The Added Value of Digital HIV Risk Assessment Tools & the Role of Context in Predicting New HIV Infections in South Africa: A Bayesian Analysis of Trial Data

Cindy Leung Soo

Department of Epidemiology, Biostatistics and Occupational Health

McGill University, Montreal

July 2022

A thesis submitted to McGill University in partial fulfillment of the requirements of the degree of

Master of Science in Epidemiology

Table of Contents

Abstract

Background. Testing uptake for human immunodeficiency virus (HIV) remains unequal in South Africa with people of lower socioeconomic status (SES) being less likely to test for HIV. In 2017-2018, HIVSmart!, an app-based HIV self-testing programme was evaluated in townships of South Africa in a quasi-randomised trial. The HIVSmart! app contained a risk staging tool which had not yet been validated. It is also unclear whether the risk of acquiring HIV in South African townships differs across contextual variables such as education, employment, dwelling and income. Understanding the impact of contextual factors on HIV infection and including them into an HIV risk assessment tool could improve testing uptake among those most affected by HIV.

Objectives. In this manuscript-based thesis, I used secondary data from the HIVSmart! quasi-randomised trial to: 1) evaluate the HIVSmart! digital risk levels; 2) develop and validate an HIV risk staging model for South African township populations and; 3) quantify the association between contextual variables and new HIV infections while accounting for individual and behavioural factors.

Methods. In this thesis, I evaluated the HIVSmart! risk levels through logistic regression. Using Bayesian predictive projection, I identified predictors of HIV to construct a risk assessment model which I validated in external data. Next, I evaluated the performance of the model when combined with a digital HIV self-testing programme. Finally, I estimated, through multivariate logistic regression, the impact of contextual factors on HIV in South African townships.

Results. My analyses included 1535 participants from the self-testing arm and 1560 from the conventional testing arm of the HIVSmart! trial. The original HIVSmart! risk levels that focused on behavioural risk factors were associated with HIV infection (odds ratio (OR), 89% credible interval (CrI): 1.15, 0.98 – 1.35) but had overall modest predictive ability (area

under the receiver operating characteristic curve (AUC-ROC), 0.62). I developed a model with an improved performance (AUC-ROC, 89% CrI: 0.71, 0.68 – 0.71) which included sociodemographic and behavioural predictors such as being unmarried, HIV testing history, having had sex with a partner living with HIV, dwelling situation and education. The HIV risk assessment model had a sensitivity of 91.1% and a specificity of 13.2%. Combined with results from a digital self-testing programme, the resulting specificity of the model was improved to 96.1% while the sensitivity remained relatively unchanged at 90.9%. Further, living in hostels/informal dwellings, or having lower levels of education increased the odds of HIV (adjusted OR, 89% CrI: 1.33, 1.05 - 1.67 and 1.75, 1.24 – 2.50, respectively), after controlling for individual and behavioural factors. The likelihood of having recently sought HIV testing, however, did not differ according to socioeconomic factors.

Discussion. This is the first validated HIV risk assessment tool to be developed specifically for South African township populations and one the few that utilises contextual variables. Integrating social determinants of health such as education and dwelling situation in risk staging tools acknowledges the impact of health inequities. Moreover, offering HIV risk assessment and self-testing through a digital platform such as HIVSmart! can improve testing uptake among those disproportionately impacted by HIV. My results also suggest that while people living in informal housing and having lower levels of education are more impacted by HIV, current HIV testing and prevention services may be insufficient to address health inequities as they are not linking more people of lower SES to HIV testing.

Conclusion. Offering an HIV risk self-assessment tool through a digital platform can be an important strategy to effectively communicate personalised information on HIV risk to a broader audience and further encourage those who are at high risk to seek HIV self-testing. However, focusing on public health policies alone will likely be insufficient to address the impact of social determinants of health on the burden of HIV; social policies to improve housing and promote general education are warranted in the fight towards the elimination of HIV in South Africa.

Résumé

Contexte. Le recours au dépistage du virus de l'immunodéficience humaine (VIH) demeure inégal en Afrique du Sud : les individus avec un statut socio-économique (SSE) faible étant moins susceptibles de se faire dépister pour le VIH. En 2017-2018, HIVSmart!, un programme d'auto-dépistage du VIH assisté par une application digitale, a été évalué dans des townships en Afrique du Sud dans le cadre d'une étude quasi-randomisée contrôlée. L'application HIVSmart! comprenait une cote de risque du VIH qui n'avait pas encore été validée. Il est également difficile de savoir si le risque de contracter le VIH dans les townships sud-africains diffère selon des facteurs circonstanciels tels que l'éducation, l'emploi, le type de logement et le revenu. Mieux comprendre l'impact des facteurs circonstanciels sur le VIH et les inclure dans un outil d'évaluation du risque de contracter le VIH pourrait améliorer le recours au dépistage parmi les plus touché.es par le VIH.

Objectifs. Dans ce mémoire par articles, j'ai utilisé des données secondaires de l'étude quasi-randomisée contrôlée du HIVSmart! pour : 1) évaluer l'exactitude de la cote de risque de l'application HIVSmart! ; 2) développer et valider un modèle pour l'évaluation du risque du VIH pour les populations des townships sud-africains et ; 3) quantifier l'association entre les variables circonstancielles et les nouvelles infections au VIH en tenant compte des facteurs de risque individuels et comportementaux.

Méthodes. Dans ce mémoire, j'ai évalué la cote de risque du HIVSmart! par la régression logistique. Ayant recours à la projection prédictive bayésienne, j'ai identifié des prédicteurs du VIH afin de construire un modèle d'évaluation de risque que j'ai validé avec des données externes. J'ai par la suite évalué la performance du modèle en combinaison avec un programme d'auto-dépistage digital pour le VIH. Enfin, j'ai estimé l'impact des facteurs circonstanciels sur le VIH dans les townships sud-africains par la régression logistique multivariée.

Résultats. Mes analyses incluent 1535 participants du groupe d'auto-dépistage et 1560 du groupe de dépistage conventionnel de l'étude HIVSmart!. La cote de risque du HIVSmart! axée sur les facteurs de risque comportementaux était associée au VIH (rapport de cote (OR), intervalle de crédibilité (ICr) de 89 % : 1,15, 0,98 - 1,35) mais avait une capacité prédictive modeste (aire sous la courbe récepteur-opérateur (AUC-ROC), 0,62). J'ai développé un modèle ayant une performance améliorée (AUC-ROC, 89 % CrI : 0,71, 0,68 - 0,71) qui comprenait des prédicteurs sociodémographiques et comportementaux tels que le célibat, les antécédents de dépistage du VIH, la présence de relations sexuelles avec un partenaire vivant avec le VIH, le type de logement et l'éducation. Le modèle d'évaluation du risque de contracter le VIH avait une sensibilité de 91,1 % et une spécificité de 13,2 %. En combinant les résultats d'un programme d'auto-dépistage digital, la spécificité du modèle a été améliorée à 96,1 % tandis que la sensibilité est restée relativement semblable à 90,9 %. De plus, le fait de vivre dans des foyers/logements informels, ou d'avoir un niveau de scolarisation inférieur avait augmenté le risque de contracter le VIH (OR corrigé, 89 % CrI : 1,33, 1,05 - 1,67 et 1,75, 1,24 - 2,50, respectivement), après avoir contrôlé pour les facteurs individuels et comportementaux. Toutefois, la probabilité d'avoir été récemment testé pour le VIH ne différait pas selon les facteurs socioéconomiques.

Discussion. Il s'agit du premier outil ayant été validé pour l'évaluation du risque de contracter le VIH et ayant été développé spécifiquement pour les populations des townships sud-africains ainsi que l'un des rares outils qui utilise des variables circonstancielles. L'intégration des déterminants sociaux de la santé tels que l'éducation et le type de logement dans les outils de stadification du risque reconnaît l'impact des iniquités de la santé. De plus, proposer une évaluation du risque du VIH et un auto-dépistage via une plateforme digitale telle que HIVSmart! peut améliorer le recours au dépistage parmi les personnes disproportionnellement touchées par le VIH. Mes résultats suggèrent également que même si les personnes vivant dans des logements informels et ayant un niveau inférieur de scolarisation sont les plus touchées par le VIH, les services actuels de dépistage et de prévention du VIH peuvent être insuffisants pour remédier aux

iniquités de la santé, car ils ne relient pas davantage de personnes avec un SSE inférieur au dépistage du VIH.

Conclusion. Offrir un outil d'auto-évaluation du risque de VIH à travers une plateforme digitale peut être une stratégie importante pour communiquer le risque de contracter le VIH de manière efficace à un public plus large et d'encourager davantage les personnes à haut risque à recourir à l'auto-dépistage du VIH. Par contre, se concentrer uniquement sur les politiques de santé publique ne sera probablement pas suffisant pour remédier à l'impact des déterminants sociaux de la santé sur le fardeau des infections au VIH : des politiques sociales visant à améliorer les logements et à promouvoir la scolarisation sont nécessaires dans la lutte pour l'élimination du VIH en Afrique du Sud.

Acknowledgements

I would first like to express my sincere gratitude to my supervisor Dr. Nitika Pant Pai for welcoming me into her lab and trusting me with this unique and exciting project. Thank you as well for sharing your expertise and offering me countless opportunities to grow as a young researcher.

My deepest thanks to my co-supervisor Dr. Sahir Bhatnagar for his expert guidance on the data analysis and invaluable inputs at all stages of this project. Thank you for always being readily available to answer my questions.

I am also thankful to Dr. Susan J. Bartlett, who served on my supervisory committee, for her valuable feedback and support throughout this project.

I am grateful to Janet Fu, from the Bhatnagar lab, for her indispensable help with the data analysis. Thank you as well to members of the Pant Pai lab, particularly Olivia Vaikla, for keeping things running smoothy and for proofreading the manuscripts, and Fio Vialard, for being a reliable and authentic colleague as well as climbing partner.

I also wish to acknowledge all research collaborators and participants of the HIVSmart! trial, without whom this work would have never seen the light.

Finally, I am profoundly thankful for the unwavering support of my family and friends. I am grateful to my parents, for being the exemplification of hard and smart work, and to my partner, with whom I was able to share this journey.

Contribution of Authors

Cindy Leung Soo (CLS) performed the literature review and wrote all parts of this manuscript-based thesis. Dr. Nitika Pant Pai (NP) reviewed and provided feedback on all thesis chapters. Dr. Sahir Bhatnagar (SB) and Dr. Susan J. Bartlett (SJB) contributed crucial inputs to this thesis project.

For manuscripts 1 and 2, CLS analysed the data and wrote the first draft of the manuscripts. SB provided critical support in the data analysis. SB, SJB and NP critically revised the first draft of the manuscript. NP, Dr. Aliasgar Esmail (AE) and Dr. Keertan Dheda (KD) conceptualised and designed the original HIVSmart! research study.

List of Tables

List of Figures

List of Abbreviations and Acronyms

AGYW	Adolescent girls and young women
AIC	Akaike information criterion
AIDS	Acquired immunodeficiency syndrome
ANOVA	Analysis of variance
ART	Antiretroviral therapy
ARV	Antiretroviral drug
AUC-ROC	Area under the receiver operating characteristic curve
BIC	Bayesian information criterion
CICT	Client-initiated counseling and testing
COVID-19	Coronavirus disease
CrI	Credible interval
ER	Error rate
HIPPA	Health Insurance Portability and Accountability Act
HIV	Human immunodeficiency virus
HIVST	HIV self-testing
IQR	Interquartile range
LASSO	Least absolute shrinkage and selection operator
LMIC	Low and middle income county
MICE	Multiple imputation by chained equations
MSM	Men who have sex with men
NPV	Negative predictive value
OR	Odds ratio
PICT	Provider-initiated counselling and testing
PLWH	People living with HIV
PPV	Positive predictive value
PrEP	Pre-exposure prophylaxis
PWID	People who inject drugs
RCT	Randomised control trial
RDT	Rapid diagnostic test
SD	Standard deviation
SES	Socioeconomic status
STI	Sexually transmitted infection
TB	Tuberculosis
UK	United Kingdom
UNAIDS	Joint United Nations Programme on HIV/AIDS
USA	United States of America
WHO	World Health Organization

Chapter 1: Introduction

South Africa has the largest human immunodeficiency virus (HIV) burden in the world with 7.8 million people affected by the virus, 8% of which are unaware of their seropositive status[1]. Within the country, the distribution of HIV remains unequal; while the national HIV prevalence is 11%, the prevalence among people living in townships, urban settlements originally designated during apartheid for non-whites only, is 21%[2].

To end HIV by 2030, ensuring high testing uptake and diagnosing people living with undetected HIV (PLWH) is the first and most challenging step in reaching the UNAIDS 95-95-95 targets (95% of PLWH tested and diagnosed, 95% of those diagnosed positive are treated and 95% of those on treatment are virally supressed)[3,4].

HIV self-testing, a process by which a person performs their own HIV rapid test and interprets their result in the comfort of their own private space[5], was recommended by the World Health Organization (WHO) in 2016 as a complement to conventional testing[6]. Self-testing is predicted to be one of the most impactful and cost-effective strategies to increasing testing uptake by being convenient, private and highly acceptable[7,8]. Following the WHO pre-qualification of an HIV self-testing kit, the South African national Department of Health recommended self-testing to "improve HIV testing among the historically HIV under-tested, test-averse and hard to reach groups"[9,10]. Self-tests for HIV are today available on the South African market as well as through several community-based programmes[9].

Self-testing can be either supervised or unsupervised; in supervised self-testing, a healthcare worker is present with the tester to provide assistance if needed, while in unsupervised self-testing, individuals are tested at home or any private space of their choosing with little support[8]. The self-testing process involves performing an HIV rapid diagnostic test (RDT) and proactively seeking linkage to counselling, confirmatory testing or treatment[11]. Digital platforms can facilitate the HIV self-testing process by guiding result interpretation and linking individuals to care[12]. Additionally, digital platforms can be linked to an online or phone-based service which can provide assistance in the conduct and interpretation of a test, as well as link testers to a clinic[11].

HIVSmart! is one of the world's first app-based digital HIV self-testing programmes, which was developed into a programme of care for townships populations[13]. Funded by the Governments of Canada and South Africa, this app-based programme was evaluated in township populations of Cape Town in a quasi-randomised control trial (RCT) for impact and effectiveness[13]. HIVSmart! is an Android/iOS phone, tablet and web-based platform, with a secure dashboard, cloud server and a peer worker-based 24/7 counselling service[14-16]. It is available in many languages including Xhosa, Afrikaans, Swahili, Zulu, English, and French.

In the quasi-RCT, conducted over an 18 months period, our team has demonstrated that supervised and unsupervised HIV self-testing aided by a digital programme were effective in linking and retaining individuals in care, detecting new HIV infections, and encouraging testers to refer friends and family to HIV self-testing[13]. Qualitative findings from the trial further indicated that participants valued the convenience of self-testing with a digital platform[14,17]. HIVSmart! has been announced for adoption and scale up in Fast-Track Cities worldwide as a tool to end the HIV epidemic[18].

Despite the advent of decentralised testing options, testing uptake in South Africa remains well below targets with individuals of lower socioeconomic status (SES) or with low levels of education being less likely to seek HIV testing[19,20]. Stigma and discrimination are important barriers to testing which could be mitigated by targeted health education programmes[21]. Another barrier to HIV testing is that people who are at high risk of HIV do not perceive themselves to be at risk and thus, are less likely to seek testing[22-24]. Bridging this difference between perceived and actual HIV risk is crucial as it could help discourage risky behaviour, and perhaps increase the uptake of HIV testing as well as improve treatment adherence[25].

Risk staging tools for HIV can conveniently provide a personalised and objective assessment of HIV risk, at once raising awareness with respect to HIV risk factors and helping individuals form a more accurate portrait of their risk of HIV. Findings from qualitative studies indicated that individuals were more comfortable answering questions about sensitive topics such as sexual behaviours through a self-administered questionnaire

rather than through a healthcare worker that they did not know[17,26]. Others highlighted that risk staging tools could help initiate a discussion about HIV risk factors with their health provider[17,26].

Many risk staging tools for HIV have been developed and validated[27]. However, a majority of them were targeted to specific populations, developed for low prevalence settings or to be used in clinical settings for triage[27]. While invaluable for medical practice, these risk staging tools have limited potential in reaching and encouraging HIV testing among individuals who do not seek routine HIV testing, or do not avail to testing due to the perceived stigma and discrimination of showing up in healthcare facilities.

In 2021, the WHO published their Global Strategy on Digital Health in which they "advocate for people-centered health systems that are enabled by digital health"[28]. In light of the COVID-19 pandemic, the value of digital innovations for providing people-centered care is being increasingly recognised, and digital risk staging tools have gained in relevance. Digital risk staging offers many advantages compared to paper-based approaches; they can ensure privacy and confidentiality through self-administration, they can be made accessible outside of clinical settings, and most crucially, they can facilitate linkages to self-testing options.

Rationale

HIV testing services have been shown to be most cost-effective and impactful in reaching undiagnosed people living with HIV when targeted to people or areas with the greatest HIV burden[27]. Indeed, a validated digital HIV risk assessment tool relevant for township populations of South Africa could help link high risk individuals to self-testing options by empowering people with personalised information about their HIV risk. However, there is little evidence on the implementation of risk staging in combination with a digital self-testing programme. Additionally, there are currently no HIV risk assessment tool developed for township populations of South Africa[27].

The HIVSmart! application, recently evaluated in South Africa, contained an HIV risk self-assessment questionnaire which was piloted among participants[13], but all

participants proceeded to testing regardless of their risk levels[14]. While qualitative findings indicated that participants enjoyed interacting with the risk assessment tool[14,17], the accuracy of the HIVSmart! risk levels in predicting new infections was not evaluated.

Understanding the impact of contextual variables on predicting new HIV infection and incorporating them into a risk staging model could help improve the predictive accuracy of the risk assessment tool, and facilitate targeted HIV self-testing initiatives to vulnerable populations that are at greater risk of HIV. Indeed, a digital HIV risk assessment tool could be used as an educational tool to encourage people to self-stage their risk and consider getting tested in the comfort of their home, a clinic based kiosk or any private space, assisted by a app-based programme that could be operated out of their mobile devices.

In South Africa, the unequal burden of HIV parallels socioeconomic inequities; studies have indicated that individuals who are socially and economically disadvantaged are at increased risk of HIV[29-32] and some have attributed this to an increase in high risk behaviours among those of lower SES[33,34].

As we scale up HIV prevention interventions to improve testing uptake among people who are most affected by HIV, we also ought to better understand how socioeconomic indicators impact HIV acquisition and detection of new infection. In the South African context, it remains unclear the extent to which contextual factors such as education, employment status, dwelling situation, and income impact the likelihood of HIV infection, after accounting for behavioural factors. Evaluating the role of contextual factors on HIV infection in South African township populations could also help inform social policies aimed at mitigating health inequities.

Objectives

Thus, there are three specific objectives to this thesis:

1. To determine the accuracy of the HIVSmart! digital risk level in predicting new HIV infections in township populations of South Africa;

2. To develop a risk staging model for South African township populations and validate it in combination with a digital self-testing programme; and
3. To quantify the association between contextual variables (subdistrict of residence, education, dwelling situation, employment and income) and detection of new HIV infections while accounting for individual and behavioural factors.

I hypothesize that the HIVSmart! risk level are correlated with new HIV infections, and that incorporating contextual variables such as education, employment status, dwelling situation and income would improve its predictive accuracy. I further hypothesize that the subdistrict of residence of participants would be associated with new HIV infections, and that individuals of lower SES would have increased likelihood of HIV after controlling for individual and behavioural factors.

Chapter 2: Literature Review

In part I of this literature review, I contextualise the HIV epidemic in South Africa by providing a brief background on HIV infection and epidemiology, as well as the history of the South African response to the HIV epidemic. Then, in part II, I review the current evidence on determinants of HIV in South Africa. In part III, I discuss the HIV care pathway, and highlight the current barriers to the utilization of HIV services in South Africa. Finally, in part IV of this literature review, I discuss key strategies to mitigate the barriers to the utilization of HIV services, and review the current evidence on HIV self-testing, digital innovations supporting HIV services, and HIV risk assessment tools.

Part I: HIV Infection and Epidemiology

HIV infection

Human immunodeficiency virus (HIV) is a virus that infects immune cells. If untreated, an HIV infection can lead to acquired immunodeficiency syndrome (AIDS). While there are currently no cures for HIV/AIDS, treatment can allow people living with HIV (PLWH) to live long and healthy lives[35].

HIV can be transmitted through sex, sharing needles, syringes or other drug injection equipment, contact with infected blood as well as perinatally[35]. Flu-like symptoms generally appear within four weeks of an exposure with HIV which corresponds to the acute phase of the HIV infection[35]. Upon initial infection, the virus establishes itself in lymphoid organs from which it disseminates throughout the lymphoid system[36]. A strong immune response ensues but typically fails to completely eliminate the virus[36]. As a result, HIV is able to establish a persistent infection in the lymphoid organs where the virus continues to replicate. The resolution of symptoms indicates the transition from the acute phase of the infection to the chronic or latent phase[36].

The chronic phase of HIV infection is characterised by a lack of symptoms during which HIV remains active and replicating. This leads to the progressive depletion of CD4 T cells, the destruction of lymphoid tissue and ultimately, to extensive immune

dysfunction[36]. Without appropriate treatment, the HIV infection will progress to AIDS, the advance stage of the disease[35].

AIDS is characterised by profound immunosuppression which leaves the individual vulnerable to opportunistic infection. The median time from an initial HIV infection to AIDS, if untreated, is eight to ten years, but in about 10% to 15% of infected individuals, the progression to AIDS will occur within two or three years, and in less than 5% progression of disease will not occur for an extended period of time[36]. The life expectancy for a person with AIDS is three years[35].

Antiretroviral therapy (ART) can stop the progression of a latent HIV infection to AIDS and considerably extent the life expectancy of a person living with HIV[37]. Today, a person living with HIV who is on ART can expect to live almost as long as person who is HIV negative, depending on how early after initial infection they have accessed treatment[38]. Early detection and treatment of HIV is therefore key to enhance survival on ART. Furthermore, sustained treatment can lead to undetectable viral loads and render a latent HIV infection untransmissible[39].

Tuberculosis (TB) is the leading cause of death among people living with HIV[3]. In South Africa, TB is the primary co-infection and cause of death for those affected with HIV[40]. There are two primary ways in which TB and HIV interact biologically. At the individual level, the HIV virus attacks the immune cells which keep the TB bacteria latent, allowing the tuberculosis infection to progress to an active infection[41]. At a population level, the transmission of HIV and progression of disease increases the degree and proportion of immunosuppression in a population, making more people vulnerable to a tuberculosis infection[42]. Antiretroviral therapy can effectively limit both the spread of HIV and TB; between 2010 and 2019, South Africa saw a 77% reduction in tuberculosis deaths among PLWH largely thanks to improved access to ART[3].

HIV epidemiology

Acquired immunodeficiency syndrome (AIDS) was first identified in the United States (USA) in 1981 in young gay men and thus was initially thought to only affect men who have

sex with men (MSM), but reports of the disease in people who inject drugs (PWID), haemophiliacs and Haitian migrants to the USA emerged to counter the initial narative[43]. Since the clinical profile of AIDS took several years to develop following initial HIV infection, identifying HIV as the virus causing AIDS was challenging. The link between HIV and AIDS became more obvious as HIV was repeatedly isolated in AIDS patients[43].

From the early phase of the epidemic in the 1980s, the epidemiology of HIV markedly differed depending on the setting; in many industrialised countries, HIV cases were limited to MSM, PWID and their sexual partners, while in some developing countries in sub-Saharan Africa, the Caribbean, and Latin America, heterosexual transmission predominated[44].

History of HIV in South Africa

South Africa has one of the highest HIV burden in the world with an estimated 19.1% of the population living with HIV in 2020 (Figure 1B)[45]. The South African government, with the help of international partners, spends over US$ 2.5 billion annually on the world's largest HIV programme[1]. The government response to the HIV epidemic in South Africa has however not always been this committed; the early phase of the epidemic, in the 80s and 90s, was overlooked by the apartheid government, the Mandela government subsequently struggled with their response following a series of controversies, and, as the HIV epidemic continued to grow in the early 2000s, the Mbeki government's support for unorthodox AIDS denialist theories further undermined the response to the epidemic. As a result, HIV was able to spread in South Africa relatively unabated for 25 years. In 2009, a change of government brought urgency and renewed focus to the HIV epidemic, and the South African response took a drastic turn[43,46].

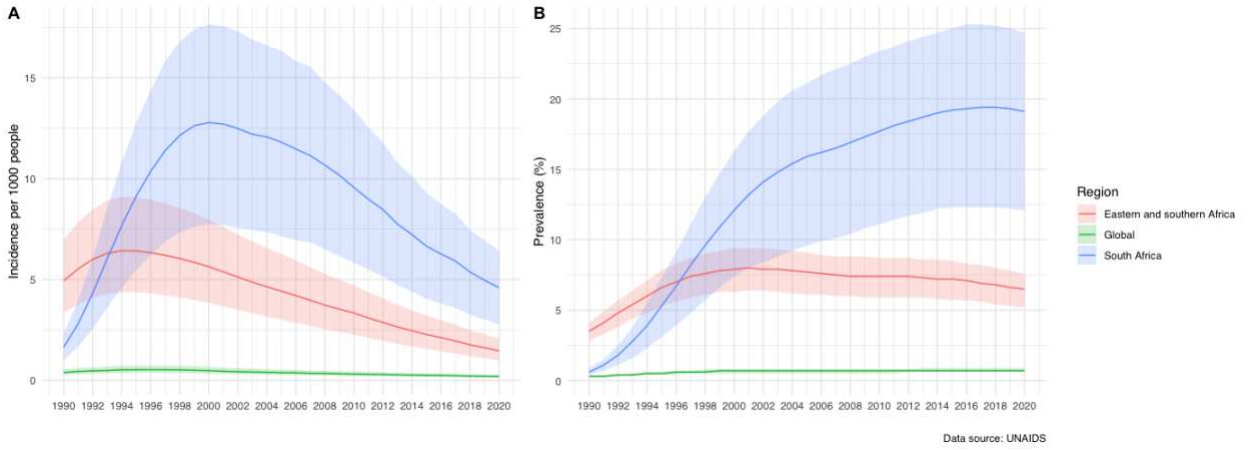


Figure 1. (A) Incidence of HIV per 1000 people and (B) prevalence of HIV (%) in South Africa, Eastern and southern Africa and Globally from 1990 to 2020.
The data from this graph was retrieved from the UNAIDS website[47].

The early phase of the AIDS epidemic in South Africa was limited to MSM, haemophiliacs, and foreign African mine workers[43,46]. Haemophiliacs were early victims of the epidemic as they routinely received unsafe blood transfusions derived from hundreds of blood donations. Following the implementation of routine HIV screening in donated blood, a safe blood supply was secured in 1985[43]. Heterosexual transmission quickly came to dominate as the main mode of transmission which gave rise, through perinatal transmission, to a concomitant HIV epidemic among newborns and young children in the early 1990s[43]. The apartheid government responded to the growing HIV epidemic by distributing condoms and providing "safe-sex" education which was seen as tainted, judgemental and stigmatizing[43,46].

In 1994, South Africa's first democratically elected government, led by Nelson Mandela, adopted the National AIDS Convention of South Africa Strategic Plan and prioritised it as one of the Reconstruction and Development Programme Lead Projects[43]. However, scandals involving a cabinet support of an industrial solvent as a cure for AIDS and the government's refusal to fund antiretrovirals (ARVs) to prevent mother to child transmission of HIV subsequently undermined the government's response[23]. During Mandela's time in office, the prevalence of HIV more than doubled, going from 3.9% in 1994 to 9.6% in 1998 (Figure 1B).

The second South African democratic elections in 1999 elected the Mbeki government whose support for AIDS denialist theories heavily undermined its response to the epidemic[43]. Despite trials demonstrating the effectiveness of antiretroviral drugs in preventing perinatal transmission of HIV, the South African government refused to provide them to pregnant women[46]. This decision was challenged in court and in 2001, the Constitutional Court ruled against the government, instructing it to provide ARVs to all pregnant women living with HIV[43]. The number of new HIV infections in South Africa peaked around that time with an estimated 520,000 new infections per year in 2000 and 2001 (Figure 2A)[45].

In an unexpected turnaround, Mbeki's government decided in 2003 to expand HIV treatment beyond pregnant women and to offer ARVs in public health services across the country[43]. While antiretroviral therapies were offered for free as of 2004, relatively few were able to access the drugs due to complex bureaucratic procedures making them difficult to procure for public health facilities[23]. Towards the end of Mbeki's time in office in 2008, South Africa had 5.3 million individuals living with HIV and only 588,000, representing 35% of the then eligible population, were receiving treatment (Figure 2B)[48].

Following a change in government in 2009, the new appointed Minister of Health Dr. Aaron Motsoaledi set the tone of the response to the HIV epidemic for the years to come. Motsoaledi began his mandate by acknowledging the chilling impact of HIV/AIDS on South Africa and by openly citing the extensive data on HIV as well as critical reviews of the national health system[46]. The government drastically increased funding to expand antiretroviral therapy access, scale up programmes for the prevention of mother-to-child transmission, promote the integration of HIV and TB treatment, and launch a massive HIV prevention campaign[48]. To facilitate access to treatment, bureaucracies and accreditation processes for health facilities to procure ARVs were stripped down, and nurses were trained to provide treatment to patients[46]. The South African government used international benchmarks to pressure drug manufacturers to reduce the prices of ARVs leading to a 40% decrease in the average cost of the drugs between 2010 and 2014[46]. Largely as a result of the expanded access to ART, the life expectancy in South Africa jumped from 53 years in 2004

to 64 years in 2020[49,50], with the number of AIDS-related deaths which peaked in 2006 with 270,000 deaths, being reduced to 83,000 deaths per year in 2020 (Figure 2C).

However, the legacy of many years of inadequate response to the HIV epidemic can still be felt today as South Africa continues to harbour the largest HIV epidemic in the world. In KwaZulu-Natal, the country's most affected province, a 15 year-old girl in some communities still has an 80% risk of acquiring HIV in her lifetime[49].

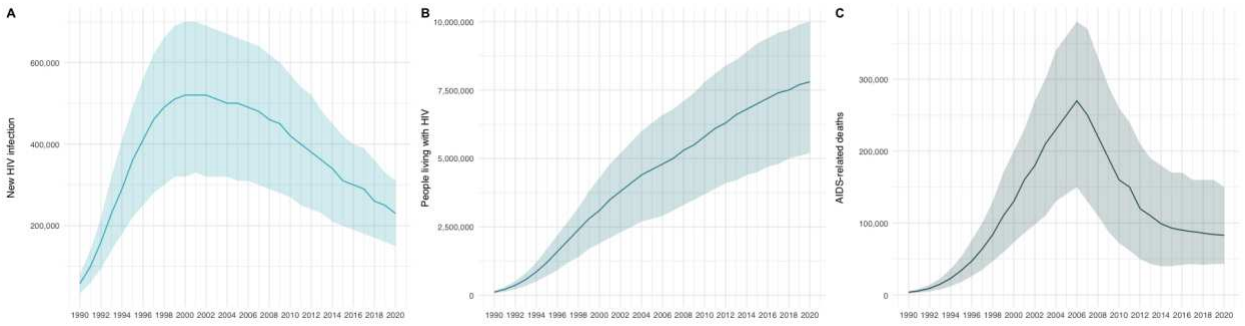


Figure 2. (A) Number of new HIV infection, (B) people living with HIV, (C) AIDS-related deaths in South Africa from 1990 to 2020.
The data from this graph was retrieved from the UNAIDS website[47].

Part II: Determinants of HIV in South Africa

The burden of diseases is unequally distributed in South Africa with economically disadvantaged and rural provinces often showing relatively poorer health and slower progress since the 1990s[51]. The concurrent HIV and TB epidemics are major contributors to the overall burden of disease and remain important drivers of population health trends. While inequalities in mortality and life expectancy in South Africa widened between 1990 and 2007, these health inequities have been on a downward trend since 2007[51]. Nevertheless, the burden of HIV remains unequally distributed across sociodemographic factors as well as geographically[52].

Young women are typically more vulnerable to sexually transmitted infections (STIs) and their complications relative their male counterparts due to a greater biological suceptibility[53], but also to gender-based violence which has been shown to increase the risk of acquiring HIV[3]. The effects of gender-based violence are particularly impactful in sub-

Saharan Africa where adolescent girls and young women (AGYW) account for a quarter of HIV infections despite representing only 10% of the population[3]. According to a UNAIDS report, 30.3% of South African women aged 15-49 reported having experienced physical or sexual violence from a male intimate partner in the past year[1]. Overall, South African women are disproportionately affected by HIV with 20.6% of women living with HIV compared to 11.5% of men affected by the virus[52].

Previous studies at testing sites in sub-Saharan Africa indicate that individuals who are unmarried are generally more likely to be found HIV positive than married individuals[53,54]. It has been suggested that this could be due to increased access to sexual and reproductive care services for married women[53]. A study found that in South Africa, unmarried or women who were not cohabiting with a partner had 2 to 3 times the risk of HIV[55].

Sexual behaviours can certainly impact the risk of HIV as it is a sexually transmitted infection. The risk of acquiring HIV from a single sexual contact varies depending on the infectiousness of the HIV positive individual and the susceptibility of their sexual partner, but it is generally low relative to other SITs[56]. Repeatedly engaging in high risk sexual behaviours can however increase the likelihood of acquiring HIV. High risk sexual behaviours include unprotected sex, having multiple sexual partners, and having sex with a commercial sex worker[32]. While alcohol consumption during sex has been associated with an increased risk of HIV infection[57], it has been suggested that it is due to intersecting socioeconomic factors such as migrancy[32,58].

Migrancy is very common in South Africa as people migrate from rural to urban areas in search for work[59]. The urbanisation process is often unplanned, leading to the creating of "informal settlements" generally defined as "areas containing unregulated and unplanned dwellings where inhabitants lack secure tenure and thus generally lack access to adequate basic services prior to government intervention"[59]. In South Africa, informal settlements can be traced back to the apartheid regime which confined a majority of the population of colour to poor and underserviced living environments[59].

The evidence on the impact of socioeconomic status (SES) on HIV in sub-Saharan Africa has been mixed. Studies conducted earlier in the epidemic associated higher risk of HIV to individuals of higher SES while more recent studies report an opposite relationship[60]. It has been argued that in the early years of the HIV epidemic, the relatively rich and educated were more vulnerable to HIV, having higher rates of partner change due to greater autonomy and mobility[60]. The scarcity of information on HIV and how to prevent infection further added to the challenge of reducing transmission in the early days of the epidemic[60]. As the epidemic progressed and information about how to prevent HIV became more readily available, it has been suggested that those of higher SES adopted safer sexual practices[60,61]. People of lower SES however remain vulnerable to stressors such as poverty or food insecurity which increased sexual risk taking and placed them at higher risk of infection[60]. Indeed, findings from Booysen and Summerton indicated that individuals who are economically disadvantaged were more likely to engage in risky sexual practices and were less likely to be knowledgeable about HIV[33,34].

In South Africa, recent studies indicated that educational attainment has a protective effect on HIV. A 2017 study conducted in the South African province of Western Cape found a 10% decrease in HIV prevalence for every additional year of schooling[62]. Moreover, in the province of KwaZulu-Natal, every year of education decreased the risk of HIV infection by 7%, after adjusting for sex, age and other sociodemographic factors[29]. It has been suggested that education can empower individuals to adopt HIV risk-reducing behaviours through an improved ability to comprehend health promotion messages[29,60]. Housing conditions also appeared to be associated with HIV infection; in South Africa, individuals living in informal dwellings had consistently greater risk of acquiring HIV relative to those living in formal urban areas[63,64].

While previous studies consistently report on the protective effect of education and stable dwelling situation, there has been no consistent association between employment status and HIV infection[32,62]. It also remains unclear how wealth affects the risk of HIV; a large multi-country study found no statistical association between wealth and HIV status, after controlling for urban residence, age, education and sexual behaviours[65]. Another

multi-country study in sub-Saharan Africa found that in wealthier areas, the relatively poor were most affected by HIV whereas in poorer settings, the relatively wealthy were most affected[66]. This indicated that individuals of median wealth were the most impacted by the HIV epidemic.

Altogether, the evidence on how socioeconomic status affect HIV infection is still evolving. Recent evidence appear to more strongly support a protective effect of education on HIV acquisition and highlight the harmful effects of living in informal housing in terms of HIV infection. The effects of employment, income or wealth on HIV, however, remains unclear. Many have previously hypothesized that socioeconomic stressors could compel individuals of lower SES to adopt riskier sexual behaviours including transactional sex[33,34,60], but in the face of the constantly evolving HIV epidemic there is little evidence on how socioeconomic indicators affect HIV when controlling for behavioural and demographic factors.

Part III: HIV Care Pathway

HIV prevention

Demographic realities place a particular urgency in ramping up prevention efforts for HIV in sub-Saharan Africa; in the coming years, a large cohort of young people will enter adolescence and young adulthood and become sexually active, increasing their risk of acquiring HIV[67]. Successfully tackling the HIV transmission requires a combination of structural, behavioural and biomedical approaches such as regular testing and counseling for HIV and other STIs, voluntary medical male circumcision, comprehensive sexual education, pre-exposure prophylaxis (PrEP) and antiretroviral drugs[37].

One of the most cost-effective intervention to limit HIV infection is condom use[37]. It is estimated that the increase use of condoms has averted 117 million HIV infections since 1990[3]. However, consistent condom usage has been difficult to achieve in some contexts, especially among women in low and middle income countries (LMICs)[1]. Requesting condom use is viewed as a sign of mistrust and suspicion of affair and is thus impractical within a marriage or committed relationship[68]. Additionally, women who experience

intimate partner violence are often not able to negotiate condom use. Indeed, when deciding whether to adopt certain HIV prevention behaviours such as condom use, individuals often have to balance competing priorities including preserving a relationship, earning money, being a good parent, and conforming to accepted social norms[68].

Another effective strategy to combat the spread and impact of HIV is to combat perinatal transmission of the virus through mother-to-child HIV transmission interventions[69]. Indeed, the risk of mother-to-child transmission of HIV can be considerably reduced by providing ART during and after pregancy[70]. In South Africa, the prevention of mother-to-child transmission programmes have reduced the annual number of HIV infected newborns by 70% between 2010 and 2020[3].

Comprehensive sexual education in schools is another strategy that has been implemented in South Africa to prevent HIV and other STIs. While there is little evidence regarding the effectiveness of these programmes in reducing the incidence of HIV, they have been effective in promoting risk reduction behaviour such as condom use[71].

Daily oral pre-exposure prophylaxis are drugs which have been shown to be well-tolerated and efficacious at preventing HIV infection[72]. PrEP is one of the more costly prevention interventions[69], but can be made more cost-effective if targeted to high-risk populations[73]. Despite this, much of its rollout has been concentrated in a limited number of countries, notably, the United States, Kenya and South Africa, and in 2020, only 28% of the target 3 million high-risk individuals in LMIC had access to PrEP[1,37]. While targeted programmes that offer routine HIV testing have been largely successful in integrating PrEP in their provision of services, primary care settings for the general population have not been able to effectively implement this service due to the high patient flow and time constraints[74].

Overall, preventing new HIV infections has been challenging with the annual number of new infections amongst adults having barely changed in the past four years[3], yet modelling findings suggest that scaling-up prevention interventions will be very impactful in accelerating the decline of HIV incidence and that innovating current programmes will further improve their public health impact[75].

HIV testing

HIV testing services aim to deliver a diagnosis, and to facilitate access and uptake of HIV prevention treatment and care. In the past 15 years, the global scale-up of HIV testing services has had a substantial impact. Indeed, in South Africa, the proportion of people living with HIV who were unaware of their status has declined from over 80% in the early 2000s to 8% in 2020[1,76]. Rapid diagnostic tests (RDTs) have played a major role in improving access to HIV testing through increasing task-sharing which allowed lay providers to deliver HIV testing services outside of laboratory settings[3]. Nevertheless, gaps in testing remain today, especially among key populations, as well as men, adolescents and young adults[5]. To end the HIV epidemic, reaching out to these populations is imperative.

The implementation of provider-initiated counseling and testing (PICT), where providers are encouraged to offer HIV testing to all patients presenting at health facilities, has been very successful in improving HIV testing uptake[77]. However, findings from a survey indicated that PICT is unpopular in South Africa with only 11.6% of participants preferring this practice compared to 66.1% preferring client-initiated counseling and testing (CICT)[78]. Many have reported that they felt pressured into being tested for HIV or were not given the opportunity to decline while other indicated that they believed opting out of testing would compromise the care that they received[78]. While PICT has successfully improved testing uptake, its impact on finding undiagnosed PLWH and linking them to care is unclear as reports suggest that a large proportion of testers recruited through PICT do not return to collect their test results[78].

In LMIC, women are more likely to get tested for HIV then men, largely thanks to the success in integrating HIV testing into reproductive health services including antenatal care[6]. As a consequence, men in sub-Saharan African countries are less likely to know their HIV status and typically have higher HIV mortality rate than women[6,76,79]. In South Africa, HIV testing also appears to be higher among married or cohabiting individuals. In addition, men with a secondary or tertiary level of education were twice as likely to test for HIV[80].

Repeat testing is an important strategy to address the HIV epidemic as it often permits early detection of the virus and early initiation of ART. Indeed, repeat testers have been shown to have higher CD4 T cell counts when an HIV infection is diagnosed, an earlier start of antiretroviral therapy and reduced mortality[81]. Early detection of the virus through repeat testing may also reduce transmission of the virus through engaging more people in ART which lowers viral load and infectivity. In South Africa, repeat testing was more common among younger individuals, women and those who knew a person living with HIV[82]. The rate of repeat testing in South Africa appears to vary considerably according to the setting with studies reporting between 26% and 71% of people seeking HIV testing had previously tested for the virus[82]. Furthermore, knowledge of one's HIV status has been associated with reduced risk behaviours[56]. As such, early diagnosis of people living with HIV will likely also lead to a reduction of onward transmission.

HIV treatment

HIV can be treated using a combination of three or more antiretroviral drugs (ARVs) which serve to supress viral replication during the chronic phase of the infection and allow the body to regain the capacity to fight opportunistic infections[37]. With antiretroviral therapies (ART), the progression of and HIV infection to AIDS and the death of HIV-infected patients is significantly delayed. ART is however not a cure; it cannot rid the body of latent HIV and treatment is required for life to control the infection[43].

Furthermore, individuals on ART who are virally supressed (HIV RNA copies < 1000/ml) cannot transmit HIV[39,83], making early initiation of ART for infected individuals an important strategy to mitigate the spread of HIV[3]. This has had a significant impact on HIV transmission as a study by the Africa Centre found that those living in communities where 30-40% of PLWH were treated were 38% less likely to acquire HIV that those living in communities with fewer than 10% coverage in HIV treatment[84].

In the past, provision of ART was delayed until a person living with HIV had CD4 cell counts that were low enough to be deemed eligible for treatment[85]. With the mounting evidence on the benefits of early initiation of ART and its potential in preventing HIV

transmission, the WHO amended its recommendations in 2015 to recommend that all persons diagnosed with HIV be offered ART regardless of CD4 levels[86].

Over the past 20 years, the global rollout of ART has saved an estimated 16.6 million lives and contributed to a 47% decreased in AIDS-related deaths since 2010[3]. In 2020, it is estimated that 73% of PLWH were on treatment[37]. Despite the promise of ART, even the most successful antiretroviral therapies cannot fully prevent all clinical event, especially when initiated in advanced HIV disease. In approximately 10-15% of individuals who initiate therapy in advance stages of the HIV infection, progression to AIDS can still occur[43]. Thus, early diagnosis of HIV remains key to reduce AIDS-related mortality. Incomplete adherence to ARTs is the primary cause of failure to achieve viral suppression. Although studies have demonstrated that adherence is possible when patients are supported, poor adherence to HIV treatment remains of particular challenge in the developing world[43].

Barriers to utilization of HIV services

While engagement in HIV care in South Africa has improved in recent years[4], there remain numerous barriers to the utilisation of HIV prevention, testing and treatment services as well as high heterogenicity in the uptake of HIV services (Table 1)[82].

In South Africa, 20% of women and 16% of men live in areas where the nearest healthcare facility is over 30 minutes away[52]. Additional structural barriers include a lack of human resources, long queues, a lack in confidential spaces, and negative patient-provider relationships (Table 1)[87-89].

At the individual level, people may be unaware of available HIV services, lack the resources to travel frequently, fear positive results, or worry about lack of confidentiality, stigma or a violent reaction of a partner if they seek HIV prevention and care services[87-90]. A study among HIV testers found that those who believed the test results to be confidential were significantly more likely to accept testing despite their held beliefs on HIV[91].

Additionally, low risk perception can also contribute to an underutilisation of services as well as decreased adherence to treatment[87,88]. HIV risk perception is complex; at the individual level, it may reflect people's knowledge or beliefs about HIV, whether they

know a friend living with HIV, and the trust that they place in their romantic relationships[22,68,92]. A systematic review indicated that risk perception of HIV decreased as trust grew in relationships which is also often accompanied by an abandonment of condom use as a sign of love and commitment[68]. Consequently, as suggested by many studies, people often do not have an accurate perception of their risk of HIV[22-24]. Bridging this gap in risk perception could encourage high risk individuals to seek testing.

Several of these barriers to utilisation of HIV services can also intersect with sociodemographic factors. In South Africa, misconceptions with respect to the cause of HIV/AIDS intersects with poverty[93].

Table 1. Barriers and suggested solutions to improved utilisation HIV services in South Africa

Barriers	Solutions
Structural 1. Limited access to healthcare facilities in underserved areas[52] 2. Staff shortages and long queues[87-89] 3. Lack of confidential spaces[87,88] 4. Negative patient-provider relationship[87,88] **Individual** 1. Lack of awareness or understanding of available interventions[82,88,89] 2. Fear of an HIV positive result[89,90] 3. Anticipated HIV-related stigma when seeking healthcare services[87-90] 4. Actual or anticipated violent reaction from male partners (women)[89] 5. Concerns about lack of confidentiality and privacy[78,82] 6. Low HIV risk perception[87,88] 7. HIV medication side effects[87,88]	1. Improving access to decentralised HIV testing options (home-based testing, self-testing) 2. Providing health education programmes and community-based health promotion activities[89] 3. Changing social norms related to HIV risk and protective behaviours[94] 4. Promoting social support from family[89] 5. Offering an HIV risk self-assessment tool to inform HIV risk[95] 6. Using digital platforms to support HIV self-testing[12] or improve treatment adherence[96]

Part IV: Strategies to improve the utilisation of HIV services

Self-testing

An key strategy to improve uptake of HIV testing is HIV self-testing (HIVST). HIVST is a process by which a person collects their own specimen (either oral fluid or blood), performs an HIV rapid test and interprets their result[5]. HIVST has been deemed safe, accurate and an effective tool to reach individuals who have reduced access to HIV testing services[5,6]. An over the counter self-administered test for HIV has been approved since 2012 by the Food and Drug Administration of the USA[97], and since 2016, the WHO has recommended the implementation of HIVST as a strategy to expand access to HIV testing services[6]. A positive result from and HIV self-test requires confirmatory testing. As such, HIVST should be regarded as tools for triage with the strong potential to alleviate the burden at HIV testing facilities[6].

In qualitative studies, HIV self-testing was found to be convenient, private and confidential, and generally preferred over conventional HIV testing[8,98,99]. Acceptability of HIVST has been found to be consistently high[8,100] with many studies indicating that HIVST can double the uptake of testing compared to conventional testing[101,102]. Importantly, despite initial concerns regarding linkages to care[103], HIVST has been shown to achieve linkages rates which are at least as high if not better than conventional testing, irrespective of whether or not support for linkages was provided[12,101,102]. HIV positivity rates were also similar for HIVST while reported misuse of test kits, social harms or other adverse events have been rare[8,101,102]. Social harms such as intimate partner violence resulting from a positive test would not be unique to HIV self-testing although the context or setting of the test may alter these experiences[102].

It is estimated that the cost of HIV RDT for self-testing varies between US$7.50 and US$43 in high income countries and US$3 to US$16 in low and middle income countries which remain greater than facility-based rapid diagnostic testing (US$0.50 to US$11 per test kit)[6]. Despite the higher cost of self-testing strategies compared to conventional testing, modelling analyses suggest that, given its high uptake and potential to reach individuals

who would not otherwise have access to testing, HIVST will be one of the most cost-effective and impactful strategy to reduce HIV morbidity and mortality in the South African context[7].

A recent systematic review found that all HIVST distribution strategies showed higher HIV testing uptake compared to facility-based testing while having similar HIV positivity and linkages[104]. Testing uptake was highest when self-testing kits were distributed by antenatal clients to their partners suggesting an avenue for increasing HIV testing among men who do not routinely attend health services[104].

Indeed, HIVST can be an important strategy to encourage HIV testing among men who are generally less likely than women to be aware of their HIV status[76]. Qualitative studies report high acceptability of HIVST among men irrespective of their age, marital status, education level, socioeconomic status or prior awareness of HIVST[95]. Men primarily valued the privacy and confidentiality that self-testing offers[105], indicating that it was a welcomed alternative for those who wished to escape the visibility of seeking facility-based testing[106]. Men also preferred HIVST for its efficiency and convenience, characteristics that were particularly valued by those who worked long hours and found it challenging to fit a visit to the clinic in their schedule[105].

Overall, HIV self-testing is an key strategy to mitigate important barriers to the utilisation of HIV testing services such as large travel distances, long wait times, the visibility and anticipated stigma associated with attending sexual health clinics, and concerns about lack of privacy and confidenciality[107,108].

Digital innovations to support HIV services

Another way to address barriers to the utilization of HIV services is to leverage digital tools to support HIV prevention, testing and treatment. The COVID-19 pandemic has accelerated the digitization of provision of health services making the use of digital tools to close gaps in HIV service delivery all the more relevant.

Digital tools can play an important role in HIV prevention: they can be used for surveillance, to amplify prevention messaging, and to increase uptake of PrEP as well as

access to testing services[109]. Digital platforms can improve access to HIV services through providing a platform to request an HIV self-testing kit, and can facilitate communication and support from healthcare providers[109]. Moreover, social media-based prevention interventions were found to be effective in reducing risky sexual behaviours[110].

Digital platforms can also aid in the process of self-testing. A recent systematic review found that the use of digital platforms to support HIVST was acceptable to a majority of participants, and that social media and app-based programmes were generally preferred over web-based and SMS interventions[12]. Moreover, digital tools improved the uptake of HIVST with social-media based distributions of self-testing kits being effective at reaching high-risk population[12]. While there were initial concerns regarding the ability of self-testing programmes to link individuals to care and counseling[103], recent evidence suggest that HIV self-testing supported by digital tools can potentially improve linkages to care rates compared to conventional HIV testing[12]. Finally, digital HIV self-testing options were appealing to end users and were deemed effective in reaching vulnerable populations[12]. This evidence is key as the control efforts for the HIV epidemic shift to target vulnerable populations worldwide.

The evidence regarding the contribution of digital tools to support the provision of treatment for HIV is particularly robust. A collection of reviews, systematic reviews and meta-analyses concluded that digital interventions, specifically mobile health programmes, significantly improve ART adherence[111]. The strongest benefit was identified for programmes with two-way messaging which allowed for patients to interact with care providers[111]. This suggest that digital interventions are most impactful when they are used as links to the care pathway and a network of health providers rather than a stand-alone resource.

In 2019, 93% of the South African population was covered by 4G/LTE , making South Africa one of the most connected country in sub-Saharan Africa[112,113]. A digital divide is however present; individuals without secondary education or below the median income are approximately half as likely to be connected to the internet[113]. However, in the wake of the COVID-19 pandemic, the South African government has announced plans to connect every

community to the internet by 2024 and to provide free data to low-income households[112]. A key advantage of utilising digital platforms for health interventions is that once people are connected to the internet, digital technologies are very scalable[114].

Providing internet access to all South Africans is crucial for South Africa's national digital health strategy in which the government affirms its commitment to leverage "the potential of digital health technologies to improve the quality and coverage of healthcare, increase access to services and skills, and promote positive changes in health behaviours"[115]. Among the identified priorities, the government indicated its commitment to scale up mobile health initiatives intended to expand health promotion coverage[115].

HIV risk assessment tools

Finding those who are at high risk of HIV infection is particularly challenging in sub-Saharan Africa where the HIV epidemic affects a large proportion of the general population through heterosexual transmission[95]. Providing a risk assessment tool through a digital platform to the general population can help individuals better understand their risk level for HIV, and be a discrete, convenient and low-cost strategy to increase testing uptake among those who are at highest risk of infection. HIV risk assessment tools consist of guided questionnaires which tally specific risk factors into a global risk level and often use a combination of demographics, clinical examination findings, and risk behaviours to provide a risk assessment[27]. Providing individuals with a personalised assessment of their risk could encourage them to consider getting an HIV self-test which would discreetly link them to care if tested positive.

In low prevalence areas, targeted screening for the general population has been shown to be cost-effective; a study in France estimated that the incremental cost-effectiveness ratio of targeted screening vs non-targeted was US$1,476 per new HIV diagnosis[116]. Similarly, a study in Spain estimated that targeted screening using a risk assessment tool would save US$3,282 per person diagnosed with HIV compared with universal screening[117].

For high prevalence areas, on the other hand, using risk assessment tools to restrict access to HIV screening could lead to many undiagnosed people living with HIV being missed[27]. Rather than using these tools to limit access to testing, it has been suggested that risk staging tools be used to prompt discussions with respect to HIV prevention in order to encourage HIV testing among those who would not have otherwise tested[95].

A randomised control trial in China found that individuals who interacted with an HIV risk assessment tool had reduced their high-risk sexual behaviours, indicated increased willingness to test for HIV, and had higher engagement with general HIV prevention material relative to those who were only offered non-personalised HIV educational material[118]. Findings from this study indicate that allowing people to assess their own risk of HIV can ultimately lead to risk reduction behaviour and increased testing uptake.

Digital platforms could further make self-assessment of HIV risk more accessible and more acceptable. A qualitative study by Jones et al. found that a digital risk assessment tool was preferred over the paper-based version[119]. Further, some participants indicated that they would feel more comfortable filling the questionnaire than discussing discomforting topics, such as sexual behaviours, with a provider they did not know[119]. Others suggested that self-assessing their own HIV risk would help prompt a discussion about their risk behaviours with their provider[119].

Existing risk assessment tools for HIV. We surveyed 41 existing risk staging tools for HIV from the published literature (see Appendix, Table 2). A majority (28/41) of them were targeted to specific populations including 14 intended for men who have sex with men (MSM), 5 for paediatric populations[120–124], 4 for adult African women[55,125-128] and 2 for adolescent girls and young women (AGYW)[129,130]. Close to half (20/41) of the HIV risk assessment tools were developed for populations in sub-Saharan Africa (see Appendix, Table 2). Seven of the surveyed tools were developed to identify acute HIV infection[131–137]. Tools designed to identify acute and early HIV infection tend to rely more heavily on clinical signs and symptoms of HIV and be better at identifying acute HIV cases. On the

other hand, these tools would likely miss people with a latent HIV infection who typically do not exhibit symptoms.

Included predictors. While a majority of risk assessment tools included individual, clinical or behavioural predictors of HIV, only a minority utilised information on contextual factors such as social determinants of health (see Appendix, Table 2). Of the surveyed risk assessment tools for HIV, 63% (26/41) included individual predictors such as age, sex or marital status, 78% (32/41) included clinical predictors such as signs or symptoms for HIV or presence of other STIs, 71% (29/41) included behavioural predictors related to drug use or sex, and 29% (12/41) included socioeconomic indicators such as education level, employment status or income level (see Appendix, Table 2). Additionally, three risk assessment tools included data on HIV prevalence[57,126,138].

Six out of the 13 non-targeted risk assessment tools included contextual variables[57,138-143]. Including socioeconomic indicators is particularly relevant for non-targeted risk assessment tools as it acknowledges the different burden of HIV for individuals of different socioeconomic status. Moreover, in a community or outpatient setting, it may be easier to gather information on contextual factors than on clinical predictors such as presence of other sexually transmitted infections.

Development methods. Just under half (19/41) of the surveyed development studies reported how they delt with missing values: 14 analysed complete cases only, 4 imputed their missing data, and 1 encoded the missing values as 0 (see Appendix, Table 2).

To develop risk assessment models, 11 of development studies relied on prior knowledge of HIV or HIV epidemiology to select predictors, including one that relied solely on data from the literature and expert consultations[144], and one that based their tool on WHO recommendations[124]. A majority (24/41) of the studies selected variables for inclusion in their risk assessment model based on statistical association; many of which (19/24) defined specific p-value thresholds (see Appendix, Table 2).

Additionally, 18 studies used forward selection, backward elimination or stepwise selection approaches for model selection, with 10 studies using p-values as a stopping rule and 8 using Akaike information criterion (AIC) or Bayesian information criterion (BIC) to select a final model (see Appendix, Table 2).

The forward selection method starts with an empty model and progressively adds predictor variables based on a selected criteria. Predictors are added until none of the remaining variables meet the criteria for entry into the model[145]. The backward elimination approach starts with a full model with all variables, and predictors are progressively removed until none of the variables in the model meet the criteria for elimination[145]. In stepwise selection, predictors can be added or removed at every step to optimise a given criterion[145].

Akaike information criterion and Bayesian information criterion are such criterion which can aid in model selection when comparing models of different sizes; a preferred model is one with a smaller AIC or BIC. Both criterion are composed of two terms; one which is a measure of fit and another which penalizes the addition of variables. BIC will typically select smaller models than AIC.

Three development studies used least absolute shrinkage and selection operator (LASSO)[138,141,142], a method for variable selection, and one study identified candidate models for evaluating medical notes through random forest classifiers[146].

Validation. Of the 41 surveyed HIV risk assessment tools, 15 were externally validated (see Appendix, Table 2). An additional 3 were internally validated using a testing set, 3 were internally cross-validated and 2 were internally validated by bootstrapping (see Appendix, Table 2). The remaining 18 tools have only been internally validated, that is evaluated in the same data used to develop the model (see Appendix, Table 2).

Most validation studies reported the area under the receiver operating characteristic curve (AUC-ROC), a measure of how well a model can discriminate between cases and non-cases. An AUC-ROC value greater than 0.5 indicates that a model is better than random and a value of 1 indicates that a model can perfectly discriminate between HIV positives

and negatives. Many validation studies also reported the sensitivity and specificity of the tools.

Only two of the non-targeted risk staging tools for the general population in high prevalence settings were externally validated; one developed to identify acute HIV infection had an AUC-ROC of 0.79[134] and one for South African populations had an AUC-ROC of 0.68 for women and 0.71 for men[138]. Among externally validated tools for the general population in low prevalence settings, the AUC-ROC ranged between 0.70 and 0.84 (see Appendix, Table 2). The VOICE risk score, developed to assess the risk of HIV infection among African women, was externally validated six times with AUC-ROC ranging from 0.55 to 0.69 (see Appendix, Table 2).

Overall, two of the best performing externally validated tools were developed for screening in the general population of North America; the Denver Risk Score had an AUC-ROC of 0.80 when validated in a Canadian cohort[147], and a tool developed using electronic health records and machine learning had a performance of 0.84 in external data[142]. Other well performing tools during external validation were developed to assess HIV risk among MSM in China (AUC-ROC of 0.827)[148] or to identify cases of acute HIV infection among MSM (AUC-ROC of 0.88)[149].

Sensitivities of surveyed tools ranged from 42% to 96% and specificities from 6% to 97% (see Appendix, Table 2). The sensitivity and specificity of a risk assessment tool can vary depending on the cut-off value that defines high vs low risk. Given the high cost of missing a person living with HIV, choosing a cut-off value which favours a high sensitivity over a high specificity would be preferred. The chosen cut-off values could also be adapted depending on the local context: lower thresholds for high prevalence settings and higher thresholds for low prevalence settings[95].

Unpublished HIV risk staging tools. Ong et al. surveyed 80 unpublished risk assessment tools for HIV and found that a majority of them were being used in Africa, including 4 in South Africa[150]. The setting these tools were deployed in was not specified for a majority of the tools, but among those where the setting was known, most were being used in hospital

or community settings[150]. A quarter of the tools were developed for use in the general population and most were administered by providers or unspecified[150]. About 40% of the unpublished tools were developed in collaboration with the Ministries of Health or a funding agency while a minority were informed using studies from the published literature[150]. While close to half were reportedly evaluated, less than a quarter reported a positive outcome such as improved positivity yield or decreased congestion in clinics[150].

Summary

In the first part of this literature review, I have summarised the history of the HIV epidemic in South Africa. Many years and many administrations of inadequate response to the growing HIV epidemic has left the country with a unique burden of HIV. The HIV response took a sharp turn in 2009 with the election of a new government and since then has been exceptionally aggressive. As a result, South Africa today has one of the largest HIV prevention and treatment programme.

In the second part of this review, I have examined the evidence on how demographic, behavioural and socioeconomic factors impact the risk of acquiring HIV. While the impact of high-risk behaviours on HIV infection has been consistent, straightforward and well characterised, the impact of socioeconomic status on HIV has evolved with the epidemic. In the earlier days of the HIV epidemic, individuals of higher SES seemed to be most impacted, as the epidemic matured, this relationship appeared to shift. In light of the dynamicity of the epidemic, it is today unclear to what extent socioeconomic indicators impact HIV infection after controlling for behaviours.

In the third part of this literature review, I have discussed the HIV care pathway, and highlighted the gaps in service delivery and barriers to the utilization of HIV services in South Africa. While there are many HIV prevention initiatives, they have seen little success in reducing HIV transmission. In terms of testing, South Africa has made enormous progress, yet testing uptake remains below targets and unequal. Provision of treatment for HIV has been very successful in extending the life expectancy of South Africans, but retention in care might still be a challenge for many populations. Finally, many barriers

impede utilization of services, including fear of stigma and discrimination, worry about the lack of confidentiality and long queues at health facilities.

In the final part of this review, I have highlighted three solutions for addressing barriers to the utilization of services. First, self-testing is a highly acceptable, private and convenient strategy to improve testing uptake, especially among those who might not otherwise seek testing. Second, leveraging digital tools can improve patient engagement and experience at all stages of the care pathway, by improving the reach of HIV prevention, facilitating distribution of self-tests and linkages to care, and by improving adherence to HIV treatment. Finally, risk assessment tools for HIV can tackle low risk perception and encourage high risk individuals to test for HIV, but many of the surveyed existing risk staging tools are not suitable for township populations of South Africa. As such, the development and implementation of a risk staging tool for HIV that is relevant for South African township populations can be useful in achieving the first UNAIDS 95-95-95 target of diagnosing 95% of people living with HIV.

Chapter 3: An HIV Risk Assessment Tool for South African Township Populations

Preface to manuscript 1

In Chapter 2, we found evidence that testing uptake remains inadequate in South Africa, particularly among certain groups such as men and individuals of lower socioeconomic status (SES). Indeed, several barriers such as fear of stigma and discrimination, concerns about the lack of privacy and confidentiality, and long queues at health clinics impede access to HIV services. Many studies have also indicated that individuals who are vulnerable to HIV often underestimate their HIV risk, and are thus less likely to seek testing.

Through my literature review, I have found evidence that offering a digital tool for individuals to self-assess their risk of HIV reduced risk taking behaviours and increased self-reported willingness to seek testing. Further, qualitative findings indicated that people generally preferred learning about HIV risk factors through a digital risk self-assessment questionnaire, without fear of judgement, than discussing discomforting topics such as sexual behaviours with a health provider. As such, a risk assessment tool targeted to township populations of South Africa would be very relevant to raise awareness with respect to HIV risk factors and potentially increase testing uptake.

In Chapter 2, I have reviewed 41 existing risk staging tools for HIV and found none specifically targeted to township populations of South Africa. Moreover, none of the surveyed validation studies evaluated the added value of their risk assessment tool by examining it in combination with existing testing programmes.

As such, in manuscript 1, we set out to evaluate an existing risk staging questionnaire piloted during the HIVSmart! trial which had not yet been assessed for predictive accuracy, identify a parsimonious set of sociodemographic and behavioural predictors of HIV, construct a risk assessment model for South African township populations, and validate the model in combination with a digital HIV self-testing programme.

This manuscript has been prepared for submission at the *Journal of Acquired Immune Deficiency Syndrome*.

Development of a Digital HIV Risk Assessment Tool for South African Township Populations and Validation in Combination with an App-based HIV Self-testing programme: A Bayesian Approach

Authors

Cindy Leung Soo[1], cindy.leungsoo@mail.mcgill.ca
Sahir Bhatnagar[1], PhD, sahir.bhatnagar@mcgill.ca
Susan J. Bartlett[2,3], PhD, susan.bartlett@mcgill.ca
Aliasgar Esmail[4], MD, ali.esmail@uct.ac.za
Keertan Dheda[4], MBBS PhD, keertan.dheda@uct.ac.za
Nitika Pant Pai[2,3], MD, MPH, PhD, nitika.pai@mcgill.ca

Affiliations

[1]Department of Epidemiology, Biostatistics, and Occupational Health, McGill University, Montreal, Quebec, Canada.
[2]Department of Medicine, McGill University, Montreal, Quebec, Canada.
[3]Research Institute of the McGill University Health Centre, Montreal, Quebec, Canada.
[4]Centre for Lung Infection and Immunity, Division of Pulmonology, UCT Lung Institute and Department of Medicine, University of Cape Town, Cape Town, South Africa.

Corresponding Author

Nitika Pant Pai, MD, MPH, PhD
Department of Medicine, McGill University
Research Institute of the McGill University Health Centre
5252 boul de Maisonneuve Ouest
Montreal, Quebec, Canada
H4A 3S5
(514) 934-1934 x44729
nitika.pai@mcgill.ca

Meetings

None

Sources of support

This work was supported by operating grants 0732–05 and 0710–05 from Grand Challenges Canada (funded by Government of Canada). As part of comatching fund requirements, grant support was provided by: (1) the Government of South Africa, through its Medical Research Council SHIP programme and the Department of Science and Technology and (2) the McGill University Health Centre Foundation. Kit donation and support was provided by (3) OraSure Technologies Inc, Bethlehem.

Dr Pant Pai also acknowledges the support of Fonds de recherche du Quebec – Santé Research Scholar Awards (Junior and Senior), and Canadian Institutes of Health Research grants (HHP 137872, PJT 153149, HBR 422155) for the duration of the project. KD acknowledges funding from the SA MRC (RFA-EMU-02-2017), EDCTP (TMA-2015SF-1043, TMA-1051-TESAII, TMA-CDF2015), UK Medical Research Council (MR/S03563X/1) and the Wellcome Trust (MR/S027777/1).

Conflicts of interest

We declare no competing interests.

Running head

Development of a Digital HIV Risk Score

Keywords

HIV; Risk score; Risk assessment; South Africa; Bayesian; Social determinants

Abstract

Background. Low risk perception is an important barrier to the utilisation of HIV services. In this context, offering an online platform for people to assess their risk of HIV and inform their decision to test can be impactful in increasing testing uptake. In a quasi-randomised trial in 2017-18, we evaluated HIVSmart!, an app-based HIV self-testing program in township populations of South Africa. The HIVSmart! app, contained a risk self-assessment tool which had not yet been explored for its performance. We aimed to 1) determine the accuracy of the HIVSmart! digital risk levels; 2) identify a parsimonious set of predictors of HIV; and 3) develop a risk staging model for South African township populations and validate it in combination with a digital self-testing program.

Methods. We assessed the relationship between the HIVSmart! risk levels and HIV status using logistic regression. Using predictive projection, a Bayesian variable selection approach, we identified predictors of HIV to construct a risk assessment model which was validated in external data as well as in combination with a digital HIV self-testing program.

Results. Our analyses included 3095 participants from the HIVSmart! trial (self-testing arm: n=1535, conventional testing arm: n=1560). The original HIVSmart! risk levels that focused predominantly on behavioural risk factors were associated with HIV infection (posterior median odds ratio (OR), 89% credible interval (CrI): 1.15, 0.98 – 1.35). We further identified a model of five predictors that performed best during external validation (AUC-ROC, 89% CrI: 0.71, 0.68 – 0.72). Predictors were: being unmarried, HIV testing history, having had sex with a partner living with HIV, dwelling situation, and education. With these predictors, the sensitivity of our HIV risk staging model was 91.0% (89% CrI: 89.1% – 92.7%) and the specificity was 13.2% (89% CrI: 8.5% – 19.8%). Combined with a digital HIV self-testing program, the resulting specificity of the risk staging model increased to 91.6%

(89% CrI: 95.9% – 96.4%) and the sensitivity remained similar at 90.9% (89% CrI: 89.1% – 92.6%).

Conclusion. This is the first validated HIV risk assessment tool to be developed for South African township populations and the first study to have evaluated the added value of the risk assessment tool by considering its implementation with a digital self-testing program. A digital HIV risk assessment tool, if implemented in South Africa's national HIV program, could raise awareness with respect to risk factors for HIV in a non-judgemental manner and encourage HIV testing among those who do not routinely seek testing.

Introduction

South Africa has one of the highest human immunodeficiency virus (HIV) burden in the world with an estimated 7.8 million people affected by the virus, among which 8% are unaware of their seropositive status[1]. Within South Africa, the distribution of HIV remains unequal, often paralleling existing socioeconomic divides; South African townships have a prevalence that is nearly twice as high as the national average[2].

Increasing testing uptake amongst high-risk groups is the first step to achieving the UNAIDS 95-95-95 targets (95% tested, 95% on treatment, 95% virally supressed) for HIV elimination[3]. While South Africa has achieved tremendous progress in expanding access to HIV testing in the past two decades, testing uptake remains suboptimal among certain populations such as men and those of lower socioeconomic status (SES)[4,5]. Decentralised testing services such as home-based and self-testing are predicted to be one of the most impactful strategies to improve uptake by increasing access in a convenient and private manner[6,7].

Recognising the potential of decentralised testing services, the World Health Organization (WHO) recommended, in 2016, the implementation of HIV self-testing (HIVST) as a complement to facility-based testing[8]. By 2018, 59 countries globally had implemented HIVST and an additional 53 were developing policies for self-testing[9]. In South Africa, HIVST kits are currently available on the market as well as through several community-based programs[10].

Making use of digital platforms to guide HIV self-testing can further improve accuracy of results, increase referrals and facilitate linkages to care[11,12]. In 2017-2018, HIVSmart! an app-based self-testing program was evaluated in South African township populations in a quasi-randomised trial. HIVSmart! is an Android/iOS app-based self-testing program, with a secure dashboard, cloud server and a peer worker-based 24/7 counselling service[13–15]. The program demonstrated that both supervised and unsupervised self-testing aided by digital tools were feasible, and could link individuals to care just as well as conventional testing[12]. Additionally, the HIVSmart! self-testing program was

referred more often than conventional testing and qualitative findings indicated that participants valued the convenience of HIVST with a digital platform[12,13,16].

Health inequities are apparent in the HIV epidemic in South Africa with people of lower SES being more likely to be affected by HIV[17-19], and yet, generally having poorer testing uptake[20]. As decentralised HIV testing services such as HIVSmart! are being scaled up, it is imperative to ensure that they reach those most affected by the epidemic. Previous publications have indicated that individuals who are at high risk of HIV often do not perceive themselves to be at risk of infection[21,22]. Health education and raising awareness with respect to HIV risk factors can address the disconnect between risk perception and actual risk, and may help improve testing uptake as well as increase the frequency of testing[23].

Yun et al. found, in a randomised controlled trial in China, that high-risk individuals who interacted with a digital HIV risk assessment tool had higher engagement with HIV prevention material, had reduced their sexual risk behaviours, and indicated increased willingness to seek testing[24]. These findings indicate that offering a platform for people to assess their risk of HIV and inform their decision to test can increase testing uptake. While previous risk staging tools for HIV have been developed for sub-Saharan African populations[25], we found none that were specifically targeted to township populations of South Africa.

Part of the HIVSmart! application was an HIV risk self-assessment questionnaire which was piloted among participants[12]. In follow-up interviews, participants indicated enjoying interacting with the risk assessment tool through the app and that it helped them better understand risk factors for HIV, without fear of judgement[26]. However, the accuracy of the HIVSmart! risk levels in predicting HIV infection had not yet been explored and all participants proceeded to self-testing regardless of their risk levels.

Thus, using secondary data from our quasi-randomised control trial, we set out to evaluate the following three objectives: 1) to determine the accuracy of the HIVSmart! digital risk levels in predicting new HIV infections; 2) to identify a parsimonious set of individual, contextual and/or behavioural factors that can predict HIV infection in

township populations; and 3) to develop a risk staging model for South African township populations and validate it in combination with HIVSmart!, an app-based HIV self-testing program.

Methods

Design and setting

We conducted a secondary data analysis using data from the HIVSmart! trial in Cape Town, South Africa[12]. Participants were at least 18 years with unknown HIV status at recruitment, with access to a smartphone or tablet who presented at community outreach clinics for HIV testing. Participants were recruited from townships of Cape Town and offered either HIV self-testing or conventional testing, depending on which strategy their clinic was randomly allocated to. Further details have been published elsewhere[12].

In the conventional and self-testing arm, upon consent, a research nurse collected sociodemographic data from all participants, including: individual factors (age, sex (female/male), marital status (married / not married), and comorbidities (i.e., tuberculosis (TB), other lung infection, diabetes, hypertension, asthma)), and contextual factors (post-secondary education (yes/no), type of dwelling (hostel or informal / house or other), employment status (working part-time or full-time / unemployed or retired), and monthly income (less than 3000 rand / 3000-6000 rand / 6001-9000 rand / over 9000 rand)).

Through the HIVSmart! risk assessment questionnaire, participants in the self-testing arm provided information on behavioural risk factors, testing history (tested in the last six months / not tested in the last 6 months), and HIV exposure in the last 6 months (yes/no/abstain) prior to testing, while participants in the conventional testing arm answered similar questions through nurse-led interviews.

Behavioural risk factors included history of drug injection, sexual activity as well as sexual risk factors such as condomless sex, sex with sex workers, sex with multiple partners, sex with a partner living with HIV, and sex under the influence of alcohol and/or drugs (yes/no/abstain).

A risk level for HIV (HIVSmart! risk level) was computed for each participant in the trial based on answers in the risk assessment questionnaire. The HIVSmart! risk level was unweighted and computed according to the algorithm described in Table 1, such that the maximum risk level was 12. The risk levels were automatically computed and shared with participants in the self-testing arm through the HIVSmart! app prior to testing. While there was no specific cut-off value for the determination of a "high-risk individual", a person was deemed to be "high-risk" if they engaged in any of they engaged in any of the risk factors or if they were exposed to HIV (questions 5-7). We retroactively computed the risk level for participants in the conventional testing arm using answers that were obtained through nurse-led interviews. Qualitative interviews were also conducted in parallel to document the risk perception of participants[13,16,26].

The ascertainment of new HIV infections was done using an oral HIV self-test (OraQuick advance HIV-1/2, OraSure Technologies Inc, USA) in the self-testing arm and blood-based rapid test in the conventional testing arm. HIV test results were confirmed using a blood-based p24 antigen rapid test, and an laboratory HIVRNA test. For the analyses, unless otherwise specified, we used the confirmed test results to define an HIV case.

Statistical Analysis

We imputed missing data using multiple imputation by chained equations (MICE) with five imputed datasets. Since MICE relies on other variables to impute a missing value[27,28], we excluded observations with more than 20% missing values. We conducted a sensitivity analysis where all observations were included during the imputation process, and found that there was no significant impact on estimated coefficients and credible intervals (CrI), but that it resulted in some divergent transitions during modelling (data not shown). Where the number of missing values was low (≤ 5%), we conducted complete case analysis[28,29].

The prior distributions for age and sex were informed by estimates from the extant literature[18]; we used a normal prior with mean of log(0.99541) and standard deviation (SD)

of 1 for age, and a normal prior of mean log(1.83141) and SD of 1 for sex[18]. We used normal null priors with SD of 1 for all other estimated coefficients, meaning that 95% of the prior distributions were consistent with an odds ratio of 0.14 to 7.39. In sensitivity analyses we investigated the impact of using weaker priors (with SD of 2), and null priors for age and sex, and found that the study conclusions were not significantly impacted (data not shown).

To assess the accuracy of the HIVSmart! risk levels in predicting new HIV infections, we computed the area under the receiver operating characteristic curve (AUC-ROC) of the risk levels in the self-testing and conventional testing arm. The AUC-ROC is measure of how well a model can discriminate between cases and non-cases with values greater than 0.5 indicating that a model is better than random at predicting HIV status and values of 1 indicating that a model can perfectly discriminate between HIV positives and negatives. We also computed the sensitivity and the specificity of the HIVSmart! risk levels at every cut-off value between 3 and 8 inclusive. Then, using Bayesian logistic regression, we assessed the relationship between the HIVSmart! risk levels and HIV status, and reported the posterior median odds ratios (OR) with 89% CrI.

To identify predictors of HIV infection, we used predictive projection feature selection, a novel approach for variable selection within the Bayesian framework which offers a good trade-off between sparsity and predictive accuracy and is robust to overfitting[30,31]. First, we fit a full model with all 20 individual, contextual and behavioural predictors for HIV. Then, we used projection to identify submodels that approximate the predictions of the full model[31]. We evaluated the smallest eight submodels in training and testing data, and used the validation set to select the sparsest submodel with acceptable predictive accuracy. Since we selected our model by assessing its performance in the validation set, our final model is less likely to overfit the training data.

We reported the median AUC-ROC, with 89% CrI for all submodels and the full model, computed using expected values for HIV status based on 2000 draws from the posterior predictive distributions. We also report the error rate, sensitivity, specificity, positive predictive value (PPV) and negative predictive value (NPV) computed using 2000 draws from the posterior predictive distributions.

Using sensitivity and specificity data from the self-testing arm of the HIVSmart! trial, we computed the sequential sensitivity and specificity of the risk assessment models combined with the HIVSmart digital self-testing program. Sequential testing is a process by which two tests are performed one after the other; if an individual is deemed positive in the first test (in this case, the HIV risk assessment model) they progress to the second test (in this case, HIV self-testing guided by HIVSmart!) that determines whether an individual is deemed "negative" or "positive" for HIV. The sequential sensitivity/specificity is the overall sensitivity/specificity of this two-step screening process. Of note, a positive result from a self-test would generally not be considered an HIV diagnosis and would require further confirmatory testing.

We chose to report the 89% CrI rather than the 95% CrI since the 95% CrI can be unstable when the effective sample sizes are low (< 10,000)[32]. Additionally, it has been suggested that 89% intervals be used when computing credible intervals to avoid confusion with confidence intervals[33]. However, where relevant, both the 89% CrI and the 95% CrI are plotted in the figures.

All analyses were performed using R version 4.0.2[34]. Bayesian modelling was performed using the *brms* package[35], multiple imputation was performed using the *mice* package[36], and predictive projection feature selection was performed with the *projpred* package[37]. For measures of performance, we used the *performance* package[38] and for data wrangling and visualisation, the *tidyverse* package[39].

Ethics approval

The original quasi-randomised trial was approved by the Institutional review board of the Research Institute of the McGill University Health Center and the University of Cape Town, and research permits were obtained from the City of Cape Town. All participants provided written informed consent. Ethics approval for the secondary data analyses was obtained as part of an extension to the primary study.

Results

Determining the accuracy of the HIVSmart! digital risk levels in predicting new HIV infections in township populations of South Africa

Out of 1535 participants, 63 (4%) were missing an HIVSmart! digital risk level in the self-testing arm of the trial (training data) and 18 (1%) were missing the HIV status. Overall, 80 (5%) participants were missing at least one value. In the conventional arm, there were 17 (1%) participants out of 1560 that were missing data for HIV status. Since the HIVSmart! digital risk score was computed retroactively for the conventional testing arm, there were no missing values for this variable. Since the number of missing cases were relatively low, we use complete cases only for these analyses.

The AUC-ROC was lower in the self-testing arm data than in the conventional testing data (AUC-ROC, 0.52 and 0.62 respectively) (Figure 1A). In both datasets, the optimal cut-off for the HIVSmart! digital risk level was 5.5; at this cut-off, the sensitivity was 81.1% in the self-testing arm data and 82.3% in the conventional arm data while the specificity was 25.0% in the self-testing arm data and 40.6% in the conventional arm data (Figure 1B). We estimated that every additional level of the HIVSmart! risk assessment tool increased the odds of HIV infection by 15% (OR, 89% CrI: 1.15, 0.98 – 1.35) in the self-testing data.

Identifying a parsimonious set of individual, contextual and/or behavioural predictors of HIV

Out of the 1535 observations in the self-testing arm of the trial, 174 (11%) were missing at least one value of which 81 (5%) were missing values for more than 20% of variables and were thus excluded from analysis. The remaining missing values were imputed five times, and in total, 1454 observations from the self-testing arm were included in the analysis. In the conventional arm data, 38 (2%) out of 1560 observations were missing at least one value but none were missing more than 20% of variables. As such, we imputed all missing values five times and conserved the original sample size of 1560 participants.

Our reference model of 20 predictors converged (Rhat < 1.05) without any divergent transitions. We assessed the posterior distributions which revealed no apparent outliers and all Pareto K values, being below 0.5, were deemed satisfactory. We identified using the predictive projection approach a sparse model of five sociodemographic and behavioural predictors which performed better (AUC-ROC, 89% CrI: 0.71, 0.68 – 0.72) than the HIVSmart! digital risk levels (Figure 2).

The included variables were marital status, HIV testing history, having had sex with an HIV infected partner, dwelling type, and education level. With these variables, the AUC-ROC of the model was 0.71 (89% CrI: 0.68 – 0.72). The median error rate was 14.2% (89% CrI: 12.6% – 15.8%), the sensitivity was 91.0% (89% CrI: 89.1% – 92.7%), and the specificity was 13.2% (89% CrI: 8.5% – 19.8%) in the testing data (Table 2). At an HIV prevalence of 10%, the NPV was 93.2% (89% CrI: 89.2% – 95.3%) and the PPV was 10.5% (89% CrI: 9.9% – 11.2%) in the testing data.

Validating the HIV risk assessment model it in combination with a digital self-testing program

The oral self-testing program aided by the HIVSmart! application had on overall sensitivity of 99.93% and specificity of 95.52%. We computed the sequential sensitivity and specificity of our risk staging tool followed by the HIV self-testing and found the resulting specificity to be significantly improved (sequential specificity, 89% CrI: 96.1%, 95.9% – 96.4%) while the sensitivity was not significantly impacted (sequential sensitivity, 89% CrI: 90.9%, 89.1% – 92.6%) in external data (Table 2).

All variables included in our selected model were associated with HIV infection. Individuals who had not tested for HIV in the previous 6 months had increased odds of HIV (OR, 89% CrI: 1.94, 1.43 - 2.65). Additionally, individuals who were unmarried (OR, 89% CrI: 2.24, 1.51 - 3.37) or who had sex with an HIV infected partner were over two times more likely (OR, 89% CrI: 2.96, 1.61 - 5.24) to be HIV positive. Socioeconomic indicators also impacted HIV infection with those not having post-secondary education (OR, 89% CrI: 2.02, 1.29 - 3.26) and living in hostels or informal dwelling (OR, 89% CrI: 1.69, 1.25 - 2.29)

being more likely to be HIV positive compared to those with post-secondary education and those living in a house, respectively (Figure 3).

Discussion

We developed and validated a prediction model to assess the risk of HIV in township populations of South Africa that includes five sociodemographic and behavioural predictors of HIV. This risk assessment tool is, to our knowledge, the first to be developed for township populations in South Africa. During validation, our predictive model of five variables outperformed the original HIVSmart! digital risk staging tool and had similar predictive ability than the full model containing all predictors.

While the initial digital HIVSmart! risk levels, which did not include any contextual predictors and predominantly included behavioural risk factors, had a modest predictive accuracy (AUC-ROC of 0.52 in the self-testing arm and 0.62 in the conventional testing arm), we found that incorporating information on socioeconomic factors improved the accuracy of the risk predictions (AUC-ROC of 0.68 in the training data and 0.71 in the testing data). These results suggest that incorporating information on socioeconomic indicators can improve the accuracy of HIV risk assessment models and that risk staging for HIV should expand beyond behavioural risk factors.

The incorporation of behavioural risk factors, especially with regards to sex and drug use, is common amongst existing risk staging tools[40], but less common is the use of contextual factors related to social determinants of health. In the sub-Saharan African context, we found few other published risk assessment tools for HIV that incorporated contextual factors. Among those that we identified, two were targeted to young women in South Africa[41-43], and incorporated variables on financial dependence and school absence, and three were non-targeted risk assessment tools developed for use in Kenya which included information on occupation[44], Uganda with a variable for education[45], and rural South Africa with variables for education, place of residence and SES quintile[46]. Incorporating contextual factors such as education and dwelling situation in an HIV risk assessment tool not only acknowledges the differential impact of the HIV epidemic for

populations of lower SES, but can also help mitigate these health inequities by referring them to testing.

Our selected HIV risk assessment model included two contextual factors, dwelling situation and education level, two individual factors, HIV testing history and marital status, as well as one behavioural factor, having had sex with a person living with HIV, all of which were associated with HIV. Studies using machine learning techniques previously identified several social predictors of HIV in sub-Saharan Africa, including education level, dwelling situation and wealth[47,48], highlighting the value of using contextual information in predicting HIV. Our study, which uses Bayesian predictive projection to identify HIV predictors, similarly found educational attainment and dwelling situation to be important predictors of HIV status. On the other hand, employment status and monthly income were not a useful predictors of HIV.

In terms of performance, our HIV risk assessment tool is comparable to other externally or cross-validated non-targeted HIV risk assessment models intended for use in different sub-Saharan populations such as Kenya (AUC-ROC of 0.69)[44], Uganda (AUC-ROC of 0.69 in men and 0.67 in women)[45] and rural South Africa (AUC-ROC of 0.71 in men and 0.68 in women)[46].

Interestingly, the performance of our model was better in the testing data, corresponding to the conventional testing arm of the trial, relative to the training data, corresponding to the self-testing arm. This might be explained by the fact that a greater proportion of participants in the self-testing arm of the trial were referred to self-testing by other participants (16.7% in self-testing vs 3.1% in conventional testing arm)[12]; those who were referred by others were probably less likely to have recently sought HIV testing yet reasonably similar in their likelihood of having HIV, making previous HIV testing history a poorer predictor of HIV status in the self-testing arm compared to the conventional testing arm, and the overall predictive performance of the models better in the testing data.

While in low prevalence areas, targeting HIV testing to high-risk individuals can be cost-effective[49,50], restricting access to testing in high prevalence settings such as South Africa is discouraged given the high cost of missing a person living with HIV[40]. It has

instead been suggested that HIV risk assessment tools be used to prompt offer of testing as well as discussions with respect to HIV risk factors[25].

The sensitivity and specificity of the oral HIV self-testing program aided by the HIVSmart! application was high (sensitivity of 99.93% and specificity of 95.52%) when evaluated in township populations of South Africa. This sensitivity estimate is higher than that reported by the test kit manufacturer (91.7%)[51] indicating that digital innovations that support self-testing and its interpretation can minimize user errors and improve the sensitivity of self-testing.

For those who are seeking to test for HIV, self-testing is an convenient, private and confidential alternative to conventional HIV testing[7,52,53]. However, for those who might still be reluctant to test or who might not be aware of self-testing options, a digital risk assessment tool for HIV can help encourage high-risk individuals to test for HIV and facilitate linkages to self-testing options.

The sensitivity of our risk assessment model was relatively high in testing data (sensitivity: 91.0%, 89.1% – 92.7%) indicating that it would appropriately capture a majority of people living with HIV (PLWH) into the higher risk category and refer them to testing. In modelling the sequential sensitivity and specificity of our risk assessment models coupled with a digital HIV self-testing program, we found the resulting sequential specificity of our model to be significantly improved (sequential specificity: 96.1%, 95.9% – 96.4%) in the testing data while the sensitivity remained relatively similar (sequential sensitivity: 90.9%, 89.1% – 92.6%). This suggests that combining self-assessment of HIV risk with digital self-testing could be an effective strategy to find for undiagnosed PLWH and link them to care. Indeed, those who self-tested with HIVSmart! showed high rates of linkages to care, at least similar, if not better, than those achieved by conventional testing[12]. While there has been significant improvement in improving linkages to care and treatment retention in South Africa, ensuring high HIV testing uptake and finding PLWH who are unaware of their status remain important challenges, particularly among individuals of lower socioeconomic status[5,20].

Several studies have reported discrepancies between individuals' perceived risk of HIV and their actual risk[21,22]; bridging this gap is an important step in increasing testing uptake as several meta-analyses have indicated that the more people perceive themselves to be at risk of a disease, the greater their motivation to engage in healthy behaviours[54,55]. Stigma and misconceptions about HIV are another important barrier to the uptake of testing[56,57]. Dzinamarira et al. highlight the value of discussing HIV etiology and transmission in health education programs to combat misconceptions related to HIV and increase testing uptake[57].

An HIV risk assessment tool offered through a digital platform is an effective way to convey personalised information on HIV risk and raise awareness with respect to risk factors for HIV. Indeed, our risk staging tool can inform a decision to test in high prevalence settings by educating people on the pertinence of self-testing in a non-judgemental manner. In follow-up interviews, participants from the HIVSmart! trial indicated that they appreciated the digital risk assessment questionnaire as it allowed them to reflect on their risk factors without fear of judgement[26].

Digital platforms can further improve accessibility and acceptability of HIV self-assessment tools. Jones et al. found in their qualitative study that participants had a marked preference for a digital risk assessment tool compared to the paper-based alternative[58]. Some participants also indicated that they felt more comfortable filling the questionnaire than discussing discomforting topics, such as sexual behaviours, with a provider they did not know, while others indicated that the digital risk assessment questionnaire would encourage them to ask questions about risk behaviours to their provider[58].

Furthermore, a randomised control study in China recently found that individuals who were provided with a digital HIV risk assessment tool had higher engagement with other HIV prevention material, had reduced their high-risk behaviours, and were more likely to indicate intention to seek testing relative to those who only received general educational information on HIV[24]. Altogether, these studies suggest that HIV risk self-assessment can help spark discussions about HIV risk, encourage risk reduction behaviour,

and increase HIV testing among high-risk individuals, especially in settings where people might be reluctant to be forthcoming to healthcare workers about their risk factors[59-61].

Indeed, many healthcare facilities in South Africa have incorporated provider-initiated counseling and testing (PICT), where providers are encouraged to offer HIV testing to all patients presenting at health facilities. As a result, sexual and reproductive health services in Cape Town often require presenting individuals to test for HIV despite not having specifically sought this service. This strategy has been extremely successful in improving HIV testing uptake[62], but it is rather unpopular with only 11.6% of surveyed South Africans preferring this practice compared to 66.1% preferring client-initiated counseling and testing[63]. Under PICT policies, individuals reported having felt pressured into being tested for HIV and not being given the opportunity to decline[63], occasionally leading to strained relationships between health providers and community members who feel arbitrarily targeted[16,64].

Here, we suggest a digital HIV risk assessment tool that offers a non-judgemental and non-paternalistic way of increasing HIV testing uptake in high prevalence areas by empowering individuals to assess their own need for testing and raising awareness of HIV risk factors. Referring individuals presenting at clinics to a risk assessment tool rather than requiring them to test for HIV could help ease tensions between healthcare workers and community members all the while being a possibly more cost-effective approach than PICT. Indeed, task-shifting personalised communication of HIV risk and promotion of testing services to a digital platform has the potential to reach a wider audience, and facilitate the offer of home testing options, pre-test and post-test counselling to those most vulnerable to HIV[12,65].

Strengths. A main strength of the present study is its novelty in using contextual factors for HIV risk assessment among township populations of South Africa. Utilizing information on sociodemographic factors in a community-based risk staging tool not only acknowledges the impact of socioeconomic position on the risk of HIV, but also leverages information that is easily accessible outside of clinical settings.

Second, our study demonstrated that a digital HIV self-testing program can be highly accurate in high prevalence settings, and that an HIV risk assessment tool combined with self-testing can further screen individuals who might not otherwise test for HIV, and maintain high sensitivity and specificity. The high accuracy of these digital strategies calls for the integration of digital diagnostics, including apps, in the implementation self-testing initiatives.

A third strength of this study relates to the approach we used to identify predictors of HIV. Bayesian predictive projection has been shown to be more robust to overfitting than traditional frequentist approaches[30,31]. Furthermore, we selected our final model using a validation set which further limits the risk of overfitting to the training data. Using a Bayesian framework in our analyses also allowed us to incorporate prior information on the relationship between age and sex, and HIV, and to account for the uncertainty with respect to missing data through multiple imputation. Finally, another advantage of the Bayesian framework is that it provides uncertainty estimates for parameters that are often credited as having a more intuitive interpretation than frequentist approaches, referring to probability distributions rather than likelihood[33].

Limitations. Our findings should be considered in light of the limitations associated with our present study. First, there might be measurement bias. In the self-testing arm, many individuals chose to abstain from answering certain questions related to their sexual practices but very few chose to abstain from answering similar questions in the conventional testing arm during nurse-led interviews. Participants in the conventional testing arm might have felt compelled to answer every question during nurse-led interviews, and we cannot exclude the possibility of inaccurate reporting or social desirability bias their answers. Second, since we used secondary data collected from townships of Cape Town, our risk assessment tool may not be generalizable to other settings. Third, broad strata within our categorical variables could have led to some residual confounding. Finally, while multiple imputation can help address bias due to missing data, some selection bias from the recruitment process in the quasi-randomised control trial may remain[27,66].

Conclusion

In this study, we have developed an HIV risk assessment model for township populations of South Africa and validated it in combination with a digital HIV self-testing program. To our knowledge, this is the first validated HIV risk staging tool to be developed for South Africa township populations. It is also the first study to have evaluated the added value of the risk assessment tool by considering its implementation with a digital self-testing program. Indeed, our digital HIV risk assessment tool could encourage HIV testing among those who do not routinely seek testing by offering a personalised assessment of HIV risk and facilitating linkages to self-testing options. Furthermore, it can raise awareness with respect to risk factors for HIV in a non-judgemental manner and help spark risk reduction discussion between patients and providers. Finally, our risk assessment tool for HIV is one of the few to utilise information on social determinants of health with the aim of improving testing uptake among those disproportionately affected by HIV and mitigating the impact of health inequities.

Acknowledgements

We would like to extend our deepest acknowledgements and thanks to all study participants for their participation and time. We also wish to thank staff members (health care workers, peer counsellors, and nurses) of the University of Cape Town and at the RI-MUHC for their dedication to the project.

References

1. UNAIDS Data 2021. Geneva: Joint United Nations Programme on HIV/AIDS;2021.
2. Cain D, Pitpitan EV, Eaton L, et al. Collective efficacy and HIV prevention in South African townships. *J Community Health.* 2013;38(5):885-893.
3. UNAIDS. Confronting Inequalities: Leasson for pandemic responses from 40 years of AIDS. Geneva: Joint United Nations Program on HIV/AIDS;2021.
4. Johnson LF, Rehle TM, Jooste S, Bekker LG. Rates of HIV testing and diagnosis in South Africa: successes and challenges. *Aids.* 2015;29(11):1401-1409.
5. Jooste S, Mabaso M, Taylor M, North A, Tadokera R, Simbayi L. Trends and determinants of ever having tested for HIV among youth and adults in South Africa from 2005-2017: Results from four repeated cross-sectional nationally representative household-based HIV prevalence, incidence, and behaviour surveys. *PloS one.* 2020;15(5):e0232883-e0232883.
6. Johnson LF, van Rensburg C, Govathson C, Meyer-Rath G. Optimal HIV testing strategies for South Africa: a model-based evaluation of population-level impact and cost-effectiveness. *Scientific Reports.* 2019;9(1):12621.
7. Pant Pai N, Sharma J, Shivkumar S, et al. Supervised and unsupervised self-testing for HIV in high- and low-risk populations: a systematic review. *PLoS Med.* 2013;10(4):e1001414.
8. World Health O. Guidelines on HIV self-testing and partner notification: supplement to consolidated guidelines on HIV testing services. Geneva: World Health Organization; 2016.
9. Hamilton A, Thompson N, Choko AT, et al. HIV Self-Testing Uptake and Intervention Strategies Among Men in Sub-Saharan Africa: A Systematic Review. *Front Public Health.* 2021;9:594298.
10. Majam M. Progress on the scaling up of HIV testing in South Africa through varied distribution models using the oral HIV self-test kit. *Oral Diseases.* 2020;26(S1):137-140.
11. McGuire M, de Waal A, Karellis A, et al. HIV self-testing with digital supports as the new paradigm: A systematic review of global evidence (2010-2021). *EClinicalMedicine.* 2021;39:101059-101059.
12. Pai N, Esmail A, Saha Chaudhuri P, et al. Impact of a personalised, digital, HIV self-testing app-based program on linkages and new infections in the township populations of South Africa. *BMJ Glob Health.* 2021;6(9).
13. Janssen R, Engel N, Esmail A, et al. Alone But Supported: A Qualitative Study of an HIV Self-testing App in an Observational Cohort Study in South Africa. *AIDS Behav.* 2020;24(2):467-474.
14. Pant Pai N, Smallwood M, Desjardins L, Goyette A, Vassal A, Thomas R. Can an HIVSmart! app-optimized self-testing strategy be operationalized in Canada? IDWeek; 2017; San Diego, USA.
15. Pant Pai N, Smallwood M, Desjardins L, et al. An Unsupervised Smart App-Optimized HIV Self-Testing Program in Montreal, Canada: Cross-Sectional Study. *J Med Internet Res.* 2018;20(11):e10258.

16. Janssen R, Krumeich J, Esmail A, et al. Moments of Uncertainty: Exploring How an App-Based Oral HIV Self-Testing Strategy Fits in Communities 'Living Under' HIV Risk. *Medicine Anthropology Theory.* 2021;8(2):1-24.
17. Mabaso M, Makola L, Naidoo I, Mlangeni LL, Jooste S, Simbayi L. HIV prevalence in South Africa through gender and racial lenses: results from the 2012 population-based national household survey. *Int J Equity Health.* 2019;18(1):167.
18. Bärnighausen T, Hosegood V, Timaeus IM, Newell ML. The socioeconomic determinants of HIV incidence: evidence from a longitudinal, population-based study in rural South Africa. *Aids.* 2007;21 Suppl 7(Suppl 7):S29-38.
19. Bunyasi EW, Coetzee DJ. Relationship between socioeconomic status and HIV infection: findings from a survey in the Free State and Western Cape Provinces of South Africa. *BMJ Open.* 2017;7(11):e016232.
20. Jooste S, Mabaso M, Taylor M, North A, Shean Y, Simbayi LC. Socio-economic differences in the uptake of HIV testing and associated factors in South Africa. *BMC Public Health.* 2021;21(1):1591.
21. Price JT, Rosenberg NE, Vansia D, et al. Predictors of HIV, HIV Risk Perception, and HIV Worry Among Adolescent Girls and Young Women in Lilongwe, Malawi. *J Acquir Immune Defic Syndr.* 2018;77(1):53-63.
22. Wilton J, Kain T, Fowler S, et al. Use of an HIV-risk screening tool to identify optimal candidates for PrEP scale-up among men who have sex with men in Toronto, Canada: disconnect between objective and subjective HIV risk. *J Int AIDS Soc.* 2016;19(1):20777.
23. Regan S, Losina E, Chetty S, et al. Factors associated with self-reported repeat HIV testing after a negative result in Durban, South Africa. *PLoS One.* 2013;8(4):e62362.
24. Yun K, Chu Z, Zhang J, et al. Mobile Phone Intervention Based on an HIV Risk Prediction Tool for HIV Prevention Among Men Who Have Sex With Men in China: Randomized Controlled Trial. *JMIR Mhealth Uhealth.* 2021;9(4):e19511.
25. Jia KM, Eilerts H, Edun O, et al. Risk scores for predicting HIV incidence among adult heterosexual populations in sub-Saharan Africa: a systematic review and meta-analysis. *Journal of the International AIDS Society.* 2022;25(1):e25861.
26. Janssen R, Engel N, Pant Pai N, et al. 'You're only there on the phone'? A qualitative exploration of community, affect and agential capacity in HIV self-testing using a smartphone app. *Sociol Health Illn.* 2021;43(3):591-606.
27. Azur MJ, Stuart EA, Frangakis C, Leaf PJ. Multiple imputation by chained equations: what is it and how does it work? *Int J Methods Psychiatr Res.* 2011;20(1):40-49.
28. Schafer JL, Graham JW. Missing data: our view of the state of the art. *Psychological methods.* 2002;7(2):147-177.
29. Jakobsen JC, Gluud C, Wetterslev J, Winkel P. When and how should multiple imputation be used for handling missing data in randomised clinical trials - a practical guide with flowcharts. *BMC medical research methodology.* 2017;17(1):162.
30. Bartonicek A, Wickham SR, Pat N, Conner TS. The value of Bayesian predictive projection for variable selection: an example of selecting lifestyle predictors of young adult well-being. *BMC Public Health.* 2021;21(1):695.
31. Piironen J, Paasiniemi M, Vehtari A. Projective inference in high-dimensional problems: Prediction and feature selection. *Electronic Journal of Statistics.* 2020;14(1).

32. Kruschke JK. Doing Bayesian data analysis : a tutorial with R, JAGS, and Stan. In: 2nd edition ed. Amsterdam: Academic Press; 2015.
33. McElreath R, O'Reilly for Higher E, Safari aORMC. Statistical Rethinking, 2nd Edition. In: 2nd edition. ed.: Chapman and Hall/CRC; 2020: https://ezproxy.torontopubliclibrary.ca/login?url=https://learning.oreilly.com/library/view/-/9780429639142/?ar.
34. R Core Team (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/
35. Bürkner P-C. Advanced Bayesian Multilevel Modeling with the R Package brms. *The R Journal.* 2018;10(1):395-411.
36. Buuren Sv, Groothuis-Oudshoorn K. mice: Multivariate Imputation by Chained Equations in R. 2011;45(3):1-67.
37. Juho Piironen, Markus Paasiniemi, Alejandro Catalina, Aki Vehtari (2020). projpred: Projection Predictive Feature Selection. R package version 2.0.2. https://CRAN.R-project.org/package=projpred
38. Lüdecke D, Ben-Shachar MS, Patil I, Waggoner P, Makowski D. performance: An R package for assessment, comparison and testing of statistical models. *Journal of Open Source Software.* 2021;6(60).
39. Wickham H, Averick M, Bryan J, et al. Welcome to the tidyverse. *Journal of Open Source Software.* 2019;4(43):1686.
40. Ong JJ, Coulthard K, Quinn C, et al. Risk-Based Screening Tools to Optimise HIV Testing Services: a Systematic Review. *Current HIV/AIDS Reports.* 2022;19(2):154-165.
41. Ayton SG, Pavlicova M, Tamir H, Abdool Karim Q. Development of a prognostic tool exploring female adolescent risk for HIV prevention and PrEP in rural South Africa, a generalised epidemic setting. *Sex Transm Infect.* 2020;96(1):47-54.
42. Balkus JE, Brown E, Palanee T, et al. An Empiric HIV Risk Scoring Tool to Predict HIV-1 Acquisition in African Women. *J Acquir Immune Defic Syndr.* 2016;72(3):333-343.
43. Giovenco D, Pettifor A, MacPhail C, et al. Assessing risk for HIV infection among adolescent girls in South Africa: an evaluation of the VOICE risk score (HPTN 068). *J Int AIDS Soc.* 2019;22(7):e25359.
44. Muttai H, Guyah B, Musingila P, et al. Development and Validation of a Sociodemographic and Behavioral Characteristics-Based Risk-Score Algorithm for Targeting HIV Testing Among Adults in Kenya. *AIDS Behav.* 2021;25(2):297-310.
45. Kagaayi J, Gray RH, Whalen C, et al. Indices to Measure Risk of HIV Acquisition in Rakai, Uganda. *PLoS ONE.* 2014;9(4).
46. Roberts DA, Cuadros D, Vandormael A, et al. Predicting the Risk of Hiv-1 Acquisition in Rural South Africa Using Geospatial Data. *Clin Infect Dis.* 2022.
47. Mutai CK, McSharry PE, Ngaruye I, Musabanganji E. Use of machine learning techniques to identify HIV predictors for screening in sub-Saharan Africa. *BMC Med Res Methodol.* 2021;21(1):159.
48. Orel E, Esra R, Estill J, et al. Prediction of HIV status based on socio-behavioural characteristics in East and Southern Africa. *PLOS ONE.* 2022;17(3):e0264429.

49. Leblanc J, Hejblum G, Costagliola D, et al. Targeted HIV Screening in Eight Emergency Departments: The DICI-VIH Cluster-Randomized Two-Period Crossover Trial. *Ann Emerg Med.* 2018;72(1):41-53.e49.
50. Domínguez-Berjón MF, Pichiule-Castañeda M, García-Riolobos MC, et al. A feasibility study for 3 strategies promoting HIV testing in primary health care in Madrid, Spain (ESTVIH project). *J Eval Clin Pract.* 2017;23(6):1408-1414.
51. OraSure Technologies Inc. OraQuick In-Home HIV Test, Package Insert. https://www.fda.gov/media/83607/download. Published 2012. Accessed June 7th 2022.
52. Harichund C, Moshabela M, Kunene P, Abdool Karim Q. Acceptability of HIV self-testing among men and women in KwaZulu-Natal, South Africa. *AIDS Care.* 2019;31(2):186-192.
53. Lippman SA, Lane T, Rabede O, et al. High Acceptability and Increased HIV-Testing Frequency After Introduction of HIV Self-Testing and Network Distribution Among South African MSM. *J Acquir Immune Defic Syndr.* 2018;77(3):279-287.
54. Floyd DL, Prentice-Dunn S, Rogers RW. A Meta-Analysis of Research on Protection Motivation Theory. *Journal of Applied Social Psychology.* 2000;30(2):407-429.
55. Milne S, Sheeran P, Orbell S. Prediction and Intervention in Health-Related Behavior: A Meta-Analytic Review of Protection Motivation Theory. *Journal of Applied Social Psychology.* 2000;30(1):106-143.
56. Chimoyi L, Tshuma N, Muloongo K, Setswe G, Sarfo B, Nyasulu PS. HIV-related knowledge, perceptions, attitudes, and utilisation of HIV counselling and testing: a venue-based intercept commuter population survey in the inner city of Johannesburg, South Africa. *Glob Health Action.* 2015;8:26950.
57. Dzinamarira T, Mulindabigwi A, Mashamba-Thompson TP. Co-creation of a health education program for improving the uptake of HIV self-testing among men in Rwanda: nominal group technique. *Heliyon.* 2020;6(10):e05378.
58. Jones J, Stephenson R, Smith DK, et al. Acceptability and willingness among men who have sex with men (MSM) to use a tablet-based HIV risk assessment in a clinical setting. *SpringerPlus.* 2014;3(1):1-7.
59. Caldwell DH, Jan G. Computerized assessment facilitates disclosure of sensitive HIV risk behaviors among African Americans entering substance abuse treatment. *The American journal of drug and alcohol abuse.* 2012;38(4):365-369.
60. Ferrand RA, Weiss HA, Nathoo K, et al. A primary care level algorithm for identifying HIV-infected adolescents in populations at high risk through mother-to-child transmission. *Tropical Medicine & International Health.* 2011;16(3):349-355.
61. Gillet C, Darling KEA, Senn N, Cavassini M, Hugli O. Targeted versus non-targeted HIV testing offered via electronic questionnaire in a Swiss emergency department: A randomized controlled study. *PloS one.* 2018;13(3):e0190767.
62. Kennedy CE, Fonner VA, Sweat MD, Okero FA, Baggaley R, O'Reilly KR. Provider-initiated HIV testing and counseling in low- and middle-income countries: a systematic review. *AIDS Behav.* 2013;17(5):1571-1590.
63. van Dyk AC. Client-initiated, provider-initiated, or self-testing for HIV: what do South Africans prefer? *J Assoc Nurses AIDS Care.* 2013;24(6):e45-56.

64. Martínez Pérez G, Cox V, Ellman T, et al. 'I Know that I Do Have HIV but Nobody Saw Me': Oral HIV Self-Testing in an Informal Settlement in South Africa. *PloS one*. 2016;11(4):e0152653.
65. Cao B, Bao H, Oppong E, et al. Digital health for sexually transmitted infection and HIV services: a global scoping review. *Curr Opin Infect Dis*. 2020;33(1):44-50.
66. Sterne JAC, White IR, Carlin JB, et al. Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. *BMJ: British Medical Journal*. 2009;339(7713):157-160.

Appendix

Figures

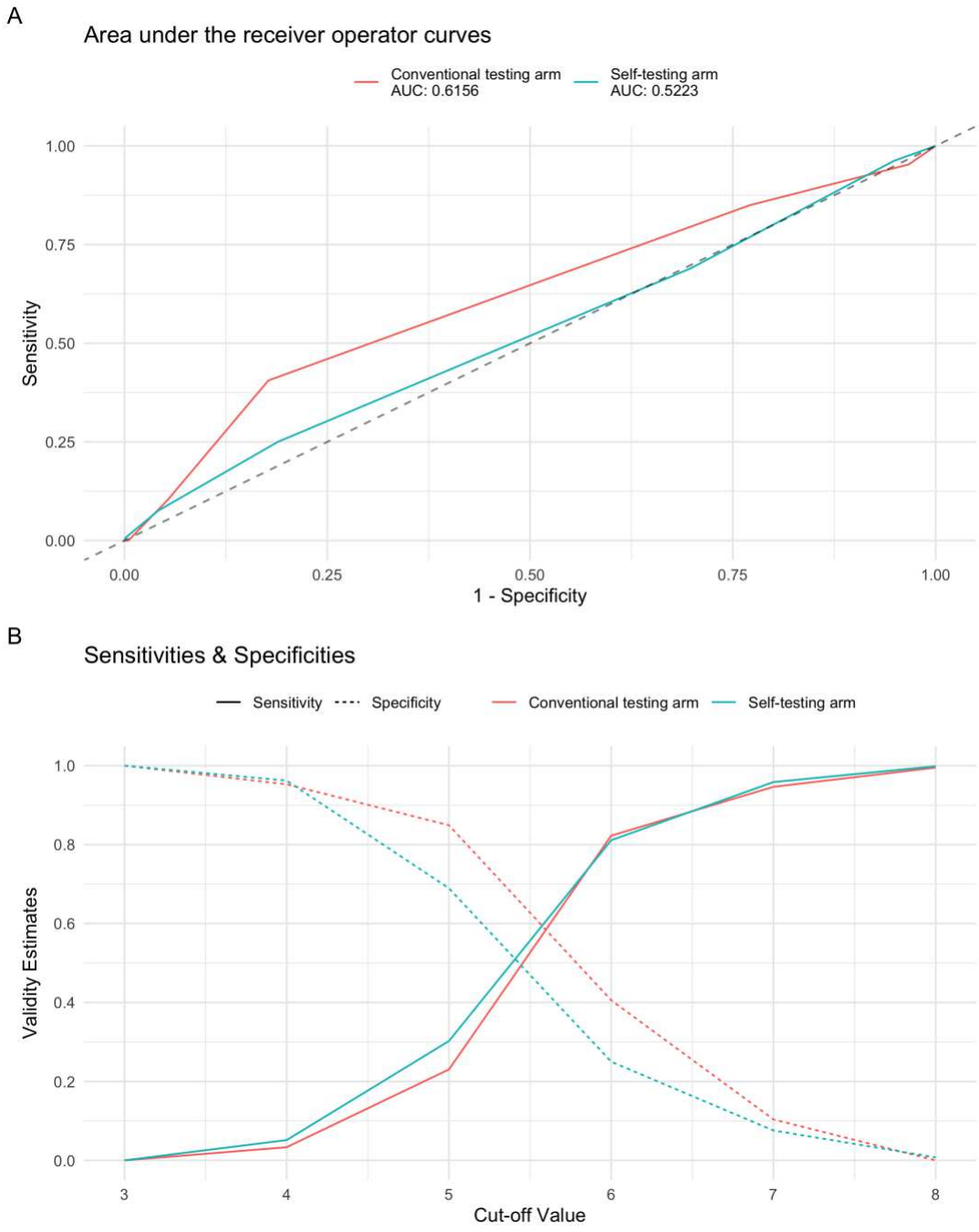


Figure 1. (A) Area under the receiver operating characteristic curve (AUC-ROC) and (B) sensitivity and specificity for the HIVSmart! digital risk level in the self-testing and the conventional testing arm data. Sensitivity and specificity values were computed at cut-offs 3 to 8 for the HIVSmart! digital risk level. A cut-off of 5.5 was deemed to be optimal, meaning that those

with a score of 6 or higher would be deemed "high risk" whereas those with a risk level of less than 6 would be deemed "low risk".

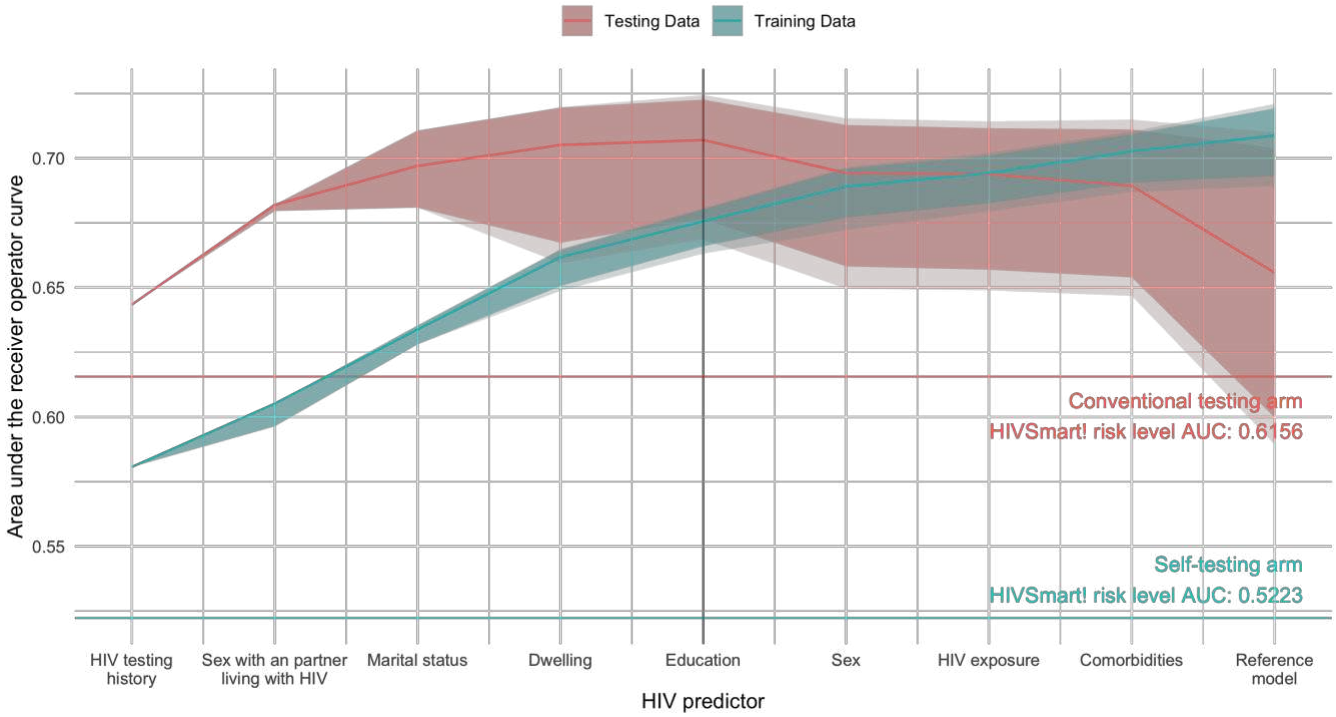


Figure 2. Area under the receiver operating characteristic curve (AUC-ROC) of submodels and reference model in the testing (N = 1560) and training (N = 1454) data. Green and pink lines represent the median AUC-ROC darker shaded areas represent the 89% credible intervals and lighter shaded areas represent the 95% intervals, computed using 2000 draws from the posterior predictive distribution. The X-axis represents the HIV predictor that was added to the submodel. The black vertical line represents the selected risk assessment model with five predictors (HIV testing history, sex with an HIV infected partner, marital status, dwelling, and education).

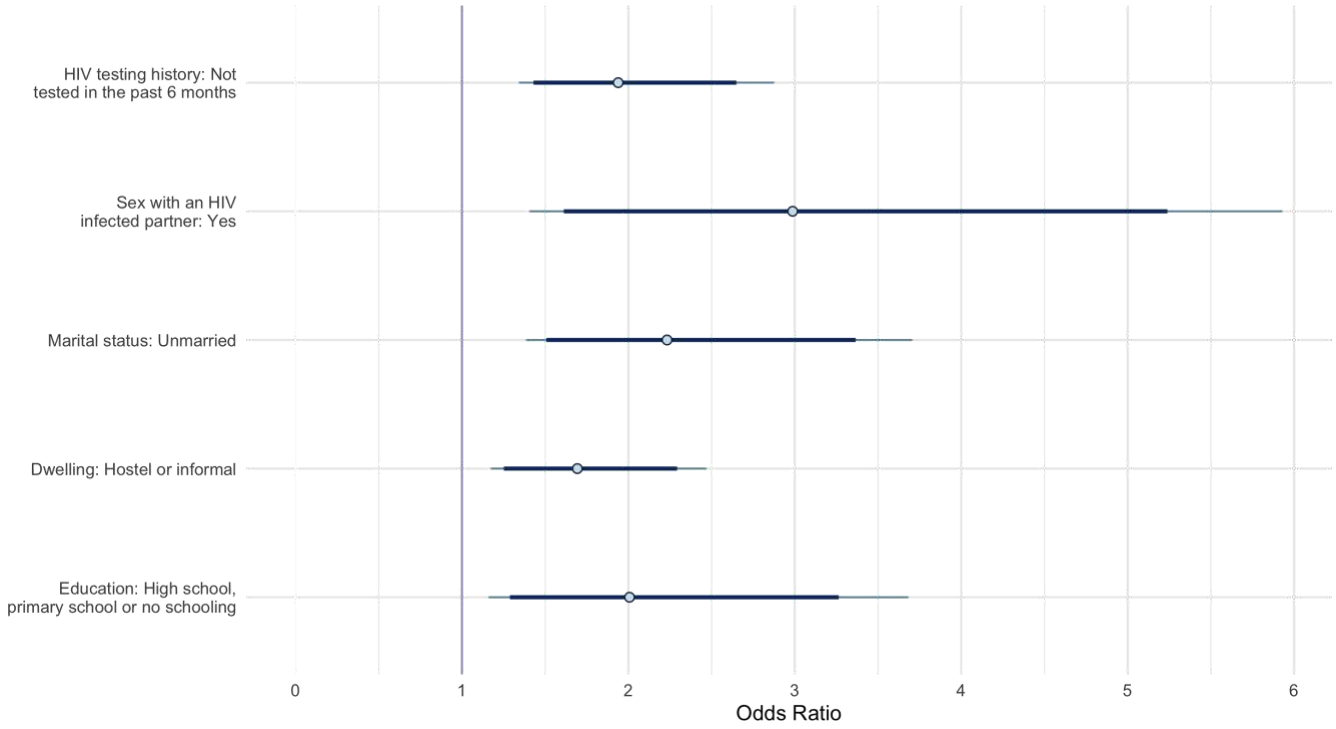


Figure 3. Median, 89% and 95% credible intervals of estimated odds ratio for HIV status for the selected risk assessment model. Points represent the posterior medians, thick segments represent the 89% credible intervals and thin segments represent the 95% intervals of the estimated odds ratios of HIV infection. Reference categories are: tested in the past 6 months for HIV testing history, no for sex with an HIV infected partner, married for marital status, house or other for dwelling, and post-secondary level education for education.

Tables

Table 1. HIVSmart! digital risk level algorithm

Question	Risk level
1. To which age group do you belong?	Any answer*: +1
2. What is your sexual orientation?	Any answer*: +1
3. Have you ever been tested for HIV?	Any answer*: +1
4. Are you sexually active?	Yes: +1
5. In the past 6 months, I have had sex… (select all that apply) o Without a condom o With multiple partners o With a commercial sex worker o With an HIV infected partner o Under the influence of alcohol o Under the influence of drugs	For each selected option: +1
6. In the past 6 months, have you used injected drugs (excluding medicine)?	Yes: +1
7. In the past 6 months, have you been exposed to HIV at your workplace or elsewhere?	Yes: +1

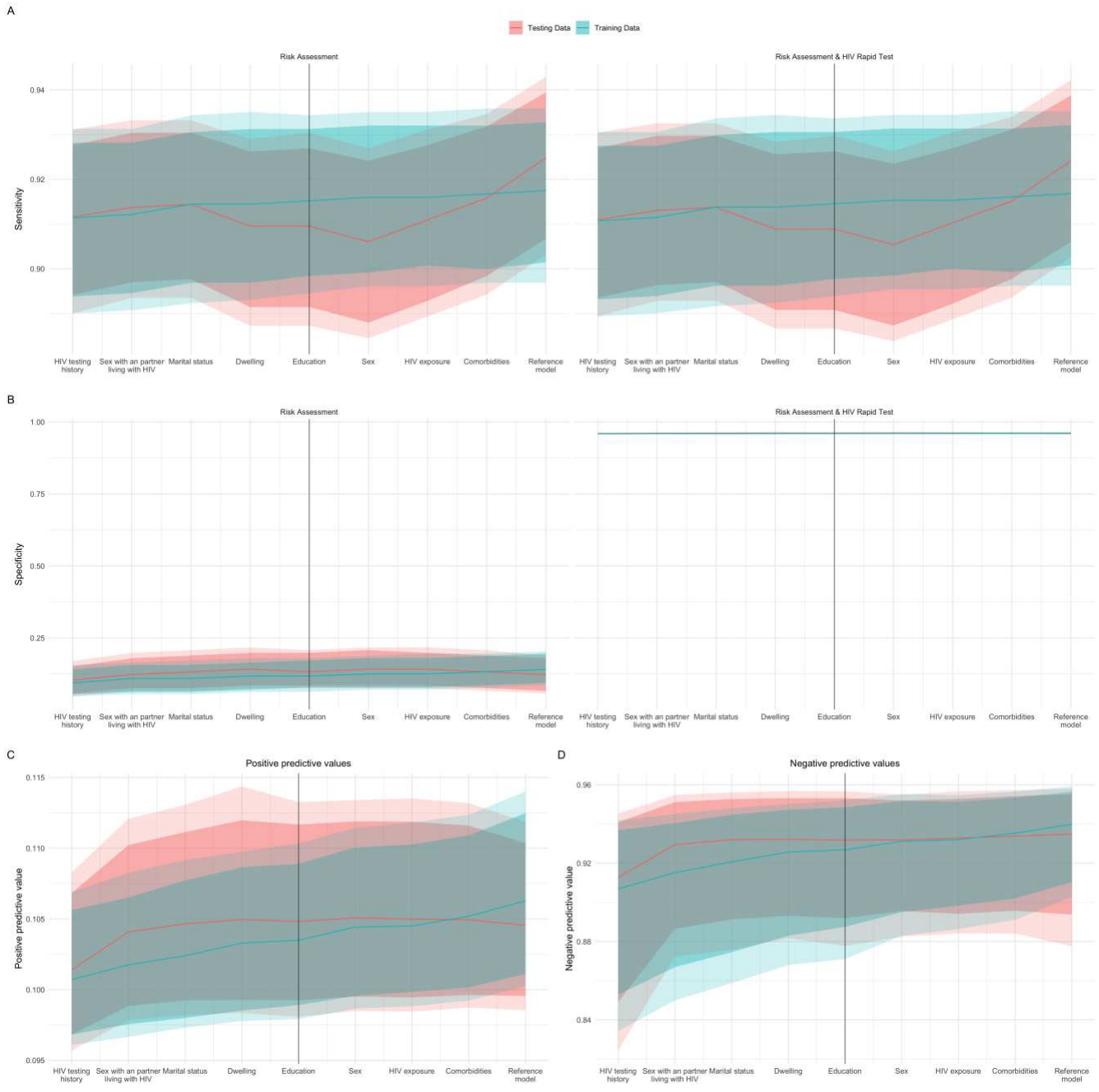
* Those who abstained from answering were given a score of 0 for the question

Table 2. Performance of selected risk assessment model in the testing (N = 1560) and training (N = 1454) data

	Median (89% CrI)	
	Testing data	Training data
AUC-ROC	0.71 (0.68 – 0.72)	0.68 (0.67 – 0.68)
Error Rate	14.2 (12.6 – 15.8)	15.4 (14.0 – 16.9)
Sensitivity	91.0 (89.1 – 92.7)	91.5 (89.9 – 93.1)
Specificity	13.2 (8.5 – 19.8)	11.7 (7.8 – 17.2)
Sequential Sensitivity	90.9 (89.1 – 92.6)	91.5 (89.8 – 93.1)
Sequential Specificity	96.1 (95.9 – 96.4)	96.0 (95.9 – 96.3)
Positive Predictive Value	10.5 (9.9 – 11.2)	10.3 (9.9 – 10.9)
Negative Predictive Value	93.2 (89.2 – 95.3)	92.7 (88.8 – 94.8)

Supplementary Digital Content

Figure



Supplementary figure 1. (A) Sensitivity, (B) Specificity, (C) Positive predictive value and (D) Negative predictive value of submodels identified through predictive projection and the reference model. Green and pink lines represent the median values, darker shaded areas represent the 89% credible intervals and lighter shaded areas represent the 95% intervals, computed using 2000 draws from the posterior distribution. The X-axis indicates the HIV predictor that was added

to the submodel. The black vertical line represents the selected risk assessment model with five predictors. For panel A and B, the plots on the left represent the sensitivity/ specificity of the risk assessment models and the plots on the right represent the sequential sensitivity/ specificity of the risk assessment models when combined with a digital self-testing program. For panel C and D, the positive and negative predictive values were computed for an HIV prevalence of 10%.

Table

Supplementary table 1. Variables selected through predictive projection

HIV predictor
1. Marital status
2. Tested in the past 6 months
3. Sex with an HIV infected partner
4. Dwelling
5. Education
6. Sex
7. HIV exposure
8. Comorbidities
9. Sex with multiple people
10. Sex with alcohol
11. Monthly income
12. Abstain from sharing sexual behaviours
13. Age
14. Sexually active
15. Drug injection
16. Sex with sex workers
17. Sex without condom
18. Sex with drugs
19. Employment status
20. Tuberculosis infection

Chapter 4: The Impact of Contextual factors on HIV

Preface to manuscript 2

In the literature review, we learnt that the relationship between socioeconomic status (SES) and HIV in sub-Saharan Africa has shifted as the epidemic matured. While individuals of higher SES seemed to be most affected in the early years of the epidemic, the latest evidence now appears to suggest that people of lower SES are disproportionately impacted by HIV.

In the first manuscript, we found that two of the most relevant predictors of HIV in South African townships were socioeconomic indicators: education level and dwelling type. By examining the posterior distributions, we further affirmed the harmful impact of low SES on HIV infection as those with a lower education level, and living in hostels or informal dwellings were significantly more likely to be living with HIV.

In Chapter 2, we explore potential explanations for this effect; many have suggested that people of lower SES are more likely to engage in risk taking behaviours and less likely to be knowledgeable about HIV, making them more vulnerable to the viral infection. Indeed, socioeconomic stressors such as poverty and food insecurity can constraint people into adopting certain high-risk sexual practices such as unprotected transactional sex.

It nevertheless remains unclear to what extent contextual factors impact the risk of HIV despite controlling for behavioural risk factors related to sex and drug use. Thus, in this second manuscript, we seek to quantify the effect of contextual factors (subdistrict of residence, education level, dwelling situation, employment, and income) on HIV infection, after controlling for behaviours.

This manuscript has been prepared for submission at the *Journal of the International AIDS Society*.

Contextual Factors Impact the Risk of HIV Infection in South African Townships

Cindy Leung Soo[1], Sahir Bhatnagar[1], Susan J. Bartlett[2,3], Aliasgar Esmail[4], Keertan Dheda[4], Nitika Pant Pai[2,3§]

[1]Department of Epidemiology, Biostatistics, and Occupational Health, McGill University, Montreal, Canada.
[2]Department of Medicine, McGill University, Montreal, Canada.
[3]Research Institute of the McGill University Health Centre, Montreal, Canada.
[4]Centre for Lung Infection and Immunity, Division of Pulmonology, UCT Lung Institute and Department of Medicine, University of Cape Town, Cape Town, South Africa.

[§]Corresponding Author: Nitika Pant Pai

5252 boul de Maisonneuve Ouest
Montreal, H4A 3S5, Canada
(514) 934-1934 x44729
nitika.pai@mcgill.ca

E-mail addresses of authors:
CLS: cindy.leungsoo@mail.mcgill.ca
SB: sahir.bhatnagar@mcgill.ca
SJB: susan.bartlett@mcgill.ca
AE : ali.esmail@uct.ac.za
KD : keertan.dheda@uct.ac.za
NP : nitika.pai@mcgill.ca

Keywords: HIV; Context; South Africa; Bayesian; Social determinants

Word count:
Abstract: 247/350
Main text: 3558/3500

Abstract

Introduction. With a prevalence almost twice as high as the national average, South African townships are particularly impacted by the HIV epidemic. Yet, it remains unclear whether the risk of acquiring HIV differs across subdistricts and what role contextual variables play in predisposing individuals to HIV infection. Our objective was to estimate the extent to which contextual factors (education level, dwelling situation, employment status, and monthly income) explain the risk of HIV in South African townships, after controlling for behavioural factors.

Methods. Using Bayesian logistic regression, we analysed secondary data from a quasi-randomised trial which recruited participants (N = 3095) from townships of Cape Town. We controlled for individual factors (age, sex marital status, testing history, HIV exposure, comorbidities and tuberculosis infection) and behavioural factors (unprotected sex, sex with multiple partners, with sex workers, with an HIV infected person, under the influence of alcohol or drugs), and accounted for the uncertainly due to missing data through multiple imputation.

Results. Individuals residing in hostels or informal dwellings, without post-secondary education, were at increased odds of HIV after controlling for subdistrict of residence, individual and behavioural factors.

Conclusions. While HIV prevention strategies have typically emphasized personal responsibility in health, our analyses showed that individual and behavioural factors alone cannot fully explain HIV infection. A greater emphasis on addressing social determinants of health is warranted to end the HIV epidemic. This implies a greater deployment of social programmes that mitigate health inequities by improving living conditions and promoting general education.

Introduction

South Africa has one of the highest HIV rates with an estimated 19.1% of the population infected with the virus (1). In high HIV prevalence countries like South Africa, the risk of acquiring an HIV infection differs significantly among geographical areas (2). South African townships, urban settlements originally designated during apartheid for non-whites only, are particularly impacted by the HIV epidemic with a prevalence that is nearly twice as high as the national average (3).

In Cape Town, one out of every five households live in informal dwellings according to the 2011 South African Census and this proportion is expected to have risen as more people move to urban areas for employment (4). Additionally, 36% of households in Cape Town live below the poverty line with less than 3500 rand of income per month and 9% have no access to on-site sanitation (5). According to the census, 24% of people living in Cape Town are unemployed, 79% have not completed secondary education, including 3% with no formal schooling, and only 4% of Cape Town residents have post-secondary education (5).

Prevention strategies for HIV have been dominated by interventions aimed at influencing knowledge, attitudes and behaviours (6). While behavioural factors play a major role in the transmission of HIV, they poorly explain health inequities (7), which are health disparities that are unnecessary, avoidable, unfair and unjust (8). By contrast, social determinants of health such as education, employment status, dwelling situation, and income can not only inform policy implementers as to where to target HIV service delivery, but may also help address health inequities and minimize the risk of HIV acquisition in the South African context.

Better understanding how contextual factors related to social determinants of health impact HIV infection can inform how public health policies should be implemented. While studies have previously reported on the negative impact of socioeconomic disadvantage on the risk of acquiring HIV in Southern Africa, it is still unclear whether the risk of HIV infection differs across socioeconomic divides after controlling for behavioural and individual factors (2, 7, 9, 10).

Furthermore, although the impact of behavioural factors, especially behaviours with respect to sexual activities, on HIV acquisition is relatively well characterized, it remains unclear to what extent contextual factors such as education, employment status, dwelling situation, and income impact the likelihood of HIV infection within townships of South Africa and whether they differ across subdistricts of Cape Town.

The objective of this study was to estimate the extent to which contextual factors (education level, dwelling situation, employment status and monthly income) explain the risk of HIV in South African townships, while controlling for individual and behavioural factors. In exploratory analyses, we also aimed to estimate the effect of contextual factors on certain sexual behaviours and the effect of individual, behavioural and contextual factors on HIV testing history.

Methods

Design and Setting

We conducted a secondary data analysis using data from a quasi-randomised trial in Cape Town, South Africa, between January 2017 and June 2018. The objective of the trial was to evaluate HIVSmart!, a digital self-testing programme for HIV with an oral self-test, on increased linkage to care, detection of new infections and increased referrals to test (11). HIVSmart! is an Android and iOS app-based self-testing programme, with a secure dashboard, Health Insurance Portability and Accountability Act (HIPAA)-compliant cloud server, and a peer worker-based 24/7 counselling service (12-14). The study was conducted at six community clinics; participants were recruited from three subdistricts of Cape Town (denoted here as subdistrict A, B and C) and offered either HIV self-testing or conventional testing.

The ascertainment of new HIV infections was done using an oral HIV self-test (OraQuick advance HIV-1/2, OraSure Technologies Inc, USA) in the self-testing arm and blood-based rapid test in the conventional testing arm. HIV test results were confirmed using a blood based p24 antigen rapid test as well as an laboratory HIVRNA test. For the analyses, we used confirmed test results to define an HIV case.

Measures

Individual factors included age, sex (female/male), marital status (married/not married), testing history (tested in the last six months / not tested in the last six months), HIV exposure in the last 6 months (yes/no/abstain) as well as selected comorbidities (i.e., tuberculosis (TB), other lung infection, diabetes, hypertension, and asthma).

Behavioural risk factors included history of drug injection, and sexual risk factors such as unprotected sex, sex with sex workers, sex with multiple partners, sex with HIV infected individuals, and sex under the influence of alcohol and/or drugs (yes/no/abstain).

Contextual variables assessed included post-secondary education (yes/no), type of dwelling (hostel or informal / house), employment status (working part-time or full-time / unemployed or retired), and monthly income (less than 3000 rand / 3000-6000 rand / 6001-9000 rand / over 9000 rand).

Measured variables were selected based on knowledge of HIV epidemiology during the design of the HIVSmart! trial.

Statistical analysis

Variables were summarised using means, standard deviation (SD), and interquartile range (IQR) for continuous variables, with an ANOVA p-value. Categorical variables were summarised by count and frequency with a chi-square p-value.

We conducted a majority of the analyses using Bayesian statistics which offer two key advantages compared to frequentist approaches: it permits the incorporation of prior information and provides more intuitive and meaningful inferences.

We used Bayesian logistic regression to estimate the effect of subdistrict of residence, individual (age, sex, marital status, testing history, HIV exposure, presence of comorbidities or TB), contextual (education level, dwelling situation, employment status and monthly income), and behavioural (drug injection, sexually active, unprotected sex, sex with sex workers, sex with multiple partners, sex with HIV infected partner, sex under the influence

of alcohol or drugs) factors on the risk of acquiring HIV. We report the posterior medians as well as the 89% and 95% credible intervals (CrI) of the adjusted odds ratios.

Priors for age and sex were informed by a study by Bärnighausen et al. (2). The prior for age was set as normal prior with a mean of log(0.9954) and a standard deviation of 1. The prior for sex was set as a normal prior with a mean of log(1.8314) and an SD of 1. The prior for the beta coefficient of all other variables was specified as a normal prior with a mean of 0 and an SD of 1. We investigated, through sensitivity analyses, the impact of using weaker priors (with SD of 2), as well as null priors for age and sex, and found that the study conclusions were not significantly impacted (data not shown).

Missing data was imputed using multiple imputation by chained equations with five imputed datasets. Results were pooled from the posterior draws of the five submodels to account for the additional uncertainty caused by the imputation procedure (13, 14). Analyses were performed in R version 4.0.2 (15). Multiple imputation was performed using the *mice* package (16) and Bayesian modelling was performed using the *brms* package (17). We used the *tidyverse* package (18) for data wrangling and visualisation and the *arsenal* package (19) to create summary tables.

Ethics approval

The primary quasi-randomised study was approved by the Institutional Review Board of the Research Institute of the McGill University Health Centre and the University of Cape Town, and research permits were obtained from the City of Cape Town. All participants provided written informed consent. Ethics approvals for secondary data analyses were obtained as part of an extension to the primary study.

Results

Demographics

In total, 3095 participants were recruited (11) of which 212 (7%) were missing at least one variable and 81 (3%) were missing more than five variables. 893 (29%) participants were from subdistrict A, 1023 (33%) from subdistrict B and 1112 (36%) from subdistrict C of Cape

Town (Table 1 1). 64 (2%) participants did not specify their subdistrict of residence. Participants were on average 29 years old, predominantly female (70%), and unmarried (71%). Close to half (48%) had tested for HIV in the past 6 months but were undiagnosed and not on treatment at the time of testing.

Differences across townships

Participants from subdistrict A were on average younger while participants from subdistrict C were slightly older (Table 1). As compared with subdistrict C, a smaller proportion of participants from subdistrict A and subdistrict B were unmarried (% unmarried, 77.0%, 68.7%, 68.2%, respectively). Participants from subdistrict B were less likely to have tested in the past 6 months.

Other factors which differed across townships are frequency of comorbidities, whether engaging in unprotected sex or sex under the influence of alcohol, dwelling situation, employment situation, and monthly income (Table 1).

Table 1. Baseline characteristics of participants in subdistricts of Cape Town

	Subdistrict A (N=893)	Subdistrict B (N=1023)	Subdistrict C (N=1112)	Overall (N=3095)	p value
Missing subdistrict				64	
Individual Factors					
Age					
Missing	0	0	1	1	
Mean (SD)	28 (8)	29 (8)	29 (9)	29 (9)	< 0.001
IQR	17 - 67	17 - 67	17 - 76	17 - 76	
Sex					
Missing	0	1	1	2	
Male	264 (29.6%)	300 (29.4%)	336 (30.2%)	919 (29.7%)	0.896
Female	629 (70.4%)	722 (70.6%)	775 (69.8%)	2174 (70.3%)	
Marital status					
Missing	11	29	36	81	
Unmarried	606 (68.7%)	678 (68.2%)	828 (77.0%)	2153 (71.4%)	< 0.001
Previous HIV test					
Missing	0	1	0	64	

Has not tested in the past 6 months	464 (52.0%)	611 (59.8%)	494 (44.4%)	1573 (51.9%)	< 0.001
In the past 6 months, have you been exposed to HIV?					
Missing	0	1	0	64	
Yes	43 (4.8%)	63 (6.2%)	73 (6.6%)	179 (5.9%)	0.529
Abstain	13 (1.5%)	17 (1.7%)	19 (1.7%)	49 (1.6%)	
Have you ever been diagnosed with tuberculosis?					
Yes	68 (7.6%)	70 (6.8%)	106 (9.5%)	247 (8.0%)	0.063
Have you ever been diagnosed with any of these illnesses: lung infection, diabetes, hypertension, asthma?					
Yes	77 (8.6%)	136 (13.3%)	85 (7.6%)	305 (9.9%)	< 0.001
Behavioural Factors					
Are you sexually active?					
Missing	1	3	7	74	
Yes	811 (90.9%)	906 (88.8%)	983 (89.0%)	2704 (89.5%)	0.142
Abstain	24 (2.7%)	22 (2.2%)	36 (3.3%)	82 (2.7%)	
In the past 6 months, have you injected drugs (excluding medicine)?					
Missing	0	1	0	64	
Yes	23 (2.6%)	34 (3.3%)	43 (3.9%)	100 (3.3%)	0.167
Abstain	3 (0.3%)	10 (1.0%)	12 (1.1%)	25 (0.8%)	
In the past 6 months, I have had sex... (select all that applies)					
Missing	5	5	8	81	
Without a condom	566 (63.7%)	716 (70.3%)	811 (73.5%)	2097 (69.6%)	< 0.001
With multiple partners	83 (9.3%)	113 (11.1%)	107 (9.7%)	303 (10.1%)	0.391
With a commercial sex worker	5 (0.6%)	6 (0.6%)	13 (1.2%)	24 (0.8%)	0.203
With an HIV infected partner	15 (1.7%)	28 (2.8%)	34 (3.1%)	77 (2.6%)	0.132
Under the influence of alcohol	84 (9.5%)	129 (12.7%)	103 (9.3%)	317 (10.5%)	0.021
Under the influence of drugs (e.g. marijuana, cocaine, heroin, etc.)	12 (1.4%)	29 (2.8%)	23 (2.1%)	64 (2.1%)	0.077
Abstain	13 (1.5%)	23 (2.3%)	23 (2.1%)	59 (2.0%)	0.428

Contextual Factors					
What is your highest level of education?					
Graduate, undergraduate or college	173 (19.4%)	207 (20.2%)	199 (17.9%)	593 (19.2%)	0.38
High school, primary school or no schooling	720 (80.6%)	816 (79.8%)	913 (82.1%)	2502 (80.8%)	
What sort of dwelling do you live in?					
Formal house or other	489 (54.8%)	457 (44.7%)	584 (52.5%)	1568 (50.7%)	< 0.001
Hostel or informal dwelling	404 (45.2%)	566 (55.3%)	528 (47.5%)	1527 (49.3%)	
What is your work situation?					
Employed (part-time or full-time)	349 (39.1%)	468 (45.7%)	470 (42.3%)	1288 (41.6%)	0.013
Not employed or retired	544 (60.9%)	555 (54.3%)	642 (57.7%)	1807 (58.4%)	
What is your monthly income?					
Missing	0	0	3	66	
>9000 R	20 (2.2%)	32 (3.1%)	56 (5.0%)	108 (3.6%)	< 0.001
6001-9000 R	28 (3.1%)	50 (4.9%)	29 (2.6%)	107 (3.5%)	
3000-6000 R	135 (15.1%)	171 (16.7%)	150 (13.5%)	457 (15.1%)	
<3000 R	710 (79.5%)	770 (75.3%)	874 (78.8%)	2357 (77.8%)	

The odds of HIV infection differed across subdistrict of Cape Town, South Africa

We found that those residing in subdistrict B and C were at increased odds of being HIV positive compared to residents of subdistrict A in unadjusted analyses (Supplementary table). This difference persisted even after controlling for individual, behavioural, and contextual factors, although the null effect was within the 89% probability bounds of the odds ratio for subdistrict C (aOR, 89% CrI: 1.54, 1.16 - 2.07 and 1.21, 0.90 - 1.64, respectively) (Figure 1).

Sex, marital status, having pre-existing conditions, testing history and previous HIV exposure affected the odds of new HIV infection

Female participants were at increased odds of being HIV positive compared to male participants (aOR, 89% CrI: 1.44, 1.11 – 1.89), after controlling for all other individual, behavioural, and contextual factors as well as subdistrict of residence. Additionally, in our adjusted analyses, those who were unmarried or who had not tested in the previous 6 months had at least twice the odds of having an HIV infection (aOR, 89% CrI: 2.07, 1.56 – 2.76 and 2.51, 1.97 – 3.23, respectively). In adjusted analyses, individuals co-infected with tuberculosis had similar odds while those with other comorbidities had lower odds of testing positive for HIV (aOR, 89% CrI: 0.39, 0.24 – 0.63). Finally, those who indicated having been exposed to HIV in the past 6 months were also at increased odds of HIV infection, in adjusted analyses (Figure 1).

Engaging in unprotected sex, sex with multiple partners, or sex with an HIV infected partner increased the odds of HIV infection

Having had sex with an HIV infected partner was associated with the highest adjusted odds ratio of infection (aOR, 89% CrI: 6.43, 4.22 – 9.74), after controlling for individual, behavioural, and contextual factors. Other behavioural factors associated with greater odds of HIV infection included having engaged in unprotected sex, having had sex with multiple partners, and having chosen not to disclose sexual behaviours. By contrast, those who reported having had sex under the influence of drugs were at lower odds of being HIV positive in adjusted analyses (aOR, 89% CrI: 0.47, 0.20 – 1.02) (Figure 1).

Less stable housing situation and lower education level increased the odds of HIV infection

In adjusted analyses, individuals living in hostels or informal dwellings were at greater odds of being HIV positive (aOR, 89% CrI: 1.34, 1.07 - 1.68) relative to those living in a house. In addition, individuals with lower education also had increased odds of HIV infection (aOR,

89% CrI: 1.82, 1.29 - 2.61) relative to those having achieved post-secondary education or higher in our adjusted analysis (Figure 1).

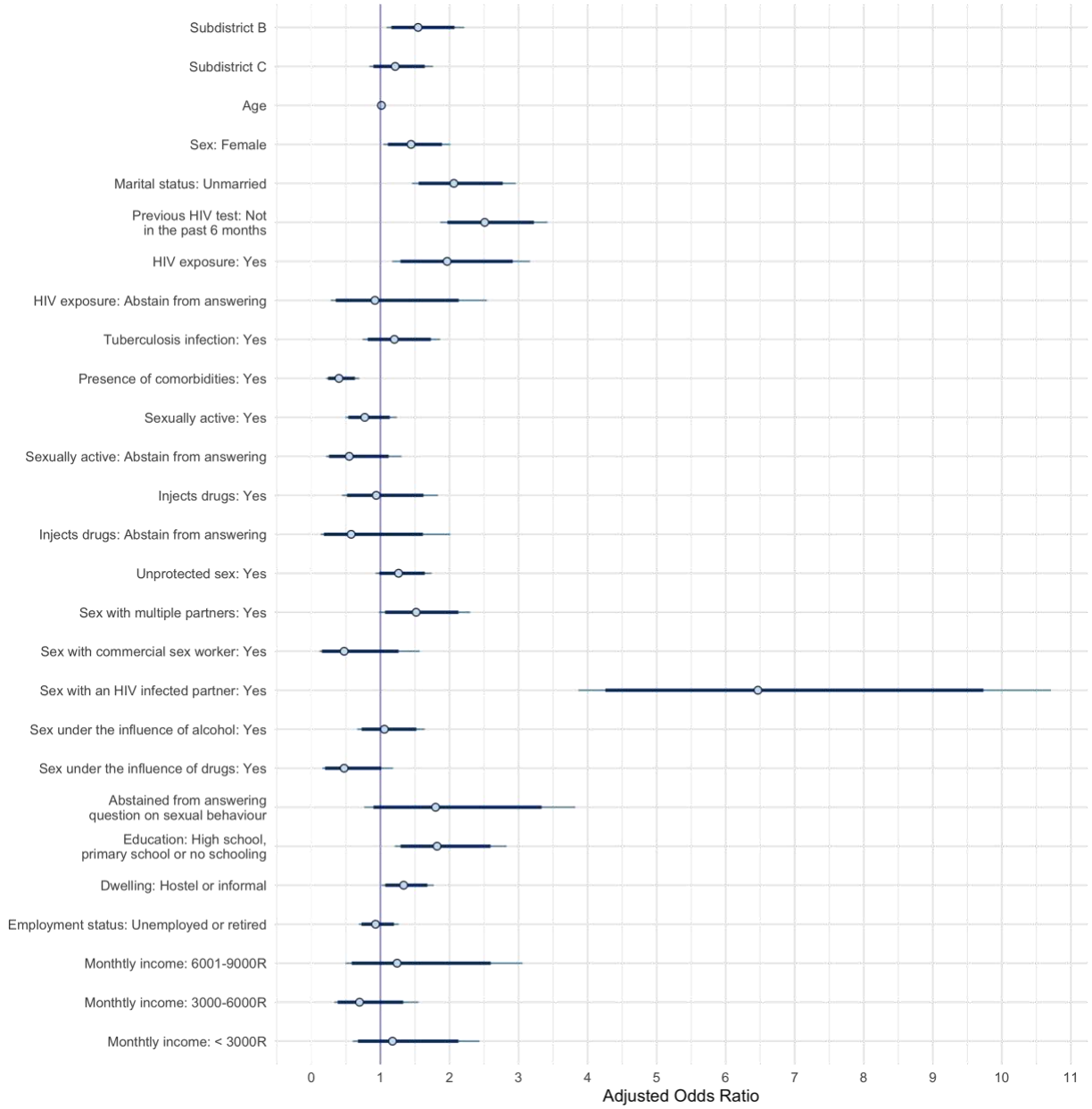


Figure 1. Bayesian logistic regression model posterior intervals of adjusted odds ratios of HIV infection. Points represent the posterior medians, thick segments represent the 89% credible intervals and thin segments represent the 95% intervals of the adjusted odds ratios.

Contextual factors impacted the odds of engaging in sexual behaviours associated with HIV infection

In exploratory analyses, we investigated whether contextual factors impacted sexual behavioural factors associated with HIV infection, specifically engaging in unprotected sex, sex with multiple people, or sex with a person living with HIV, after controlling for age and sex.

According to our analyses, individuals in the 3000 – 6000R per month income level were at increased odds of engaging in unprotect sex compared to those in the highest income level (> 9000R per month), after adjusting for age, sex, and other contextual factors (Figure 2). We did not find a difference in sexual behaviours for other income strata.

Our adjusted analyses also revealed increased odds of engaging in sex with multiple partners for those with a less stable housing situation and those who are not currently employed (Figure 2).

Finally, individuals without post-secondary education were over two times more likely to engage in sex with an HIV infected person in adjusted analyses (Figure 2).

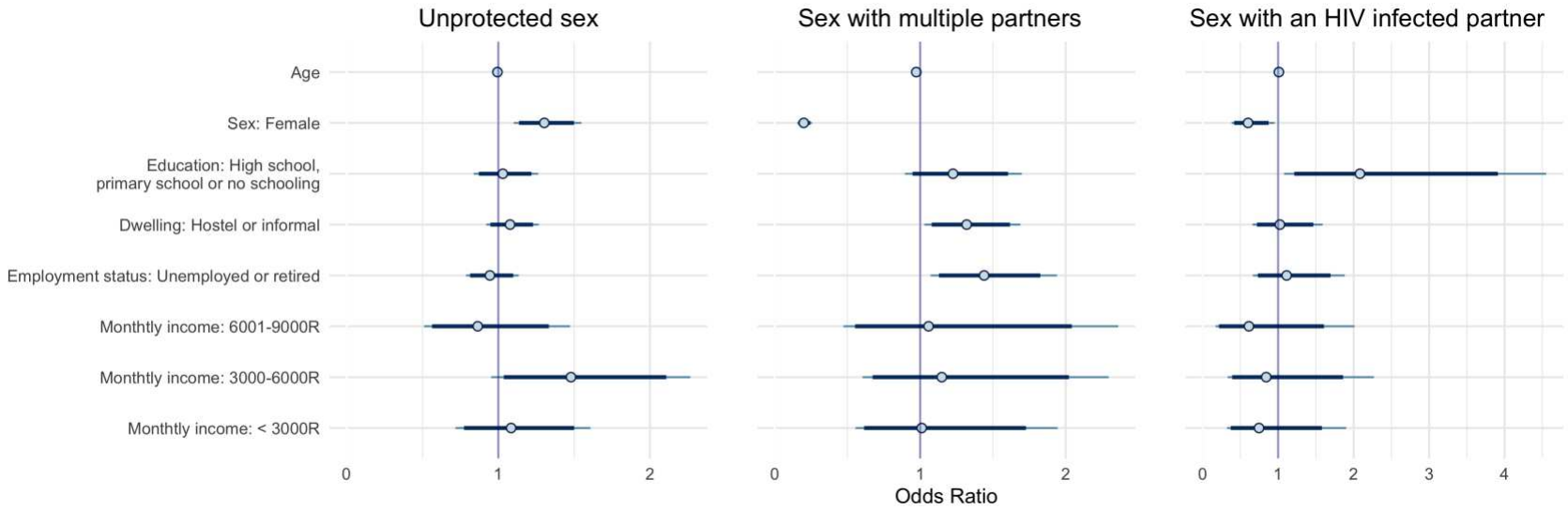


Figure 2. Bayesian logistic regression model posterior intervals of adjusted odds ratios of the impact of contextual factors on select behavioural factors. Points represent the posterior medians, thick segments represent the 89% credible intervals and thin segments represent the 95% intervals of the adjusted odds ratios.

Contextual factors did not impact the likelihood of having been tested for HIV in the past 6 months

Given the strong association we observed between HIV infection and testing history, in exploratory analyses, we examined the relationship between individual, behavioural, and contextual factors and having sought HIV testing in the past 6 months.

In unadjusted analyses, individuals residing in subdistrict B were less likely to have been tested in the past 6 months compared to residents of subdistrict A (OR, 89% CrI: 0.74, 0.64 – 0.86) . By contrast, those from subdistrict C of Cape Town were more likely to have sought HIV testing the past 6 months relative to subdistrict A residents (OR, 89% CrI: 1.36, 1.18 – 1.56).

Each additional year of age decreased the odds of having been tested for HIV in the past 6 months by 2% (Figure 3). Additionally, we found that women were approximately twice as likely as men to have been tested for HIV in the past 6 months. Individuals infected with tuberculosis were also more likely to have recently sought HIV testing.

On the other hand, individuals with comorbidities or who were married were less likely to have recently been tested for HIV. Behavioural factors associated with lower odds of testing in the past 6 months include having engaged in unprotected sex, having had sex with multiple partners, an HIV infected person or under the influence of drugs.

Finally, those without post-secondary education were slightly less likely to have sought testing in the past 6 months compared to those with post-secondary education, although the null effect was within the 89% probability bounds for the estimated odds ratio (Figure 3).

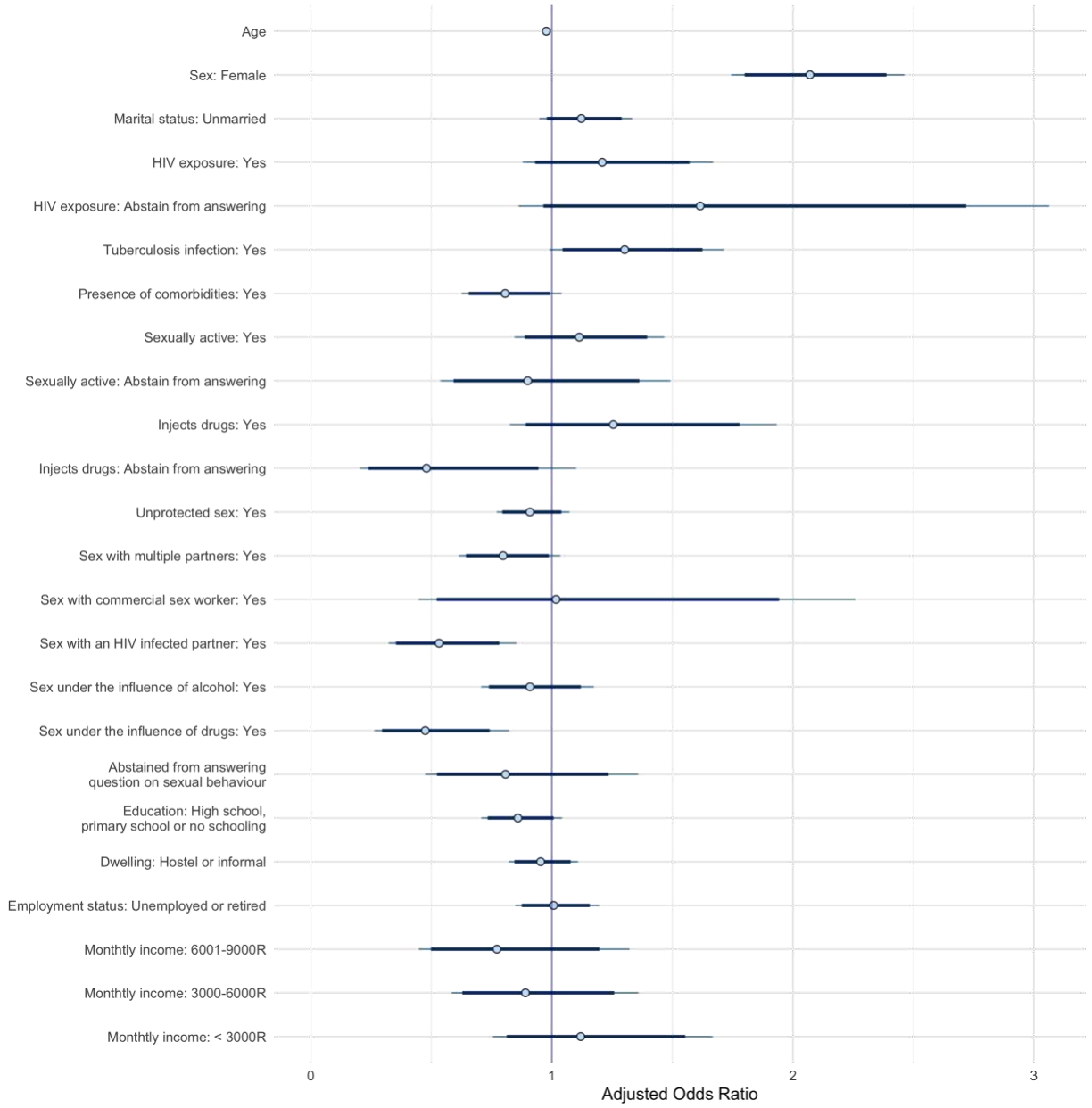


Figure 3. Bayesian logistic regression model posterior intervals of adjusted odds ratios for having tested in the past 6 months. Points represent the posterior medians, thick segments represent the 89% credible intervals and thin segments represent the 95% intervals of the adjusted odds ratios.

Discussion

In these analyses, we demonstrated that contextual factors such as less stable housing, lower education level, and subdistrict of residence impact the odds of HIV infection even after controlling for individual and behavioural factors. In addition, contextual factors associated with lower socioeconomic status increase the odds of engaging in sexual behaviours associated with HIV infection suggesting a potential pathway for how contextual factors impact the risk of HIV infection. Finally, individuals of lower socioeconomic status were not more or less likely to have been tested in the previous six months.

Consistent with previous reports, women (20-22), unmarried individuals (20, 21), and individuals who had not tested for HIV in the previous six months were also at increased odds of HIV infection in our adjusted analyses. Furthermore, our adjusted analyses showed that engaging in unprotected sex, sex with multiple partners, and sex with an HIV infected partner led to an increase in the odds of HIV infection, in line with what had previously been reported (20).

HIV prevention strategies in South Africa and globally have very much been focused on addressing individual risk of HIV through encouraging use of condoms, promoting sexual education, offering voluntary medical male circumcision, and targeting pre-exposure prophylaxis to those deemed at highest risk (23, 24). While key populations (people who inject drugs, men who have sex with men, sex workers, etc.) account for 65% of HIV infections globally, they represent a minority (39%) of infections in sub-Saharan Africa (24). Results from our study indicate that individual and behavioural factors do not fully explain HIV infection; contextual factors associated with socioeconomic status also impact the odds of HIV in townships of South Africa.

The contribution of contextual factors in explaining incidence of HIV has been overshadowed by the emphasis put on individual and behavioural factors in prevention initiatives, and a greater focus on addressing social determinants of health is warranted. Indeed, on the topic of public health policies, Geoffrey Rose famously argued that population strategies aimed at targeting population determinants of disease should be

favoured over high risk strategies as they address the root causes of incidence of disease, and they are behaviourally appropriate in that they do not discriminate across individual susceptibility (25).

The high risk strategy, focused on key populations and behavioural risk factors, has been a dominant approach for HIV prevention programmes, but they poorly address the underlying social determinants of health. On the other hand, population strategies addressing contextual factors by promoting more stable housing and higher educational attainment can have a positive impact on the incidence of HIV by shifting the disease curve of the entire population. As Rose points out, this approach is not only radical in its attempt to alter the underlying causes of disease but also non-discriminatory (25) which is particularly welcomed in the context of HIV, where discrimination and stigma towards people living with HIV have been important impediments to the progress towards elimination of the disease (24).

Other studies have previously reported on the impact of social determinants on risk of HIV (2, 6, 26, 27). One study found that those living in "urban informal areas" had over twice the odds of HIV prevalence compared to those in "urban formal areas" (22), consistent with our own findings that less stable housing increased the odds of being HIV positive even after controlling for individual and behavioural factors.

Additionally, in line with what our adjusted analyses showed, a number of studies have previously reported that educational attainment reduces the risk of HIV infection (2, 26, 27). It has been suggested that education can empower individuals to engage with prevention initiatives and adopt risk reducing behaviour more effectively (2). The results of our exploratory analyses would seem to support this mechanism as we found the odds of engaging in sex with an HIV infected individual was higher for those with lower levels of education after controlling for age, sex, and other contextual factors. On the other hand, educational attainment was not significantly associated with other behavioural risk factors for HIV, namely engaging in unprotected sex and sex with multiple partners.

While we found no significant association between employment status and HIV infection, consistent with previous studies (10, 26), our exploratory analyses found that

being not being employed did increase the odds of engaging in sex with multiple partners which was associated with increased odds of HIV infection. Indeed, several socioeconomic stressors stemming or intersecting with unemployment such as poverty and food insecurity can constrain individuals to engage in risky health behaviours including unprotected transactional sex (28) making these individuals particularly vulnerable to HIV infection and further deepening the impact of health inequities.

Finally, we found no significant association between contextual factors (education level, dwelling situation, employment status, and monthly income) and the odds of having tested for HIV in the previous six months, after controlling individual and behavioural factors. By contrast, other studies have reported an association between lower socioeconomic status and a decreased likelihood of HIV testing (22, 29, 30), but had not controlled for behavioural factors related to sex or drug injection in their analyses.

Our exploratory analyses did however reveal an association between HIV status and township of residence, as well as several individual and behavioural factors. This suggests that access to testing or attitudes toward testing are not necessarily dictated by contextual factors related to social determinants of health but rather by people's place of residence, individual, and behavioural factors. With this in mind, improving access to HIV testing and prevention services may not be sufficient in addressing the impact of contextual factors such as informal housing situations and lower levels of education on HIV infection.

Limitations. Several limitations are associated with our present study. First, there might be residual confounding due to broad strata within our categorical variables, if the effects of a confounding variable remain within strata. Moreover, since this study relied on secondary data, information on possible confounding factors such as knowledge of or attitudes toward HIV were not available to us. Additionally, some variables had few observations per strata, possibly leading to some imprecision in the reported credible intervals.

Second, although imputing data can help address bias due to missing data, it is plausible that some selection bias from the recruitment process remains (13, 14). We

assessed the pattern in missingness and noticed that observations with more than five missing values seem to be primarily missing values for variables which stemmed from the HIVSmart! application questions. This suggests that some responses might have failed to properly save and that the missing at random assumption, required by the data imputation process, is appropriate.

Third, although participants were given the option to abstain from answering some questions to mitigate this bias, there might nevertheless be measurement bias due to inaccurate reporting or social desirably bias.

Finally, since we used secondary data collected from townships of Cape Town, the results of our analyses may not generalise to other settings. We also recognise that the secondary data which was available to us may not have been well suited to detect differences in HIV infection across factors which were distributed relatively homogenously in our study population. Replicating these analyses within a population with greater heterogeneity in contextual factors could help shed a better light on the impact of context on HIV infection.

Conclusions

In conclusion, we have shown that contextual factors such as less stable housing, lower education level, and subdistrict of residence impact the odds of HIV infection even after controlling for individual and behavioural factors. While HIV prevention strategies have typically emphasised personal responsibility in health, our analyses show that individual and behavioural factors alone cannot fully explain HIV infection. Furthermore, contextual factors associated with lower socioeconomic position increased the odds of engaging in sexual behaviours associated with HIV infection suggesting that contextual factors may impact the risk of HIV infection through certain behaviours. Finally, individuals of lower socioeconomic position did not differ from those of higher socioeconomic position in their likelihood of having sought HIV testing in the previous six months indicating that an increased focus on promoting HIV testing and prevention services may be insufficient to mitigate the impact of contextual factors on the risk of HIV infection. Altogether, our

results stress the importance of moving beyond intervention which focus on influencing individual behaviours and tackle health inequities through a population strategy. Indeed, improving housing and promoting general education might be very impactful interventions in the fight towards the elimination of HIV and health inequities.

Competing interests

We declare no competing interests.

Authors' contributions

CLS analysed the data and wrote the first draft of the manuscript. SB provided critical support in the data analysis. SB, SJB and NP critically revised the first draft of the manuscript. NP, AE and KD conceptualised and designed the original research study.

Funding

This work was supported by operating grants 0732–05 and 0710–05 from Grand Challenges Canada (funded by Government of Canada). As part of comatching fund requirements, grant support was provided by: (1) the Government of South Africa, through its Medical Research Council SHIP programme and the Department of Science and Technology and (2) the McGill University Health Centre Foundation. Kit donation and support was provided by (3) OraSure Technologies Inc, Bethlehem. Dr Pant Pai also acknowledges the support of Fonds de recherche du Quebec – Santé Research Scholar Awards (Junior and Senior), and Canadian Institutes of Health Research grants (HHP 137872, PJT 153149, HBR 422155) for the duration of the project. KD acknowledges funding from the SA MRC (RFA-EMU-02-2017), EDCTP (TMA-2015SF-1043, TMA-1051-TESAII, TMA-CDF2015), UK Medical Research Council (MR/S03563X/1) and the Wellcome Trust (MR/S027777/1).

Acknowledgements

We acknowledge and thank our study participants for their participation and time. We also thank our staff members (health care workers, peer counsellors, and nurses) of the University of Cape Town as well as our staff members at the RI-MUHC for their dedication to the project.

References

1. UNAIDS. South Africa Country Factsheet: UNAIDS; 2022 [Available from: https://www.unaids.org/en/regionscountries/countries/southafrica.
2. Bärnighausen T, Hosegood V, Timaeus IM, Newell ML. The socioeconomic determinants of HIV incidence: evidence from a longitudinal, population-based study in rural South Africa. Aids. 2007;21 Suppl 7(Suppl 7):S29-38.
3. Cain D, Pitpitan EV, Eaton L, Carey KB, Carey MP, Mehlomakulu V, et al. Collective efficacy and HIV prevention in South African townships. J Community Health. 2013;38(5):885-93.
4. City of Cape Town. About Informal Housing: City of Cape Town; 2022 [Available from: https://www.capetown.gov.za/Family%20and%20home/Residential-property-and-houses/Informal-housing/About-informal-housing.
5. Statistics South Africa. City of Cape Town Statistics South Africa: Department: Statistics South Africa, Republic of South Africa; 2011 [Available from: http://www.statssa.gov.za/?page_id=993&id=city-of-cape-town-municipality.
6. Naidoo P, Chirinda W, McHunu G, Swartz S, Anderson J. Social and structural factors associated with vulnerability to HIV infection among young adults in South Africa. Psychol Health Med. 2015;20(3):369-79.
7. Djiadeu P, Smith MDR, Kushwaha S, Odhiambo AJ, Absalom D, Husbands W, et al. Social, Clinical, and Behavioral Determinants of HIV Infection and HIV Testing among Black Men in Toronto, Ontario: A Classification and Regression Tree Analysis. J Int Assoc Provid AIDS Care. 2020;19:2325958220934613.
8. Whitehead M. The concepts and principles of equity and health. Int J Health Serv. 1992;22(3):429-45.
9. Nyindo M. Complementary factors contributing to the rapid spread of HIV-I in sub-Saharan Africa: a review. East Afr Med J. 2005;82(1):40-6.
10. Johnson L, Budlender D. HIV risk factors: a review of demographic, socioeconomic, biomedical and behavioural determinants of HIV prevalence in South Africa. . Cape Town: University of Cape Town; 2002.
11. Pai N, Esmail A, Saha Chaudhuri P, Oelofse S, Pretorius M, Marathe G, et al. Impact of a personalised, digital, HIV self-testing app-based program on linkages and new infections in the township populations of South Africa. BMJ Glob Health. 2021;6(9).
12. Janssen R, Engel N, Esmail A, Oelofse S, Krumeich A, Dheda K, et al. Alone But Supported: A Qualitative Study of an HIV Self-testing App in an Observational Cohort Study in South Africa. AIDS Behav. 2020;24(2):467-74.
13. Azur MJ, Stuart EA, Frangakis C, Leaf PJ. Multiple imputation by chained equations: what is it and how does it work? Int J Methods Psychiatr Res. 2011;20(1):40-9.
14. Sterne JAC, White IR, Carlin JB, Spratt M, Royston P, Kenward MG, et al. Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. BMJ: British Medical Journal. 2009;339(7713):157-60.
15. R Core Team. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2020.

16. Buuren Sv, Groothuis-Oudshoorn K. mice: Multivariate Imputation by Chained Equations in R. 2011;45(3):1-67.
17. Bürkner P-C. Advanced Bayesian Multilevel Modeling with the R Package brms. The R Journal. 2018;10(1):395-411.
18. Wickham H, Averick M, Bryan J, Chang W, McGowan LDA, François R, et al. Welcome to the tidyverse. Journal of Open Source Software. 2019;4(43):1686.
19. Heinzen E, Sinnwell J, Atkinson E, Gunderson T, Dougherty G. arsenal: An Arsenal of 'R' Functions for Large-Scale Statistical Summaries. 2021.
20. Mabaso M, Makola L, Naidoo I, Mlangeni LL, Jooste S, Simbayi L. HIV prevalence in South Africa through gender and racial lenses: results from the 2012 population-based national household survey. Int J Equity Health. 2019;18(1):167.
21. Mba C. Sexual Behaviour and the Risks of HIV/AIDS and other STDs among Young People in Sub-Saharan Africa: a Review. Research Review of the Institute of African Studies. 2004;19.
22. Wabiri N, Taffa N. Socio-economic inequality and HIV in South Africa. BMC Public Health. 2013;13:1037.
23. WHO. HIV/AIDS Geneva: World Health Organization; 2021 [updated November 30, 2021. Available from: https://www.who.int/news-room/fact-sheets/detail/hiv-aids.
24. UNAIDS. Confronting Inequalities: Leasson for pandemic responses from 40 years of AIDS. Geneva: Joint United Nations Program on HIV/AIDS; 2021.
25. Rose G. Sick individuals and sick populations. INTERNATIONAL JOURNAL OF EPIDEMIOLOGY. 2001;30(Part 3):427-32.
26. Bunyasi EW, Coetzee DJ. Relationship between socioeconomic status and HIV infection: findings from a survey in the Free State and Western Cape Provinces of South Africa. BMJ Open. 2017;7(11):e016232.
27. Yousufzai M. Income Status and Education as Predictors of HIV Transmission in South Africa. NEXUS: The Canadian Student Journal of Anthropology. 2017;25.
28. Gillespie S, Kadiyala S, Greener R. Is poverty or wealth driving HIV transmission? AIDS. 2007;21.
29. Chimoyi L, Tshuma N, Muloongo K, Setswe G, Sarfo B, Nyasulu PS. HIV-related knowledge, perceptions, attitudes, and utilisation of HIV counselling and testing: a venue-based intercept commuter population survey in the inner city of Johannesburg, South Africa. Glob Health Action. 2015;8:26950.
30. Jooste S, Mabaso M, Taylor M, North A, Tadokera R, Simbayi L. Trends and determinants of ever having tested for HIV among youth and adults in South Africa from 2005-2017: Results from four repeated cross-sectional nationally representative household-based HIV prevalence, incidence, and behaviour surveys. PloS one. 2020;15(5):e0232883-e.

Appendix. Supplementary Tables

Supplementary table 1. Posterior median, 89% and 95% credible intervals (CrI) of the unadjusted odds ratios of the impact of subdistrict of residence on new HIV infections

	Odds Ratio	89% CrI	95% CrI
Township			
Subdistrict B	1.71	1.30 - 2.26	1.22 - 2.39
Subdistrict C	1.36	1.02 - 1.79	0.97 - 1.91

Supplementary table 2. Posterior median, 89% and 95% CrI of the adjusted odds ratios of the impact of individual, behavioural and contextual factors on new HIV infections

	Adjusted Odds Ratio	89% CrI	95% CrI
Township			
Subdistrict B	1.54	1.16 - 2.07	1.08 - 2.21
Subdistrict C	1.21	0.90 - 1.64	0.84 - 1.76
Individual Factors			
Age	1.01	1.00 - 1.03	1.00 - 1.03
Sex			
Female	1.44	1.11 - 1.89	1.05 - 2.02
Marital status			
Unmarried	2.07	1.56 - 2.76	1.47 - 2.96
Previous HIV test			
Not in the past 6 months	2.51	1.97 - 3.23	1.86 - 3.43
HIV exposure			
Yes	1.96	1.30 - 2.93	1.17 - 3.22
Abstain from answering	0.90	0.35 - 2.14	0.28 - 2.52
Tuberculosis infection			
Yes	1.20	0.81 - 1.74	0.74 - 1.88
Presence of comorbidities			
Yes	0.39	0.24 - 0.63	0.21 - 0.69
Behavioural Factors			
Sexually active			
Yes	0.78	0.54 - 1.14	0.50 - 1.24
Abstain from answering	0.54	0.26 - 1.12	0.21 - 1.30
Drug injection			
Yes	0.93	0.51 - 1.63	0.45 - 1.83
Abstain from answering	0.56	0.18 - 1.61	0.14 - 2.01
Sexual behaviour in the past 6 months			
Unprotected sex	1.26	0.98 - 1.64	0.93 - 1.75
Sex with multiple partners	1.51	1.07 - 2.12	0.98 - 2.29
Sex with a commercial sex worker	0.46	0.16 - 1.26	0.12 - 1.55
Sex with an HIV infected partner	6.43	4.22 - 9.74	3.81 - 10.68
Sex under the influence of alcohol	1.06	0.72 - 1.52	0.66 - 1.65

Sex under the influence of drugs	0.47	0.20 - 1.02	0.16 - 1.19
Abstained from answering question on sexual behaviour	1.78	0.91 - 3.34	0.77 - 3.83
Contextual Factors			
Education			
High school, primary school or no schooling	1.82	1.29 - 2.61	1.20 - 2.84
Dwelling			
Hostel or informal dwelling	1.34	1.07 - 1.68	1.01 - 1.77
Employment status			
Unemployed or retired	0.93	0.72 - 1.20	0.68 - 1.27
Monthly income			
6001-9000 R	1.23	0.59 - 2.56	0.49 - 3.01
3000-6000 R	0.71	0.38 - 1.33	0.33 - 1.54
<3000 R	1.18	0.67 - 2.12	0.60 - 2.45

Supplementary table 3. Posterior median, 89% and 95% CrI of the adjusted odds ratios of the impact of contextual factors on select behavioural factors

	Odds Ratio	89% CrI	95% CrI
Engaging in unprotected sex			
Age	0.99	0.99 – 1.00	0.99 – 1.00
Sex			
Female	1.30	1.14 - 1.50	1.10 - 1.55
Education			
High school, primary school or no schooling	1.03	0.87 - 1.22	0.84 - 1.26
Dwelling			
Hostel or informal dwelling	1.08	0.95 - 1.23	0.92 - 1.27
Employment status			
Unemployed or retired	0.94	0.81 - 1.10	0.79 - 1.13
Monthly income			
6001-9000 R	0.87	0.56 - 1.33	0.51 - 1.47
3000-6000 R	1.48	1.04 - 2.11	0.95 - 2.27
<3000 R	1.08	0.77 - 1.50	0.72 - 1.61
Engaging in sex with multiple partners			
Age	0.97	0.96 - 0.99	0.96 - 0.99
Sex			
Female	0.20	0.16 - 0.24	0.15 - 0.26
Education			
High school, primary school or no schooling	1.23	0.95 - 1.60	0.9 - 1.70
Dwelling			
Hostel or informal dwelling	1.32	1.08 - 1.62	1.03 - 1.69
Employment status			
Unemployed or retired	1.44	1.13 - 1.83	1.07 - 1.94
Monthly income	1.06	0.55 - 2.04	0.47 - 2.36

6001-9000 R			
3000-6000 R	1.16	0.67 - 2.02	0.60 - 2.30
<3000 R	1.02	0.61 - 1.73	0.56 - 1.95
Engaging in sex with an HIV infected partner			
Age	1.01	0.99 - 1.03	0.98 - 1.03
Sex			
Female	0.60	0.42 - 0.87	0.38 - 0.96
Education			
High school, primary school or no schooling	2.12	1.21 - 3.91	1.08 - 4.55
Dwelling			
Hostel or informal dwelling	1.02	0.72 - 1.47	0.66 - 1.59
Employment status			
Unemployed or retired	1.11	0.73 - 1.69	0.67 - 1.88
Monthly income			
6001-9000 R	0.61	0.22 - 1.61	0.17 - 2.01
3000-6000 R	0.85	0.39 - 1.86	0.33 - 2.27
<3000 R	0.76	0.37 - 1.58	0.32 - 1.90

Supplementary table 4. Posterior median, 89% and 95% CrI of the unadjusted odds ratios of the impact of subdistrict of residence on having tested for HIV in the past 6 months

	Odds Ratio	89% CrI	95% CrI
Township			
Subdistrict B	0.74	0.64 - 0.86	0.62 - 0.89
Subdistrict C	1.36	1.18 - 1.56	1.14 - 1.61

Supplementary table 5. Posterior median, 89% and 95% CrI of the adjusted odds ratios of the impact of individual, behavioural and contextual factors on having been tested for HIV in the past 6 months

	Adjusted Odds Ratio	89% CrI	95% CrI
Individual Factors			
Age	0.98	0.97 - 0.98	0.97 - 0.99
Sex			
Female	2.07	1.80 - 2.39	1.74 - 2.46
Marital status			
Unmarried	1.12	0.98 - 1.29	0.95 - 1.33
Previous HIV test			
Not in the past 6 months	1.21	0.93 - 1.57	0.88 - 1.67
HIV exposure			
Yes	1.62	0.96 - 2.72	0.86 - 3.06
Abstain from answering	1.30	1.04 - 1.62	0.99 - 1.71
Tuberculosis infection			
Yes	0.81	0.66 - 0.99	0.63 - 1.04
Presence of comorbidities	0.98	0.97 - 0.98	0.97 - 0.99

Yes			
Behavioural Factors			
Sexually active			
Yes	1.11	0.89 - 1.40	0.85 - 1.47
Abstain from answering	0.90	0.59 - 1.36	0.54 - 1.49
Drug injection			
Yes	1.26	0.89 - 1.78	0.83 - 1.93
Abstain from answering	0.48	0.24 - 0.94	0.20 - 1.10
Sexual behaviour in the past 6 months			
Unprotected sex	0.91	0.79 - 1.04	0.77 - 1.07
Sex with multiple partners	0.80	0.64 - 0.99	0.61 - 1.04
Sex with a commercial sex worker	1.01	0.52 - 1.94	0.45 - 2.26
Sex with an HIV infected partner	0.53	0.35 - 0.78	0.32 - 0.85
Sex under the influence of alcohol	0.91	0.74 - 1.12	0.71 - 1.17
Sex under the influence of drugs	0.47	0.30 - 0.74	0.26 - 0.82
Abstained from answering question on sexual behaviour	0.81	0.52 - 1.23	0.48 - 1.36
Contextual Factors			
Education			
High school, primary school or no schooling	0.86	0.73 - 1.01	0.71 - 1.04
Dwelling			
Hostel or informal dwelling	0.95	0.84 - 1.08	0.82 - 1.11
Employment status			
Unemployed or retired	1.01	0.88 - 1.16	0.85 - 1.20
Monthly income			
6001-9000 R	0.77	0.50 - 1.20	0.45 - 1.32
3000-6000 R	0.89	0.63 - 1.26	0.58 - 1.36
<3000 R	1.12	0.81 - 1.55	0.76 - 1.67

Chapter 5: Discussion

Summary of the findings

South Africa's initial response to the HIV epidemic and early inaccessibility of treatment for HIV for sub-Saharan African populations left the country of 59.3 million with the largest HIV burden in the world with nearly one fifth of its population affected by the virus[45]. In the past 15 years however, the South African government has been exceptionally aggressive in its response to the epidemic, evidenced by the increase in diagnosed and treated individuals, and the drastic improvement in life expectancy. There however remains place for progress as an estimated 8% of people living with HIV (PLWH) are still unaware of their status.

The burden of HIV remains unequally distributed in South Africa with people living in South African townships being more likely to be affected by the virus[2]. The impact of socioeconomic status (SES) on HIV has evolved with the epidemic; although individuals of higher SES appeared to be most impacted by HIV in the early years of the epidemic, this relationship seemed to have shifted as the epidemic progressed[60]. As such, today, it remains unclear to what extent socioeconomic indicators impact HIV infection after controlling for behaviours.

Through my literature review, I have also identified a number of structural and individual barriers to the utilisation of HIV prevention, testing and care services[52]. Many PLWH in South Africa live in underserved areas with limited access to healthcare facilities[52]. Staff shortages and long queues further stress access to HIV services[87-89]. At the individual level, some worry about breaches in confidentiality[82] and the stigma associated with HIV[87-89] while others are simply unaware of the services that are available to them. Finally, low risk perception may further impede testing uptake.

As such, providing individuals a self-administered risk assessment tool can be impactful in encouraging, in a non-judgemental manner, high-risk individuals to test for HIV[150] by tackling low risk perception. I have reviewed 41 existing risk staging tools for HIV from the published literature, and have identified few non-targeted risk staging tools for populations in Sub-Saharan Africa and none specifically for South African township

populations. Ong et al. reported on the use of unpublished risk staging tools in various settings, of which, less than half had been evaluated, and less than a fifth were known to have been beneficial[150]. Furthermore, none of the existing HIV risk assessment tools had been evaluated in combination with an HIV self-testing strategy.

As described in Chapter 1, the overall aim of this thesis was three-fold; using secondary data from the quasi-randomised trial, I aimed to:

1. Determine the accuracy of the HIVSmart! digital risk level in predicting new HIV infections in township populations of South Africa
2. Develop a risk staging model for South African township populations and validate it in combination with a digital self-testing programme
3. Quantify the association between contextual variables (subdistrict of residence, education, dwelling situation, employment and income) and new HIV infections while accounting for individual and behavioural factors

I had hypothesized that the HIVSmart! risk level would be correlated with new HIV infections and that incorporating contextual variables such as education, employment status, dwelling situation and income would improve its predictive accuracy. In manuscript 1, I evaluated the predictive accuracy of the HIVSmart! digital risk staging tool, which only incorporated information on behavioural risk factors related to drug use and sex. While the HIVSmart! risk levels were indeed associated with HIV (OR, 89% CrI: 1.15, 0.98 – 1.35), they poorly predicted HIV (AUC-ROC, 0.62).

Using predictive projection, a Bayesian approach for variable selection, I subsequently identified a parsimonious model of five sociodemographic and behavioural predictors of HIV with and improve predictive ability (AUC-ROC, 89% CrI: 0.71, 0.68 – 0.71). Consistent with my hypothesis, two contextual factors, education and dwelling situation, were useful in predicting HIV and thus included in my final risk assessment model. To my knowledge, this HIV risk assessment tool is the first to be developed for township populations in South Africa. Additionally, the risk staging model had a high sensitivity

(90.9%, 89.1% – 92.6%) and specificity (96.1%, 95.9% – 96.4%) when combined with a digital self-testing programme, suggesting that this combined approach can be an impactful strategy to screen for undiagnosed people living with HIV.

Furthermore, I had also hypothesised that the subdistrict of residence would be associated with new HIV infections and that individuals of lower SES would have increased likelihood of HIV after controlling for individual and behavioural factors. In manuscript 2, I evaluated this hypothesis and found that contextual factors such as less stable housing, lower education level and subdistrict of residence indeed impacted the odds of HIV infection after controlling for individual and behavioural factors. In addition, contextual factors associated with lower socioeconomic status increased the odds of engaging in sexual behaviours associated with HIV infection suggesting that contextual factors may impact the risk of HIV through certain behaviours. Finally, individuals of lower SES did not differ from those of higher SES in their likelihood of having sought HIV testing in the previous six months indicating that an increased focus on promoting HIV testing and prevention services may be insufficient to mitigate the impact of contextual factors on the risk of HIV.

Discussion of the methods

Dealing with missing data. In my analyses, I delt with missing data through multiple imputation by chained equations (MICE) with five imputed datasets. Multiple imputation involves filling in the missing values multiple times to create many "complete" datasets, and unlike single imputation, this approach accounts for the statistical uncertainly in the imputations[151]. MICE imputes missing values using information on how the variables in the data relate to each other and based on the other observed values for a given participant[151]. The MICE procedure operates under the missing at random assumption meaning that the mechanism underlying the missingness is only dependent on the observed data such that the observed data can be used to predict the unknown data[152]. Imputing under the missing at random assumption implies that after accounting for all the variables included in the imputation model, the remaining missingness is completely at random[151,153].

In my literature review of existing risk staging tools for HIV, I found that out of the 41 surveyed development studies, only 19 reported how they had delt with missing data, of which a majority conducted complete case analysis while only 4 used multiple imputation methods to address missingness. Conducting complete case analysis can be appropriate in the case where the proportion of missing values is low (below 5% as a general threshold), when only dependent values are missing or when the missing completely at random assumption is plausible, meaning that the pattern of missingness is neither dependent on the observed data nor the missing data[152]. The missing completely at random assumption is however rarely met with certainty[152]. In the case of my analyses, approximately 6.8% of the observations were missing at least one independent variable, making multiple imputation an appropriate strategy to handle missing data[152].

The Bayesian framework. I conducted the majority of the analyses included in this thesis using Bayesian statistics. A key difference between frequentist and Bayesian statistics is that in the frequentist view, parameters, which are the effects (ex. odds ratio) that we aim to estimate, are considered to be fixed yet unknown, and the data is considered random; frequentist methods are thus concerned with estimating the "true effects"[154]. By contrast, in the Bayesian framework, the probability of the parameters is considered uncertain and thus random, as such, Bayesian methods are concerned with making probability statements about parameters[154,155]. A key feature of Bayesian statistics is that, to derive inferences about the parameters, information is drawn from two sources: the prior distribution and the observed data[155].

Using Bayesian statistics for my analyses offered two key benefits. First, it permitted the integration of prior information to derive inferences about the parameters, and second, using Bayesian approaches provided more intuitive and meaningful inferences[155,156].

In my analyses, I used priors for age and sex which were informed by a longitudinal study by Bärnighausen et al.[29] Incorporating prior knowledge to compute the probability of the parameters implied that we could make use of information that was available outside of the data. In the case of the effect of sex on HIV infection, for example, it was already

understood from previous reports that women were more likely to be affected by HIV than men[29,52]. In the Bayesian framework for inference, as well as my analyses, this prior information of the effect of sex on HIV infection is considered along with the information contained in the observed data.

Bayesian statistics also offers inferences which are generally considered more intuitive and meaningful[155,156]. Whereas frequentist approaches will not indicate how likely the null or alternative hypothesis is, Bayesian analysis will directly provide a probability distribution for a parameter[155]. For instance, my analyses on the impact of educational attainment on HIV would indicate that there is an 89% probability that individuals without post-secondary education are between 29% and 161% more likely than those with post-secondary education to be living with HIV, controlling for individual and behavioural factors. The Bayesian interpretation of uncertainty is thus less convoluted and more intuitive than the frequentist interpretation[155,156].

Predictive projection feature selection. I used predictive projection feature selection, a novel approach for variable selection within the Bayesian framework, to identify key predictors of HIV in the South African context. The main advantages of the predictive projection approach are that it offers a good trade-off between sparsity and predictive accuracy, it is robust to overfitting and it produces posterior distributions which directly provide uncertainty estimates for parameters[157,158].

My review of existing HIV risk staging tools revealed that a majority of development studies used statistical association with HIV infection to select variables for inclusion in their risk assessment model. Additionally, out of 41 surveyed studies, 18 used a frequentist model selection method such as forward, backward or stepwise selection with 10 studies directly specifying a p-value as a stopping rule. However, as highlighted by Heinze and Dunkler, variable selection techniques relying on statistical association have little backing from statistical theory and can lead to systematic biases[159]. Indeed, Heinze and Dunkler point out that eliminating weak effects based on statistical insignificance can result in bias from omitting a relevant confounder[159]. They thus conclude by recommending that variable

selection be avoided if possible and that expert background knowledge be used to restrict the number of initial candidate variables[159].

Having parsimonious models is however desirable as they are often easier to interpret[157] and, in the context of an HIV risk assessment tool development, a parsimonious model limits the cost of future data collection. For these reasons, performing variable selection remains pertinent. Of 18 development studies that used a frequentist model selection method (forward, backward or stepwise selection), 8 used either Akaike information criterion (AIC) or Bayesian information criterion (BIC) to select a final model. Moreover, three development studies used the least absolute shrinkage and selection operator (LASSO) approach[138,141,142] a popular machine learning method for variable selection[157]. These two frequentist approaches are widely used for variable selection, however, as indicated by Bartonicek et al., neither can simultaneously easily offer predictive power and inference[157]. Furthermore, stepwise selection is liable to overfitting, limiting its predictive power[157].

In my analyses, I used predictive projection to select a parsimonious set of variables. This variable selection approach does not rely on statistical association to identify a sparse model and unlike many frequentist approaches, it is robust to overfitting[157]. Predictive projection works in two steps: (1) a reference model with all relevant predictor variables is fit, and (2) using projection, smaller submodels are fit to approximate the predictions of the reference model[158].

The candidate variables included in my reference model were preselected, during the design of the HIVSmart! trial, based on knowledge of HIV epidemiology, and included a combination of individual, behavioural and contextual variables. Using predictive projection, I then identified smaller models which predicted HIV status as accurately as the reference model in external data. The smallest submodel with acceptable predictive ability in the validation set was then selected. Using a validation set to select the final model further limited the possibility of overfitting the training data. Additionally, by examining the posterior distributions of the included variables in the final model, we are also able to

assess how each predictor was associated with HIV infection, thus obtaining predictive power as well as inference through this approach for variable selection.

Implications of the findings

Avoiding paternalism. Testing uptake in South Africa remains well below target levels with uptake being even lower among some groups such as men, unmarried individuals and individuals of lower socioeconomic status or with lower levels educational attainment[6,19].

While provider-initiated counseling and testing (PICT) for HIV is routine practice in South Africa and has been very successful in increasing HIV testing uptake, this practice is unpopular among South Africans[78]. Some report having felt coerced into testing and not being offered the opportunity to opt out, while others believed that HIV testing was mandatory to access other health services or that refusing testing would have compromised the care they were seeking[78,106,160]. Consequently, testers recruited through PICT often did not return to collect their test results and a majority indicated that they would not seek HIV testing in the near future[78].

Overall, these findings indicate that people are not receptive to paternalistic approaches which makes them feel disempowered. The unpopularity of PICT might be attributable to way it was implemented rather than the policy itself, but despite this, alternative approaches to improve testing uptake, link individuals to care, and that are generally preferred, exist. Indeed, individuals have repeatedly indicated a preference for initiating HIV testing and counseling on their own terms either through facility-based testing or self-testing[78].

In manuscript 1, I propose an alternative way to increase testing uptake: referring individuals to a HIV risk self-assessment tool and allowing them to assess their own need for testing. Offering a digital HIV risk screening tool and the possibility to self-test at home to individuals presenting at clinics is a less paternalistic approach than PICT. Finding from manuscript 1 indicated that risk staging followed by self-testing, guided by an application, could screen for undiagnosed PLWH with high sensitivity and specificity. In addition, task-

shifting personalised communication of HIV risk and promotion of testing services to a digital platform could be more cost-effective than PICT.

Qualitative studies on HIVSmart! have also indicated that participants appreciated the judgement free aspect of answering risk assessment questions through the HIVSmart! app, and that they felt like they could be more honest than when being questioned by a healthcare worker[14,17]. Some, however, did express concerns that the questions did not fully capture the complexity of their situation and was not an accurate depiction of their risk[17]. Nevertheless, risk assessment tools could serve as a starting point to spark discussions about HIV and testing between individuals and providers[26]. HIVSmart! supports a toll-free counseling line which would allow individuals to discuss to a counselor while maintaining their anonymity[13].

An interesting feature of the risk assessment tool could be to accompany risk assessment questions with an explanation as to why this question is being posed; this could help communicate how certain factors can contribute to an increased risk of HIV. Meta-analyses have indicated that the more people perceive themselves to be at risk of a disease, the greater their motivation to engage in healthy behaviours[161,162].

Leveraging digital tools. In 2019, 93% of the South African population was covered by 4G/LTE and the usage of international bandwidth has more than doubled between 2016 and 2018[112]. Access to cellphones in South Africa jumped from 76% in 2011 to 90% in 2018[112]. Furthermore, whereas only 33% of households had at least one person connected to the internet in 2011, this proportion was 64.7% in 2018[112].

Despite the recent progress in improving digital infrastructure and access, the South African government acknowledged that digital access remained poor and unequal[112]. Indeed, the cost of data is high (US$ 6.40 – 7.70 per GB) in South Africa, and inaccessible to many[112]. Consequently, individuals of lower socioeconomic status are less likely to own a cellphone or have internet access[113].

Recognising the relevance of improving internet use, skills and digital awareness, the South African government recently announced their intention to connect every

community to the internet by 2024 and to provide data for free to low-income households[112]. These policies aimed at improving digital access are key to the implementation of South Africa's National Digital Health Strategy[115] and to the scaling up of digital initiatives to support healthcare services[114].

As highlighted in my review of the literature, digital tools can support the provision of HIV services by connecting people HIV prevention programmes, facilitating the distribution of self-tests, linking people to care, and improving adherence to HIV treatment. In my first manuscript, I evaluated the sensitivity and specificity of HIVSmart!, a self-testing programme guided by an application, in township populations of South Africa, and found both measures of validity to be high (sensitivity of 99.93% and specificity of 95.52%). This sensitivity estimate was higher than the sensitivity reported by the manufacturer of the self-test (91.7%)[163], suggesting that offering digital supports to guide testing and interpretation can minimize user errors and improve the sensitivity of self-testing.

The WHO recommends that HIV testing be universally available and accessible in high prevalence settings[5]. As such, for those who might still be reluctant to seek testing in South African townships, I developed and validated a digital risk assessment tool for HIV which can inform a person's decision to test. The sensitivity of the risk staging tool was high in testing data (sensitivity: 91.0%, 89.1% – 92.7%) meaning that it would appropriately capture a majority of people living with HIV into the higher risk category and refer them to testing.

This risk staging tool combined with self-testing can be a powerful strategy to tackle low risk perception, a key barrier to testing uptake, by providing a convenient and private way to assess one's risk of HIV and guiding individuals through the self-testing process. When the HIV risk assessment model was evaluated in combination with the HIVSmart! self-testing programme, the resulting sequential sensitivity was 90.9% (89.1% – 92.6%) and the sequential specificity, 96.1% (95.9% – 96.4%), indicating that they would be useful tools to screen for undiagnosed people living with HIV.

Moreover, digital tools such as HIVSmart! can be particularly appealing to men who have lower testing rates, are generally more reluctant to attending healthcare facilities[6,76,79],

and yet, have higher rates of smartphone ownership and internet usage[164]. Furthermore, as reported by a systematic review, the power dynamic within a romantic relationship can often leave women powerless in deciding prevention interventions[68], making targeting HIV prevention to men all the more relevant to mitigate HIV transmission.

Digital tools are however not without potential pitfalls. Among the possible harms that have been highlighted, is the possibility for data breach, that is, a breach of security which results in the "accidental or unlawful destruction, loss, alteration, unauthorised disclosure of or access to personal data"[165]. Data breaches are unfortunately not uncommon in the health sector and can be particularly impactful; in 2019, the HIV status of over 14,000 people was revealed in a breach[165]. These types of breaches can have devastating repercussions for the people affected especially in communities where people fear the stigma and discrimination associated with a positive status. Breaches in confidentiality can also erode the trust that a community places in its health institutions and possibly reduce the utilisation of services. As such, scaling up digital interventions should not be done without the appropriate data protection measures. In the case of HIVSmart!, the application was housed on a Health Insurance Portability and Accountability Act (HIPAA)-compliant cloud server which protects private health information[13,18].

Finally, while there are concerns associated with the use of digital platforms to support provision of healthcare and the possible equity implications with scaling up interventions given the persisting digital divide, many remain optimistic about the digitization trends, contending that digital health can have a positive influence on health equity[166].

Tackling health inequities. Perhaps one of the most commonly cited definitions of health inequities is one developed by Margaret Whitehead who contends that health inequities are "differences [in health] which are unnecessary, avoidable, but in addition, are considered unfair and unjust"[167,168]. A broad interpretation of this definition would define many health disparities, particularly differences apparent across sociodemographic factors, as health inequities. The unequal distribution of the HIV burden between men and

women, across living situation, and varying educational attainment would all be considered health inequities as they are unnecessary, avoidable, unfair and unjust.

On the other hand, some may argue that discrepancies in health outcomes stemming from personal lifestyle choices such as engaging in risky sexual practices are not considered health inequities since the concept of inequity implies a moral wrong and making different life choices leading to different outcomes is not intrinsically wrong[168]. However, engaging in risky health behaviours is not always a choice freely made; poverty and food insecurity, for example, can compel individuals to engage in unprotected transactional sex[60], and health differences arising from such situations would likely fit the definition of health inequities, being unnecessary, avoidable, unfair and unjust.

Considering how health inequity is defined is important when designing and implementing health or social policies. One strategy to mitigate health inequities is to ensure equitable access to healthcare. Equity in healthcare means that there is both horizontal equity, equal treatment of those with equal needs, and vertical equity, those with greater needs are appropriately offered more healthcare interventions[169].

In my first manuscript, I developed a risk assessment tool for HIV which incorporates sociodemographic information and validated it in combination with a digital self-testing programme. This risk staging tool is intended to improve knowledge of HIV risk factors as well as raise awareness with respect to existing self-testing options in the aim of providing access to HIV testing to those who may face barriers to testing such as fear of stigma and discrimination. In addition, this digital HIV risk assessment tool incorporates contextual information on dwelling situation and educational attainment.

In my review of existing risk staging tools, only 12 out of the 41 surveyed tools included social variables such as education, income level, employment status and living situation in their risk prediction model. Including socioeconomic indicators in risk staging models for HIV implies that the tool will account for the disproportionate burden of HIV for people in different context and reflect it in the personalised risk assessments, thus appropriately prioritising those with greater needs for HIV services, in line with the principle of vertical equity in healthcare[169].

Another strategy to address inequities in health is to emphasize equal opportunity to access *health* itself and not merely *healthcare*[170]. Evidenced by the findings in manuscript 2, personal health intersects with socioeconomic status in a manner that can be independent of behaviours.

The principle of health equity contends that anyone, irrespective of their gender, background or socioeconomic status, should have equal access to health[170]. This does not necessarily imply that everyone should have the same health outcomes, but rather that everyone should be given equal opportunity to be healthy should they freely choose so.

Our analyses indicate that this currently not the case in township populations of South Africa: despite accounting for behavioural choices and individual factors, people with different levels of education and with different living arrangements have different burdens of HIV. HIV disproportionately affects those who are less educated and those living in informal housing or hostels. These findings stress the need for social policies which aim to improve education, living conditions as well as possibly other intersecting contextual factors including employment and income instability in the aims of allowing everyone equal opportunity to be healthy.

Poor sanitation and confined living spaces lead to a number of health threats which unequally affect those living in informal dwellings. Crowded spaces with poor ventilation make people living with HIV particularly vulnerable to co-infections such as tuberculosis, the primary cause of death in South Africa and for PLWH[40,171].

One study found that intimate partner violence is prevalent in households living in informal dwellings[172]. As such women are less likely to negotiate safe sex and are at increased risk of acquiring HIV. Futhermore, women living with HIV who are victims of intimate partner violence are also less likely to adhere to treatment[172]. These acts of violence can be attributed to stressors and the lack of dignity of informal living arrangements as well as intersecting factors such as unemployment; one qualitative study found that young men living in informal dwellings who are financially insecure resort to violence to achieve respect, masculinity, and social position[173].

Moreover, in urban informal settlements, men having multiple sexual partnerships is normalised, and seen as proof of desirability and masculinity. As such, men often use their sexual "success" as a way of earning respect among their peers[173]. Indeed, our findings indicated that men living in informal dwellings or who were unemployed were more likely to have multiple sexual partners. A recurrent theme from qualitative findings was the lack of dignity arising from living in urban informal settlements and being financially insecure which compelled young men to re-establish their dignity through violence and multiple sexual relationships, increasing the risk of HIV transmission[173].

People living in informal settlements often also face intersecting challenges such as deprivation in food, housing and healthcare which have been shown to increase the risk of engaging in high-risk sexual behaviours[174]. Indeed, the economic inequality that prevails in urban areas of South Africa creates opportunities for transactional sex between wealthier men and poorer women increasing the risk of HIV transmission[66]. Evidence show that women who face financial difficulties are twice as likely to engage in high-risk sexual behaviours[174]. In our adjusted analyses however, monthly income was not associated with HIV infection; it is possible that we were not able to detect an effect since we had adjusted for other intersecting factors such as education and dwelling situation.

Urban informal settlements are characterised by a lack of basic services such as water, sanitation and refuse removal, creating unhygienic environments and increasing the vulnerability to gendered violence and crime, particularly at night[175]. Further, access to basic healthcare services is often inadequate[175]; our analyses show that while people living in informal dwellings were more likely to be living with HIV, they were not more likely to have recently sought testing suggesting the need for improving access to testing and testing uptake in urban informal settlements.

Our findings add to the mounting body of evidence on how informal housing situations impact HIV infection. Despite this, there remains a lack of evidence on how improving housing situations can affect health and wellbeing, specifically HIV rates, in South Africa. Stronger evidence on the impact of housing situations on the HIV epidemic could help further advocate for housing upgrades. Nevertheless, even if improving housing

infrastructures does not directly affect health outcomes, it remains extremely valuable on its own.

Social policies are also less stigmatizing and discriminatory than targeting high-risk individuals and as Geoffrey Rose famously argued, they are radical in that they "deal with the root of the problem" and its effects are not "palliative and temporary"[176]. Further, a systematic review on HIV prevention intervention noted that the most successful interventions were those which "[moved beyond] individual-level measures of knowledge and psychosocial factors to address social and structural factors underlying HIV risk", and highlighted that programmes which focused on gender, sexual coercion, alcohol and economic risk had the greatest impact[94]. These findings stress the importance of moving beyond prevention interventions focused on influencing individual behaviours to also address social determinants of health which fuel health inequities.

Chapter 6: Conclusion

In conclusion, in this thesis, I set out to: (1) determine the accuracy of the HIVSmart! digital risk level in predicting new HIV infections in township populations of South Africa; (2) develop a risk staging model for South African township populations and validate it in combination with a digital self-testing programme, and (3) quantify the association between contextual variables and new HIV infections while accounting for individual and behavioural factors.

In manuscript 1, I evaluated the original HIVSmart! risk levels that focused on behavioural risk factors through Bayesian logistic regression and found that they were correlated with HIV infection. However, their predictive ability was modest.

To address my second objective, I used predictive projection, a variable selection approach within the Bayesian framework, to identify sociodemographic and behavioural predictors of HIV infection. I identified a model of five predictors (being unmarried, HIV testing history, having had sex with a partner living with HIV, dwelling situation, and education) which performed better than the original HIVSmart! risk levels. Additionally, when evaluated in combination with a digital self-testing programme, the sparse risk staging model showed high sensitivity and high specificity, indicating that digital HIV risk self-assessment followed by HIV self-testing can be an important strategy to screen for undiagnosed people living with HIV.

In manuscript 2, I evaluated, through Bayesian multivariate logistic regression, the extent to which contextual variables (subdistrict of residence, education, dwelling situation, employment and income) impacted HIV infection, after controlling for individual and behavioural factors. My findings indicated that, in the context of South African townships, individuals of lower educational attainment or living in informal dwellings or hostels were more likely to be affected by HIV than those with post-secondary education or living in a house, controlling for individual and behavioural factors.

Moreover, in exploratory analyses, I found that certain indicators associated with lower SES increased the odds of engaging in sexual behaviours associated with HIV infection suggesting that contextual factors may impact the risk of HIV through certain

behaviours. Finally, socioeconomic indicators did not significantly impact the likelihood of having recently tested for HIV.

Thus, HIV risk self-assessment tool combined with self-testing can help communicate personalised information on HIV risk to a broad audience, tackle low risk perception. and effectively screen for undiagnosed people living with HIV. However, public health interventions alone will unlikely be sufficient to mitigate the impact of social determinants of health on the burden of HIV. Social policies aimed at improving living arrangements and promoting general education are warranted in the fight towards the elimination of HIV and health inequities in South Africa.

References

1. *UNAIDS Data 2021.* Geneva: Joint United Nations Programme on HIV/AIDS;2021.
2. Cain D, Pitpitan EV, Eaton L, et al. Collective efficacy and HIV prevention in South African townships. *J Community Health.* 2013;38(5):885-893.
3. UNAIDS. *Confronting Inequalities: Leasson for pandemic responses from 40 years of AIDS.* Geneva: Joint United Nations Program on HIV/AIDS;2021.
4. Lippman SA, El Ayadi AM, Grignon JS, et al. Improvements in the South African HIV care cascade: findings on 90-90-90 targets from successive population-representative surveys in North West Province. *Journal of the International AIDS Society.* 2019;22(6):e25295.
5. *Consolidated guidelines on HIV testing services, 2019.* Geneva: World Health Organization;2020.
6. World Health O. *Guidelines on HIV self-testing and partner notification: supplement to consolidated guidelines on HIV testing services.* Geneva: World Health Organization; 2016.
7. Johnson LF, van Rensburg C, Govathson C, Meyer-Rath G. Optimal HIV testing strategies for South Africa: a model-based evaluation of population-level impact and cost-effectiveness. *Scientific Reports.* 2019;9(1):12621.
8. Pant Pai N, Sharma J, Shivkumar S, et al. Supervised and unsupervised self-testing for HIV in high- and low-risk populations: a systematic review. *PLoS Med.* 2013;10(4):e1001414.
9. Majam M. Progress on the scaling up of HIV testing in South Africa through varied distribution models using the oral HIV self-test kit. *Oral Diseases.* 2020;26(S1):137-140.
10. (2018). National HIV self-screening guidelines 2018. Department of Health Republic of South Africa Pretoria,
11. Pant Pai N, Behlim T, Abrahams L, et al. Will an unsupervised self-testing strategy for HIV work in health care workers of South Africa? A cross sectional pilot feasibility study. *PLoS One.* 2013;8(11):e79772.
12. McGuire M, de Waal A, Karellis A, et al. HIV self-testing with digital supports as the new paradigm: A systematic review of global evidence (2010-2021). *EClinicalMedicine.* 2021;39:101059-101059.
13. Pai N, Esmail A, Saha Chaudhuri P, et al. Impact of a personalised, digital, HIV self-testing app-based program on linkages and new infections in the township populations of South Africa. *BMJ Glob Health.* 2021;6(9).
14. Janssen R, Engel N, Esmail A, et al. Alone But Supported: A Qualitative Study of an HIV Self-testing App in an Observational Cohort Study in South Africa. *AIDS Behav.* 2020;24(2):467-474.
15. Pant Pai N, Smallwood M, Desjardins L, Goyette A, Vassal A, Thomas R. Can an HIVSmart! app-optimized self-testing strategy be operationalized in Canada? IDWeek; 2017; San Diego, USA.

16. Pant Pai N, Smallwood M, Desjardins L, et al. An Unsupervised Smart App-Optimized HIV Self-Testing Program in Montreal, Canada: Cross-Sectional Study. *J Med Internet Res.* 2018;20(11):e10258.
17. Janssen R, Engel N, Pant Pai N, et al. 'You're only there on the phone'? A qualitative exploration of community, affect and agential capacity in HIV self-testing using a smartphone app. *Sociol Health Illn.* 2021;43(3):591-606.
18. (2017). RI-MUHC scientist takes HIV self-testing global with HIVSmart!™. McGill University Health Centre. https://muhc.ca/newsroom/news/ri-muhc-scientist-takes-hiv-self-testing-global-hivsmart%E2%84%A2
19. Jooste S, Mabaso M, Taylor M, North A, Shean Y, Simbayi LC. Socio-economic differences in the uptake of HIV testing and associated factors in South Africa. *BMC Public Health.* 2021;21(1):1591.
20. Maughan-Brown B, Lloyd N, Bor J, Venkataramani AS. Changes in self-reported HIV testing during South Africa's 2010/2011 national testing campaign: gains and shortfalls. *J Int AIDS Soc.* 2016;19(1):20658.
21. Dzinamarira T, Mulindabigwi A, Mashamba-Thompson TP. Co-creation of a health education program for improving the uptake of HIV self-testing among men in Rwanda: nominal group technique. *Heliyon.* 2020;6(10):e05378.
22. Maughan-Brown B, Venkataramani AS. Accuracy and determinants of perceived HIV risk among young women in South Africa. *BMC Public Health.* 2017;18(1):42.
23. Price JT, Rosenberg NE, Vansia D, et al. Predictors of HIV, HIV Risk Perception, and HIV Worry Among Adolescent Girls and Young Women in Lilongwe, Malawi. *J Acquir Immune Defic Syndr.* 2018;77(1):53-63.
24. Wilton J, Kain T, Fowler S, et al. Use of an HIV-risk screening tool to identify optimal candidates for PrEP scale-up among men who have sex with men in Toronto, Canada: disconnect between objective and subjective HIV risk. *J Int AIDS Soc.* 2016;19(1):20777.
25. Govere SM, Galagan S, Tlou B, Mashamba-Thompson T, Bassett IV, Drain PK. Effect of perceived HIV risk on initiation of antiretroviral therapy during the universal test and treat era in South Africa. *BMC infectious diseases.* 2021;21(1):976.
26. John SA, Petroll AE, Walsh JL, Quinn KG, Patel VV, Grov C. Reducing the Discussion Divide by Digital Questionnaires in Health Care Settings: Disruptive Innovation for HIV Testing and PrEP Screening. *J Acquir Immune Defic Syndr.* 2020;85(3):302-308.
27. Ong JJ, Coulthard K, Quinn C, et al. Risk-Based Screening Tools to Optimise HIV Testing Services: a Systematic Review. *Current HIV/AIDS Reports.* 2022;19(2):154-165.
28. *Global strategy on digital health 2020-2025.* Geneva: World Health Organization;2021. 978-92-4-002092-4.
29. Bärnighausen T, Hosegood V, Timaeus IM, Newell ML. The socioeconomic determinants of HIV incidence: evidence from a longitudinal, population-based study in rural South Africa. *Aids.* 2007;21 Suppl 7(Suppl 7):S29-38.
30. Djiadeu P, Smith MDR, Kushwaha S, et al. Social, Clinical, and Behavioral Determinants of HIV Infection and HIV Testing among Black Men in Toronto,

Ontario: A Classification and Regression Tree Analysis. *J Int Assoc Provid AIDS Care.* 2020;19:2325958220934613.
31. Nyindo M. Complementary factors contributing to the rapid spread of HIV-I in sub-Saharan Africa: a review. *East Afr Med J.* 2005;82(1):40-46.
32. Johnson L, Budlender D. *HIV risk factors: a review of demographic, socioeconomic, biomedical and behavioural determinants of HIV prevalence in South Africa. .* Cape Town: University of Cape Town;2002.
33. Booysen FlR, Summerton J. Poverty, risky sexual behaviour, and vulnerability to HIV infection: evidence from South Africa. *Journal of health, population, and nutrition.* 2002;20(4):285-288.
34. le R Booysen F. HIV/AIDS, poverty and risky sexual behaviour in South Africa. *African Journal of AIDS Research.* 2004;3(1):57-67.
35. About HIV. Centers for Disease Control and Prevention. HIV Basics Web site. https://www.cdc.gov/hiv/basics/whatishiv.html. Published 2021. Updated June 1. Accessed2022.
36. Pantaleo G, Fauci AS. Immunopathogenesis of HIV infection. *Annu Rev Microbiol.* 1996;50:825-854.
37. WHO. HIV/AIDS. World Health Organization. https://www.who.int/news-room/fact-sheets/detail/hiv-aids. Published 2021. Updated November 30, 2021. Accessed2022.
38. Wandeler G, Johnson LF, Egger M. Trends in life expectancy of HIV-positive adults on antiretroviral therapy across the globe: comparisons with general population. *Current opinion in HIV and AIDS.* 2016;11(5):492-500.
39. Calabrese SK, Mayer KH, Marcus JL. Prioritising pleasure and correcting misinformation in the era of U=U. *The lancet HIV.* 2021;8(3):e175-e180.
40. (2017). Let our actions count: South Africa's national strategic plan for HIV, TB and STIs 2017–2022. SANAC Pretoria,
41. Bulled N, Singer M. In the shadow of HIV & TB: A commentary on the COVID epidemic in South Africa. *Glob Public Health.* 2020;15(8):1231-1243.
42. McShane H. Co-infection with HIV and TB: double trouble. *Int J STD AIDS.* 2005;16(2):95-100; quiz 101.
43. Abdool Karim SS, Abdool Karim Q. HIV/AIDS in South Africa. In: 2nd ed. ed. Leiden: Cambridge University Press; 2010: https://public.ebookcentral.proquest.com/choice/publicfullrecord.aspx?p=691908.
44. Mann JM, Chin J, Piot P, Quinn T. The international epidemiology of AIDS. *Sci Am.* 1988;259(4):82-89.
45. UNAIDS. South Africa Country Factsheet. UNAIDS. https://www.unaids.org/en/regionscountries/countries/southafrica. Published 2022. Accessed June 7th 2022.
46. Simelela N, Venter WD, Pillay Y, Barron P. A Political and Social History of HIV in South Africa. *Curr HIV/AIDS Rep.* 2015;12(2):256-261.
47. HIV estimates with uncertainty bounds 1990-Present. 2021. https://www.unaids.org/en/resources/documents/2021/HIV_estimates_with_uncertainty_bounds_1990-present. Accessed May 2022.

48. Mayosi BM, Lawn JE, van Niekerk A, et al. Health in South Africa: changes and challenges since 2009. *Lancet (London, England).* 2012;380(9858):2029-2043.
49. Nordling L. South Africa ushers in a new era for HIV. *Nature.* 2016;535(7611):214-217.
50. Life expectancy at birth, total (years) - South Africa. 2020. https://data.worldbank.org/indicator/SP.DYN.LEoo.IN?locations=ZA. Accessed June 6th 2022.
51. Achoki T, Sartorius B, Watkins D, et al. Health trends, inequalities and opportunities in South Africa's provinces, 1990-2019: findings from the Global Burden of Disease 2019 Study. *Journal of epidemiology and community health.* 2022.
52. Kim H, Tanser F, Tomita A, Vandormael A, Cuadros DF. Beyond HIV prevalence: identifying people living with HIV within underserved areas in South Africa. *BMJ Global Health.* 2021;6(4):e004089.
53. Mba C. Sexual Behaviour and the Risks of HIV/AIDS and other STDs among Young People in Sub-Saharan Africa: a Review. *Research Review of the Institute of African Studies.* 2004;19.
54. Shisana O, Risher K, Celentano DD, et al. Does marital status matter in an HIV hyperendemic country? Findings from the 2012 South African National HIV Prevalence, Incidence and Behaviour Survey. *AIDS Care.* 2016;28(2):234-241.
55. Wand H, Reddy T, Naidoo S, et al. A Simple Risk Prediction Algorithm for HIV Transmission: Results from HIV Prevention Trials in KwaZulu Natal, South Africa (2002-2012). *AIDS Behav.* 2018;22(1):325-336.
56. Dosekun O, Fox J. An overview of the relative risks of different sexual behaviours on HIV transmission. *Current Opinion in HIV and AIDS.* 2010;5(4):291-297.
57. Kagaayi J, Gray RH, Whalen C, et al. Indices to Measure Risk of HIV Acquisition in Rakai, Uganda. *PLoS ONE.* 2014;9(4).
58. Kalichman SC, Simbayi LC, Jooste S, Cherry C, Cain D. Poverty-related stressors and HIV/AIDS transmission risks in two South African communities. *J Urban Health.* 2005;82(2):237-249.
59. Weimann A, Oni T. A Systematised Review of the Health Impact of Urban Informal Settlements and Implications for Upgrading Interventions in South Africa, a Rapidly Urbanising Middle-Income Country. *International Journal of Environmental Research and Public Health.* 2019;16(19):3608.
60. Gillespie S, Kadiyala S, Greener R. Is poverty or wealth driving HIV transmission? *AIDS.* 2007;21.
61. Hargreaves JR, Bonell CP, Boler T, et al. Systematic review exploring time trends in the association between educational attainment and risk of HIV infection in sub-Saharan Africa. *AIDS.* 2008;22(3).
62. Bunyasi EW, Coetzee DJ. Relationship between socioeconomic status and HIV infection: findings from a survey in the Free State and Western Cape Provinces of South Africa. *BMJ Open.* 2017;7(11):e016232.
63. Wabiri N, Taffa N. Socio-economic inequality and HIV in South Africa. *BMC Public Health.* 2013;13:1037.

64. Andrew G, Tarylee R, Kristin D, Rachel J. HIV-Prevalence in South Africa by settlement type: A repeat population-based cross-sectional analysis of men and women. *PLOS ONE.* 2020;15(3). doi:10.1371/journal.pone.0230105.
65. Mishra V, Assche SB, Greener R, et al. HIV infection does not disproportionately affect the poorer in sub-Saharan Africa. *Aids.* 2007;21 Suppl 7:S17-28.
66. Fox AM. THE HIV–POVERTY THESIS RE-EXAMINED: POVERTY, WEALTH OR INEQUALITY AS A SOCIAL DETERMINANT OF HIV INFECTION IN SUB-SAHARAN AFRICA? *Journal of Biosocial Science.* 2012;44(4):459-480.
67. Isbell MT, Kilonzo N, Mugurungi O, Bekker L-G. We neglect primary HIV prevention at our peril. *The Lancet HIV.* 2016;3(7):e284-e285.
68. Warren EA, Paterson P, Schulz WS, et al. Risk perception and the influence on uptake and use of biomedical prevention interventions for HIV in sub-Saharan Africa: A systematic literature review. *PLOS ONE.* 2018;13(6):e0198680.
69. Sarkar S, Corso P, Ebrahim-Zadeh S, Kim P, Charania S, Wall K. Cost-effectiveness of HIV Prevention Interventions in Sub-Saharan Africa: A Systematic Review. *EClinicalMedicine.* 2019;10:10-31.
70. Maartens G, Celum C, Lewin SR. HIV infection: epidemiology, pathogenesis, treatment, and prevention. *Lancet.* 2014;384(9939):258-271.
71. Sani AS, Abraham C, Denford S, Ball S. School-based sexual health education interventions to prevent STI/HIV in sub-Saharan Africa: a systematic review and meta-analysis. *BMC Public Health.* 2016;16(1):1069.
72. Nikolopoulos GK, Christaki E, Paraskevis D, Bonovas S. Pre-Exposure Prophylaxis for HIV: Evidence and Perspectives. *Curr Pharm Des.* 2017;23(18):2579-2591.
73. Abbas UL, Anderson RM, Mellors JW. Potential Impact of Antiretroviral Chemoprophylaxis on HIV-1 Transmission in Resource-Limited Settings. *PLOS ONE.* 2007;2(9):e875.
74. Venter WDF. Pre-exposure Prophylaxis: The Delivery Challenge. *Front Public Health.* 2018;6:188.
75. Smith JA, Anderson S-J, Harris KL, et al. Maximising HIV prevention by balancing the opportunities of today with the promises of tomorrow: a modelling study. *The Lancet HIV.* 2016;3(7):e289-e296.
76. Johnson LF, Rehle TM, Jooste S, Bekker LG. Rates of HIV testing and diagnosis in South Africa: successes and challenges. *Aids.* 2015;29(11):1401-1409.
77. Kennedy CE, Fonner VA, Sweat MD, Okero FA, Baggaley R, O'Reilly KR. Provider-initiated HIV testing and counseling in low- and middle-income countries: a systematic review. *AIDS Behav.* 2013;17(5):1571-1590.
78. van Dyk AC. Client-initiated, provider-initiated, or self-testing for HIV: what do South Africans prefer? *J Assoc Nurses AIDS Care.* 2013;24(6):e45-56.
79. Musheke M, Ntalasha H, Gari S, et al. A systematic review of qualitative findings on factors enabling and deterring uptake of HIV testing in Sub-Saharan Africa. *BMC Public Health.* 2013;13:220.
80. Chimoyi L, Tshuma N, Muloongo K, Setswe G, Sarfo B, Nyasulu PS. HIV-related knowledge, perceptions, attitudes, and utilisation of HIV counselling and testing: a

venue-based intercept commuter population survey in the inner city of Johannesburg, South Africa. *Glob Health Action.* 2015;8:26950.
81. Gras L, van Sighem A, Bezemer D, Smit C, Wit F, de Wolf F. Lower mortality and earlier start of combination antiretroviral therapy in patients tested repeatedly for HIV than in those with a positive first test. *Aids.* 2011;25(6):813.
82. Regan S, Losina E, Chetty S, et al. Factors associated with self-reported repeat HIV testing after a negative result in Durban, South Africa. *PLoS One.* 2013;8(4):e62362.
83. Levi J, Raymond A, Pozniak A, Vernazza P, Kohler P, Hill A. Can the UNAIDS 90-90-90 target be achieved? A systematic analysis of national HIV treatment cascades. *BMJ Global Health.* 2016;1(2):e000010.
84. Tanser F, Bärnighausen T, Grapsa E, Zaidi J, Newell M-L. High coverage of ART associated with decline in risk of HIV acquisition in rural KwaZulu-Natal, South Africa. *Science (New York, NY).* 2013;339(6122):966-971.
85. Fox MP, Rosen S. A new cascade of HIV care for the era of "treat all". *PLoS medicine.* 2017;14(4):e1002268-e1002268.
86. World Health O. *Consolidated guidelines on the use of antiretroviral drugs for treating and preventing HIV infection: recommendations for a public health approach.* 2nd ed. Geneva: World Health Organization; 2016.
87. Musheke M, Ntalasha H, Gari S, et al. A systematic review of qualitative findings on factors enabling and deterring uptake of HIV testing in Sub-Saharan Africa. *BMC Public Health.* 2013;13(1):220.
88. Sharma M, Ying R, Tarr G, Barnabas R. Systematic review and meta-analysis of community and facility-based HIV testing to address linkage to care gaps in sub-Saharan Africa. *Nature.* 2015;528(7580):S77-S85.
89. Chimoyi L, Chikovore J, Musenge E, et al. Understanding factors influencing utilization of HIV prevention and treatment services among patients and providers in a heterogeneous setting: A qualitative study from South Africa. *PLOS Global Public Health.* 2022;2(2):e0000132.
90. van Heerden A, Harris DM, van Rooyen H, et al. Perceived mHealth barriers and benefits for home-based HIV testing and counseling and other care: Qualitative findings from health officials, community health workers, and persons living with HIV in South Africa. *Social science & medicine (1982).* 2017;183:97-105.
91. Ryan S, Hahn E, Rao A, et al. The impact of HIV knowledge and attitudes on HIV testing acceptance among patients in an emergency department in the Eastern Cape, South Africa. *BMC Public Health.* 2020;20(1):1066.
92. Tenkorang EY, Maticka-Tyndale E, Rajulton F. A multi-level analysis of risk perception, poverty and sexual risk-taking among young people in Cape Town, South Africa. *Health & Place.* 2011;17(2):525-535.
93. Kalichman SC, Simbayi L. Traditional beliefs about the cause of AIDS and AIDS-related stigma in South Africa. *AIDS care.* 2004;16(5):572-580.
94. Harrison A, Newell M-L, Imrie J, Hoddinott G. HIV prevention for South African youth: which interventions work? A systematic review of current evidence. *BMC Public Health.* 2010;10(1):102.

95. Jia KM, Eilerts H, Edun O, et al. Risk scores for predicting HIV incidence among adult heterosexual populations in sub-Saharan Africa: a systematic review and meta-analysis. *Journal of the International AIDS Society.* 2022;25(1):e25861.
96. Manby L, Aicken C, Delgrange M, Bailey JV. Effectiveness of eHealth Interventions for HIV Prevention and Management in Sub-Saharan Africa: Systematic Review and Meta-analyses. *AIDS and Behavior.* 2022;26(2):457-469.
97. Roehr B. FDA approves "instant" HIV home test. *BMJ : British Medical Journal.* 2012;345:e4636.
98. Harichund C, Moshabela M, Kunene P, Abdool Karim Q. Acceptability of HIV self-testing among men and women in KwaZulu-Natal, South Africa. *AIDS Care.* 2019;31(2):186-192.
99. Lippman SA, Lane T, Rabede O, et al. High Acceptability and Increased HIV-Testing Frequency After Introduction of HIV Self-Testing and Network Distribution Among South African MSM. *J Acquir Immune Defic Syndr.* 2018;77(3):279-287.
100. Krause J, Subklew-Sehume F, Kenyon C, Colebunders R. Acceptability of HIV self-testing: a systematic literature review. *BMC Public Health.* 2013;13:735.
101. Jamil MS, Eshun-Wilson I, Witzel TC, et al. Examining the effects of HIV self-testing compared to standard HIV testing services in the general population: A systematic review and meta-analysis. *EClinicalMedicine.* 2021;38:100991.
102. Njau B, Damian DJ, Abdullahi L, Boulle A, Mathews C. The effects of HIV self-testing on the uptake of HIV testing, linkage to antiretroviral treatment and social harms among adults in Africa: A systematic review and meta-analysis. *PLoS One.* 2021;16(1):e0245498.
103. Brown AN, Djimeu EW, Cameron DB. A review of the evidence of harm from self-tests. *AIDS Behav.* 2014;18 Suppl 4(Suppl 4):S445-449.
104. Eshun-Wilson I, Jamil MS, Witzel TC, et al. A Systematic Review and Network Meta-analyses to Assess the Effectiveness of Human Immunodeficiency Virus (HIV) Self-testing Distribution Strategies. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America.* 2021;73(4):e1018-e1028.
105. Hlongwa M, Mashamba-Thompson T, Makhunga S, Murararaneza C, Hlongwana K. Men's perspectives on HIV self-testing in sub-Saharan Africa: a systematic review and meta-synthesis. *BMC Public Health.* 2020;20(1):66.
106. Martínez Pérez G, Cox V, Ellman T, et al. 'I Know that I Do Have HIV but Nobody Saw Me': Oral HIV Self-Testing in an Informal Settlement in South Africa. *PloS one.* 2016;11(4):e0152653.
107. Tahlil KM, Ong JJ, Rosenberg NE, et al. Verification of HIV Self-Testing Use and Results: A Global Systematic Review. *AIDS Patient Care STDS.* 2020;34(4):147-156.
108. Njau B, Covin C, Lisasi E, et al. A systematic review of qualitative evidence on factors enabling and deterring uptake of HIV self-testing in Africa. *BMC Public Health.* 2019;19(1):1289.
109. Cao B, Bao H, Oppong E, et al. Digital health for sexually transmitted infection and HIV services: a global scoping review. *Curr Opin Infect Dis.* 2020;33(1):44-50.

110. Daher J, Vijh R, Linthwaite B, et al. Do digital innovations for HIV and sexually transmitted infections work? Results from a systematic review (1996-2017). *BMJ Open.* 2017;7(11):e017604.
111. Jongbloed K, Parmar S, van der Kop M, Spittal PM, Lester RT. Recent Evidence for Emerging Digital Technologies to Support Global HIV Engagement in Care. *Curr HIV/AIDS Rep.* 2015;12(4):451-461.
112. (2022). South Africa's National Infrastructure Plan 2050. Department of Public Works and Infrastructure, South Africa. https://www.gov.za/sites/default/files/gcis_document/202203/46033gon1874.pdf
113. *Internet Connectivity Seen as Having Positive Impact on Life in Sub-Saharan Africa.* Pew Research Center;2018.
114. Simoni JM, Kutner BA, Horvath KJ. Opportunities and Challenges of Digital Technology for HIV Treatment and Prevention. *Curr HIV/AIDS Rep.* 2015;12(4):437-440.
115. (2019). National Digital Health Strategy for South Africa. National Department of Health, Pretoria, South Africa. https://www.health.gov.za/wp-content/uploads/2020/11/national-digital-strategy-for-south-africa-2019-2024-b.pdf
116. Leblanc J, Hejblum G, Costagliola D, et al. Targeted HIV Screening in Eight Emergency Departments: The DICI-VIH Cluster-Randomized Two-Period Crossover Trial. *Ann Emerg Med.* 2018;72(1):41-53.e49.
117. Domínguez-Berjón MF, Pichiule-Castañeda M, García-Riolobos MC, et al. A feasibility study for 3 strategies promoting HIV testing in primary health care in Madrid, Spain (ESTVIH project). *J Eval Clin Pract.* 2017;23(6):1408-1414.
118. Yun K, Chu Z, Zhang J, et al. Mobile Phone Intervention Based on an HIV Risk Prediction Tool for HIV Prevention Among Men Who Have Sex With Men in China: Randomized Controlled Trial. *JMIR Mhealth Uhealth.* 2021;9(4):e19511.
119. Jones J, Stephenson R, Smith DK, et al. Acceptability and willingness among men who have sex with men (MSM) to use a tablet-based HIV risk assessment in a clinical setting. *SpringerPlus.* 2014;3(1):1-7.
120. Allison WE, Kiromat M, Vince J, et al. Development of a clinical algorithm to prioritise HIV testing of hospitalised paediatric patients in a low resource moderate prevalence setting. *Arch Dis Child.* 2011;96(1):67-72.
121. Antelman G, Gill MM, Jahanpour O, et al. Balancing HIV testing efficiency with HIV case-identification among children and adolescents (2-19 years) using an HIV risk screening approach in Tanzania. *PLoS One.* 2021;16(5):e0251247.
122. Du Plessis NM, Muller CJB, Avenant T, Pepper MS, Goga AE. An Early Infant HIV Risk Score for Targeted HIV Testing at Birth. *Pediatrics.* 2019;143(6).
123. Ferrand RA, Weiss HA, Nathoo K, et al. A primary care level algorithm for identifying HIV-infected adolescents in populations at high risk through mother-to-child transmission. *Tropical Medicine & International Health.* 2011;16(3):349-355.
124. Mafaune HW, Sacks E, Chadambuka A, et al. Effectiveness of Maternal Transmission Risk Stratification in Identification of Infants for HIV Birth Testing: Lessons From Zimbabwe. *J Acquir Immune Defic Syndr.* 2020;84 Suppl 1(1):S28-s33.

125. Balkus JE, Brown E, Palanee T, et al. An Empiric HIV Risk Scoring Tool to Predict HIV-1 Acquisition in African Women. *J Acquir Immune Defic Syndr.* 2016;72(3):333-343.
126. Peebles K, Palanee-Phillips T, Balkus JE, et al. Age-Specific Risk Scores Do Not Improve HIV-1 Prediction Among Women in South Africa. *J Acquir Immune Defic Syndr.* 2020;85(2):156-164.
127. Pintye J, Drake AL, Kinuthia J, et al. A Risk Assessment Tool for Identifying Pregnant and Postpartum Women Who May Benefit From Preexposure Prophylaxis. *Clin Infect Dis.* 2017;64(6):751-758.
128. Wand H, Ramjee G. Assessing and evaluating the combined impact of behavioural and biological risk factors for HIV seroconversion in a cohort of South African women. *AIDS Care.* 2012;24(9):1155-1162.
129. Ayton SG, Pavlicova M, Tamir H, Abdool Karim Q. Development of a prognostic tool exploring female adolescent risk for HIV prevention and PrEP in rural South Africa, a generalised epidemic setting. *Sex Transm Infect.* 2020;96(1):47-54.
130. Rosenberg NE, Kudowa E, Price JT, et al. Identifying Adolescent Girls and Young Women at High Risk for HIV Acquisition: A Risk Assessment Tool From the Girl Power-Malawi Study. *Sex Transm Dis.* 2020;47(11):760-766.
131. Dijkstra M, de Bree GJ, Stolte IG, et al. Development and validation of a risk score to assist screening for acute HIV-1 infection among men who have sex with men. *BMC Infect Dis.* 2017;17(1):425.
132. Facente SN, Pilcher CD, Hartogensis WE, et al. Performance of risk-based criteria for targeting acute HIV screening in San Francisco. *PLoS One.* 2011;6(7):e21813.
133. Lin TC, Gianella S, Tenenbaum T, Little SJ, Hoenigl M. A Simple Symptom Score for Acute Human Immunodeficiency Virus Infection in a San Diego Community-Based Screening Program. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America.* 2018;67(1):105-111.
134. Powers KA, Miller WC, Pilcher CD, et al. Improved detection of acute HIV-1 infection in sub-Saharan Africa: development of a risk score algorithm. *Aids.* 2007;21(16):2237-2242.
135. Sanders EJ, Wahome E, Powers KA, et al. Targeted screening of at-risk adults for acute HIV-1 infection in sub-Saharan Africa. *Aids.* 2015;29 Suppl 3(0 3):S221-230.
136. Wahome E, Fegan G, Okuku HS, et al. Evaluation of an empiric risk screening score to identify acute and early HIV-1 infection among MSM in Coastal Kenya. *Aids.* 2013;27(13):2163-2166.
137. Hoenigl M, Weibel N, Mehta SR, et al. Development and validation of the San Diego Early Test Score to predict acute and early HIV infection risk in men who have sex with men. *Clin Infect Dis.* 2015;61(3):468-475.
138. Roberts DA, Cuadros D, Vandormael A, et al. Predicting the Risk of Hiv-1 Acquisition in Rural South Africa Using Geospatial Data. *Clin Infect Dis.* 2022.
139. Haukoos JS, Lyons MS, Lindsell CJ, et al. Derivation and validation of the Denver Human Immunodeficiency Virus (HIV) risk score for targeted HIV screening. *Am J Epidemiol.* 2012;175(8):838-846.

140. Kabapy AF, Shatat HZ, Abd El-Wahab EW. Identifying factors increasing the risk of acquiring HIV among Egyptians to construct a consensus web-based tool for HIV risk assessment. *Curr Med Res Opin.* 2021;37(6):973-984.
141. Krakower DS, Gruber S, Hsu K, et al. Development and validation of an automated HIV prediction algorithm to identify candidates for pre-exposure prophylaxis: a modelling study. *The lancet HIV.* 2019;6(10):e696-e704.
142. Marcus JL, Hurley LB, Krakower DS, Alexeeff S, Silverberg MJ, Volk JE. Use of electronic health record data and machine learning to identify candidates for HIV pre-exposure prophylaxis: a modelling study. *The lancet HIV.* 2019;6(10):e688-e695.
143. Muttai H, Guyah B, Musingila P, et al. Development and Validation of a Sociodemographic and Behavioral Characteristics-Based Risk-Score Algorithm for Targeting HIV Testing Among Adults in Kenya. *AIDS Behav.* 2021;25(2):297-310.
144. Li LL, Jiang Z, Song WL, Ding YY, Xu J, He N. [Development of HIV infection risk assessment tool for men who have sex with men based on Delphi method]. *Zhonghua Liu Xing Bing Xue Za Zhi.* 2017;38(10):1426-1430.
145. Bursac Z, Gauss CH, Williams DK, Hosmer DW. Purposeful selection of variables in logistic regression. *Source Code for Biology and Medicine.* 2008;3(1):17.
146. Feller DJ, Zucker J, Yin MT, Gordon P, Elhadad N. Using Clinical Notes and Natural Language Processing for Automated HIV Risk Assessment. *J Acquir Immune Defic Syndr.* 2018;77(2):160-166.
147. Falasinnu T, Gustafson P, Gilbert M, Shoveller J. P144 Validation of the denver HIV risk score for targeting HIV screening in vancouver, british columbia. *Sexually Transmitted Infections.* 2015;91(Suppl 1):A63.62-A63.
148. Zheng M, He J, Yuan Z, et al. Risk assessment and identification of HIV infection among men who have sex with men: a cross-sectional study in Southwest China. *BMJ Open.* 2020;10(11):e039557.
149. Lin TC, Dijkstra M, De Bree GJ, Schim van der Loeff MF, Hoenigl M. Brief Report: The Amsterdam Symptom and Risk-Based Score Predicts for Acute HIV Infection in Men Who Have Sex With Men in San Diego. *J Acquir Immune Defic Syndr.* 2018;79(2):e52-e55.
150. Ong J, Coulthard K, Quinn C, et al. Risk-based screening tools to optimise HIV testing services: a systematic review. 2021.
151. Azur MJ, Stuart EA, Frangakis C, Leaf PJ. Multiple imputation by chained equations: what is it and how does it work? *Int J Methods Psychiatr Res.* 2011;20(1):40-49.
152. Jakobsen JC, Gluud C, Wetterslev J, Winkel P. When and how should multiple imputation be used for handling missing data in randomised clinical trials - a practical guide with flowcharts. *BMC medical research methodology.* 2017;17(1):162.
153. Schafer JL, Graham JW. Missing data: our view of the state of the art. *Psychological methods.* 2002;7(2):147-177.
154. Makowski D, Ben-Shachar MS, Lüdecke D. bayestestR: Describing effects and their uncertainty, existence and significance within the Bayesian framework. *Journal of Open Source Software.* 2019;4(40):1541.
155. O'Hagan A. Bayesian statistics: principles and benefits. *Frontis.* 2004:31-45.

156. McElreath R, O'Reilly for Higher E, Safari aORMC. Statistical Rethinking, 2nd Edition. In: 2nd edition. ed.: Chapman and Hall/CRC; 2020: https://ezproxy.torontopubliclibrary.ca/login?url=https://learning.oreilly.com/library/view/-/9780429639142/?ar.
157. Bartonicek A, Wickham SR, Pat N, Conner TS. The value of Bayesian predictive projection for variable selection: an example of selecting lifestyle predictors of young adult well-being. *BMC Public Health.* 2021;21(1):695.
158. Piironen J, Paasiniemi M, Vehtari A. Projective inference in high-dimensional problems: Prediction and feature selection. *Electronic Journal of Statistics.* 2020;14(1).
159. Heinze G, Dunkler D. Five myths about variable selection. *Transpl Int.* 2017;30(1):6-10.
160. Janssen R, Krumeich J, Esmail A, et al. Moments of Uncertainty: Exploring How an App-Based Oral HIV Self-Testing Strategy Fits in Communities 'Living Under' HIV Risk. *Medicine Anthropology Theory.* 2021;8(2):1-24.
161. Floyd DL, Prentice-Dunn S, Rogers RW. A Meta-Analysis of Research on Protection Motivation Theory. *Journal of Applied Social Psychology.* 2000;30(2):407-429.
162. Milne S, Sheeran P, Orbell S. Prediction and Intervention in Health-Related Behavior: A Meta-Analytic Review of Protection Motivation Theory. *Journal of Applied Social Psychology.* 2000;30(1):106-143.
163. OraSure Technologies Inc. OraQuick In-Home HIV Test, Package Insert. https://www.fda.gov/media/83607/download. Published 2012. Accessed June 7th 2022.
164. *The Mobile Gender Gap Report 2021.* London: GSMA Connected Women;2021.
165. Sun N, Esom K, Dhaliwal M, Amon JJ. Human Rights and Digital Health Technologies. *Health Hum Rights.* 2020;22(2):21-32.
166. Sinha C, Schryer-Roy A-M. Digital health, gender and health equity: invisible imperatives. *Journal of Public Health.* 2018;40(suppl_2):ii1-ii5.
167. Whitehead M. The concepts and principles of equity and health. *Int J Health Serv.* 1992;22(3):429-445.
168. Smith MJ. Health Equity in Public Health: Clarifying our Commitment. *Public Health Ethics.* 2015;8(2):173-184.
169. McIntyre D, Gilson L. Redressing dis-advantage: promoting vertical equity within South Africa. *Health Care Anal.* 2000;8(3):235-258.
170. Bradley SH. The ethics and politics of addressing health inequalities. *Clin Med (Lond).* 2021;21(2):147-149.
171. Bekker LG, Wood R. The changing natural history of tuberculosis and HIV coinfection in an urban area of hyperendemicity. *Clin Infect Dis.* 2010;50 Suppl 3:S208-214.
172. Gibbs A. Tackling gender inequalities and intimate partner violence in the response to HIV: moving towards effective interventions in Southern and Eastern Africa. *Afr J AIDS Res.* 2016;15(2):141-148.

173. Gibbs A, Sikweyiya Y, Jewkes R. 'Men value their dignity': securing respect and identity construction in urban informal settlements in South Africa. *Glob Health Action.* 2014;7:23676.
174. Kamndaya M, Thomas L, Vearey J, Sartorius B, Kazembe L. Material deprivation affects high sexual risk behavior among young people in urban slums, South Africa. *Journal of urban health : bulletin of the New York Academy of Medicine.* 2014;91(3):581-591.
175. Vearey J, Palmary I, Thomas L, Nunez L, Drimie S. Urban health in Johannesburg: the importance of place in understanding intra-urban inequalities in a context of migration and HIV. *Health Place.* 2010;16(4):694-702.
176. Rose G. Sick individuals and sick populations. *INTERNATIONAL JOURNAL OF EPIDEMIOLOGY.* 2001;30(Part 3):427-432.
177. Haukoos JS, Hopkins E, Bucossi MM, et al. Brief report: Validation of a quantitative HIV risk prediction tool using a national HIV testing cohort. *J Acquir Immune Defic Syndr.* 2015;68(5):599-603.
178. Hsieh YH, Haukoos JS, Rothman RE. Validation of an abbreviated version of the Denver HIV risk score for prediction of HIV infection in an urban ED. *Am J Emerg Med.* 2014;32(7):775-779.
179. Damery S, Nichols L, Holder R, et al. Assessing the predictive value of HIV indicator conditions in general practice: a case-control study using the THIN database. *Br J Gen Pract.* 2013;63(611):e370-377.
180. Leal L, Torres B, León A, et al. Predictive Factors for HIV Seroconversion Among Individuals Attending a Specialized Center After an HIV Risk Exposure: A Case-Control Study. *AIDS Res Hum Retroviruses.* 2016;32(10-11):1016-1021.
181. Balkus JE, Brown ER, Palanee-Phillips T, et al. Performance of a Validated Risk Score to Predict HIV-1 Acquisition Among African Women Participating in a Trial of the Dapivirine Vaginal Ring. *J Acquir Immune Defic Syndr.* 2018;77(1):e8-e10.
182. Castor D, Burgess EK, Yende-Zuma N, Heck CJ, Abdool Karim Q. Age-Restriction of a Validated Risk Scoring Tool Better Predicts HIV Acquisition in South African Women: CAPRISA 004. *AIDS and Behavior.* 2022.
183. Giovenco D, Pettifor A, MacPhail C, et al. Assessing risk for HIV infection among adolescent girls in South Africa: an evaluation of the VOICE risk score (HPTN 068). *J Int AIDS Soc.* 2019;22(7):e25359.
184. Bandason T, McHugh G, Dauya E, et al. Validation of a screening tool to identify older children living with HIV in primary care facilities in high HIV prevalence settings. *Aids.* 2016;30(5):779-785.
185. Bandason T, Dauya E, Dakshina S, et al. Screening tool to identify adolescents living with HIV in a community setting in Zimbabwe: A validation study. *PLoS One.* 2018;13(10):e0204891.
186. Moucheraud C, Chasweka D, Nyirenda M, Schooley A, Dovel K, Hoffman RM. Simple Screening Tool to Help Identify High-Risk Children for Targeted HIV Testing in Malawian Inpatient Wards. *J Acquir Immune Defic Syndr.* 2018;79(3):352-357.

187. Beymer MR, Weiss RE, Sugar CA, et al. Are Centers for Disease Control and Prevention Guidelines for Preexposure Prophylaxis Specific Enough? Formulation of a Personalized HIV Risk Score for Pre-Exposure Prophylaxis Initiation. *Sex Transm Dis.* 2017;44(1):48-56.
188. Dijkstra M, Lin TC, de Bree GJ, Hoenigl M, Schim van der Loeff MF. Validation of the San Diego Early Test Score for Early Human Immunodeficiency Virus Infection Among Amsterdam Men Who Have Sex With Men. *Clinical infectious diseases.* 2020;70(10):2228-2230.
189. Tordoff DM, Barbee LA, Khosropour CM, Hughes JP, Golden MR. Derivation and Validation of an HIV Risk Prediction Score Among Gay, Bisexual, and Other Men Who Have Sex With Men to Inform PrEP Initiation in an STD Clinic Setting. *J Acquir Immune Defic Syndr.* 2020;85(3):263-271.
190. Jones J, Hoenigl M, Siegler AJ, Sullivan PS, Little S, Rosenberg E. Assessing the Performance of 3 Human Immunodeficiency Virus Incidence Risk Scores in a Cohort of Black and White Men Who Have Sex With Men in the South. *Sex Transm Dis.* 2017;44(5):297-302.
191. Luo Q, Huang X, Li L, et al. External validation of a prediction tool to estimate the risk of human immunodeficiency virus infection amongst men who have sex with men. *Medicine (Baltimore).* 2019;98(29):e16375.
192. Menza TW, Hughes JP, Celum CL, Golden MR. Prediction of HIV acquisition among men who have sex with men. *Sex Transm Dis.* 2009;36(9):547-555.
193. Smith DK, Pals SL, Herbst JH, Shinde S, Carey JW. Development of a clinical screening index predictive of incident HIV infection among men who have sex with men in the United States. *J Acquir Immune Defic Syndr.* 2012;60(4):421-427.
194. Scott H, Vittinghoff E, Irvin R, et al. Development and Validation of the Personalized Sexual Health Promotion (SexPro) HIV Risk Prediction Model for Men Who Have Sex with Men in the United States. *AIDS Behav.* 2020;24(1):274-283.
195. Wahome E, Thiong'o AN, Mwashigadi G, et al. An Empiric Risk Score to Guide PrEP Targeting Among MSM in Coastal Kenya. *AIDS Behav.* 2018;22(Suppl 1):35-44.
196. Yin L, Zhao Y, Peratikos MB, et al. Risk Prediction Score for HIV Infection: Development and Internal Validation with Cross-Sectional Data from Men Who Have Sex with Men in China. *AIDS Behav.* 2018;22(7):2267-2276.
197. Smith DK, Pan Y, Rose CE, et al. A Brief Screening Tool to Assess the Risk of Contracting HIV Infection Among Active Injection Drug Users. *J Addict Med.* 2015;9(3):226-232.
198. Kahle EM, Hughes JP, Lingappa JR, et al. An empiric risk scoring tool for identifying high-risk heterosexual HIV-1-serodiscordant couples for targeted HIV-1 prevention. *J Acquir Immune Defic Syndr.* 2013;62(3):339-347.
199. Willcox AC, Richardson BA, Shafi J, et al. Derivation of an HIV Risk Score for African Women Who Engage in Sex Work. *AIDS Behav.* 2021;25(10):3292-3302.

Appendix

Table 2. Existing risk assessment tools for HIV (N = 41)

Development				Validation		
Author (Year)* Name of the tool (if any)	Sample size Country Population	Included predictors I[a] C[b] B[c] S[d] Included social predictors and Other predictors	Dealing with missing data Development method	Validation type Author (Year) Internal Cross-validated or Bootstrap Training/ testing split External	Sample size Country Population (if different from derivation cohort)	AUC-ROC Sensitivity (Sens) Specificity (Spec) (95% confidence intervals)
Non-targeted risk staging tools						
Haukoos (2012)[139] **Denver Risk Score**	92,635 USA General adult population	I C B S Race/ ethnicity Testing history for HIV	Complete case analysis Multiple imputation Variable selection based on knowledge of HIV epidemiology and statistical association	External Haukoos (2015)[177] External Falasinnu (2015)[147] Internal	4,830,941 USA 47,175 Canada 92,635	0.77 0.80 (0.79-0.81) Sens: 96% 0.75 (0.70-0.78)

[a] Shaded indicates Individual predictors (age, sex, marital status, etc.) were included in the risk assessment tool
[b] Shaded indicates Clinical predictors (signs and symptoms, comorbidities, presence of other STI) were included in the risk assessment tool
[c] Shaded indicates Behavioural predictors (sexual or related to drug use) were included in the risk assessment tool
[d] Shaded indicates Social predictors (education/ income level, employment status, living situation) were included in the risk assessment tool
*Asterisk indicates that risk assessment tool was developed to predict acute HIV infection

			Risk scores derived from beta coefficients of odds ratios	Cross-validation	USA	
				External	15,184	0.70 (0.65-0.75)
				Abbreviated Denver Risk Score Hseieh (2014)[178]	USA	
Kabapy (2021)[140]	600 Egypt General adult population	I C B S Place of residence Literacy Employment Homelessness Incarceration Sexual abuse in childhood Testing history for HIV	Variable selection based on knowledge of HIV epidemiology and statistical association Risk scores derived from beta coefficients of odds ratios	Internal	600 Egypt	0.998–1.000
Kagaayi (2014)[57]	13,280 Uganda General adult population	I C B S Education HIV prevalence	Multiple imputation Backward model selection with AIC as stopping rule	Internal Bootstrap resampling	13,280 Uganda	Men 0.69 (0.66-0.73) Women 0.67 (0.64-0.70)
Muttai (2021)[143]	19,458 Kenya	I C B S	Complete case analysis	Internal Cross-validation	19,458 Kenya	0.69 (0.60-0.77)

	General adult population	Occupation Testing history for HIV	Initial variable selection based on knowledge of HIV epidemiology Stepwise model selection with AIC as stopping rule			
Roberts (2022)[138]	9,623 South Africa (rural) General adult population	I C B S Education Employment Socioeconomic status quintile HIV prevalence Population prevalence of detectable viremia Place of residence	Multiple imputation Cox proportional hazards with LASSO penalties Model selection using 10-fold cross validated mean AUC	External	9,933	Men 0.71 Sens: 77% Women 0.68 Sens: 65%
Damery (2013)[179]	3,515 UK Patients	I C B S	Complete case analysis Backward stepwise model selection with a p-value < 0.2 threshold	Internal	3,515 UK	0.66 Sens: 42% Spec: 80%
Feller (2018)[146]	724 USA	I C B S	Natural language processing to decode medical notes	Internal 10-fold cross-validation	724 USA	F1: 73.3 (72.1-74.5)

	Patients	Testing history for HIV	Developed candidate models using mutual information criteria random forest classifiers			
Krakower (2019)[141]	1,155,966 USA Patients	I C B S Race Testing history for HIV	Developed candidate models using various machine learning techniques Validated using cross-validated AUC Selected models developed through LASSO	External	33,404 USA	0.77 (0.74-0.79)
				Internal Prospective validation	537,257 USA	0.91 (0.81–1.00)
				Internal Cross-validation	1,155,966 USA	0.86 (0.82-0.90)
Leal (2016)[180]	138 Spain Patients	I C B S	Initial variable selection based on clinical significance Backward stepwise model selection	Internal	138 Spain	0.78 (0.69-0.84) Sens: 64% Spec: 74%
Marcus (2019)[142] **Marcus tool**	3,750,664 USA Patients MSM	I C B S Neighbourhood deprivation index Place of residence (urban)	Coded as 0 Model selection using LASSO	External	606,701 USA	0.84 (0.80-0.89) Sens: 59.10% Spec: 97.8%
Facente (2011)[132]*	12,622 USA	I C B S	Complete case analysis	Internal	12,622 USA	0.669 Sens: 83.3%

	STI clinic attendees		Initial variable selection based on statistical significance (p-value < 0.2) Backward model selection			(67.2–93.6) Spec: 50.4% (49.5–51.3)
Powers (2007)[134]* **UMRSS**	1,448 Malawi STI clinic attendees	I C B S	Complete case analysis Backward stepwise model selection with a p-value < 0.15 threshold	External Wahome (2013)[136] Internal	6,531 visits Kenya 1,448 Malawi	0.79 Sens: 75.3% Spec: 76.4% 0.89
Sander (2015)[135]*	19,971 visits Kenya Malawi South Africa STI clinic attendees MSM Sex workers Women	I C B S	Initial variable selection based on statistical association in univariate models (p-value < 0.15) Model selection based on statistical association (p-value < 0.05)	Internal Internal Internal Internal	19,971 visits Kenya Malawi South Africa 7,735 visits Kenya Female sex workers 6,531 visits Kenya MSM 860 visits	0.78 0.77 Sens: 58.4% Spec: 84.8% 0.89 Sens: 90.0% Spec: 74.1% 0.83

					Malawi STI clinic attendees	Sens: 92.9% Spec: 50.5%
				Internal	4,845 visits South Africa Women	0.61 Sens: 18.2% Spec: 95.4%
Targeted risk staging tools						
Adult women						
Balkus (2016)[125] **VOICE risk score**	4,834 South Africa Uganda Zimbabwe United States Kenya Tanzania Women	I C B S Earns her own income Financial/ material support from partner	Complete case analysis Initial variable selection based on statistical association in univariate models (p-value < 0.05) Stepwise model selection based on AIC	External Rosenberg (2020)[130]	795 Malawi	0.64 (0.52–0.75)
				External Burgress (2017)	1,115 South Africa	0.56 (0.50-0.62) Sens: 96% Spec: 7%
				External Balkus (2018)[181]	2,539 Malawi South Africa	0.69 (0.64-0.74) Sens: 91% Spec: 36%
				External Peebles (2020)[126]	5,573 South Africa	0.61 (0.58-0.65)
				External	444	0.66 (0.54-0.74)

				Castor (2022)[182]	South Africa	Sens: 63% Spec: 57%
				Internal Cross-validation	4,834 South Africa Uganda Zimbabwe United States Kenya Tanzania	0.70 (0.65-0.75) Sens: 58% Spec: 71%
				External Giovenco (2019)[183]	2,178 South Africa AGYW	0.55 (0.44-0.65) Sens: 85% Spec: 6%
Peebles (2020)[126]	5,573 South Africa Women	I C B S Province of residence HIV prevalence	Complete case analysis Stepwise model selection based on AIC	Internal	5,573 South Africa	18-24 years old: 0.64 (0.60-0.67) 25-35 years old: 0.68 (0.62-0.73)
Wand (2012)[128]	1,010 South Africa Women	I C B S	Backward model selection Risk scores derived from beta coefficients	Internal 2:1 training/ testing split	475 South Africa	0.79 (0.70-0.81) Sens: 90% Spec: 36%
Wand (2018)[55]	6,018 South Africa Women	I C B S	Backward model selection with a p-value < 0.05 threshold	Internal 2:1 training/ testing split	2,964 South Africa	0.71

			Risk scores derived from beta coefficients of hazard ratios			
Pintye (2017)[127]	1,304 Kenya Pregnant and Postpartum Women	I C B S	Stepwise model selection based on AIC	Internal Cross-validation	1,304 Kenya	0.73 (0.57-0.90)
Adolescent girls and young women (AGYW)						
Ayton (2020)[129] **Ayton tool**	1,049 South Africa AGYW	I C B S Testing history for HIV HIV knowledge Financial dependence School absence	Variable selection using exploratory factor and latent class analysis Model selection based on BIC	Internal	1,049 South Africa	Sens: 60% (32-84) Spec: 58% (55-61)
Rosenberg (2020)[130]	795 Malawi AGYW	I C B S	Backward stepwise model selection with a p-value < 0.15 threshold Risk scores derived from beta coefficients	Internal	795 Malawi	0.79 (0.69-0.89)
Paediatric populations (Adolescents, Children and Infants)						
Antelman (2021)[121]	21,008	I C B S	Variable selection based on knowledge	Internal	21,008	Sens: 85.3% (74.6-92.7)

	Tanzania Adolescent and children	Whether biological parents died	of HIV epidemiology, statistical association (p-value < 0.15) and AIC/BIC		Tanzania	Spec: 44.2% (43.5-44.9)
Ferrand (2011)[123] **Bandason tool**	254 Zimbabwe Adolescent and children	I C B S Whether biological parents died	Initial variable selection based on statistical association (p-value < 0.1) Backward stepwise model selection with a p-value < 0.05 threshold	External Bandason (2016)[184]	9,568 Zimbabwe	0.73 (0.72-0.75) Sens: 80.4% (76.5-84.0) Spec: 66.3% (65.3-67.2)
				External Bandason (2018)[185]	5,384 Zimbabwe	0.65 (0.60-0.72) Sens: 56.3% (44.0-68.1) Spec: 75.1% (73.9-76.3)
				External Modified Bandason tool Moucheraud (2018)[186]	8,602 Malawi Children and infants	0.7116 Sens: 84.4% (77.0-91.9) Spec: 39.6% (38.5-40.6)
				Internal 1:1 training/ testing split	254 Zimbabwe	Sens: 74.0% (64-82) Spec: 80% (71-87)
Allison (2011)[120]	487	I C B S	Forward stepwise model selection with	Internal	487	Sens: 96.30% Spec: 25%

	Papua New Guinea Infants		a p-value < 0.05 threshold		Papua New Guinea	
Du Plessis (2019)[122]	1,759 South Africa Infants	I C B S	Variable selection based on statistical association and BIC	Internal	1,759 South Africa	Sens: 80%
Mafaune (2020)[124]	No development dataset Zimbabwe Infants	I C B S	Modified version of the WHO recommendation on HIV birth testing	External	1,970 Zimbabwe	Sens: 62.1% (44.4-79.7) Spec: 87.2% (85.7-88.7)
Men who have sex with men (MSM)						
Beymer (2017)[187]	9,481 USA MSM	I C B S	Complete case analysis Variable selection based on knowledge of HIV epidemiology Risk scores derived from beta coefficients of hazard ratios	Internal	9,481 USA	Sens: 74.60% Spec: 50.20%
Dijkstra (2017)[131]*	1,562 Netherlands	I C B S	Complete case analysis	External	3,751 USA	0.78 (0.74-0.82) Sens: 56.2% (47.5-64.7)

Amsterdam Score	MSM		Variable selection based on knowledge of HIV epidemiology and statistical association			Spec: 88.8% (88.4-89.1)
				External Lin (2018)[149]	757 USA	0.88 (0.84-0.91) Sens: 78.2% Spec: 81.0%
Hoenigl (2015)[137]*	5,568 USA	I C B S	Complete case analysis	External Dijkstra (2020)[188]	1,071 Netherlands	0.701 (0.639-0.762) Sens: 54.0% Spec: 77.9%
San Diego Early Test Score (SDET)	MSM		Forward stepwise model selection with p-value < 0.05 threshold	Internal 2:1 training/ testing split	2,758 USA	0.703 (0.625-0.781) Sens: 60% Spec: 77%
				External Tordoff (2020)[189]	13,527 USA	0.63 (0.59-0.67)
				External Jones (2017)[190]	562 USA	0.55 (0.44-0.66) Sens: 25% (11.5-43.4) Spec: 83.9% (80.5-87.0)
Li (2017)[144] **HIV RISK Assessment Tool**	No development dataset China	I C B S	Variable selection based on literature review and specialist consultations (Delphi method)	External Luo (2019)[191]	1,442 China	0.63 (0.60-0.67)
				External	1330	0.827

	MSM			Zheng (2020)[148]	Southwest China	Sens: 78.5% Spec: 74.7%
Lin (2018)[133]*	673	I C B S	Variable selection based on statistical association (p-value < 0.2)	Internal 2:1 training/ testing split	325	0.851 (0.780-0.922)
	USA				USA	
San Diego Symptom Score (SDSS)	MSM		Risk scores derived from beta coefficients of odds ratios			Sens: 72% Spec: 96%
Menza (2009)[192]	1,903	I C B S	Variable selection based on knowledge of HIV epidemiology and statistical association	External	2,081	0.67 (0.60-0.75)
	USA				USA	
Menza score	MSM	Race		External	562	0.51 (0.41-0.60)
			Risk scores derived from beta coefficients of hazard ratios	Jones (2017)[190]	USA	Sens: 62.5% (43.7-78.9) Spec: 41.1% (36.9-45.5)
				External	13,527	0.66 (0.62-0.70)
				Tordoff (2020)[189]	USA	
Smith (2012)[193]	4,386	I C B S	Backward model selection with p-value < 0.05 threshold	External	3,368	0.721
	USA				USA	Sens: 84% Spec: 45%
HIRI-MSM	MSM		Risk scores derived from beta coefficients	External	13,527	0.61 (0.57-0.65)
					USA	

				Tordoff (2020)[189]		
				External Jones (2017)[190]	562 USA	0.62 (0.52-0.72) Sens: 62.5% (43.7-78.9) Spec: 56.7% (52.4-61.0)
				External (VAX0004 cohort) Scott (2020)[194]	4,643 USA	0.740
				External (HPTN cohort) Scott (2020)[194]	1,553 USA	0.751
				External (HVTN cohort) Scott (2020)[194]	2,504 USA	0.735
Tordoff (2020)[189] **Seattle PrEP Score**	13,527 37,814 visits USA	I C B S	Stepwise model selection with AIC as a stopping rule Risk scores derived from beta	Internal 2:1 training/ testing split	9,234 18,908 visits	0.60 (0.54-0.66) Sens: 46.3% Spec: 69%

	MSM		coefficients of hazard ratios			
Wahome (2013)[136]* **CDRSS**	6,531 visits Kenya MSM	I C B S	Complete case analysis Variable selection based on statistical association (p-value < 0.05) Risk scores derived from beta coefficients of odds ratios	Internal	6,531 visits Kenya	0.85 Sens: 80.8% Spec: 76.0%
Wahome (2018)[195]	753 Kenya MSM	I C B S	Variable selection based on statistical association (p-value < 0.1) Risk scores derived from beta coefficients of odds ratios	Internal	753 Kenya	0.76 (0.71-0.8) Sens: 97.90% Spec: 16.90%
Yin (2018)[196]	3,588 China MSM	I C B S Being a resident of Beijing Years living in Beijing	Complete case analysis Initial variable selection using shrinkage of estimated regression coefficients and penalized maximum	Internal Bootstrap	3,588 China	0.71

			likelihood estimation Backward model selection with p-value < 0.25 threshold			
Yun (2019)[118]	667 China MSM	I C B S	Multiple imputation Initial variable selection based on statistical association in univariate model (p-value < 0.25) Backward model selection based on AIC Variables with p-value < 0.05 retained	Internal 2:1 training/ testing split	332 China	0.60 (0.45-0.74)
Scott (2020)[194]	4,295 USA MSM (inclusive of black MSM)	I C B S Black race/ Latino ethnicity	Initial variable selection based on knowledge of HIV epidemiology Model validation using cross-validated AUC and measures of goodness of fit	External (VAX0004 cohort)	4,643 USA	0.73 Sens: 64.4% Spec: 67.4%
				External (HPTN cohort)	1,553 USA	0.71 Sens: 100% Spec: 0%
				External (HVTN cohort)	2504 USA	0.72 Sens: 75.4% Spec: 51.8%

Other populations						
Smith (2015)[197] **ARCH-IDU**	1,904 USA People who inject drugs	I C B S	Complete case analysis Initial variable selection based on statistical association in univariate models (p-value < 0.20) Backward model selection with a p-value < 0.05 threshold	Internal	1,904 USA	0.72 Sens: 86.20% Spec: 42.50%
Khale (2013)[198]	3,297 couples Botswana Kenya Rwanda South Africa Tanzania Uganda Zambia Serodicordant couples	I C B S	Complete case analysis Stepwise model selection based on AIC	External (Partners in PrEP cohort) External (Couples Observational Study cohort)	1,499 couples Kenya Uganda 476 couples Kampala Uganda South Africa	0.70 (0.64-0.76) Sens: 55% 0.76 (0.70-0.83) Sens: 80%
Wilcox (2021)[199]	660 Kenya Sex workers	I C B S	Complete case analysis Initial variable selection based on	Internal	660 Kenya	0.67 Sens: 53% Spec: 76%

	statistical association in univariate models (p-value < 0.15) Forward model selection	

[a] Shaded indicates Individual predictors (age, sex, marital status, etc.) were included in the risk assessment tool
[b] Shaded indicates Clinical predictors (signs and symptoms, comorbidities, presence of other STI) were included in the risk assessment tool
[c] Shaded indicates Behavioural predictors (sexual or related to drug use) were included in the risk assessment tool
[d] Shaded indicates Social predictors (education/ income level, employment status, living situation) were included in the risk assessment tool
*Asterisk indicates that risk assessment tool was developed to predict acute HIV infection