Americans Based on Phylogenetic Analysis by Population Level and Ultra-Deep DNA Sequencing.* PLoS One, 2014. **9**(2): p. e89611.

167. Ng, K.T., L.Y. Ong, S.H. Lim, Y. Takebe, A. Kamarulzaman, and K.K. Tee, *Evolutionary history of HIV-1 subtype B and CRF01_AE transmission clusters among men who have sex with men (MSM) in Kuala Lumpur, Malaysia.* PLoS One, 2013. **8**(6): p. e67286.

168. Hue, S., J.P. Clewley, P.A. Cane, and D. Pillay, *HIV-1 pol gene variation is sufficient for reconstruction of transmissions in the era of antiretroviral therapy.* AIDS, 2004. **18**(5): p. 719-728.

169. Antoniadou, Z.A., I. Kousiappa, L. Skoura, D. Pilalas, S. Metallidis, P. Nicolaidis, . . . L.G. Kostrikis, *Short Communication: Molecular Epidemiology of HIV Type 1 Infection in Northern Greece (2009-2010): Evidence of a Transmission Cluster of HIV Type 1 Subtype A1 Drug-Resistant Strains Among Men Who Have Sex with Men.* AIDS Res Hum Retroviruses, 2014. **30**(3): p. 225-232.

170. Buskin, S.E., G.M. Ellis, G.G. Pepper, L.M. Frenkel, S.A. Pergam, G.S. Gottlieb, . . . R.W. Wood, *Transmission Cluster of Multiclass Highly Drug-Resistant HIV-1 Among 9 Men Who Have Sex With Men in Seattle/King County, WA, 2005–2007.* J Acquir Immune Defic Syndr, 2008. **49**(2): p. 205-211.

171. Little, S.J., S.L. Kosakovsky Pond, C.M. Anderson, J.A. Young, J.O. Wertheim, S.R. Mehta, . . . D.M. Smith, *Using HIV networks to inform real time prevention interventions.* PLoS One, 2014. **9**(6): p. e98443.

172. Dennis, A.M., J.T. Herbeck, A.L. Brown, P. Kellam, T. de Oliveira, D. Pillay, . . . M.S. Cohen, *Phylogenetic studies of transmission dynamics in generalized HIV epidemics: an essential tool where the burden is greatest?* J Acquir Immune Defic Syndr, 2014. **67**(2): p. 181-195.

173. Hall, H.I., E.L. Frazier, P. Rhodes, D.R. Holtgrave, C. Furlow-Parmley, T. Tang, . . . J. Skarbinski. *FRLBX05 - Oral Abstract: Continuum of HIV care: differences in care and treatment by sex and race/ethnicity in the United States*. in *AIDS 2012*. 2012. Washington, D.C.

174. Bradley, H., H.I. Hall, R.J. Wolitski, M.M. Van Handel, A.E. Stone, M. LaFlam, . . . L. Valleroy, *Vital Signs: HIV Diagnosis, Care, and Treatment Among Persons Living with HIV — United States, 2011.* MMWR, 2014. **63**(47): p. 1113-1117.

175. Singh, S., H. Bradley, X. Hu, J. Skarbinski, H.I. Hall, and A. Lansky, *Men Living with Diagnosed HIV Who Have Sex with Men: Progress Along the Continuum of HIV Care — United States, 2010.* MMWR, 2014. **63**(38): p. 829-833.

176. Lubelchek, R.J., S.C. Hoehnen, A.L. Hotton, S.L. Kincaid, D.E. Barker, and A.L. French, *Transmission clustering among newly diagnosed HIV patients in Chicago, 2008 to 2011: using phylogenetics to expand knowledge of regional HIV transmission patterns.* J Acquir Immune Defic Syndr, 2015. **68**(1): p. 46-54.

177. Castor, D., A. Low, T. Evering, S. Karmon, B. Davis, A. Figueroa, . . . M. Markowitz, *Transmitted drug resistance and phylogenetic relationships among acute and early HIV-1-infected individuals in New York City.* J Acquir Immune Defic Syndr, 2012. **61**(1): p. 1-8.

178. Brenner, B., M.A. Wainberg, and M. Roger, *Phylogenetic inferences on HIV-1 transmission: implications for the design of prevention and treatment interventions.* AIDS, 2013. **27**(7): p. 1045-1057.

179. Chandler, R.K., S.Y. Kahana, B. Fletcher, D. Jones, M.S. Finger, W.M. Aklin, . . . C. Webb, *Data Collection and Harmonization in HIV Research: The Seek, Test, Treat, and Retain Initiative at the National Institute on Drug Abuse.* Am J Public Health, 2015. **105**(12): p. 2416-2422.

180. UNAIDS, *90-90-90: An ambitious treatment target to help end the AIDS epidemic*. 2014, Joint United Nations Programme on HIV/AIDS. p. 1-40.

181. Gardner, E.M., M.P. McLees, J.F. Steiner, C. Del Rio, and W.J. Burman, *The spectrum of engagement in HIV care and its relevance to test-and-treat strategies for prevention of HIV infection.* Clin Infect Dis, 2011. **52**(6): p. 793-800.

182. Hurt, C.B., S.I. McCoy, J.D. Kuruc, J. Nelson, M. Kerkau, S. Fiscus, . . . J.J. Eron, Jr., *Transmitted Antiretroviral Drug Resistance among Acute and Recent HIV Infections in North Carolina, 1998 to 2007.* Antivir Ther, 2009. **14**(5): p. 673-678.

183. Pasquale, D.K., I.A. Doherty, L.A. Sampson, S. Hue, P.A. Leone, J. Sebastian, . . . A.M. Dennis, *Leveraging Phylogenetics to Understand HIV Transmission and Partner Notification Networks.* J Acquir Immune Defic Syndr, Accepted 2018.

184. Drummond, A.J., M. Suchard, D. Xie, and A. Rambaut, *Bayesian phylogenies with BEAUti and the BEAST 1.7.* Mol Biol Evol, 2012. **29**: p. 1969-1973.

185. Cui, J., *QIC program and model selection in GEE analyses.* The Stata Journal, 2007. **7**(2): p. 209-220.

186. Pan, W., *Akaike's information criterion in generalized estimating equations.* Biometrics, 2001. **57**(1): p. 120-125.

187. StataCorp. 2011. *Stata Statistical Software: Release 12*. [12.1] College Station, TX: StataCorp LP.

188. Moore, Z.S., S. McCoy, J. Kuruc, M. Hilton, and P. Leone, *Number of named partners and number of partners newly diagnosed with HIV infection identified by persons with acute versus established HIV infection.* J Acquir Immune Defic Syndr, 2009. **52**(4): p. 509-513.

189. Campsmith, M.L., P.H. Rhodes, I. Hall, and T.A. Green, *Undiagnosed HIV prevalence among adults and adolescents in the United States at the end of 2006.* J Acquir Immune Defic Syndr, 2010. **53**(5): p. 619-624.

190. Skarbinski, J., E. Rosenberg, G. Paz-Bailey, H.I. Hall, C.E. Rose, A.H. Viall, . . . J.H. Mermin, *Human immunodeficiency virus transmission at each step of the care continuum in the United States.* JAMA Intern Med, 2015. **175**(4): p. 588-596.

191. Hall, H.I., Q. An, T. Tang, R. Song, M. Chen, T. Green, and J. Kang, *Prevalence of Diagnosed and Undiagnosed HIV Infection — United States, 2008–2012.* MMWR, 2015. **64**(24): p. 657-662.

192. and, C.f.D.C. and Prevention, *Monitoring selected national HIV prevention and care objectives by using HIV surveillance data—United States and 6 dependent areas—*

2013, in *HIV Surveillance Supplemental Report*. 2015, National Center for HIV/AIDS, Viral Hepatitis, STD, and TB Prevention,: Atlanta, GA. p. 70.

193. Miller, W.C., C.R. Lesko, and K.A. Powers, *The HIV care cascade: simple concept, complex realization.* Sex Transm Dis, 2014. **41**(1): p. 41-42.

194. Powers, K.A. and W.C. Miller, *Critical Review: Building on the HIV Cascade: A Complementary "HIV States and Transitions" Framework for Describing HIV Diagnosis, Care, and Treatment at the Population Level.* J Acquir Immune Defic Syndr, 2015. **69**(3): p. 341-347.

195. Buskin, S.E., J.B. Kent, J.C. Dombrowski, and M.R. Golden, *Migration distorts surveillance estimates of engagement in care: results of public health investigations of persons who appear to be out of HIV care.* Sex Transm Dis, 2014. **41**(1): p. 35-40.

196. Cohen, M.S., Y.Q. Chen, M. McCauley, T. Gamble, M.C. Hosseinipour, N. Kumarasamy, . . . T.R. Fleming, *Prevention of HIV-1 infection with early antiretroviral therapy.* N Engl J Med, 2011. **365**(6): p. 493-505.

197. Rodger, A.J., V. Cambiano, T. Bruun, P. Vernazza, S. Collins, J. van Lunzen, . . . P.S. Group, *Sexual Activity Without Condoms and Risk of HIV Transmission in Serodifferent Couples When the HIV-Positive Partner Is Using Suppressive Antiretroviral Therapy.* JAMA, 2016. **316**(2): p. 171-181.

198. Danon, L., A.P. Ford, T. House, C.P. Jewell, M.J. Keeling, G.O. Roberts, . . . M.C. Vernon, *Networks and the epidemiology of infectious disease.* Interdiscip Perspect Infect Dis, 2011. **2011**: p. 284909.

199. Garnett, G.P., J.P. Hughes, R.M. Anderson, B.P. Stoner, S.O. Aral, W.L. Whittington, . . . K.K. Holmes, *Sexual mixing patterns of patients attending sexually transmitted diseases clinics.* Sex Transm Dis, 1996. **23**(3): p. 248-257.

200. Rothenberg, R. and S.Q. Muth, *Large-network concepts and small-network characteristics: fixed and variable factors.* Sex Transm Dis, 2007. **34**(8): p. 604-612.

201. USDA Economic Research Service. *Rural-Urban Continuum Codes*. 2012 5-July-2012 [cited 2013 10-Feb-2013]; Available from: http://www.ers.usda.gov/data-products/rural-urban-continuum-codes.aspx.

202. U. S. Census Bureau Population Division. *American FactFinder, North Carolina, 2012*. Annual Estimates of the Resident Population: April 1, 2010 to July 1, 2012: 2012 Population Estimates 2012 [cited 2017 31-Dec-2017]; Available from: https://factfinder.census.gov/faces/tableservices/jsf/pages/productview.xhtml?src=bkmk.

203. NC Department of Health and Human Services, *2014 North Carolina HIV/STD Surveillance Report*. 2015, NC Division of Public Health: Raleigh, NC. p. 98.

204. Dziak, J.J., D.L. Coffman, S.T. Lanza, and R. Li, *Sensitivity and Specificity of Information Criteria*, in *Technical Report Series*. 2012, The Methodology Center, The Pennsylvania State University: State College, PA. p. 31.

205. Fox, J., P.J. White, N. Macdonald, J. Weber, M. McClure, S. Fidler, and H. Ward, *Reductions in HIV transmission risk behaviour following diagnosis of primary HIV infection: a cohort of high-risk men who have sex with men.* HIV Med, 2009. **10**(7): p. 432-438.

206. Crepaz, N., C.M. Lyles, R.J. Wolitski, W.F. Passin, S.M. Rama, J.H. Herbst, . . . R. Stall, *Do prevention interventions reduce HIV risk behaviours among people living with HIV? A meta-analytic review of controlled trials.* AIDS, 2006. **20**(2): p. 143-157.

207. McKinstry, L.A., A. Zerbe, B. Hanscom, J. Farrior, A.E. Kurth, J. Stanton, . . . W.M. El-Sadr, *A Randomized-Controlled Trial of Computer-based Prevention Counseling for HIV-Positive Persons (HPTN 065).* J AIDS Clin Res, 2017. **8**(7).

208. Metsch, L.R., D.J. Feaster, L. Gooden, B.R. Schackman, T. Matheson, M. Das, . . . G.N. Colfax, *Effect of risk-reduction counseling with rapid HIV testing on risk of acquiring sexually transmitted infections: the AWARE randomized clinical trial.* JAMA, 2013. **310**(16): p. 1701-1710.

209. Doherty, I.A., D.K. Pasquale, A.K. Liebman, A.A. Adimora, and P.A. Leone, *Sexual Networks of Syphilis and HIV Among Heterosexuals, MSM, and Bisexual Men in North Carolina*, in *19th Annual International Society for Sexually Transmitted Diseases Research Meeting*. 2011: Quebec City, Quebec, Canada.

210. Fujimoto, K., C.A. Flash, L.M. Kuhns, J.Y. Kim, and J.A. Schneider, *Social networks as drivers of syphilis and HIV infection among young men who have sex with men.* Sex Transm Infect, 2018.

211. Almquist, Z.W., *Random errors in egocentric networks.* Soc Networks, 2012. **34**(4): p. 493-505.

212. Smith, J.A. and J. Moody, *Structural Effects of Network Sampling Coverage I: Nodes Missing at Random.* Soc Networks, 2013. **35**(4).

213. Kitsak, M., L.K. Gallos, S. Havlin, F. Liljeros, L. Muchnik, H.E. Stanley, and H.A. Makse, *Identification of influential spreaders in complex networks.* Nature Physics, 2010. **6**(11): p. 888-893.

214. Castellano, C. and R. Pastor-Satorras, *Competing activation mechanisms in epidemics on networks.* Sci Rep, 2012. **2**: p. 371.

215. Cope, A.B., A.M. Crooks, T. Chin, J.D. Kuruc, K.S. McGee, J.J. Eron, . . . C.L. Gay, *Incident sexually transmitted infection as a biomarker for high-risk sexual behavior after diagnosis of acute HIV.* Sex Transm Dis, 2014. **41**(7): p. 447-452.

216. Mobley, V.L. and E. Foust, *Modernization of North Carolina's HIV control measures*. 2018, NC Department of Health and Human Services,: Raleigh, NC.

217. Bullo, M., E. Samoff, N. Dzialowy Adams, B. Wheeler, K.A. Powers, and A.B. Cope. *HIV Viral Suppression Assessed by Two Metrics in North Carolina during 2016*. in *CROI*. 2018. Boston, MA.

218. Mayer, K.H. and K.K. Venkatesh, *Interactions of HIV, other sexually transmitted diseases, and genital tract inflammation facilitating local pathogen transmission and acquisition.* Am J Reprod Immunol, 2011. **65**(3): p. 308-316.

219. Vallabhaneni, S., J.J. McConnell, L. Loeb, W. Hartogensis, F.M. Hecht, R.M. Grant, and C.D. Pilcher, *Changes in seroadaptive practices from before to after diagnosis of recent HIV infection among men who have sex with men.* PLoS One, 2013. **8**(2): p. e55397.

220. Helleringer, S. and H.P. Kohler, *Sexual network structure and the spread of HIV in Africa: evidence from Likoma Island, Malawi.* AIDS, 2007. **21**(17): p. 2323-2332.

221. Li, J.Z., R. Paredes, H.J. Ribaudo, M.J. Kozal, E.S. Svarovskaia, J.A. Johnson, . . . D.R. Kuritzkes, *Impact of minority nonnucleoside reverse transcriptase inhibitor resistance mutations on resistance genotype after virologic failure.* J Infect Dis, 2013. **207**(6): p. 893-897.

222. Metzner, K., *The significance of minority drug-resistant quasispecies*, in *Antiretroviral Resistance in Clinical Practice*, A.M. Geretti, Editor. 2006, Mediscript: London.

223. Metzner, K.J., S.G. Giulieri, S.A. Knoepfel, P. Rauch, P. Burgisser, S. Yerly, . . . M. Cavassini, *Minority quasispecies of drug-resistant HIV-1 that lead to early therapy failure in treatment-naive and -adherent patients.* Clin Infect Dis, 2009. **48**(2): p. 239-247.

224. Lataillade, M., J. Chiarella, R. Yang, S. Schnittman, V. Wirtz, J. Uy, . . . M. Kozal, *Prevalence and Clinical Significance of HIV Drug Resistance Mutations by Ultra-Deep Sequencing in Antiretroviral-Naı¨ve Subjects in the CASTLE Study.* PLoS One, 2010. **5**(6): p. e10952.

225. Balaji, A.B., A.M. Oster, A.H. Viall, J.D. Heffelfinger, L.A. Mena, and C.A. Toledo, *Role flexing: how community, religion, and family shape the experiences of young black men who have sex with men.* AIDS Patient Care STDS, 2012. **26**(12): p. 730-737.

226. Maulsby, C., G. Millett, K. Lindsey, R. Kelley, K. Johnson, D. Montoya, and D. Holtgrave, *HIV among Black men who have sex with men (MSM) in the United States: a review of the literature.* AIDS Behav, 2014. **18**(1): p. 10-25.

227. Clarke, T.C., T. Norris, and J.S. Schiller, *Early Release of Selected Estimates Based on Data From the 2016 National Health Interview Survey*, in *National Health Interview Survey Early Release Program*, C.f.D.C.a. Prevention, Editor. 2017, Division of Health Interview Statistics, National Center for Health Statistics.

228. Bertolli, J., P. Morse Garland, E.E. Valverde, L. Beer, J.L. Fagan, C. Hart, and Never in Care Pilot Project Team, *Missed Connections: HIV-Infected People Never in Care.* Public Health Reports, 2013. **128**: p. 117-126.

229. Sweeney, P., L.I. Gardner, K. Buchacz, P. Morse Garland, M.J. Mugavero, J.T. Bosshart, . . . J. Bertolli, *Shifting the Paradigm: Using HIV Surveillance Data as a Foundation for Improving HIV Care and Preventing HIV Infection.* The Milbank Quaterly, 2013. **91**(3): p. 558-603.

230. Gray, R.R., A.J. Tatem, S. Lamers, W. Hou, O. Laeyendecker, D. Serwadda, . . . M. Salemi, *Spatial phylodynamics of HIV-1 epidemic emergence in east Africa.* AIDS, 2009. **23**(14): p. F9-F17.

231. Faria, N.R., I. Hodges-Mameletzis, J.C. Silva, B. Rodes, S. Erasmus, S. Paolucci, . . . P. Lemey, *Phylogeographical footprint of colonial history in the global dispersal of human immunodeficiency virus type 2 group A.* J Gen Virol, 2012. **93**(Pt 4): p. 889-899.

232. Lai, A., M. Ciccozzi, M. Franzetti, F.R. Simonetti, G. Bozzi, F. Binda, . . . G. Zehender, *Local and global spatio-temporal dynamics of HIV-1 subtype F1.* J Med Virol, 2014. **86**(2): p. 186-192.

233. Afonso, J.M., M.G. Morgado, and G. Bello, *Evidence of multiple introductions of HIV-1 subtype C in Angola.* Infect Genet Evol, 2012. **12**(7): p. 1458-1465.

234. Afonso, J.M., G. Bello, M.L. Guimaraes, M. Sojka, and M.G. Morgado, *HIV-1 genetic diversity and transmitted drug resistance mutations among patients from the North, Central and South regions of Angola.* PLoS One, 2012. **7**(8): p. e42996.

235. U.S. Census Bureau. *State and County QuickFacts: Wake County, NC*. 2015 05-February-2015 [cited 2015 21-March-2015]; Data derived from Population Estimates,

American Community Survey, Census of Population and Housing, State and County Housing Unit Estimates, County Business Patterns, Nonemployer Statistics, Economic Census, Survey of Business Owners, Building Permits]. Available from: http://quickfacts.census.gov/qfd/states/37/37183.html.

236. Shattock, R.J. and J.P. Moore, *Inhibiting sexual transmission of HIV-1 infection.* Nat Rev Microbiol, 2003. **1**(1): p. 25-34.

237. Machado, J.R., M.V. da Silva, C.L. Cavellani, M.A. dos Reis, M.L.G. dos Reis Monteiro, V. de Paula Antunes Teixeira, and R.R.M. Corrêa, *Mucosal Immunity in the Female Genital Tract, HIV/AIDS.* Biomed Res Int, 2014. **2014**: p. 20.

238. Rodriguez-Garcia, M., M.V. Patel, and C.R. Wira, *Innate and adaptive anti-HIV immune responses in the female reproductive tract.* J Reprod Immunol, 2013. **97**(1): p. 74-84.

239. Carias, A.M., S. McCoombe, M. McRaven, M. Anderson, N. Galloway, N. Vandergrift, . . . T.J. Hope, *Defining the Interaction of HIV-1 with the Mucosal Barriers of the Female Reproductive Tract.* J Virol, 2013. **87**(21): p. 11388-11400.

240. Bobardt, M.D., U. Chatterji, S. Selvarajah, B. Van der Schueren, G. David, B. Kahn, and P.A. Gallay, *Cell-free human immunodeficiency virus type 1 transcytosis through primary genital epithelial cells.* J Virol, 2007. **81**(1): p. 395-405.

241. Shen, R., E.R. Drelichman, D. Bimczok, C. Ochsenbauer, J.C. Kappes, J.A. Cannon, . . . P.D. Smith, *GP41-specific antibody blocks cell-free HIV type 1 transcytosis through human rectal mucosa and model colonic epithelium.* J Immunol, 2010. **184**(7): p. 3648-3655.

242. Nunes, R., B. Sarmento, and J. das Neves, *Formulation and delivery of anti-HIV rectal microbicides: Advances and challenges.* Journal of Controlled Release, 2014. **194**: p. 278-294.

243. Keele, B.F. and J.D. Estes, *Barriers to mucosal transmission of immunodeficiency viruses.* Blood, 2011. **118**(4): p. 839-846.

244. Hickey, D.K., M.V. Patel, J.V. Fahey, and C.R. Wira, *Innate and adaptive immunity at mucosal surfaces of the female reproductive tract: stratification and integration of immune protection against the transmission of sexually transmitted infections.* J Reprod Immunol, 2011. **88**(2): p. 185-194.

245. Kaushic, C., *The role of the local microenvironment in regulating susceptibility and immune responses to sexually transmitted viruses in the female genital tract.* J Reprod Immunol, 2009. **83**(1-2): p. 168-172.

246. Nazli, A., O. Chan, W.N. Dobson-Belaire, M. Ouellet, M.J. Tremblay, S.D. Gray-Owen, . . . C. Kaushic, *Exposure to HIV-1 directly impairs mucosal epithelial barrier integrity allowing microbial translocation.* PLoS Pathog, 2010. **6**(4): p. e1000852.

247. Southern, P.J., *Missing out on the biology of heterosexual HIV-1 transmission.* Trends Microbiol, 2013. **21**(5): p. 245-252.

248. Muwanga, F., *HIV and STDs: how are they linked?* Afr Health, 1995. **17**(3): p. 40.

249. Augenbraun, M.H. and W.M. McCormack, *Sexually transmitted diseases in HIV-infected persons.* Infect Dis Clin North Am, 1994. **8**(2): p. 439-448.

250. Gadkari, D.A., T.C. Quinn, R.R. Gangakhedkar, S.M. Mehendale, A.D. Divekar, A.R. Risbud, . . . R.C. Bollinger, *HIV-1 DNA shedding in genital ulcers and its associated risk factors in Pune, India.* J Acquir Immune Defic Syndr Hum Retrovirol, 1998. **18**(3): p. 277-281.

251. McClelland, R.S., L. Lavreys, C. Katingima, J. Overbaugh, V. Chohan, K. Mandaliya, . . . J.M. Baeten, *Contribution of HIV-1 infection to acquisition of sexually transmitted disease: a 10-year prospective study.* J Infect Dis, 2005. **191**(3): p. 333-338.

252. Muiru, A.N., B.L. Guthrie, R. Bosire, M. Merkel, A.Y. Liu, R.Y. Choi, . . . C. Farquhar, *Incident HSV-2 infections are common among HIV-1-discordant couples.* J Infect Dis, 2013. **208**(7): p. 1093-1101.

253. Phiri, S., S. Zadrozny, H.A. Weiss, F. Martinson, N. Nyirenda, C.Y. Chen, . . . I.F. Hoffman, *Etiology of genital ulcer disease and association with HIV infection in Malawi.* Sex Transm Dis, 2013. **40**(12): p. 923-928.

254. Chen, C.Y., R.C. Ballard, C.M. Beck-Sague, Y. Dangor, F. Radebe, S. Schmid, . . . S.A. Morse, *Human immunodeficiency virus infection and genital ulcer disease in South Africa: the herpetic connection.* Sex Transm Dis, 2000. **27**(1): p. 21-29.

255. Fox, J., P.J. White, J. Weber, G.P. Garnett, H. Ward, and S. Fidler, *Quantifying sexual exposure to HIV within an HIV-serodiscordant relationship: development of an algorithm.* AIDS, 2011. **25**(8): p. 1065-1082.

256. Sanchez, J., J.R. Lama, J. Peinado, A. Paredes, A. Lucchetti, K. Russell, . . . J.L. Sebastian, *High HIV and ulcerative sexually transmitted infection incidence estimates among men who have sex with men in Peru: awaiting for an effective preventive intervention.* J Acquir Immune Defic Syndr, 2009. **51 Suppl 1**: p. S47-51.

257. Cachay, E.R., S.D. Frost, A.F. Poon, D. Looney, S.M. Rostami, M.E. Pacold, . . . D.M. Smith, *Herpes simplex virus type 2 acquisition during recent HIV infection does not influence plasma HIV levels.* J Acquir Immune Defic Syndr, 2008. **47**(5): p. 592-596.

258. Cachay, E.R., S.D. Frost, D.D. Richman, D.M. Smith, and S.J. Little, *Herpes simplex virus type 2 infection does not influence viral dynamics during early HIV-1 infection.* J Infect Dis, 2007. **195**(9): p. 1270-1277.

259. Lawrence, P., D. Portran, R. Terrasse, S. Palle, T. Olivier, J. Fantini, . . . O. Delezay, *Selective transmigration of monocyte-associated HIV-1 across a human cervical monolayer and its modulation by seminal plasma.* AIDS, 2012. **26**(7): p. 785-796.

260. Keele, B.F., E.E. Giorgi, J.F. Salazar-Gonzalez, J.M. Decker, K.T. Pham, M.G. Salazar, . . . G.M. Shaw, *Identification and characterization of transmitted and early founder virus envelopes in primary HIV-1 infection.* Proc Natl Acad Sci U S A, 2008. **105**(21): p. 7552-7557.

261. Meng, G., X. Wei, X. Wu, M.T. Sellers, J.M. Decker, Z. Moldoveanu, . . . P.D. Smith, *Primary intestinal epithelial cells selectively transfer R5 HIV-1 to CCR5+ cells.* Nature Medicine, 2002. **8**(2): p. 150-156.

262. Smith, P.D., G. Meng, J.F. Salazar-Gonzalez, and G.M. Shaw, *Macrophage HIV-1 infection and the gastrointestinal tract reservoir.* J Leukoc Biol, 2003. **74**(5): p. 642-649.

263. Davenport, M.P., J.J. Zaunders, M.D. Hazenberg, H. Schuitemaker, and R.P. van Rij, *Cell turnover and cell tropism in HIV-1 infection.* Trends Microbiol, 2002. **10**(6): p. 275-278.

264. Powers, K.A., A.C. Ghani, W.C. Miller, I.F. Hoffman, A.E. Pettifor, G. Kamanga, . . . M.S. Cohen, *The role of acute and early HIV infection in the spread of HIV and implications for transmission prevention strategies in Lilongwe, Malawi: a modelling study.* The Lancet, 2011. **378**(9787): p. 256-268.

265. Yerly, S., S. Vora, P. Rizzardi, J.-P. Chave, P.L. Vernazza, M. Flepp, . . . H.I.V.C.S. Swiss, *Acute HIV infection: impact on the spread of HIV and transmission of drug resistance.* AIDS, 2001. **15**(17): p. 2287-2292.

266. Miller, W.C., N.E. Rosenberg, S.E. Rutstein, and K.A. Powers, *Role of acute and early HIV infection in the sexual transmission of HIV.* Curr Opin HIV AIDS, 2010. **5**(4): p. 277-282.

267. Yerly, S., T. Junier, A. Gayet-Ageron, E.B. Amari, V. von Wyl, H.F. Gunthard, . . . H.I.V.C.S. Swiss, *The impact of transmission clusters on primary drug resistance in newly diagnosed HIV-1 infection.* AIDS, 2009. **23**(11): p. 1415-1423.

268. Brown, A.E., R.J. Gifford, J.P. Clewley, C. Kucherer, B. Masquelier, K. Porter, . . . D. Pillay, *Phylogenetic reconstruction of transmission events from individuals with acute HIV infection: toward more-rigorous epidemiological definitions.* J Infect Dis, 2009. **199**(3): p. 427-431.

269. Butler, D.M., D.M. Smith, E.R. Cachay, G.K. Hightower, C.T. Nugent, D.D. Richman, and S.J. Little, *Herpes simplex virus 2 serostatus and viral loads of HIV-1 in blood and semen as risk factors for HIV transmission among men who have sex with men.* AIDS, 2008. **22**(13): p. 1667-1671.

270. Lythgoe, K.A. and C. Fraser, *New insights into the evolutionary rate of HIV-1 at the within-host and epidemiological levels.* Proc Biol Sci, 2012. **279**(1741): p. 3367-3375.

271. Holmes, E.C., *The Evolution and Emergence of RNA Viruses*. Oxford Series in Ecology and Evolution, ed. P.H. Harvey and R.M. May. 2009, Croydon, UK: Oxford University Press. 254.

272. Ghosn, J., I. Pellegrin, C. Goujard, C. Deveau, J.-P. Viard, J. Galimand, . . . M.-L. Chaix, *HIV-1 resistant strains acquired at the time of primary infection massively fuel the cellular reservoir and persist for lengthy periods of time.* AIDS, 2006. **20**(2): p. 159-170.

273. Redd, A.D., A.N. Collinson-Streng, N. Chatziandreou, C.E. Mullis, O. Laeyendecker, C. Martens, . . . T.C. Quinn, *Previously transmitted HIV-1 strains are preferentially selected during subsequent sexual transmissions.* J Infect Dis, 2012. **206**(9): p. 1433-1442.

274. Alizon, S. and C. Fraser, *Within-host and between-host evolutionary rates across HIV-1 genome.* Retrovirology, 2013. **10**(49): p. 10.

275. Maldarelli, F., M. Kearney, S. Palmer, R. Stephens, J. Mican, M.A. Polis, . . . J.M. Coffin, *HIV populations are large and accumulate high genetic diversity in a nonlinear fashion.* J Virol, 2013. **87**(18): p. 10313-10323.

276. Dennis, A., D. Burns, S.H. Eshleman, and M.S. Cohen. *Phylogeny Studies in HIV Prevention Research*. in *HIV Prevention Trials Network*. 2012. Imperial College, London.

277. Li, J.Z., R. Paredes, H.J. Ribaudo, E.S. Svarovskaia, K.J. Metzner, M.J. Kozal, . . . D.R. Kuritzkes, *Low-frequency HIV-1 drug resistance mutations and risk of NNRTI-based antiretroviral treatment failure: a systematic review and pooled analysis.* JAMA, 2011. **305**(13): p. 1327-1335.

278. Metzner, K.J., P. Rauch, H. Walter, C. Boesecke, B. Zollner, H. Jessen, . . . H.J. Stellbrink, *Detection of minor populations of drug-resistant HIV-1 in acute seroconverters.* AIDS, 2005. **19**(16): p. 1819-1825.

279. Salomon, H., M.A. Wainberg, B. Brenner, Y. Quan, D. Rouleau, P. Cote, . . . Investigators of the Quebec Primary Infection Study, *Prevalence of HIV-1 resistant to antiretroviral drugs in 81 individuals newly infected by sexual contact or injecting drug use.* AIDS, 2000. **14**(2): p. F17-F23.

280. Duwe, S., M. Brunn, D. Altmann, O. Hamouda, B. Schmidt, H. Walter, . . . C. Kucherer, *Frequency of Genotypic and Phenotypic Drug-Resistant HIV-1 Among Therapy-Naive Patients of the German Seroconverter Study.* J Acquir Immune Defic Syndr, 2001. **26**: p. 266-273.

281. Brenner, B., M.A. Wainberg, H. Salomon, D. Rouleau, A. Dascal, B. Spira, . . . Investigators of the Quebec Primary Infection Study, *Resistance to antiretroviral drugs in patients with primary HIV-1 infection.* Int J Antimicrob Agents, 2000. **16**: p. 429-434.

282. Ammaranond, P., P. Cunningham, R. Oelrichs, K. Suzuki, C. Harris, L. Leas, . . . A.D. Kelleher, *Rates of transmission of antiretroviral drug resistant strains of HIV-1.* Journal of Clinical Virology, 2003. **26**(2): p. 153-161.

283. Brenner, B.G., J.P. Routy, M. Petrella, D. Moisi, M. Oliveira, M. Detorio, . . . M.A. Wainberg, *Persistence and Fitness of Multidrug-Resistant Human Immunodeficiency Virus Type 1 Acquired in Primary Infection.* J Virol, 2002. **76**(4): p. 1753-1761.

284. Butler, D.M., W. Delport, S.L. Kosakovsky Pond, M.K. Lakdawala, P.M. Cheng, S.J. Little, . . . D.M. Smith, *The origins of sexually transmitted HIV among men who have sex with men.* Sci Transl Med, 2010. **2**(18): p. 18re11.

285. Saidi, H., G. Magri, N. Nasreddine, M. Requena, and L. Belec, *R5- and X4-HIV-1 use differentially the endometrial epithelial cells HEC-1A to ensure their own spread: implication for mechanisms of sexual transmission.* Virology, 2007. **358**(1): p. 55-68.

286. Mild, M., R.R. Gray, A. Kvist, P. Lemey, M.M. Goodenow, E.M. Fenyo, . . . P. Medstrand, *High intrapatient HIV-1 evolutionary rate is associated with CCR5-to-CXCR4 coreceptor switch.* Infect Genet Evol, 2013. **19**: p. 369-377.

287. Mild, M., A. Kvist, J. Esbjornsson, I. Karlsson, E.M. Fenyo, and P. Medstrand, *Differences in molecular evolution between switch (R5 to R5X4/X4-tropic) and non-switch (R5-tropic only) HIV-1 populations during infection.* Infect Genet Evol, 2010. **10**(3): p. 356-364.

288. Poon, A.F., L.C. Swenson, E.M. Bunnik, D. Edo-Matas, H. Schuitemaker, A.B. van 't Wout, and P.R. Harrigan, *Reconstructing the dynamics of HIV evolution within hosts from serial deep sequence data.* PLoS Comput Biol, 2012. **8**(11): p. e1002753.

289. Terrasse, R., M. Memmi, S. Palle, L. Heyndrickx, G. Vanham, B. Pozzetto, and T. Bourlet, *Visualization of X4- and R5-Tropic HIV-1 Viruses Expressing Fluorescent Proteins in Human Endometrial Cells: Application to Tropism Study.* PLoS One, 2017. **12**(1): p. e0169453.

290. Herbeck, J.T., M. Rolland, Y. Liu, S. McLaughlin, J. McNevin, H. Zhao, . . . J.I. Mullins, *Demographic processes affect HIV-1 evolution in primary infection before the onset of selective processes.* J Virol, 2011. **85**(15): p. 7523-7534.

291. Fischer, W., V.V. Ganusov, E.E. Giorgi, P.T. Hraber, B.F. Keele, T. Leitner, . . . B.T. Korber, *Transmission of single HIV-1 genomes and dynamics of early immune escape revealed by ultra-deep sequencing.* PLoS One, 2010. **5**(8): p. e12303.

292. Li, H., K.J. Bar, S. Wang, J.M. Decker, Y. Chen, C. Sun, . . . G.M. Shaw, *High Multiplicity Infection by HIV-1 in Men Who Have Sex with Men.* PLoS Pathog, 2010. **6**(5): p. e1000890.

293. Cohen, M.S., G.M. Shaw, A.J. McMichael, and B.F. Haynes, *Acute HIV-1 Infection.* N Engl J Med, 2011. **364**(20): p. 1943-1954.

294. Bar, K.J., H. Li, A. Chamberland, C. Tremblay, J.P. Routy, T. Grayson, . . . G.M. Shaw, *Wide variation in the multiplicity of HIV-1 infection among injection drug users.* J Virol, 2010. **84**(12): p. 6241-6247.

295. Haaland, R.E., P.A. Hawkins, J. Salazar-Gonzalez, A. Johnson, A. Tichacek, E. Karita, . . . E. Hunter, *Inflammatory genital infections mitigate a severe genetic bottleneck in heterosexual transmission of subtype A and C HIV-1.* PLoS Pathog, 2009. **5**(1): p. e1000274.

296. Alteri, C., V. Svicher, C. Gori, R. D'Arrigo, M. Ciccozzi, F. Ceccherini-Silberstein, . . . Sendih Study Group, *Characterization of the patterns of drug-resistance mutations in newly diagnosed HIV-1 infected patients naive to the antiretroviral drugs.* BMC Infect Dis, 2009. **9**: p. 111.

297. Lee, H.Y., E.E. Giorgi, B.F. Keele, B. Gaschen, G.S. Athreya, J.F. Salazar-Gonzalez, . . . A.S. Perelson, *Modeling sequence evolution in acute HIV-1 infection.* J Theor Biol, 2009. **261**(2): p. 341-360.

298. Keele, B.F., *Identifying and characterizing recently transmitted viruses.* Curr Opin HIV AIDS, 2010. **5**(4): p. 327-334.

299. Bezemer, D., A. de Ronde, M. Prins, K. Porter, R. Gifford, D. Pillay, . . . L. van der Hoek, *Evolution of transmitted HIV-1 with drug-resistance mutations in the absence of therapy: effects on CD4+ T-cell count and HIV-1 RNA load.* Antivir Ther, 2006. **11**(2): p. 173-178.

300. Jain, V., M.C. Sucupira, P. Bacchetti, W. Hartogensis, R.S. Diaz, E.G. Kallas, . . . F.M. Hecht, *Differential persistence of transmitted HIV-1 drug resistance mutation classes.* J Infect Dis, 2011. **203**(8): p. 1174-1181.

301. Perno, C.F., A. Cenci, C. Piro, R. D'Arrigo, L. Marcon, F. Ceccherini-Silberstein, . . . A. Antinori, *HIV fitness and resistance as covariates associated with the appearance of mutations under antiretroviral treatment.* Scand J Infect Dis Suppl, 2003. **106**: p. 37-40.

302. Wagner, B.G., J.G. Garcia-Lerma, and S. Blower, *Factors limiting the transmission of HIV mutations conferring drug resistance: fitness costs and genetic bottlenecks.* Sci Rep, 2012. **2**: p. 320.

303. Richman, D.D., S.J. Little, D.M. Smith, T. Wrin, C. Petropoulos, and J.K. Wong, *HIV evolution and escape.* Trans Am Clin Climatol Assoc, 2004. **115**: p. 289-303.

304. Hirsch, M.S., H.F. Gunthard, J.M. Schapiro, F. Brun-Vezinet, B. Clotet, S.M. Hammer, . . . D.D. Richman, *Antiretroviral drug resistance testing in adult HIV-1 infection: 2008 recommendations of an International AIDS Society-USA panel.* Clin Infect Dis, 2008. **47**(2): p. 266-285.

305. Nijhuis, M., S.G. Deeks, and C.A.B. Boucher, *Implications of antiretroviral resistance on viral fitness.* Curr Opin Infect Dis, 2001. **14**: p. 23-28.

306. Deeks, S.G., *Treatment of antiretroviral-drug-resistant HIV-1 infection.* The Lancet, 2003. **362**(9400): p. 2002-2011.

307. Woodman, Z. and C. Williamson, *HIV molecular epidemiology: transmission and adaptation to human populations.* Curr Opin HIV AIDS, 2009. **4**(4): p. 247-252.

308. Grant, R.M., F.M. Hecht, M. Warmerdam, L. Liu, T. Liegler, C.J. Petropoulos, . . . J.O. Kahn, *Time Trends in Primary HIV-1 Drug Resistance Among Recently Infected Persons.* JAMA, 2002. **288**(2): p. 181-188.

309. Johnson, J.A., J.-F. Li, X. Wei, J. Lipscomb, D. Irlbeck, C. Craig, . . . W. Heneine, *Minority HIV-1 Drug Resistance Mutations Are Present in Antiretroviral Treatment–Naı¨ve Populations and Associate with Reduced Treatment Efficacy.* PLoS Medicine, 2008. **5**(7): p. 1112-1122.

310. Kuritzkes, D.R., *HIV drug resistance: New insight and updated practices.* The PRN Notebook, 2004. **9**(3): p. 9-13.

311. Oette, M., R. Kaiser, M. Daumer, R. Petch, G. Fatkenheuer, H. Carls, . . . R.S. Team, *Primary HIV Drug Resistance and Efficacy of First-Line Antiretroviral Therapy Guided by Resistance Testing.* J Acquir Immune Defic Syndr, 2006. **41**(5): p. 573-581.

312. Tang, J.W. and D. Pillay, *Transmission of HIV-1 drug resistance.* J Clin Virol, 2004. **30**(1): p. 1-10.

313. Frentz, D., D.A.M.C. van de Vijver, C.A.B. Boucher, and J. Albert, *Estimates of HIV Transmitted Drug Resistance Can be Inflated Due to Natural Sequence Polymorphisms.* J Acquir Immune Defic Syndr, 2011. **58**(5): p. e135-e137.

314. Shafer, R.W., S.-Y. Rhee, D. Pillay, V. Miller, P. Sandstrom, J.M. Schapiro, . . . D.E. Bennett, *HIV-1 protease and reverse transcriptase mutations for drug resistance surveillance.* AIDS, 2007. **21**(2): p. 215-223.

315. Bennett, D.E., R.J. Camacho, D. Otelea, D.R. Kuritzkes, H. Fleury, M. Kiuchi, . . . R.W. Shafer, *Drug resistance mutations for surveillance of transmitted HIV-1 drug-resistance: 2009 update.* PLoS One, 2009. **4**(3): p. e4724.

316. Blanco, J.L., V. Varghese, S.Y. Rhee, J.M. Gatell, and R.W. Shafer, *HIV-1 integrase inhibitor resistance and its clinical implications.* J Infect Dis, 2011. **203**(9): p. 1204-1214.

317. Re, M.C., P. Monari, I. Bon, M. Borderi, F. Chiodo, and H.I.V.E.R.I.D.S. Group, *Conflicting interpretations of the prevalence of mutations associated with drug resistance in antiviral naive HIV-1 patients with acute and chronic infection.* Int J Antimicrob Agents, 2004. **23**(2): p. 164-168.

318. Little, S.J., S.D. Frost, J.K. Wong, D.M. Smith, S.L. Pond, C.C. Ignacio, . . . D.D. Richman, *Persistence of transmitted drug resistance among subjects with primary human immunodeficiency virus infection.* J Virol, 2008. **82**(11): p. 5510-5518.

319. Van de Vijver, D.A.M.C., A.M.J. Wensing, and C.A.B. Boucher, *The Epidemiology of Transmission of Drug Resistant HIV-1*, in *HIV Sequence Compendium 2006/2007*, T. Leitner, et al., Editors. 2006, Theoretical Biology and Biophysics Group, Los Alamos National Laboratory: Los Alamos, NM. p. 17-36.

320. Supervie, V., J.G. Garcia-Lerma, W. Heneine, and S. Blower, *HIV, transmitted drug resistance, and the paradox of preexposure prophylaxis.* Proc Natl Acad Sci U S A, 2010. **107**(27): p. 12381-12386.

321. Vandamme, A.M., R.J. Camacho, F. Ceccherini-Silberstein, A. de Luca, L. Palmisano, D. Paraskevis, . . . A. Sonnerborg, *European recommendations for the clinical use of HIV drug resistance testing: 2011 update.* AIDS Rev, 2011. **13**(2): p. 77-108.

322. Little, S.J., S. Holte, J.-P. Routy, E.S. Daar, M. Markowitz, A.C. Collier, . . . D.D. Richman, *Antiretroviral-Drug Resistance Among Patients Recently Infected with HIV.* N Engl J Med, 2002. **347**(6): p. 385-394.

323. Yanik, E.L., S. Napravnik, C.B. Hurt, A. Dennis, E.B. Quinlivan, J. Sebastian, . . . J.J. Eron, Jr., *Prevalence of Transmitted Antiretroviral Drug Resistance Differs Between Acutely and Chronically HIV-Infected Patients.* J Acquir Immune Defic Syndr, 2012. **61**(2): p. 258-262.

324. Jain, V., T. Liegler, E. Vittinghoff, W. Hartogensis, P. Bacchetti, L. Poole, . . . F.M. Hecht, *Transmitted drug resistance in persons with acute/early HIV-1 in San Francisco, 2002-2009.* PLoS One, 2010. **5**(12): p. e15510.

325. von Wyl, V., T. Klimkait, S. Yerly, D. Nicca, H. Furrer, M. Cavassini, . . . T.R. Glass, *Adherence as a predictor of the development of class-specific resistance mutations: the Swiss HIV Cohort Study.* PLoS One, 2013. **8**(10): p. e77691.

326. Soeters, H.M., S. Napravnik, O.M. Zakharova, J.J. Eron, and C.B. Hurt, *Opportunities for sexual transmission of antiretroviral drug resistance among HIV-infected patients in care.* AIDS, 2013.

327. Chin-Hong, P.V., S.G. Deeks, T. Liegler, E. Hagos, M.R. Krone, R.M. Grant, and J.N. Martin, *High-risk sexual behavior in adults with genotypically proven antiretroviral-resistant HIV infection.* J Acquir Immune Defic Syndr, 2005. **40**(4): p. 463-471.

328. Kalmar, E.M., S.S. Sanabani, A. Charlys da Costa, S. Ferreira, C.C. Barreto, S. Chen, and E.C. Sabino, *Evaluation of HIV-1 resistance to antiretroviral drugs among 150 patients after six months of therapeutic interruption.* Int J STD AIDS, 2012. **23**(2): p. 120-125.

329. Kozal, M.J., K.R. Amico, J. Chiarella, T. Schreibman, D. Cornman, W. Fisher, . . . G. Friedland, *Antiretroviral resistance and high-risk transmission behavior among HIV-positive patients in clinical care.* AIDS, 2004. **18**(16): p. 2185-2189.

330. Abbas, U.L., G. Hood, A.W. Wetzel, and J.W. Mellors, *Factors Influencing the Emergence and Spread of HIV Drug Resistance Arising from Rollout of Antiretroviral Pre-Exposure Prophylaxis (PrEP).* PLoS One, 2011. **6**(4): p. e18165.

331. Tang, M.W. and R.W. Shafer, *HIV-1 Antiretroviral Resistance: Scientific Principles and Clinical Applications.* Drugs, 2012. **72**(9): p. e1-e25.