Title

Using Machine Learning to Identify Predictors of Sexually Transmitted Infections Over Time Among Young People Living With or at Risk for HIV Who Participated in ATN Protocols 147, 148, and 149

Journal

Sexually Transmitted Diseases, 50(11)

ISSN

0148-5717

Authors

Comulada, W Scott
Rotheram-Borus, Mary Jane
Arnold, Elizabeth Mayfield
et al.

Publication Date

2023-11-01

DOI

10.1097/olq.0000000000001854

Peer reviewed

HHS Public Access
Author manuscript
Sex Transm Dis. Author manuscript; available in PMC 2024 November 01.

Published in final edited form as:
Sex Transm Dis. 2023 November 01; 50(11): 739–745. doi:10.1097/OLQ.0000000000001854.

Using Machine Learning to Identify Predictors of Sexually Transmitted Infections Over Time among Young People Living with or At Risk for HIV Who Participated in ATN Protocols 147, 148, and 149

W. Scott Comulada, DrPH[1,*], **Mary Jane Rotheram-Borus, PhD**[1], **Elizabeth Mayfield Arnold, PhD**[2], **Peter Norwood, PhD**[1], **Sung-Jae Lee, PhD**[1], **Manuel A. Ocasio, PhD**[3], **Risa Flynn, BA**[4], **Karin Nielsen, MD**[5], **Robert Bolan, MD**[4], **Jeffrey Klausner, MD**[6], **Dallas Swendeman, PhD**[1], **Adolescent Medicine Trials Network (ATN) CARES Team**[1,3,5,7,8,9,10]

[1]Department of Psychiatry & Biobehavioral Sciences, David Geffen School of Medicine, University of California Los Angeles, Los Angeles, CA

[2]College of Medicine, University of Kentucky, Lexington, KY

[3]Department of Pediatrics, School of Medicine, Tulane University, New Orleans, LA

[4]Los Angeles LGBT Center, Los Angeles, CA

[5]Department of Pediatrics, David Geffen School of Medicine, University of California Los Angeles, Los Angeles, CA

[6]Department of Infectious Diseases, Keck School of Medicine, University of Southern CA

[7]Nova Southeastern University, Fort Lauderdale, FL

[8]School of Public Health, Portland State University, Portland, OR

[9]Department of Biostatistics, Fielding School of Public Health, University of California Los Angeles, Los Angeles, CA

[10]Friends Research Institute Inc, Los Angeles, CA

Abstract

Background—Sexually transmitted infections (STI) among youth aged 12-24 years old have doubled in the last 13 years, accounting for 50% of STI nationally. We need to identify predictors of STI among youth in urban HIV epicenters.

Methods—Sexual and gender minority (SGM; gay, bisexual, transgender, gender-diverse) and other youth with multiple life stressors (homelessness, incarceration, substance use, mental health disorders) were recruited from 13 sites in Los Angeles and New Orleans (N=1482). Self-reports and rapid diagnostic tests for STI, HIV, and drug use were conducted at 4-month intervals for up

[*]*Corresponding author:* W. Scott Comulada, wcomulada@mednet.ucla.edu; Tel: +1-310-794-0938; Fax: +1 (310) 794-8297.

Conflicts of Interest
The authors declare no conflicts of interest.

to 24 months. Machine learning was used to identify predictors of time until new STI (including a new HIV diagnosis).

Results—At recruitment, 23.9% of youth had a current or past STI. Over 24 months, 19.3% tested positive for a new STI. Heterosexual males had the lowest STI rate (12%); African-American youth were 23% more likely to acquire an STI compared to peers of other ethnicities. Time to STI was best predicted by attending group sex venues or parties, moderate but not high dating app use, and past STI and HIV seropositive status.

Conclusions—STI are concentrated among a subset of young people at highest risk. The best predictors of youth's risk are their sexual environments and networks. Machine learning will allow the next generation of research on predictive patterns of risk to be more robust.

Short Summary

Sexually transmitted infections among youth in high-risk settings in two U.S. cities was 19.6% over 24 months and predicted by group sex/parties, moderate dating app usage, and HIV/STI histories.

Keywords

HIV/AIDS; homelessness; Sexual and Gender Minority; machine learning; STI incidence

Introduction

Sexually transmitted infections (STI) are epidemic in the United States.[1] Youth aged 15-24 years account for 17% of the U.S. population but reflect 45.5% of incident STI and 19% of prevalent STI.[1] In 2013, the societal costs of STI were $16 billion, not including HIV.[1] To reduce personal and societal costs of STI, it is critical to know how to target youth at highest risk of acquiring and transmitting STI.

While all sexually active adolescents are at risk of acquiring STI, there are subpopulations with higher STI prevalence and incidence rates. Sexual and gender minority (SGM) youth (i.e., gay, bisexual and other men-who-have-sex-with-men [GBMSM], queer, transgender, non-binary, and gender diverse youth) are at higher risk of acquiring STI,[2] a pattern that remains throughout their adulthood.[3] In that study, the prevalence of STIs was 35% among GBMSM men, 25% among transgender women, and 15% among heterosexual cisgender men.[3] Socio-economic marginalization and risky sexual behaviors among SGM youth also increase STI risk. STI rates are higher among SGM youth who experience life stressors that are often linked to stigma and discrimination by families and employers,[4,5] and compounded by socio-economic disadvantage and marginalization experienced by racial/ethnic minority youth.[6,7]

The goal of this paper is to identify youth characteristics that predict STI acquisition among subpopulations at risk. Findings can help public health researchers target sexual health interventions to those most at risk. Towards this goal, we used machine learning (ML) [8] over standard regression methods to identify salient characteristics. Standard regression uses the same data to select characteristics and evaluate their statistical significance, which describes characteristics that correlate with but don't necessarily predict outcome levels. For example, characteristics may yield small p-values due in part to large sample sizes. ML address prediction by splitting data into training data to select models and characteristics and test data to evaluate how well models predict outcome observations in a new sample. We applied ML models[8] to data from three cohort studies conducted by the Adolescent Trials Network for HIV Medicine (ATN)[9–11] to identify characteristics that predict time to STI incidence (including new HIV infections).

Materials and Methods

Study Design.

Youth participating in three studies through the ATN CARES program[9–11] were included for analysis. The [BLINDED FOR REVIEW] served as the Institutional Review Board of record for investigators from collaborating universities (IRB#16-001372).

Recruitment and Eligibility.

Teams in Los Angeles, CA and New Orleans, LA screened youth at-risk for or living with HIV from agencies and adolescent medicine clinics. Recruitment was also done via social media dating apps[12] and referrals. Youth were eligible for enrollment either as: 1) youth living with HIV (ATN 147 and 148 for recent and established infection, respectively); or 2) youth not living with HIV who had multiple risk factors for HIV: recent STIs, condomless sex, or substance use; lifetime experience of homelessness, mental health hospitalization, or criminal justice contact; SGM; Black or Latinx; reporting sex partners living with HIV; or current PrEP use (ATN 149). SGM youth needed fewer risk factors to be eligible.

STI Testing.

Testing and treatment protocols are described in a protocol publication.[13] Briefly, at study eligibility screening, a rapid fourth generation Alere™ test for HIV infection was conducted to determine eligibility. Once participants were enrolled, STI testing included testing for Neisseria gonorrhea (NG) and Chlamydia trachomatis (CT) using the Xpert® CT/NG assay. Participants with positive NG or CT tests were offered same-day treatment per CDC recommendations. We screened for syphilis infection using the CLIA-waived rapid point-of-care fingerstick whole blood treponemal antibody test Syphilis Health Check™ (Diagnostics Direct, Stone Harbor, New Jersey). If the test was positive, we used participants' venous blood specimens collected at the same visit for rapid plasma reagin titer and *T pallidum* particle agglutination confirmatory testing. The titer was needed for diagnosis in those with a history of syphilis. We individually reviewed each potential syphilis case, made a clinical determination, and followed CDC recommendations for same-day treatment and medical provider referral when appropriate.

Measures.

STI testing and the following measures were collected at 4-month intervals for 24 months. Assessment domains and measures are described below.

Primary outcome.

The primary outcome was the time to a new STI. There were four STI included: HIV, syphilis, chlamydia, and gonorrhea. Only new infections counted towards the outcome for HIV and syphilis. For participants testing positive, their outcome measurement was number of days since baseline that their first positive test for any STI was recorded. Youth never testing positive were right-censored as the number of days from baseline to their final visit.

Predictors (see Table 1).

Sociodemographic Factors included sexual orientation and gender identity*.* Four categories summarize both *gender and sexual identity* for the ML analyses: female (Cis-gender females); gay, bisexual, and GBMSM which includes cis-gender male queer, pan, or youth with other self-labels or those reporting sex with men; heterosexual males (Hetero-Male), and transgender or gender diverse (Trans/Gender-Diverse).

Race/Ethnicity was categorized as Black, including biracial youth reporting Black or African-American; Latinx; White, non-Latinx, and Asian, Hawaiian/Pacific Islander, Native American, Alaska Native, or other.

Education was categorized as high school/GED, some higher education, and completed higher education.

Life Stressor History included any lifetime incarceration, mental health treatment, substance use treatment, and homelessness.

Trauma focused on sexual or violent trauma. Sexual trauma was a binary predictor for whether the youth was ever forced to perform a sexual act with someone more than five years older or forced to do something sexually. Violent trauma was a binary predictor for whether the youth ever had or was threatened by someone attempting to attack or rob them, had ever seen someone seriously injured or killed, had a close friend or family member murdered, or ever experienced any intimate partner violence.

Sexual Activity included always using condoms in the past 4 months, number of lifetime sex partners, lifetime sex-exchange, and lifetime group sex venue or event participation. Lifetime partners was treated as a continuous predictor, but we Winsorized it to have a maximum value 44 based on the 1.5*IQR rule[14] due to outliers. Group sex venues or events included attending commercial sex clubs or bathhouses, commercially organized sex parties, or any other private sex parties.

Dating app use was categorized as never, less than once a week, about once a week, once every couple of days, once a day, and several times a day.

Mental Health was measured for anxiety using the GAD-7[15] and for depression using the 9-item Patient Health Questionnaire (PHQ-9).[16]

Healthcare consisted of three indicators for having health insurance; having a routine healthcare provider; and participation in an HIV prevention program in their lifetime.

Drug and Alcohol Use was assessed by youth self-report of marijuana, methamphetamine, cocaine, and opiates, benzodiazepines, and party drugs. In addition, rapid diagnostic tests of all but party drugs were monitored. However, the tests' validity had very different time frames for each drug (e.g., marijuana for 8-30 days vs methamphetamine for 2-3 days). Drug use was a binary indicator for a positive response on either self-reports or rapid diagnostic tests for drug use. Poly drug use was present or not based on self-reported use or rapid diagnostic tests of two or more drugs, excluding marijuana. For alcohol use, we use the AUDIT-C score.[17]

History of STI was obtained through youth self-report of HIV and separately for Syphilis, Chlamydia, Gonorrhea, or Hepatitis-C at baseline, as well as testing positive for STI at baseline. The HIV indicator is 1 if the youth self-reported or tested positive for HIV at baseline.

Data Analysis.

Table 1 presents summary statistics for the full sample and for each of four gender-sexual identity subgroups.

We conducted a Cox proportional hazard model for days from baseline until STI. We determined the proportional hazards assumption to be reasonable by fitting a cox model using the *coxph* function in the *survival* package in R followed by visual and formal test diagnostics. Specifically, we plotted survival curves for the entire sample and then for each of the gender-sexual identity subgroups. We tested whether there was a non-random relationship between the *Schoenfield residuals* and the time-to-event using the *cox.zph* function.[18]

After determining the model was reasonable, we employed a ML variable selection algorithm. The analytic sample was based on available follow-up data (N=1370). We utilized ML rather than backward or stepwise regression to reduce overfitting, especially given the large number of predictors. We used a lasso Cox proportional hazard generalization to select the most useful predictors.[8] Lasso cox regression shrinks regression coefficients to zero as the penalty parameter increases retaining only the most important predictors in the model.

Using k-fold cross-validation,[8] we found an optimal parameter that maximizes the Harrell's concordance measure, a model-fit criteria specifically for time-to-event outcomes.[19] Cross validation split the data into training and testing groups and evaluated how well the trained model performed on the testing set.

Finally, we fit an unpenalized Cox regression model using the selected predictors to obtain unbiased hazard ratios (HR). While we present p-values, we advise caution in their interpretation due to the questionable validity of inference for ML models.[20] While not indicating statistical significance, inclusion in this model indicates that each variable is meaningful in explaining the outcome.

Results

Participants had a mean age of 21.2 years (SD= 2.2; Table 1). The sample was predominantly male (82.1%) and Black (51.2%) and Latinx (24.2%). While 23.7% were heterosexual, 45.1% identified as gay, 21.6% as bisexual, and 9.6% as other queer identities. Among participants older than 18 years, 14.7% had not finished high school.

Homelessness was experienced by 43.4% of youth. Incarceration was experienced by 22.1% overall, with heterosexual males reporting incarceration twice as often as cisgender females. Mental health treatment was received by half of the youth, differentially much higher among trans/gender-diverse youth compared to peers. Marijuana use was common across all groups (76.6%; Table 2). Party drugs cocaine were the second most common drugs taken (43.5% and 25.3%, respectively). Methamphetamine and benzodiazepines were used by about 15% of youth.

About half of youth (46.9%) reported always using condoms. About 2/3 of GBMSM and half of trans/gender-diverse youth used dating apps. If using dating apps, 78.5% of GBMSM used one at least every couple of days, as did 70% of trans/gender-diverse youth. About 1/3 of trans/gender-diverse youth reported exchanging sex for money, drugs, or places to stay, compared to 1/5 to 1/4 of other youth. Violent traumatic experiences were common across all groups. Only 14.5% of youth were living with HIV.

Figure 1 details the time-to-event curves for the overall group and each gender-sexual identity group. Just over 20% of study participants tested positive for any STI. This study focuses on the time to the first STI. The percentage of new STI in the first and second years was 13.5% and 10.3%, respectively; 7.2% of youth had at least two STI, with at least one in each year. The most STIs, relatively, occur among the Cis-GBMSM group and the least in the Hetero-Male group. Over time, 19.6% of study participants had tested positive for an STI.

Table 3 shows adjusted-HR for the lasso-selected variables from a multivariable model that retained the variables' original categories and a model that only included the collapsed categories created by the ML algorithm. The ML algorithm collapsed the gender-sexual identity variable into hetero males versus others. As the plot indicates, hetero males are less likely to test positive for STI. The estimated HR is 0.640, so at any given time point we expect hetero males to be 36.0% less likely to test positive for STI than other groups.

The algorithm also collapsed the race and ethnicity variable into Black/African American versus others; we estimate that those who identify as Black or African American are 23.6% more likely to test positive for an STI (HR=1.236). Education is collapsed into completed higher education versus others; we estimate those who completed higher education are 29.7% less likely to test positive (HR=0.703). We estimate a HR of 0.706 for those with lifetime history of mental health treatment. Next, we estimate a HR of 0.851 for those reporting violent trauma.

Baseline or previous STIs, including HIV, were related to major increases in the probability of testing positive in the study period. We estimate a HR of 1.918 for baseline or previous non-HIV STI and 1.313 for HIV. Unsurprisingly, those who attended group sex venues or events (4% of the sample) were much more likely to test positive, we estimate a HR of 1.706.

The model that retained the original response categories showed that data app users, regardless of frequency, had a higher hazard of testing positive relative to those who never used data apps (HR = 1.24 to 1.71). Interestingly, the ML algorithm collapsed the dating app use variable to less than once per week versus no use and use more frequently than once a week (HR = 1.49). Those who use dating apps less than once per week were nearly 50% more likely to test positive at any given time point relative to those who never use dating apps and more frequent dating app users.

Discussion

Results indicate subgroups of youth at increased risk for contracting STIs over time. Overall, sexual networks emerge as the most important predictors of STI acquisition among this sample. Youth who reported attending group sex events were 75% more likely to test positive for STI over time. This finding highlights the need to consider strategies beyond the individual-level interventions and address the social context of sexual networks of youth. Dating apps also emerged as a predictor of STI, however, not in the manner anticipated. Youth utilizing dating apps less than once a week were nearly 50% more likely to test positive for STI compared to youth who didn't use dating apps and more frequent users. Given that previous research has shown frequent users of dating websites and apps have higher numbers of casual partners compared to non-users,[25] our finding is surprising. As over half of adolescent GBMSM report using apps to meet sexual partners,[25] we posit that infrequent dating app users may represent a higher risk subpopulation who are less likely to adhere to dating behaviors that would reduce STI transmission. This finding underscores the need to investigate further the context in which dating apps are being used.

The majority of challenges youth experience are based on social determinants of risk:[21] this sample is primarily from low-income backgrounds who are living below the poverty line; most are SGM youth, and racial/ethnic minorities who experience disproportionate rates of stigma and discrimination.[22] GBMSM and transgender youth often search to find social networks of peers. In large urban centers youth often end up in settings placing them at greater risk of being abused, offered drugs, or having contact with the criminal justice system.[23] In this context, our findings are like studies indicating higher rates of STI associated with sexual identity and behaviors.[24]

As expected, baseline report of a STI, including HIV, is an important predictor of future STI acquisition – not a novel or unique finding. Furthermore, we have previously published that the more frequent assessment of youth for STI following the CDC's recommendations of testing for GBMSM, the lower was their prevalence of infection over 12 months.[26] Stunningly, in the current study the rate of STI was almost one in five youth. This appears to be substantially higher than the rate in the US general population of 15- to 24-year-olds, from 2018 data, even after including a more comprehensive list of STIs than our study assessed (e.g., HPV).[1] This indicates the confluence of risk among subsets of youth, making our predictive models even more important to recognize.

For youth who reported high rates of lifetime traumatic events, our predictive model revealed that a history of mental health treatment and violent trauma may be protective against STI acquisition. This finding is also unexpected as others have found that victimization is part of a syndemic associated with more sexual partners.[27] Our findings point to the need to examine the complex interplay between mental health, violent trauma, and STI risk--including past exposure to mental health intervention. Specific mental health symptoms may mediate the relationship between victimization and STI risk.[28]

This study has important strengths, including sample diversity. We recruited a large cohort of GBMSM and trans/gender-diverse youth, as well as cis-gender women and heterosexual men typically excluded from HIV prevention research in the U.S., with significant lifetime experiences of social stressors, mental health and substance use history, sexual risk behaviors, and lifetime trauma. GBMSM in this study reported higher levels of social stressors, including lifetime history of experiencing homelessness and incarceration. In addition, many youth reported histories of mental health and substance use treatment and high rates of sexual risk behaviors.

Another strength is our novel use of ML algorithms to select STI acquisition predictors. ML algorithms have advantages for developing predictive models, such as not requiring statistical inference, being data driven, and learning from data to identify complex nonlinear patterns.[29] They have been used to predict the future risk of other conditions[29,30] but no prior published research has used ML to predict STI or HIV acquisition among a large minoritized youth cohort.

Despite benefits of ML over traditional methods, they still bear limitations of traditional methods. Survival analyses account for attrition as right-censoring but do not account for intermittent missingness. Missing follow-up assessments may have reduced the accuracy of outcome measurements for time to new STI. Lasso uses the ML framework to reduce overfitting, while providing interpretable regression coefficients. While less interpretable, more complex ML methods could undercover nonlinear relationships between STI and risk factors to improve prediction in larger samples.[30]

In summary, ML yielded insights into complex links between HIV and STI-related outcomes and social stressors, mental health, trauma, sexual risk, and substance use. Key findings of the syndemic framework are its potential for examining not only synergistic individual-level risk factors but also the interactions with social systems that influence these individual-level factors and thereby shape the HIV epidemic among minoritized youth.[31s] Study findings linking sexual network factors to STI risk also suggest opportunities to intervene, such as smartphone-delivered ecological momentary interventions[32s] that could be triggered by dating app usage patterns.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

In addition to the authors for this paper, the Adolescent Medicine Trials Network (ATN) CARES Team includes: Sue Ellen Abdalian, Yvonne Bryson, Ruth Cortado, M. Isabel Fernandez, Tara Kerin, Marguerita Lightfoot, Norweeta Milburn, Wilson Ramos, Cathy Reback, Wenze Tang, Panteha Hayati Rezvan, and Robert E. Weiss.

Funding

This work was supported by the Adolescent Medicine Trials Network for HIV/AIDS Interventions (ATN) at the National Institutes of Health grant number U19HD089886. The Eunice Kennedy National Institute of Child Health and Human Development (NICHD) is the primary funder of this network, with support of the National Institute of Mental Health (NIMH), National Institute of Drug Abuse (NIDA), and National Institute on Minority Health and Health Disparities (NIMHD). This project was also supported by the NIMH-funded Center for HIV Identification, Prevention, and Treatment Services (CHIPTS) grant number P30MH058107) and the T32 Post-doctoral HIV Training Grant for Combination Prevention grant number T32MH109205.

References

1. Kreisel K, Spicknall IH, Gargano JW, et al. Sexually transmitted infections among US women and men: prevalence and incidence estimates, 2018. Sex Transm Dis 2021; 48(4): 208–214.
2. Pampati S, Johns MM, Szucs LE, et al. Sexual and gender minority youth and sexual health education: A Systematic mapping review of the literature. J Adolesc Health 2021; 68(6): 1040–1052. [PubMed: 33162290]
3. Shover CL, DeVost MA, Beymer MR, et al. Using sexual orientation and gender identity to monitor disparities in HIV, sexually transmitted infections, and viral hepatitis. Am J Public Health 2018; 108(S4): S277–S283. [PubMed: 30383431]
4. Cook RL, Comer DM, Wiesenfeld HC, et al. Alcohol and drug use and related disorders: An underrecognized health issue among adolescents and young adults attending sexually transmitted disease clinics. Sex Transm Dis 2006; 33(9): 565–570. [PubMed: 16572042]
5. Reback CJ, Fletcher JB. Elevated HIV and STI prevalence and incidence among methamphetamine-using men who have sex with men in Los Angeles County. AIDS Educ Prev 2018; 30(4): 350–356. [PubMed: 30148668]
6. Sutton MY, Anachebe NF, Lee R, Skanes H. Racial and ethnic disparities in reproductive health services and outcomes, 2020. Obstet Gynecol 2021; 137(2): 225–233. [PubMed: 33416284]
7. Williams SP, Bryant KL. Sexually transmitted infection prevalence among homeless adults in the United States: a systematic literature review. Sex Transm Dis 2018; 45(7): 494–504. [PubMed: 29465661]
8. Tibshirani R The lasso method for variable selection in the cox model. Stat Med 1997; 16(4): 385–395. [PubMed: 9044528]
9. Nielsen-Saines K, Mitchell K, Kerin T, et al. Acute HIV infection in youth: protocol for the Adolescent Trials Network 147 (ATN 147) Comprehensive Adolescent Research and Engagement Studies (CARES) Study. JMIR Res Protoc 2019; 8(1): e10807. [PubMed: 30650057]
10. Swendeman D, Harris D, Comulada WS, et al. Text-messaging, online peer support group, and coaching strategies to optimize the HIV prevention continuum for youth: protocol for a randomized controlled trial. JMIR Res Protoc 2019; 8(8): e11165. [PubMed: 31400109]
11. Arnold EM, Swendeman D, Rotheram-Borus MJ, et al. The stepped care intervention to suppress viral load in youth living with HIV: protocol for a randomized controlled trial. JMIR Res Protoc 2019; 8(2): e10791. [PubMed: 30810536]
12. Ocasio MA, Fernández MI, Joseph JM, et al. Engaging sexual and gender minority youth in HIV interventions through gay dating apps: recruitment protocol. JMIR Res Protoc 2021; 10(6): e28864. [PubMed: 34156342]
13. Shannon CL, Koussa M, Lee SJ, et al. Community-based, point-of-care sexually transmitted infection screening among high-risk adolescents in Los Angeles and New Orleans: protocol for a mixed-methods study. JMIR Res Protoc 2019; 8(3): e10795. [PubMed: 30900994]
14. Mowbray FI, Fox-Wasylyshyn SM, El-Masri MM. Univariate outliers: A conceptual overview for the nurse researcher. Can J Nurs Res 2019; 51(1): 31–37. [PubMed: 29969044]

15. Spitzer RL, Kroenke K, Williams JBW, Lowe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. Arch Intern Med 2006; 166(10): 1092–1097. [PubMed: 16717171]
16. Kroenke K, Spitzer RL. The PHQ-9: a new depression and diagnostic severity measure. Psychiatr Ann 2002; 32(9): 509–521.
17. Bradley KA, Bush KR, Epler AJ, et al. Two brief alcohol screening tests: the Alcohol Use Disorders Identification Test. Arch Intern Med 2003; 163(7): 821–829. [PubMed: 12695273]
18. Grambsch PM, Therneau TM. Proportional hazards tests and diagnostics based on weighted residuals. Biometrika 1994; 81(3): 515–526.
19. Harrell FE, Califf RM, Pryor DB, Lee KL, Rosati RA. Evaluating the yield of medical tests. JAMA 1982; 247(18): 2543–246. [PubMed: 7069920]
20. Lo A, Chernoff H, Zheng T, Lo, S-H. Why significant variables aren't automatically good predictors. Proc Natl Acad Sci U S A 2015; 112(45): 13892–13897. [PubMed: 26504198]
21. Rotheram-Borus MJ, Swendeman D, Chovnick G. The past, present and future of HIV prevention: Integrating behavioral, biomedical, and structural intervention strategies for the next generation of HIV prevention. Annu Rev Clin Psychol 2009; 5: 143–167. [PubMed: 19327028]
22. Berg RC, Carter D, Ross MW. A mixed-method study on correlates of HIV-related stigma among gay and bisexual men in the Southern United States. J Assoc Nurses AIDS Care 2017; 28(4): 532–544. [PubMed: 28279588]
23. Holloway IW, Rice E, Kipke M. Venue-based network analysis to inform HIV prevention efforts among young gay, bisexual, and other men who have sex with men. Prev Sci 2014; 15(3): 419–427. [PubMed: 24464324]
24. Op de Coul ELM, Warning TD, Koedijk FDH, Dutch STI clinics. Sexual behaviour and sexually transmitted infections in sexually transmitted infection clinic attendees in the Netherlands, 2007-2011. Int J STD AIDS 2014; 25(1): 40–51. [PubMed: 23970630]
25. Hecht J, Zlotorzynska M, Sanchez TH, Wohlfeiler D. Gay dating app users support and utilize sexual health features on apps. AIDS Behav 2022; 26(6): 2081–2090. [PubMed: 35018547]
26. Lee SJ, Ocasio MA, Goldbeck CS, et al. Community-based implementation of centers for disease control and prevention's recommended screening for sexually transmitted infections among youth at high risk for HIV infection in Los Angeles and New Orleans. Sex Transm Dis 2020; 47(7): 481–483. [PubMed: 32209954]
27. Turpin RE, Salerno JP, Rosario AD, Boekeloo B. Victimization, substance use, depression, and sexual risk in adolescent males who have sex with males: a syndemic latent profile analysis. Arch Sex Behav 2021; 50(3): 961–971. [PubMed: 32274744]
28. Ahrens KR, McCarty C, Simoni J, et al. Psychosocial pathways to sexually transmitted infection risk among youth transitioning out of foster care: Evidence from a longitudinal cohort study. J Adolesc Health 2013; 53(4): 478–485. [PubMed: 23859955]
29. Patel B, Sengupta P. Machine learning for predicting cardiac events: what does the future hold? Expert Rev Cardiovasc Ther 2020; 18(2): 77–84. [PubMed: 32066289]
30. Chen CY, Chen YF, Chen HY, et al. Artificial neural network and Cox regression models for predicting mortality after hip fracture surgery: a population-based comparison. Medicina (Kaunas) 2020; 56(5): 243. [PubMed: 32438724]

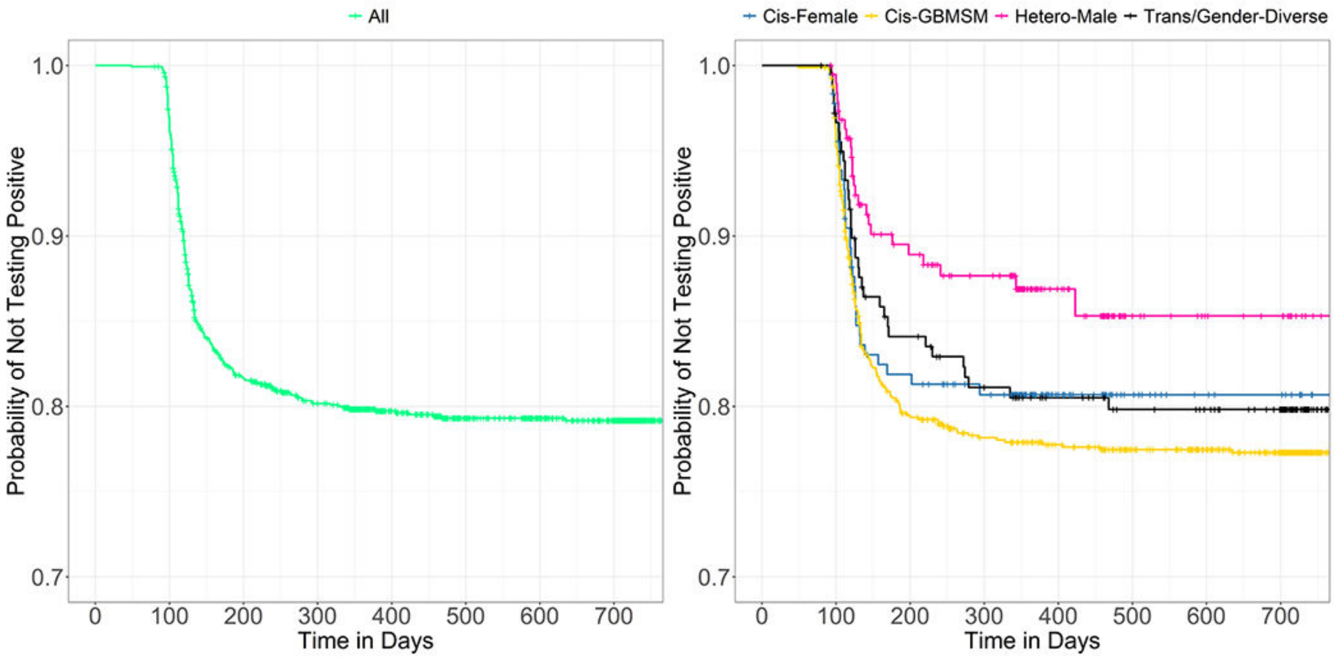

Figure 1:
Kaplan-Meier curves for time to new STI infection in the entire sample (left) and each gender-sexual identity group: cisgender female, cisgender gay, bisexual, and other men who have sex with men (GBMSM), heterosexual male, and transgender/gender-diverse (right).

Table 1

Sociodemographic characteristics and healthcare utilization of sample at baseline by sexual/gender categories: C(is)-Female, C(is)-GBMSM, H(etero)-Male, and Trans/Gender-Diverse (TGD).

Variable	C-Female (N=181)	C-GBMSM (N=819)	H-Male (N=190)	TGD (N=180)	Total (N=1370)
Age in years, mean (SD)	20.3 (2.2)	21.2 (2.1)	21.1 (2.3)	21.0 (2.2)	21.2 (2.2)
Min, median, max, interquartile range	12,20,24,3	14,21,24,3	14,21,24,4	15,21,24,4	12,21,24,4
Assessment site, n (%)					
Los Angeles	71 (39.2)	487 (59.5)	97 (51.1)	125 (69.4)	780 (56.9)
New Orleans	110 60.8)	332 (40.5)	93 (48.9)	55 (30.6)	590 (43.1)
Male sex assigned birth, n(%)	0 (0.0)	819 (100)	190 (100)	116 (64.4)	1125 (82.1)
Race/ethnicity, n(%)					
Asian	1 (0.6)	43 (5.3)	3 (1.6)	5 (2.8)	52 (3.7)
Black/African American	141 (77.9)	347 (42.4)	141 (74.2)	72 (40.0)	701 (51.2)
Hawaiian/Pacific Islander	0 (0.0)	4 (0.5)	1 (0.5)	2 (1.1)	7 (0.5)
Latinx	18 (9.9)	235 (28.7)	27 (14.2)	52 (28.9)	332 (24.2)
Native American/Alaska Native	3 (1.7)	6 (0.7)	0 (0.0)	3 (1.7)	12 (0.9)
Other	3 (1.7)	22 (2.7)	2 (1.1)	8 (4.4)	35 (2.6)
White	15 (8.3)	162 (19.8)	16 (8.4)	38 (21.1)	231 (16.9)
Sexual identity, n(%), missing=1					
Bisexual	47 (26.0)	219 (26.7)	n/a[a]	30 (16.8)	296 (21.6)
Gay/same gender loving	22 (12.2)	551 (67.3)	n/a	45 (25.1)	618 (45.1)
Heterosexual	102 (56.4)	n/a	190 (100)	32 (17.9)	324 (23.7)
Other	10 (5.5)	49 (6.0)	n/a	72 (40.2)	131 (9.6)
Education level, n(%)					
Below high school	51 (28.2)	105 (12.8)	75 (39.5)	42 (23.3)	273 (19.9)
High school/equivalent	54 (29.8)	192 (23.4)	60 (31.6)	51 (28.3)	357 (26.1)
Some college	72 (39.8)	412 (50.3)	51 (26.8)	67 (37.2)	602 (43.9)
Completed college	4 (2.2)	110 (13.4)	4 (2.1)	20 (11.1)	138 (10.1)
Employment, n(%)					

Variable	C-Female (N=181)	C-GBMSM (N=819)	H-Male (N=190)	TGD (N=180)	Total (N=1370)
Employed	59 (32.6)	444 (54.2)	82 (43.2)	68 (37.8)	653 (47.7)
Not employed	68 (37.6)	157 (19.2)	76 (40.0)	64 (35.6)	365 (26.6)
Student	54 (29.8)	218 (26.6)	32 (16.8)	48 (26.7)	352 (25.7)
Income above federal poverty level,[b] n (%)	31 (17.1)	313 (38.2)	42 (22.1)	33 (18.3)	419 (30.6)
Lifetime homelessness, n (%)	118 (65.2)	248 (30.3)	137 (72.1)	91 (50.6)	594 (43.4)
Lifetime incarceration, n (%)	44 (24.3)	128 (15.6)	94 (49.5)	37 (20.6)	303 (22.1)
Health insurance (Yes vs. No/Unsure), n (%)	135 (74.6)	638 (77.9)	131 (68.9)	149 (82.8)	1053 (76.9)
Healthcare provider, n (%)	140 (77.3)	597 (72.9)	121 (63.7)	148 (82.2)	1006 (73.4)
Lifetime participation treatment programs, n (%)					
HIV prevention	36 (19.9)	155 (18.9)	45 (23.7)	57 (31.7)	293 (21.4)
Mental health	102 (56.4)	383 (46.8)	85 (44.7)	135 (75.0)	705 (51.5)
Substance abuse	39 (21.5)	107 (13.1)	48 (25.3)	41 (22.8)	235 (17.2)

[a]Not applicable by definition

[b]$1063.33/month

Table 2

Sexual behavior and health, substance and alcohol use, and mental health measures of sample at baseline baseline by sexual/gender categories: C(is)-Female, C(is)-GBMSM, H(etero)-Male, and Trans/Gender-Diverse (TGD).

Variable	C-Female (N=181)	C-GBMSM (N=819)	H-Male (N=190)	TGD (N=180)	Total (N=1370)
Sexual behavior and health					
Always use condom, recent, n (%)	77 (39.2)	399 (45.1)	122 (58.0)	113 (58.8)	642 (46.9)
Dating app usage, n (%)					
Never	160 (88.4)	284 (34.7)	166 (87.4)	90 (50.0)	700 (51.1)
Less than once a week	4 (2.2)	61 (7.4)	2 (1.1)	17 (9.4)	84 (6.1)
About once a week	6 (3.3)	54 (6.6)	2 (1.1)	10 (5.6)	72 (5.3)
Once every couple days	4 (2.2)	104 (12.7)	4 (2.1)	14 (7.8)	126 (9.2)
Once a day	1 (0.6)	109 (13.3)	5 (2.6)	23 (12.8)	138 (10.1)
Several times a day	6 (3.3)	207 (25.3)	11 (5.8)	26 (14.4)	250 (18.2)
Lifetime sex exchange, n (%)	46 (25.4)	177 (21.6)	39 (20.5)	68 (37.8)	330 (24.1)
Sex party attendance, n (%)	0 (0.0)	39 (4.8)	5 (2.6)	9 (5.0)	53 (3.9)
Num sex partners,[a] mean (SD)	8.1 (9.3)	16.3 (14.9)	12.4 (13.2)	13.1 (14.5)	14.3 (14.3)
Current/previous non-HIV STI, n (%)	30 (16.6)	233 (28.4)	23 (12.1)	39 (21.7)	325 (23.7)
HIV, n (%)	14 (7.7)	160 (19.5)	6 (3.2)	18 (10.0)	198 (14.5)
Substance use					
Cocaine use, n (%)	26 (14.4)	224 (27.4)	43 (22.6)	53 (29.4)	346 (25.3)
Methamphetamine use, n (%)	17 (9.4)	1118 (14.4)	30 (15.8)	37 (20.6)	202 (14.7)
Marijuana use, n (%)	1114 (63.0)	656 (80.1)	138 (72.6)	141 (78.3)	1049 (76.6)
Opiates use, n (%)	40 (22.1)	139 (17.0)	60 (31.6)	49 (27.2)	288 (21.0)
Benzodiazepine use, n (%)	12 (6.6)	138 (16.8)	25 (13.2)	37 (20.6)	212 (15.5)
Party drug use, n (%)	48 (26.5)	417 (50.9)	54 (28.4)	77 (42.8)	596 (43.5)
AUDIT-C score, mean (SD)	2.4 (2.4)	3.4 (2.8)	2.1 (2.5)	2.9 (3.1)	3.0 (2.8)
Mental health					
Sexual trauma, n (%)	81 (44.8)	370 (45.2)	70 (36.8)	108 (60.0)	629 (45.9)
Violent trauma, n (%)	150 (82.9)	540 (65.9)	167 (87.9)	139 (77.2)	996 (72.7)
PHQ9, mean (SD)	8.2 (6.1)	6.7 (5.6)	5.7 (4.9)	9.7 (5.9)	7.1 (5.8)
GAD7, mean (SD)	7.4 (5.8)	6.2 (5.3)	5.3 (5.0)	8.5 (5.5)	6.5 (5.4)

[a]Winsorized at 44

Table 3

Coefficient estimates, standard errors, and p-values from multivariable Cox regression models containing predictors selected by lasso model. Results are shown for two models: 1) first retaining all predictor response categories; and 2) retaining predictor categories selected by lasso model where predictor category shown in table is compared to all other predictor categories.

	All response categories			Lasso-selected categories		
Variable	**Estimate**	**SE**	**p**	**Estimate**	**SE**	**p**
Gender-Sexual Identity						
Hetero-male	.649	.269	.11	0.640	.221	.04
Cisgender GBMSM	.933	.210	.74			
Transgender/gender-diverse	.998	.254	.99			
Cisgender female	Ref.					
Race/Ethnicity						
Black/African American	1.786	.286	.04	1.236	.125	.09
Latinx	1.456	.292	.20			
White	1.608	.311	.13			
Other racial/ethnic categories	Ref.					
Education						
Completed higher education	.641	.273	.10	0.703	.230	.12
Some higher education	.938	.175	.71			
Completed high school/equivalent	.948	.186	.77			
Less than high school	Ref.					
Dating App Use						
Less than once per week	1.705	.241	.03	1.492	.225	.07
About once a week	1.403	.262	.20			
Once every couple of days	1.239	.228	.35			
Once a day	1.246	.213	.30			
Several times a day	1.253	.182	.21			
Never	Ref.					
Mental health treatment, lifetime	.682	.129	<.01	0.706	.126	<.01
Sex party participation	1.667	.266	.05	1.751	.261	.03
Experienced violent trauma	.859	.137	.27	0.851	.135	.23
Non-HIV STI reported at study enrollment	1.900	.134	<.01	1.918	.132	<.01
HIV diagnosis reported at study enrollment	1.312	.158	.08	1.313	.156	.08