medRxiv preprint doi: https://doi.org/10.1101/2021.09.28.21264246; this version posted September 29, 2021. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY 4.0 International license .

Risk scores for predicting HIV incidence among the general population in sub-Saharan Africa: a systematic review and meta-analysis

Katherine M. Jia[1], Hallie Eilerts[2], Olanrewaju Edun[1], Kevin Lam[1], Adam Howes[3], Matthew L. Thomas[4], Jeffrey W. Eaton[1§]

[1] MRC Centre for Global Infectious Disease Analysis, School of Public Health, Imperial College London, London, United Kingdom

[2] Department of Population Health, The London School of Hygiene and Tropical Medicine, London, United Kingdom

[3] Department of Mathematics, Imperial College London, London, United Kingdom

[4] Joint Centre for Excellence in Environmental Intelligence, University of Exeter & Met Office, Exeter, United Kingdom

[§]Corresponding author: Jeffrey W. Eaton

MRC Centre for Global Infectious Disease Analysis, School of Public Health
Imperial College London
St. Mary's Hospital Campus
Norfolk Place
London W2 1PG
United Kingdom
+44 20 7594 3007
jeffrey.eaton@imperial.ac.uk

E-mail addresses of authors:

KMJ: katherine.jia19@imperial.ac.uk

HE: hallie.eilerts@lshtm.ac.uk

OE: o.edun19@imperial.ac.uk

KL: kevin.lam@imperial.ac.uk

AH: a.howes19@imperial.ac.uk

MLT: M.L.Thomas@exeter.ac.uk

JWE: jeffrey.eaton@imperial.ac.uk

Keywords: risk scores, HIV incidence, sub-Saharan Africa, adolescent girls and young women, risk factors for HIV incidence

Word count: Abstract: 346/350 words; Main text: 4424/5000 words

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/2021.09.28.21264246; this version posted September 29, 2021. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY 4.0 International license .

Abstract

Introduction: Several HIV 'risk scores' have been developed to identify individuals for prioritised HIV prevention in sub-Saharan Africa. We systematically reviewed HIV risk scores to: (i) identify factors that consistently predicted incident HIV infection, (ii) review inclusion of community-level HIV risk in predictive models, and (iii) examine predictive performance.

Methods: We systematically searched nine databases for studies developing and/or validating HIV risk scores among the general population in sub-Saharan Africa from database inception until February 15, 2021. Studies not prospectively observing seroconversion or recruiting only key populations were excluded. Record screening, data extraction, and critical appraisal were conducted in duplicate. We used random-effect meta-analysis to summarise hazard ratios and the area under the receiver-operating characteristic curve (AUC-ROC).

Results: From 1563 initial search records, we identified 14 risk scores in 13 studies. Seven studies were among sexually active women using contraception enrolled in randomised-controlled trials, three among adolescent girls and young women (AGYW), and three among cohorts enrolling both men and women. Consistently identified HIV prognostic factors among women were younger age (pooled adjusted hazard ratio: 1.62 [95% Confidence Interval: 1.17, 2.23], compared to above-25), single/not cohabiting with primary partners (2.33 [1.73, 3.13]) and having sexually transmitted infections (STIs) at baseline (HSV-2: 1.67 [1.34, 2.09]; curable STIs: 1.45 [1.17; 1.79]). Among AGYW only STIs were consistently associated with higher incidence, but studies were limited (n=3). Community-level HIV prevalence or unsuppressed viral load strongly predicted incidence but were only considered in three of 11 multi-site studies. The AUC-ROC ranged from 0.56 to 0.79 on the model development sets. Only the VOICE score was externally validated by multiple studies, with pooled AUC-ROC 0.626 [0.588, 0.663] (I^2: 64.02%).

medRxiv preprint doi: https://doi.org/10.1101/2021.09.28.21264246; this version posted September 29, 2021. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY 4.0 International license .

Conclusions: Younger age, non-cohabiting, and recent STIs were consistently identified as predicting future HIV infection. Both community HIV burden and individual factors should be considered to quantify HIV risk. However, HIV risk scores had only low-to-moderate discriminatory ability and uncertain generalizability outside of the study populations. Further evidence on the relative value of specific factors and data outside high-risk populations will help inform optimal implementation of risk scoring algorithms in HIV programmes.

PROSPERO Number: CRD42021236367

medRxiv preprint doi: https://doi.org/10.1101/2021.09.28.21264246; this version posted September 29, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY 4.0 International license .

Introduction

Efficiently identifying populations and individuals at highest risk of HIV infection and linking them to HIV prevention is essential for continued progress towards ending HIV as a public health threat [1]. This is especially the case for expensive and intensive interventions, such as pre-exposure prophylaxis (PrEP) [2, 3], and in sub-Saharan where 59% of 1.7 million global new infections in 2019 occurred and a large proportion of HIV infections occur among persons who do not identify as key populations among whom HIV is concentrated in other global regions [4].

Efficiently identifying those at high risk of acquiring HIV from the large population is essential for efficiently and effectively delivering HIV prevention. Several HIV incidence 'risk scores' have been proposed as prognostic tools for this purpose [5, 6]. HIV risk scores combine data on multiple prognostic factors into a single score, which is often compared against a threshold to classify individuals among whom specific HIV prevention services are offered or promoted. An optimal threshold maximises the share of incident infections among the higher risk group while minimizing the total proportion classified as such, but there is typically a trade-off between these. Risk scores are empirically derived using data from large-scale, longitudinal studies like HIV randomised controlled trials (RCTs) and cohort studies that collect comprehensive HIV prognostic factors spanning the behavioural, sociodemographic, partnership domains among HIV negative adults and prospectively measure HIV incidence, usually within 1 year or less after the baseline risk assessment. Generalisability is validated by applying the risk score to independently collected data and studying how well the score discriminates those who subsequently acquire HIV.

Recently, national HIV programmes have focused on prioritising interventions to geographic areas with high HIV burden [1]. This strategy is epidemiologically justified for two reasons: high HIV burden indicates previous high HIV risk, and, secondly, high HIV prevalence or unsuppressed HIV viraemia implies greater exposure to HIV infection among those currently at risk [7, 8]. This community-level determinant of exposure does not fit naturally into the

medRxiv preprint doi: https://doi.org/10.1101/2021.09.28.21264246; this version posted September 29, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY 4.0 International license .

individual-level risk framework of risk scoring. However, solely geographic prioritisation of HIV prevention is inefficient by not considering individual-level variation measured by risk scores.

Mathematical modelling has demonstrated that considering both geographic location and risk populations in prioritising of HIV prevention improves the efficiency and cost-effectiveness relative to only one dimension [9]. The new Global AIDS Strategy 2021-2026 embraces this approach—recommending that HIV prevention is prioritised for various population groups differentiated according to thresholds for the local HIV incidence [10]. For example, for adolescent girls and young women (AGYW), the strategy recommends prioritisation of services to those at high risk based on: (i) the subnational annual incidence greater than 3%, or (ii) an incidence of 1-3% and self-reported high-risk behaviours or recent sexually transmitted infection (STI) [10].

We conducted a systematic review of HIV risk score tools in sub-Saharan Africa to explore this from both perspectives. Firstly, to motivate improved modelling of HIV incidence and prioritising of HIV prevention, we sought to identify prognostic factors from the HIV risk score literature that stratify population HIV risk, and the ability of these factors to discriminate HIV incidence within a population. Secondly, we queried the extent to which HIV risk scores considered community-level HIV prevalence or population viraemia as a predictor in prognostic models for individual HIV incidence risk. Specifically, we searched literature for studies that either developed or validated a HIV incidence risk score model among the general population, and analysed the data to: (i) identify risk factors that have consistently shown strong effects on HIV incidence across different models and settings, (ii) evaluate whether community-level HIV prevalence has been considered as a determinant of HIV risk in risk score development, and (iii) examine the efficiency of risk scores in differentiating high- and low-risk individuals quantified by the area under the receiver-operating characteristic curve (AUC-ROC).

medRxiv preprint doi: https://doi.org/10.1101/2021.09.28.21264246; this version posted September 29, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY 4.0 International license .

Methods

***Search strategy*.** We searched for studies that developed and/or validated the HIV incidence risk scores among the general population of sub-Saharan Africa. Specific inclusion criteria were: (i) development and/or validation of any predictive multivariable model ("risk score") with prospectively measured HIV incidence as the main outcome (i.e., documented HIV-negative status at baseline), (ii) enrolled from the general population and (iii) conducted in sub-Saharan African countries. Studies were excluded if: (i) HIV seroconversions were not determined by a HIV negative test result at baseline followed by a positive or negative result during follow-up, (ii) study populations were key or selected populations only (men who have sex with men, female sex workers, pregnant women, serodiscordant couples, HIV-exposed infants, people who inject drugs).

Keywords, synonyms, and related terms covered the domains of "sub-Saharan Africa", "HIV/AIDS", and "risk score". The full electronic search strings for all databases are available in the Supplementary Material (Appendix I). No restrictions were imposed on the types nor years of publications; however, only publications written in English were included.

Sources of information. Nine databases were searched: MEDLINE, Embase, Global Health, PsycINFO, Maternity & Infant Care Database, CINAHL (EBSCO), Scopus, Cochrane Library, the Web of Science, on 15th February 2021.

Study selection. Titles and abstracts were independently screened by two reviewers for eligibility against the inclusion and exclusion criteria. Discrepancies were resolved by either consensus after discussion or decision of a third reviewer. After abstract screening, study full texts were reviewed for inclusion by two independent reviewers. Reasons were provided for any exclusion of studies at this stage. Again, any discrepancies in decisions or reasons were resolved through discussion or by a third reviewer.

***Data extraction and risk of bias assessment*.** Data were extracted by two independent reviewers, with discrepancies resolved through discussion. We referred to the Critical

medRxiv preprint doi: https://doi.org/10.1101/2021.09.28.21264246; this version posted September 29, 2021. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY 4.0 International license .

Appraisal and Data Extraction for Systematic Reviews of Prediction Modelling Studies (CHARMS) Checklist when creating the data extraction form (Appendix II).[11] After extraction, two reviewers assessed the risks of bias for each study independently using the Prediction Model Risk of Bias (PROBAST) assessment tool checklist,[12] under the four domains 'Population', 'Predictor', 'Outcome' and 'Analysis.' A domain where one or more criteria was/were not fulfilled might be judged as "high risk of bias," whereas a study with one (or more) domain(s) at "high risk of bias" would be judged as having an overall "high" risk of bias.

Data synthesis and reporting. Our aim was to identify significant and measurable prognostic factors that define high risk groups or individuals for prioritised HIV prevention. We first summarised the key characteristics, setting, and study population(s) of each included study, and whether it developed a risk score, externally validated a score, or both. A development study could conduct internal validation by using re-sampling methods (bootstrap or cross-validation) to estimate the AUC-ROC, or by splitting the sample into training and testing sets; external validation where the risk score was applied to a different study population than which it was originally derived can be performed in the same analysis or by others in follow-up studies. We then assessed the importance of each predictor by examining (i) the number of times it was included in the final risk prediction model of a model development study, (ii) the summary of the adjusted and unadjusted effect size estimates. Finally, we summarised the area under the receiver operating characteristic curve (AUC-ROC), proportion identified "high risk" by each score, and the corresponding HIV incidence in the high-risk group, to assess the risk scores discrimination and compared them across settings to examine generalisability.

Overall summary effect size estimates (and the 95% confidence interval) for predictors were estimated by a random effects model. Estimates were pooled for both the adjusted and unadjusted effects because adjusted effects were only available in studies that included the particular predictors in the multivariable models (due to significant univariate association), risking biasing summary estimates away from null. Between-study variance were reported

medRxiv preprint doi: https://doi.org/10.1101/2021.09.28.21264246; this version posted September 29, 2021. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY 4.0 International license .

with the I^2 statistics to evaluate the heterogeneity. Random effects meta-analysis based on the inverse variance method with Sidik-Jonkman estimator for between-study variance was done in R (version 4.0.3) [13] using the packages “meta” and “metafor” [14, 15]. Meta-analysis for AUC-ROC was performed using methods described by Zhou and colleagues in Medcalc (version 19.8) [16, 17]. Forests plots and funnel plots were created using the package “meta” and Medcalc respectively.

Protocol of this systematic review is available on PROSPERO (CRD42021236367) [18]. We referred to the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) checklist for presenting the review [19].

medRxiv preprint doi: https://doi.org/10.1101/2021.09.28.21264246; this version posted September 29, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY 4.0 International license .

Results

Database searches identified 2029 records; 466 duplicates were removed and 1563 titles and/or abstracts were screened, of which 25 studies were retained for full-text screening. One additional conference abstract was available after initial screening, adding to a total of 13 risk score literature (9 peer-reviewed articles, 2 posters, 1 editorial letter, 1 abstract) that met the inclusion criteria and were included in this review (Figure 1) [5, 6, 20-30]. Critical appraisal according to the PROBAST checklist concluded that one out of twelve models developed and two out of nine validated were of low risk of bias (Figure S1, Table S3 & S4). Inadequate adjustment for over-fitting or model optimism was common among the development studies (six out of nine). For the validation studies, inadequate sample size (four out of nine) and missing predictors (five of out nine) were common limitations. Studies also reported a high proportion of missing cases and/or loss-of-follow-up, but did not often

Study populations. Studies were conducted in South Africa (n=10), Uganda (n=4), Malawi, Zimbabwe (n=3), Kenya (n=2), Zambia (n=1), and Tanzania (n=1). Three enrolled multi-country study populations, and eight were multi-site within one country (Table 1). Incidence and risk factor data were collected before 2012 for seven studies and after 2012 for six. The majority (10 of 13) were among women only, of which three were restricted to young women under age 25 or lower; 3 included women and men aged 15–49 years or 15 years and older. The majority (10 of 13) were RCTs or quasi-experimental studies that restricted recruitment and/or eligibility to specific at-risk population groups: (i) sexually active, contraception-seeking women who attended the family-planning, STIs or research clinics (7 of 10, all RCTs) [5, 6, 20-24], or (ii) school-attending AGYW (3 of 10) [25-27]. The remaining three were large-scale cohort studies or community trials that recruited all consenting members within the communities [28-30]. Geographic locations, study periods, age groups, and settings are given in Table 1.

Factors included in HIV risk scores. Nine studies reported on development of 14 HIV risk scores, involving screening and model selection for predictors of HIV incidence (Table 1).

medRxiv preprint doi: https://doi.org/10.1101/2021.09.28.21264246; this version posted September 29, 2021. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY 4.0 International license .

Balzer used a machine learning approach which did not yield effect estimates for individual risk factors [29], and final regression results were not available for Roberts (abstract only) [30]. For the remaining studies, Table 2 reports the predictors considered for inclusion and retained in the final model for each independently developed score. The '*retainment ratio*' reports the number of times a risk factor was retained in the final score relative to the number of times it was considered as a '*candidate*' predictor, tabulated separately for risk scores for women of all ages and for AGYW only study populations.

Among four risk scores developed for women of all ages recruited from RCTs who were sexually active and using contraception, factors retained in all or three of four final models were: not being married or cohabiting with primary partner (pooled adjusted hazard-ratio [aHR] 2.33; 95% CI [1.73, 3.13]; Table S5); younger age (pooled aHR: 1.62 [1.17, 2.23]; less than 25 years old except for Peebles [6] at 27 years), and curable STIs at baseline (pooled aHR 1.45 [1.17, 1.79]) (Figure 2). Human Simplex Virus – 2 (HSV-2; pooled aHR 1.67 [1.34, 2.09]) and multiple sexual partners (Pooled aHR: 1.62 [1.27, 2.07]) were included in two of four risk scores. Other demographic, partnership, biological or community factors were either seldomly considered as candidate predictors or only retained in one or fewer risk scores (Table 2). Among unselected candidate predictors, educational attainment, employment (or earning own income) and coital frequency were considered by all four studies but not retained in any of the final models (Table S1, Supplementary Materials).

Three risk scores were developed specifically for sexually active AGYW (aged 13-24, varying across studies) (Table 2). HSV-2 was the only factor selected in all three (pooled aHR: 1.77 [1.24; 2.54]). Factors selected in two of three models were curable STIs (pooled aHR: 2.14 [1.40; 3.25]), having multiple partners (pooled aHR: 1.76 [1.19; 2.60]) and partner having other sexual partners (pooled aHR: 2.35 [0.48; 11.53]) (Figure 3). Being not married/cohabitating was not selected in any final models, unlike the models for all age women where it was selected by all models.

medRxiv preprint doi: https://doi.org/10.1101/2021.09.28.21264246; this version posted September 29, 2021. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY 4.0 International license .

In summary, marital/cohabiting status had been consistently identified and had the largest effect size estimates in studies among women in general, while for AGYW STIs had showed greatest consistency and effect sizes. Of the remaining predictors, occupation, self-perceived HIV risk, partners' occupation, having new partners, engaging in high-risk sex (e.g., under alcohol use), knowledge of partner's HIV status showed significant associations but were seldomly assessed [28].

Inclusion of community HIV prevalence. Only three of 11 multi-site studies considered community-level HIV prevalence as a covariate and all cases where considered it was selected into one or more of the final models [6, 28, 30]. In Peebles [6], compared to residing in a community with 10-15% HIV prevalence, those in a community with 16-20% prevalence had an aHR of 1.64 [1.08, 2.48], 1.71 [0.99, 2.96] for 21-25% prevalence, and 1.81 [1.03, 3.19] for 26-30%. Similarly, in Kagaayi [28], an aHR of 1.03 was associated with each percentage-point increment in community prevalence for both male [0.99,1.07] and female [1.01, 1.06]. Roberts [30] also found community HIV prevalence and unsuppressed viral load to be highly predictive, but aHRs were not available.

Predictive performance of the risk scores. We identified 14 risk scores from nine model development studies (three models developed by Balzer were considered separately) [29] (Table 3). When applied to the original data set from which it was developed, the scores had low-to-moderate AUC-ROC ranging from 0.56 to 0.79. Only the VOICE score has been externally validated in other settings. In seven validation studies with AUC-ROC estimates, the accuracy was lower (pooled AUC-ROC: 0.626 [0.588, 0.663]; I^2: 64.02%) than in the internal validation (AUC-ROC: 0.69 [0.66, 0.72]) (Table 3, Figure 4). In addition to being among different study populations, it was common for one or two predictors to be missing in external validation sets (Table 3), which may have also contributed to decreased accuracy. Regarding validation of other scores, Roberts developed their risk scores in a large-scale cohort and validated them using data collected in a subsequent time period, also showing

medRxiv preprint doi: https://doi.org/10.1101/2021.09.28.21264246; this version posted September 29, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY 4.0 International license .

moderate discriminatory power (AUC-ROC 0.68 among female and 0.72 among male; Table 3).

Several studies compared the discriminative power of combining multivariate risk scores versus single risk factors. Balkus [5] reported that marital status/cohabitation with primary partner alone yielded an AUC-ROC of 0.62 versus 0.69 for the full score, followed by age (0.60) and curable STIs (0.57). In a similar analysis, Peebles[6] found the most important predictors were age (less than 27), being married/cohabiting and the provinces of residence. Three studies [5, 6, 24] additionally provided a “modified score” that excluded the laboratory-diagnosed STIs, which are not routinely available in most settings. Removing laboratory-diagnosed STIs reduced the AUC-ROC by between one to eight percentage-points (Table 3). Roberts found that including only age, HIV prevalence, and viraemia as predictors produced an AUC-ROC of 0.65 for women compared to 0.68 when all risk factors were considered, and 0.71 for men compared to 0.72 when all risk factors were considered (Table 3).

Incidence among risk group categories. Most studies found that HIV incidence increased monotonically with the risk scores, except for Giovenco [25] (Table S6). Figure 5 shows the proportion of participants identified as high risk compared to the percentage of incident cases contributed by the high-risk group. In six of nine external validation sets of the VOICE score with such information available (Table S6), women with a VOICE score of 5 or above (having around 3 to 4 of the 7 risk factors) had incidence above 3%, the WHO-recommended threshold for PrEP prioritisation. Among studies collecting for all predictors that were intended by the VOICE score (maximum score: 11), above 60% of women scored 5 or above [5, 23, 24]. The threshold for which the observed incidence was >3% varied across populations: in South African samples, Peebles[6] found that incidence was >3% if AGYW scored 3 out of 11, while among the older sample aged 25-34 years only those scoring 6 out of 7 had incidence >3% (16.7% of the sample); in KwaZulu-Natal, Wand [21] observed >3% incidence for 88% of women enrolled in five clinical trials, while in an observational cohort only 60% (3rd quintile and above) of women had incidence >3% [30].

medRxiv preprint doi: https://doi.org/10.1101/2021.09.28.21264246; this version posted September 29, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY 4.0 International license .

Discussion

National HIV programmes in high HIV prevalence settings in sub-Saharan Africa are now considering whether and how to implement HIV risk scoring tools to support identification and prioritisation of populations and individuals for HIV prevention, especially oral PrEP and anticipated future prevention technologies. Our systematic review identified several scoring algorithms developed or validated for this purpose. Risk score development has especially focused on sexually active women of reproductive age or adolescent girls and young women. Only three studies included men and women. Among all sexually active women, younger age, not being married/cohabiting, and history of STIs (at baseline or lifetime) were consistently identified as prognostic factors. Among sexually active AGYW, history of STIs remained consistently selected, but being non-cohabiting was not consistently identified.

Risk scoring based on multiple predictors can improve efficiency in identifying individuals at higher risk of acquiring HIV compared to using individual risk factors [6, 31], but the improvement was only marginal (<0.1 increase in AUC-ROC) [31]. HIV incidence increased steadily with risk score in both development and validation studies, but the ability of risk scores to predict HIV incidence was only moderate. AUC-ROC values ranged from 0.56 to 0.79. AUC-ROC measures the discriminative power of the risk score defined as the probability that a risk score can successfully predict a HIV incident case from a case-and-control pair [32]. An AUC-ROC equal to 1 implies the model is able to perfectly discriminate those who will acquire HIV and those who do not, while 0.5 implies the model has no discriminative power. Most were lower than the AUC-ROC of scores developed for specific populations of sero-discordant couples (AUC-ROC: 0.70 [0.64, 0.76] and 0.76 [0.70,0.83] for two external validation) [33], men-having-sex-with-men in Kenya (0.76 [0.71,0.80]; derivation set) [34], and pregnant and post-partum women in Kenya (0.84 [0.72, 0.95] for derivation set; 0.73 [0.57,0.90] for internal validation set) [35]. The VOICE score was the only model externally validated by multiple studies (nine). Predictive performance of VOICE varied greatly across studies, even among those with all the predictors collected from women in

medRxiv preprint doi: https://doi.org/10.1101/2021.09.28.21264246; this version posted September 29, 2021. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY 4.0 International license .

general, which is the original intended population. Among AGYW-only populations, the discriminative power of the VOICE score is expected to be lower because one of the factors, younger age, is fulfilled by everyone in the sample. The substantial variation across settings suggests need for local validation of risk scores before implementation.

Only three of eleven multisite studies considered community-level HIV prevalence or viraemia as a prognostic factor, but all showed it being highly predictive [6, 28, 30]. This supports recommendations to consider both community-level exposure and individual factors to assess individual HIV risk and optimal prevention options. In fact, Roberts found that adding factors beyond community viraemia and age only modestly improved predictive ability, questioning the added value of potentially burdensome screening for more detailed risk behaviours [30]. In contrast, in their analysis adjusted for study sites, Balkus identified cohabitation as the most predictive factor, but additional covariates also substantially improved predictions [5]. Further data across multiple settings to adjudicate the added value of more detailed individual risk assessment will help guide HIV programme implementation strategies.

There are several possible reasons risk scores based on well-established risk factors are only moderately discriminative. First, HIV risk can change rapidly over short time intervals with life course events. Risk assessed at baseline may only be moderately predictive of an individual's actual HIV risk six to twelve months later. Second, adults with behaviour considered 'low risk', such as a single cohabiting sex partner, could still be exposed to high risk of HIV infection if their partner acquires HIV. While the HIV incidence rate among this group is relatively low, it may contribute a large proportion of total new infections, fundamentally limiting the extent to which HIV prevention can be optimised without specific, timely, and accurate information about risk among sexual partners. Third, factors included in risk scores are susceptible to reporting or measurement errors to varying degrees. Recent STI infection, identified through laboratory diagnosis in the clinical trials used for risk score development, was the most consistently identified predictive factor for HIV infection. However,

medRxiv preprint doi: https://doi.org/10.1101/2021.09.28.21264246; this version posted September 29, 2021. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY 4.0 International license .

laboratory testing for STIs is not routinely available in most low- and middle-income countries, where they were typically diagnosed through syndromic management instead. In our review, validation studies using self-reported or syndromic identified STIs [23, 24, 26] had similar accuracy (AUC-ROCs) as those using laboratory tests [6] (Table 3), but elsewhere syndromic management has consistently had only low to moderate accuracy [36-38].

Generalising and applying the risk scores and HIV incidence rate ratios reviewed here across high HIV burden settings faces several challenges. Data were disproportionately from South Africa, which has unique HIV epidemiology and low rates of marriage and cohabitation compared to neighbouring countries. Second, some data used to develop and validate risk scores were relatively old, with about half of studies completed before 2012 when HIV incidence was higher and ART coverage lower. Third, most studies were among randomised clinical trial study populations restricted to sexually active women seeking contraception, intentionally selected as relatively high-risk for testing novel HIV prevention technologies. Risk factors identified among this population are likely to generalise to a full adult population, but estimated AUC-ROC and pooled risk ratios should only be generalised with caution and appropriate adjustment. Fourth, there were subtle differences in definition and coding of risk factors across studies. This undermines the appropriateness of our pooled risk ratio estimates. In many validation studies, some selected risk factors were not available or defined differently [25, 26, 39].

More generally, developing risk scores for HIV incidence is fundamentally challenging, resulting in moderate to high assessed risk of bias using the PROBAST checklist (Table S3). As HIV infection is a relatively uncommon event, in most studies the ratio of cases observed to risk factors considered was far lower than recommended. Many studies were limited in accounting for over-fitting and model optimism, clarity about handling missing data [12]. Our review was also constrained by incomplete reporting of multivariate regression results of initial and final models in some studies. Only a few studies compared the AUC-ROC of the full models with that of individual predictors, making it difficult to draw conclusions about the

medRxiv preprint doi: https://doi.org/10.1101/2021.09.28.21264246; this version posted September 29, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY 4.0 International license .

necessity of detailed risk assessment compared to a few key characteristics—a key question for HIV programme implementation. Finally, we only focused on the general population and did not consider risk scores among key and vulnerable populations with high incidence. Other epidemiological evidence strongly supports prioritisation and provision of HIV prevention for these groups where they can be identified.

Our review identified three priorities for future studies. First, comparison of the AUC-ROCs of the full model versus individual predictors or more parsimonious models will help differentiate key predictors for identifying risk groups and prioritising resources and the relative value of factors that are more invasive or intensive to collect. Use of machine-learning techniques has also showed a potential to improving prediction accuracy and can be incorporated into some prevention interventions [40]. Second, additional risk score development and validation using recent incidence data from wider geographic settings will increase the generalisability of HIV risk scores. Finally, although in our review all AUC-ROCs in the external validation studies fell below 0.7, classified as poor discrimination by some [32], the discrimination of the risk scores may be higher when applied outside selected RCT populations. This could be explored through modelling, and further extended to study the infections averted, resources saved, and cost effectiveness of incorporating multivariable risk scores into risk stratification and prevention strategy prioritisation. Our findings inform such analyses by providing data on the incidence rate ratios and proportion of infections among each group compared to the size of the group.

Conclusion

Several risk scores have been developed for identifying individuals at increased risk for HIV among general populations in sub-Saharan Africa. For women, these have consistently identified younger age, not being married/cohabiting and STIs as risk factors. However, taken together, the programmatic benefit of implementing HIV risk scores as screening or triaging tools may be limited due to only moderate overall discriminatory ability and limited improvement compared to focusing on geographic areas with high HIV burden and basic

medRxiv preprint doi: https://doi.org/10.1101/2021.09.28.21264246; this version posted September 29, 2021. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY 4.0 International license .

demographics such as age group. The marginal benefits must be balanced with additional administrative burden for providers and consideration for whether screening questions could be perceived as stigmatizing, invasive, or demotivating for clients. Generalizability of proposed risk scores is uncertain because tools have been developed among relatively high-risk study populations and in South Africa and validations have been limited with moderate risk of bias.

medRxiv preprint doi: https://doi.org/10.1101/2021.09.28.21264246; this version posted September 29, 2021. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY 4.0 International license .

Competing interests

JWE reports grants from Bill and Melinda Gates Foundation and UNAIDS during the conduct of the study; grants from NIH, UNAIDS, and WHO and personal fees from WHO outside the submitted work. All other authors declare no competing interests.

Authors' contributions

KMJ and JWE conceptualised th review. KMJ did the initial literature search and wrote the protocol with substantial inputs from JWE. AH, HE, OE, KMJ, KL, and MT screened abstracts and full tests, extracted the data, and performed critical appraisal. KMJ performed the analysis and wrote the first draft of the manuscript with substantial inputs from JWE, OE and AH. KMJ, JWE, AH, HE, OE, KL, and MT contributed to interpretation of the results and edited the manuscript for intellectual content. All authors read and approved the final version of the manuscript.

Acknowledgments

This research was supported by the Bill & Melinda Gates Foundation (Grant numbers: OPP1190661, INV-002606, OPP1164897), UNAIDS, and the MRC Centre for Global Infectious Disease Analysis (reference MR/R015600/1), jointly funded by the UK Medical Research Council (MRC) and the UK Foreign, Commonwealth & Development Office (FCDO), under the MRC/FCDO Concordat agreement and is also part of the EDCTP2 programme supported by the European Union.

We thank Natsuko Imai for providing technical guidance and support, Adam Akullian and Allen Roberts for providing useful comments and unpublished details of their study in this review.

Additional files:

Additional file 1: Risk scores for predicting HIV incidence among general population in sub-Saharan Africa: a systematic review and meta-analysis—Supplementary Materials
File format: PDF
Description: This supplementary materials contain: (i) details of search and data extraction strategies, (ii) detailed characteristics of the data sources, (iii) summary of risk of bias assessment, (iv) summary of additional analyses of study results, and (v) PRISMA checklists.

List of abbreviations

medRxiv preprint doi: https://doi.org/10.1101/2021.09.28.21264246; this version posted September 29, 2021. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY 4.0 International license .

AGYW: adolescent girls and young women

(a)HR: (adjusted) hazard-ratio

AUC-ROC: area under the receiver operating characteristic curve

CHARMS: Critical Appraisal and Data Extraction for Systematic Reviews of Prediction Modelling Studies Checklist

HSV-2: Human Simplex Virus – 2

MMC: male medical circumcision

PrEP: pre-exposure prophylaxis

PROBAST: Prediction Model Risk of Bias assessment tool checklist

PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses

RCT: randomised controlled trial

STI: sexually transmitted infection

medRxiv preprint doi: https://doi.org/10.1101/2021.09.28.21264246; this version posted September 29, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY 4.0 International license .

References

1. The Joint United Nations Programme on HIV/AIDS (UNAIDS). End Inequalities. End AIDS. Global AIDS Strategy 2021-2026. 2021.

2. Global HIV Prevention Coalition. Implementation of the HIV Prevention 2020 Road Map. 2020 [Available from: https://www.unaids.org/en/resources/documents/2020.

3. Global HIV Prevention Coalition. Implementation of the HIV Prevention 2020 Road Map First progress report, March 2018. 2018 [Available from: https://www.unaids.org/en/resources/documents/2018/hiv-prevention-2020-road-map-first-progress-report.

4. UNAIDS. 2021 UNAIDS Global AIDS Update — Confronting inequalities — Lessons for pandemic responses from 40 years of AIDS. 2021.

5. Balkus JE, Brown E, Palanee T, Nair G, Gafoor Z, Zhang J, et al. An Empiric HIV Risk Scoring Tool to Predict HIV-1 Acquisition in African Women. Journal of Acquired Immune Deficiency Syndromes. 2016;72(3):333-43.

6. Peebles K, Palanee-Phillips T, Balkus JE, Beesham I, Makkan H, Deese J, et al. Age-Specific Risk Scores Do Not Improve HIV-1 Prediction Among Women in South Africa. Journal of acquired immune deficiency syndromes (1999). 2020;85(2):156-64.

7. Tanser F, Vandormael A, Cuadros D, Phillips AN, de Oliveira T, Tomita A, et al. Effect of population viral load on prospective HIV incidence in a hyperendemic rural African community. Science translational medicine. 2017;9(420):eaam8012.

8. Farahani M, Radin E, Saito S, Sachathep K, Hladik WA, Voetsch AC, et al. Population Viral Load, Viremia and Recent HIV-1 Infections: Findings from Population-based HIV Impact Assessments (PHIAs) in Zimbabwe, Malawi, and Zambia. J Acquir Immune Defic Syndr. 2021.

9. Anderson S-J, Cherutich P, Kilonzo N, Cremin I, Fecht D, Kimanga D, et al. Maximising the effect of combination HIV prevention through prioritisation of the people and places in greatest need: a modelling study. The Lancet. 2014;384(9939):249-56.

10. UNAIDS. Prevailing against Pandemics: by Putting People at the Centre. 2020.

11. Moons KGM, de Groot JAH, Bouwmeester W, Vergouwe Y, Mallett S, Altman DG, et al. Critical Appraisal and Data Extraction for Systematic Reviews of Prediction Modelling Studies: The CHARMS Checklist. PLOS Medicine. 2014;11(10):e1001744.

medRxiv preprint doi: https://doi.org/10.1101/2021.09.28.21264246; this version posted September 29, 2021. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY 4.0 International license .

12. Moons KGM, Wolff RF, Riley RD, Whiting PF, Westwood M, Collins GS, et al. PROBAST: A Tool to Assess Risk of Bias and Applicability of Prediction Model Studies: Explanation and Elaboration. Ann Intern Med. 2019;170(1):W1-w33.

13. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. 2020 [Available from: https://www.R-project.org/.

14. Balduzzi S, Rücker G, Schwarzer G. How to perform a meta-analysis with R: a practical tutorial. Evidence-Based Mental Health. 2019:153-60.

15. Viechtbauer W. Conducting meta-analyses in R with the metafor package. Journal of Statistical Software. 2010;3:1-48.

16. Zhou XH, Obuchowski N, McClish D. Statistical Methods in Diagnostic Medicine, Second Edition2002.

17. MedCalc. Meta-analysis: area under ROC curve 2021 [Available from: https://www.medcalc.org/manual/meta-analysis-ROC-area.php.

18. Eaton J, Thomas M, Edun O, Howes A, Eilerts H, Jia K, et al. Risk scores for predicting HIV incidence among general population in sub-Saharan Africa: a systematic review. PROSPERO 2021 CRD42021236367 2021 [Available from: https://www.crd.york.ac.uk/prospero/display_record.php?ID=CRD42021236367.

19. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ. 2021;372:n71.

20. 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):1155-62.

21. Wand H, Reddy T, Naidoo S, Moonsamy S, Siva S, Morar NS, et al. A Simple Risk Prediction Algorithm for HIV Transmission: Results from HIV Prevention Trials in KwaZulu Natal, South Africa (2002-2012). AIDS & Behavior. 2018;22(1):325-36.

22. Balkus JE, Brown ER, Palanee-Phillips T, Kiweewa FM, Mgodi N, Naidoo L, 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. Jaids. 2018;77(1):E8-E10.

medRxiv preprint doi: https://doi.org/10.1101/2021.09.28.21264246; this version posted September 29, 2021. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY 4.0 International license .

23. Burgess EK, Delany-Moretlwe S, Pisa P, Ahmed K, Sibiya S, Gama C, et al. Validation of a risk score for HIV acquisition in young African women with facts 001. Topics in Antiviral Medicine. 2017;25 (1 Supplement 1):364s-5s.

24. Burgess EK, Yende-Zuma N, Castor D, Karim QA. An age-stratified risk score to predict HIV acquisition in young South African women. Topics in Antiviral Medicine. 2018;26 (Supplement 1):419s.

25. Giovenco D, Pettifor A, MacPhail C, Kahn K, Wagner R, Piwowar-Manning E, et al. Assessing risk for HIV infection among adolescent girls in South Africa: an evaluation of the VOICE risk score (HPTN 068). Journal of the International AIDS Society. 2019;22(7):e25359 - .

26. Rosenberg NE, Kudowa E, Price JT, Pettifor A, Bekker L-G, Hosseinipour MC, et al. Identifying Adolescent Girls and Young Women at High Risk for HIV Acquisition: A Risk Assessment Tool From the Girl Power-Malawi Study. Sexually Transmitted Diseases. 2020;47(11):760-6.

27. Ayton SG, Pavlicova M, Abdool Karim Q. Identification of adolescent girls and young women for targeted HIV prevention: a new risk scoring tool in KwaZulu Natal, South Africa. Scientific reports. 2020;10(1):13017.

28. Kagaayi J, Gray RH, Whalen C, Fu P, Neuhauser D, McGrath JW, et al. Indices to measure risk of HIV acquisition in Rakai, Uganda. PLoS ONE. 2014;9(4).

29. Balzer LB, Havlir DV, Kamya MR, Chamie G, Charlebois ED, Clark TD, et al. Machine Learning to Identify Persons at High-Risk of Human Immunodeficiency Virus Acquisition in Rural Kenya and Uganda. Clinical Infectious Diseases. 2020;71(9):2326-33.

30. Roberts A, Cuadros D, Vandormael A, Gareta D, Barnabas R, Herbst K, et al. Predicting Risk of HIV Acquisition in Rural South Africa Using Geographic Data. Conference on Retroviruses and Opportunistic Infections (CROI) International Antiviral Society–USA 2021.

31. Balkus J, Palnee-Phillips T, Zhang J, Kiweewa FM, Nair G, Pather A, et al. A validated risk score to predict HIV acquisition in African women: assessing risk score performance among women who participated in the aspire trial. AIDS research and human retroviruses. 2016;32:49 - .

32. Hosmer DWJ, Lemeshow S, Sturdivant RX. Applied Logistic Regression, Third Edition: John Wiley & Sons, Inc.; 2013.

medRxiv preprint doi: https://doi.org/10.1101/2021.09.28.21264246; this version posted September 29, 2021. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY 4.0 International license .

33. Kahle EM, Hughes JP, Lingappa JR, John-Stewart G, Celum C, Nakku-Joloba E, et al. An empiric risk scoring tool for identifying high-risk heterosexual HIV-1-Serodiscordant couples for targeted HIV-1 prevention. Journal of Acquired Immune Deficiency Syndromes. 2013;62(3):339-47.

34. Wahome E, Thiong'o AN, Mwashigadi G, Chirro O, Mohamed K, Gichuru E, et al. An Empiric Risk Score to Guide PrEP Targeting Among MSM in Coastal Kenya. AIDS and Behavior. 2018;22(1):35-44.

35. Pintye J, Drake AL, Kinuthia J, Unger JA, Matemo D, Heffron RA, et al. A Risk Assessment Tool for Identifying Pregnant and Postpartum Women Who May Benefit From Preexposure Prophylaxis. Clin Infect Dis. 2017;64(6):751-8.

36. Zemouri C, Wi TE, Kiarie J, Seuc A, Mogasale V, Latif A, et al. The Performance of the Vaginal Discharge Syndromic Management in Treating Vaginal and Cervical Infection: A Systematic Review and Meta-Analysis. PLOS ONE. 2016;11(10):e0163365.

37. Barry MS, Ba Diallo A, Diadhiou M, Mall I, Gassama O, Ndiaye Guèye MD, et al. Accuracy of syndromic management in targeting vaginal and cervical infections among symptomatic women of reproductive age attending primary care clinics in Dakar, Senegal. Tropical Medicine & International Health. 2018;23(5):541-8.

38. Wi TE, Ndowa FJ, Ferreyra C, Kelly-Cirino C, Taylor MM, Toskin I, et al. Diagnosing sexually transmitted infections in resource-constrained settings: challenges and ways forward. Journal of the International AIDS Society. 2019;22(S6):e25343.

39. Balkus JE, Brown ER, Palanee-Phillips T, Matovu Kiweewa F, Mgodi N, Naidoo L, 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. Journal of acquired immune deficiency syndromes (1999). 2018;77(1):e8-e10.

40. Marcus JL, Sewell WC, Balzer LB, Krakower DS. Artificial Intelligence and Machine Learning for HIV Prevention: Emerging Approaches to Ending the Epidemic. Current HIV/AIDS Reports. 2020;17(3):171-9.

Table 1. Characteristics of the cohorts from which HIV risk scores were developed and/or validated

First author (Year)	Cohort (Study design)	Dev/ Val[§]	Year of study	Sites	Study population	Sex	Age[†]	Settings	N (Total PYs)	Incidence (per 100 PYs)
Wand (2012)	MIRA (RCT)	Dev**	2003-2006	Durban, KwaZulu-Natal, South Africa (2 sites)	18-49 yrs old, sexually active, non-pregnant, willing to use contraception	F	Mean: 27 IQR: 22-34	Family-planning clinics and other community-based organisation	1485[‡] (2162)	6.85
Wand (2018)	Multiple[¶] (RCT)	Dev**	2002-2012	KwaZulu-Natal, South Africa (multiple sites)	16+ yrs old, sexually active, non-pregnant, willing to use contraception	F	Median: 27 IQR: 22-33	Various clinical study sites	8982[‡] (11038)	7.03
Balkus (2016)	VOICE (RCT)	Dev*	2009-2011	South Africa, Uganda, and Zimbabwe	18–45 yrs old, sexually active, non-pregnant, willing to use contraception	F	Median: 24 IQR: 21-29	STD clinics, family planning clinics, and post-natal clinics, community-based locations	4834 (4348)	6.05
	HPTN 035 (RCT)	Val	2005-2009	Malawi, South Africa, Zimbabwe, and Zambia	18+ yrs old, sexually active, not (intended to be) pregnant.	F	25 (22-29)	STD clinics, family planning clinics, and post-natal clinics, community-based locations	2848 (2903)	3.38
	FEM-PrEP (RCT)	Val	2009-2011	Kenya, South Africa, and Tanzania	sexually active, non-pregnant, 18–35 yrs old women at high risk	F	23 (20-27)	Community outreach, recruitment sites, community partners, health centres, STI clinics, HIV voluntary testing and counselling centres	1804 (1231)	4.79
Balkus (2018)	ASPIRE (RCT)	Val	2012-2015	Malawi, South Africa, Uganda, and Zimbabwe	18–45 yrs old, sexually active, not (intended to be) pregnant	F	<25 (39%)	STIs, family-planning clinics	2539 (2566)	3.70
Burgess (2018)	CAPRISA 004 (RCT)	Dev	2007-2010	KwaZulu-Natal, South Africa (2 sites)	18-40 yrs old, sexually active women, non-pregnant, willing to use contraception	F	<25 (68%)	An urban and a rural CAPRISA research clinic	431 (660.7)	9.08
Burgess (2017)	FACTS 001 (RCT)	Val	2011-2014	South Africa (9 sites)	18-30 yrs old, sexually active, non-pregnant, willing to use contraception	F	Median: 23 IQR: 20-25	9 community-based clinical trial sites	1115 (1876)	4.32
Peebles (2020)	ECHO (RCT)	Dev*	2015-2018	South Africa (9 sites)	18-35 yrs old, sexually active, seeking effective contraception	F	<25 (62.1%)	9 clinics over 5 provinces	5670 (5573)	5.4 (<25) 3.4 (25+)
Giovenco (2019)	HPTN 068 (RCT)	Val	2011-2012	South Africa (single site)	school-attending 13-20 yrs old AGYW	F	Median: 15 IQR: 14-17	a random sample in the rural Bushbuckridge in Mpumalanga province	2178 (2455)	1.34
Rosenberg (2020)	Girl Power (Quasi-experi-mental)	Dev*	2016-2017	Malawi (4 sites)	sexually active, 15-24 yrs old AGYW	F	<20 (58.7%)	4 public-sector health centers in Lilongwe, Malawi	795 (672)	2.08
Ayton (2020)	CAPRISA 007 (RCT)	Val	2010-2012	South Africa (14 sites)	14-25 yrs old school-attending AGYW	F	Median: 17 IQR: 16-18	Grade 9 and 10 students in 14 schools in Vulindlela	971 (-)	1.72

medRxiv preprint doi: https://doi.org/10.1101/2021.09.28.21264246; this version posted September 29, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Kagaayi (2014)	RCCS (Cohort)	Dev*	2003-2011	Uganda (~50 communities)	sexually active, 15–49 yrs old	F	Mean: 27.0 SD: 7.8	Communities in Rakai district	7497	1.11
						M	Mean: 28.3 SD: 8.0		5783	0.98
Balzer (2020)	SEARCH (RCT)	Dev*	2018	Kenya and Uganda (16 communities)	15+ years old residents	F M	<25 (39%)	16 communities in rural Uganda and Kenya	75558 (-)	0.27-0.37
Roberts (2021)	ACDIS (Cohort)	Dev**	2012-2019	Umkanyakude district of KwaZulu-Natal, South Africa (single site)	15+ years old residents	F	-	A community in Mtubatuba town	11933‡ (28422)	4.20 (2012-15) 3.11 (2016-19)
						M	-		7623‡ (16449)	1.80 (2012-15) 1.16 (2016-19)

§ Dev: model development (marked with * if the authors did re-sampling procedures through cross-validation or bootstrapping to obtain the AUC-ROC for internal validation of their score(s), and ** if they split the sample into training and testing sets as part of their internal validation); Val: external validation of the risk score in a study population different from which the score was originally developed. Peebles [6], Burgess (2018) [24] and Rosenberg [26] developed their own risk scores while also using their samples to externally validate the VOICE score developed by Balkus [5]; AUC-ROCs are provided in Table 3.

¶ Multiple cohort studies included: MIRA, MDP 301, NCT00213083, VOICE, HPTN035. † Mean/median (IQR)

‡ Total sample size is provided here, whereas the authors split the sample into training and testing sets.

medRxiv preprint doi: https://doi.org/10.1101/2021.09.28.21264246; this version posted September 29, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

Table 2. Risk factors retained in the final models and their adjusted effects

		Women only, all ages (RCTs) ¶					AGYW only (RCT / quasi-experimental trials) ¶				General population#	
	First author (Year)	Retainment ratio	Wand (2012)	Wand (2018)	Balkus (2016)	Peebles (2020)	Retainment ratio	Peebles (2020)	Burgess (2018)	Rosenberg (2020)	Kagaayi (2014)	Kagaayi (2014)
	Cohort		MIRA	Multiple	VOICE	ECHO 25+ yrs old		ECHO <25 yrs old	CAPRISA 004 <25 yrs old	GirlsPower	RCCS female	RCCS male
Demographic	**Age (younger)**	3/4	N.I.	1.75-2.93***	1.70*	2.12*	0/3			N.I.	1.03*	0.81*
	Age of sexual debut (<16)	1/1		1.38*								
	Not married/living with partners	4/4	2.63**	3.62***	1.80*	1.85*	0/3	N.I.	N.I.	N.I.	1.72-2.08*	N.I.
	Pregnancy history						1/2	N.I.		4.55*		
	Parity (less)	1/3		2.62***	N.I.	N.I.	0/1	N.I.				
	Education (Higher)	0/4	N.I.	N.I.	N.I.	N.I.	0/2	N.I.		N.I.	0.83^	0.56*
	High-risk occupation♦										1.32*	
	Perceived HIV risk (High)										1.49*	
	Alcohol use	1/3	N.I.		1.41*	N.I.	1/2	1.45*		N.I.		
Partnership factor	**Partner >5 y older**						1/1			2.42^		
	Partner in high risk occupation											1.89*
	Fin. support from partner	1/3	N.I.		0.72*	N.I.	0/3	N.I.	N.I.	N.I.#		
	Partner has other partners:											
	Don't know vs No	1/2			1.81*	N.I.	1/1		7.56*			
	Yes vs No	1/3	N.I.		1.63*	N.I.	2/3		7.86*	N.I.		
	Yes/don't know vs No							1.31^				
	Partner has unknown HIV status											1.82*
	No. of partners†	2/4			N.I.	N.I.	2/3	1.61*	2.19*	N.I.	1	1
	2 (ref : 1)		1.34^								1.59*	1.21*
	3 (ref : 1)		1.93*									1.90*
	3+ (ref: <2)			1.61***								
	New sex partner										1.45*	
	Concurrent relationships										1.50^	

medRxiv preprint doi: https://doi.org/10.1101/2021.09.28.21264246; this version posted September 29, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

Transmission probability	**Condom use (Yes)**	1/3	N.I.	N.I.		N.I.	1/1	1.34^				
	Oral contraceptive (Yes)	0/1		N.I.								
	Inject. Contraceptive (Yes)	1/1		1.43***								
	High risk sex	1/1	1.96*	-							1.44*	1.28*
	MMC of primary partners/ MMC (male)	0/3		N.I.	N.I.	N.I.	0/2	N.I.		N.I.		0.61*
	HSV-2 [§]	2/4			1.63*	1.88*	3/3	1.51*	2.54*	1.94^		
	STIs/ genital infection[‡]	4/4	1.60**	1.57***	1.49*	2.22*	2/3	2.07*	N.I.	2.61^	1.75*	1.78*
Community	**Community type**											1.67*
	Community prevalence	0/1				N.I.	1/1	1.64*-1.81*			1.03*	1.03*
	Province	1/1				9.05*						
	Adjusted for sites (Y/N)		N	N	Y	Y		Y			Y	Y
	AUC-ROC		0.79 [§§]	0.71 [§§]	0.69 [§]	0.64 [§]		0.62 [§]	0.7	0.79	0.67 [§]	0.69 [§]
	95% CI		0.70-0.81	-	0.66-0.72	0.59-0.69		0.58-0.64	0.60-0.79	0.69-0.89	0.64-0.70	0.66-0.73

[¶] Women recruited from RCTs were sexually active and willing to use effective contraception during the study period

Balzer [29] and Roberts [30] also developed risk scores in their studies, but the final models were not available and thus details are not shown in Table. Risk factors considered by these analyses are reported in Table S2.

♦ Detailed list of high-risk occupations is provided by Kagaayi (2014) [28].

'N.I.' indicates risk factor was measured but not selected for inclusion in final model; Dark grey shaded cells indicate risk factor was not considered as a candidate predictor in the study; Light grey shaded cells in the *Retainment Ratio* column indicates risk factor not considered by any models.

* (p<0.05); ** (p<0.01); *** (p<0.0001) significant and included in the final risk score;

^ (p>0.05) not significant but included

[†] Wand (2012) [20] considered lifetime number of sex partners, while all others considered recent number of sex partners(Table S1 contains details on temporality of sexual partners for each study).Casual partnerships (none vs at least one) among sexually-active women was considered as a proxy for multiple partners in Burgess (2018) [24] study.

[‡] Methods for STIs status assessment varied by studies; details are provided in Table S2 of the Supplementary Materials.

[#] Although partners' financial support from were not included in the final model, all HIV cases (n=14) in Rosenberg [26] engaged in transactional sex.

[§] AUC-ROC was obtained through cross-validation or bootstrapping of the original derivation set for the internal validation; [§§] AUC-ROC was based on a "testing" set different from the "training" set for which the score was derived.

medRxiv preprint doi: https://doi.org/10.1101/2021.09.28.21264246; this version posted September 29, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

Table 3. Predictive performance measures

First author (Year)	Dev[¶]	Cohort (Study design)	HIV incident cases	Prediction horizon	Score developed	AUC-ROC (95% CI)	Val[¶]	Predictors available	Score externally validated	AUC-ROC (95% CI)
Wand (2012)	1 **	MIRA (RCT)	148	1 to 2 yrs	Wand-2012	0.79 (0.70, 0.81) **	-	-	-	-
Wand (2018)	1 **	Multiple § (RCT)	776	1 to 2+ yrs	Wand-2018	0.71 (-)**	1	-	-	-
Balkus (2016)	1*	VOICE (RCT)	263	1 yr	VOICE	0.69 (0.66, 0.72) *	-	-	-	-
					VOICE (Modified)	0.67 (0.64, 0.70) *	1	-	-	-
		HPTN 035 (RCT)	98	1 yr			-	6/7	VOICE	0.70 (0.65, 0.75)
		FEM-PrEP (RCT)	59	1 yr			-	6/7	VOICE	0.63 (-)
							-	4/5	VOICE (Modified)	0.58 (0.51, 0.65)
Balkus (2018)	-	ASPIRE (RCT)	95	1 yr			1	5/7	VOICE	0.69 (0.64, 0.74)
Burgess (2018)	1	CAPRISA 004 (RCT)	47 (<25) 13 (25+)	18 mo	Burgess (under 25) Burgess (under 25) (Modified)	0.70 (0.60, 0.79) 0.62 (0.56, 0.68)	1	All[#]	VOICE VOICE (under 25) VOICE (≥25)	0.66 (0.54, 0.74) 0.69 (0.60, 0.78) 0.49 (0.30, 0.63)
Burgess (2017)	-	FACTS 001 (RCT)	81	2.2 yrs			1	All[#]	VOICE	0.56 (0.50, 0.62)
Peebles (2020)	2 *	ECHO (RCT)	188 (<25) 72 (25+)	1 yr			1	All	VOICE	0.61 (0.58, 0.65)
							-	All	VOICE (Modified)	0.59 (0.56, 0.62)
					Peebles (under-25) Peebles (under-25) (Modified)	0.62 (0.58, 0.64) * 0.59 (0.55, 0.61) *	-	-	-	-
					Peebles (25+) Peebles (25+) (Modified)	0.64 (0.59, 0.69) * 0.62 (0.58, 0.67) *	-	-	-	-
Giovenco (2019)	-	HPTN 068 (RCT)	33	1 yr			1	6/7	VOICE	0.55 (0.44, 0.65)
Rosenberg (2020)	1 *	Girl Power (Quasi-experi-mental)	14	1 yr	Rosenberg	0.79 (0.69, 0.89) *	1	-	-	-
							-	All[#]	VOICE	0.64 (0.52, 0.75)
Ayton (2020)	-	CAPRISA 007 (RCT)	18	1 yr			1	4/7	VOICE	-
Kagaayi (2014)	2 *	RCCS (Cohort)	342	1 yr	Kagaayi (female)	0.67 (0.64, 0.70) *	-	-	-	-
		RCCS (Cohort)	225	1 yr	Kagaayi (male)	0.69 (0.66, 0.73) *	-	-	-	-
Balzer (2020)	3 *	SEARCH (RCT)	519	1 yr	Balzer (Risk group-based) Balzer (Model-based) Balzer (Machine-learning)	0.59 (0.55, 0.62) * 0.70 (0.68, 0.73) * 0.73 (0.71, 0.76) *	-	-	-	-

medRxiv preprint doi: https://doi.org/10.1101/2021.09.28.21264246; this version posted September 29, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

Roberts (2021)	2 **	ACDIS (Cohort)	1160	1 yr	Roberts (female) Roberts (female) (age + geographic covariates only)	0.68 ** 0.65 **	-	-	-	-
			248	1 yr	Roberts (male) Roberts (male) (age + geographic covariates only)	0.72 ** 0.71 **	-	-	-	-

¶ Number of risk score(s) developed ("Dev") / externally validated ("Val") by the study

* AUC-ROC was obtained through cross-validation or bootstrapping of the original derivation set for the internal validation; ** AUC-ROC was based on a "testing" set different from the "training" set for which the score was derived. Wand (2012) [20] and Wand (2018) [21] did a random split of data sets, while Roberts [30] split the data by time periods (2012-2015 vs 2016-2019).

Self-reported STIs histories or symptoms were used as proxies for curable STIs and HSV-2 status in Burgess (2017) [23] and Rosenberg [26], and for curable STIs in Burgess (2018) [24].

medRxiv preprint doi: https://doi.org/10.1101/2021.09.28.21264246; this version posted September 29, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

medRxiv preprint doi: https://doi.org/10.1101/2021.09.28.21264246; this version posted September 29, 2021. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY 4.0 International license .

Figures

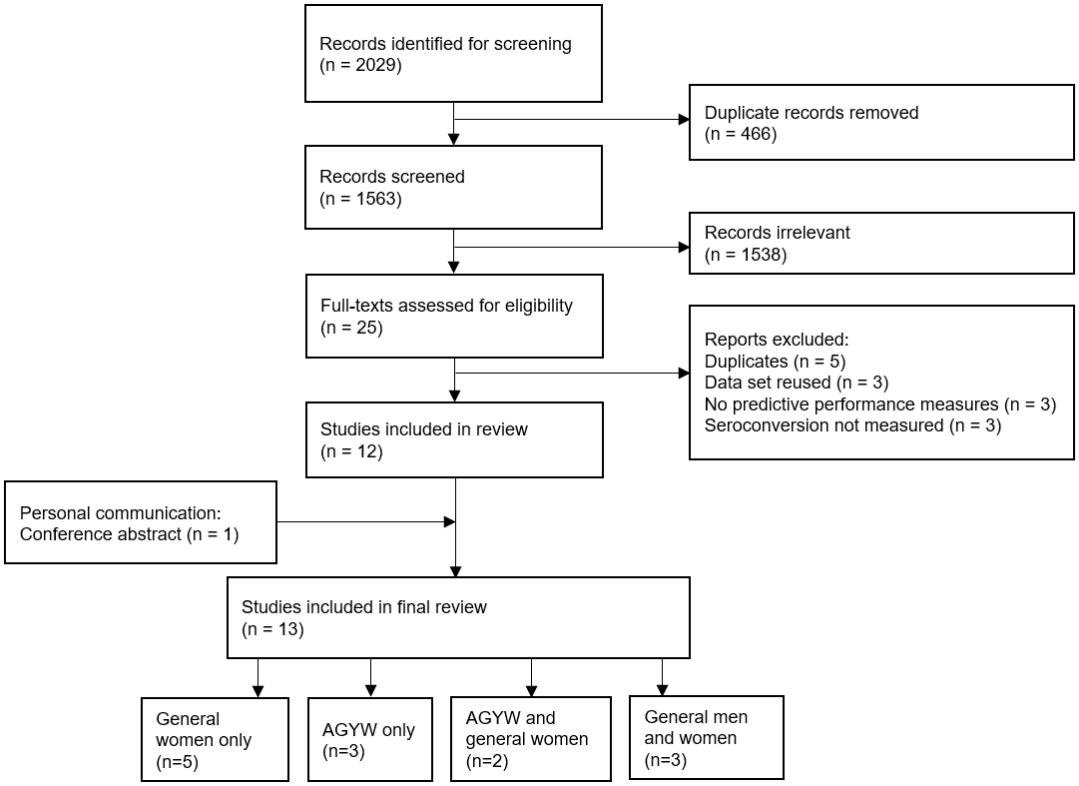

Figure 1. Study selection.

medRxiv preprint doi: https://doi.org/10.1101/2021.09.28.21264246; this version posted September 29, 2021. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY 4.0 International license .

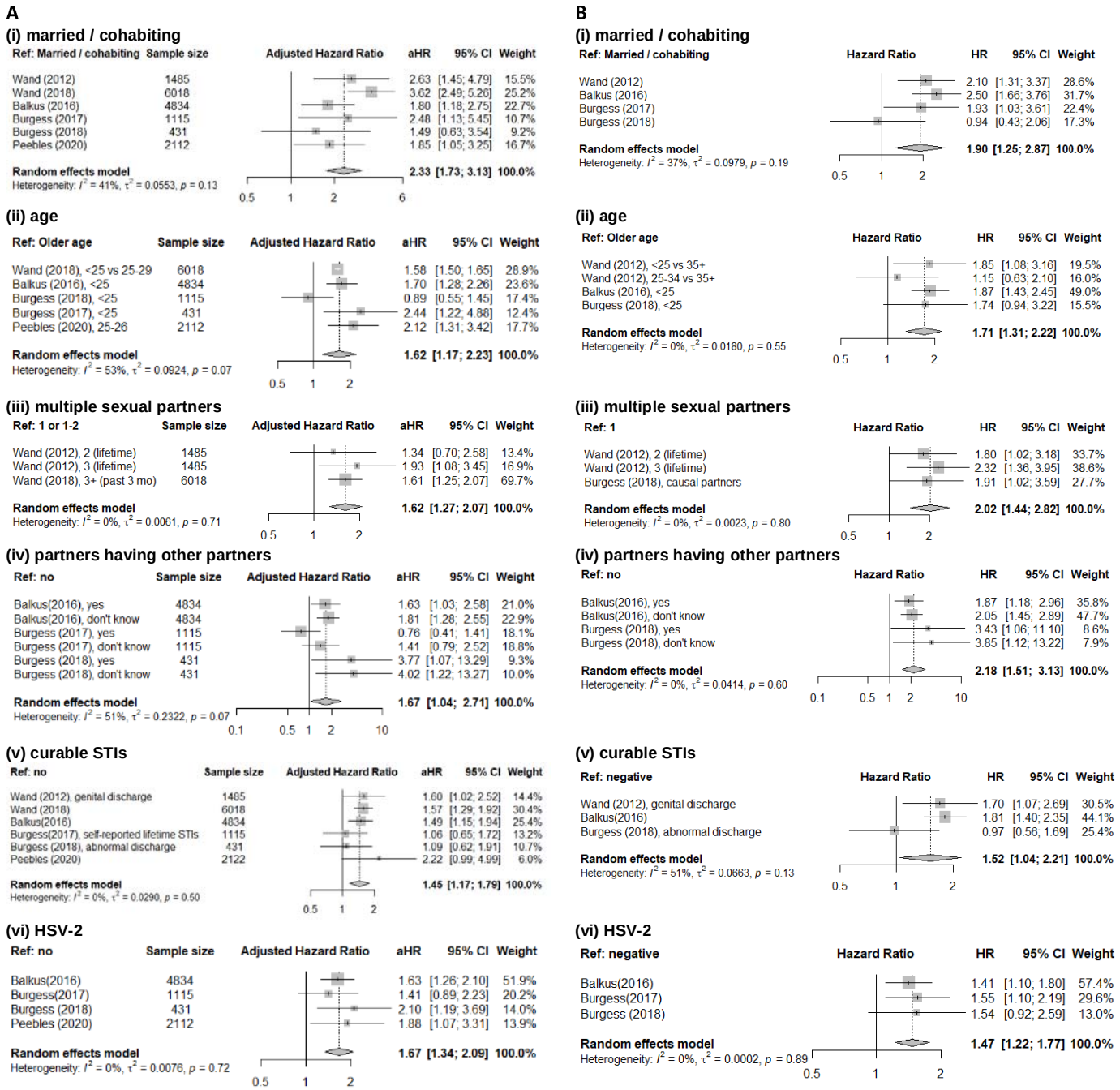

Figure 2. Forest plots of risk factor estimates among women in general. Adjusted (A) and unadjusted (B) effects were pooled together for: (i) marital / cohabiting status, (ii) age, (iii) number of sexual partners, (iv) partners having other partners, (v) curable sexually transmitted infection (STIs) and (vi) HSV-2.

medRxiv preprint doi: https://doi.org/10.1101/2021.09.28.21264246; this version posted September 29, 2021. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY 4.0 International license .

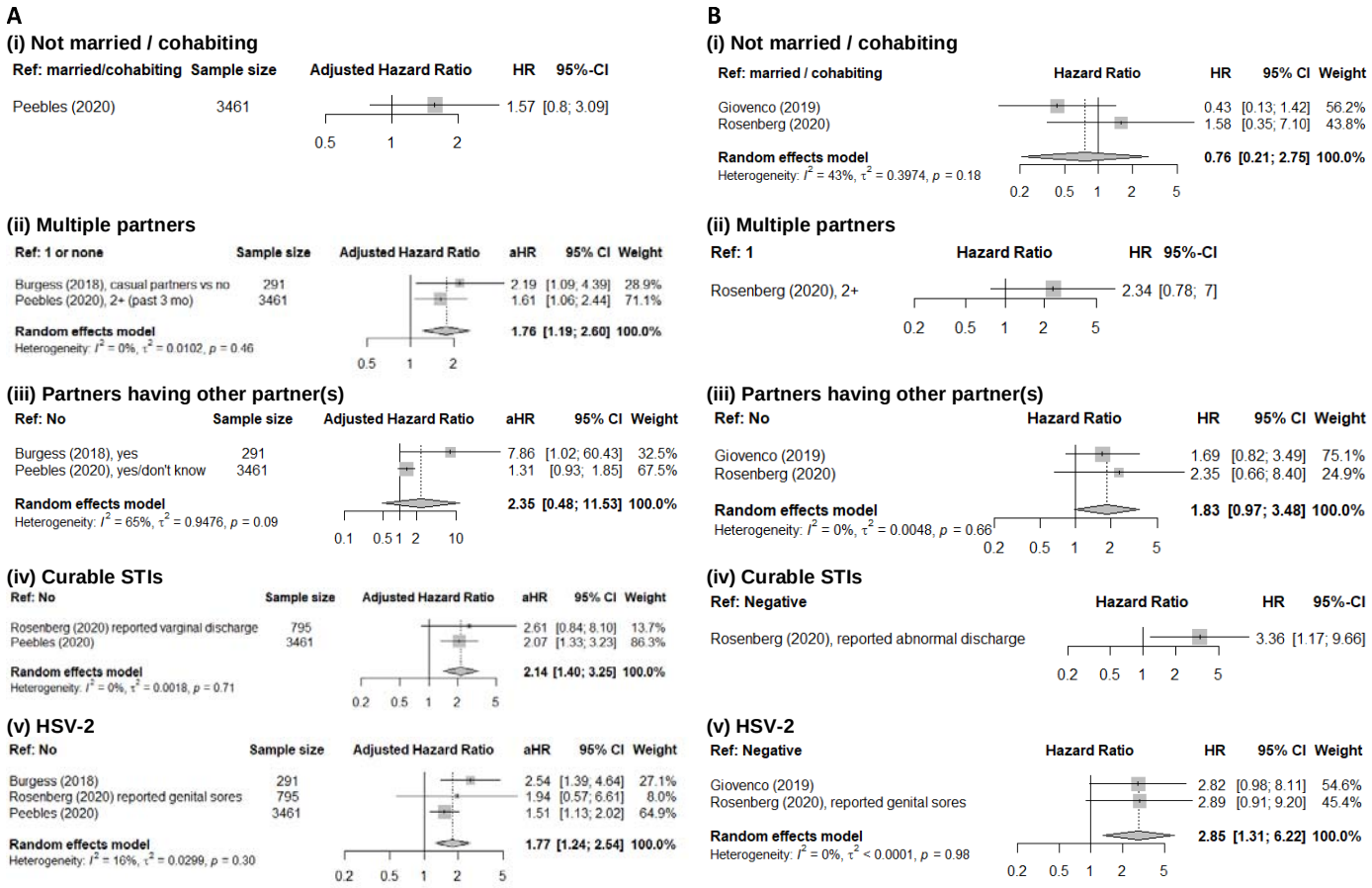

Figure 3. Forest plots of risk factor estimates among adolescent girls and young women (AGYW). Adjusted (A) and unadjusted (B) effects were pooled together for (i) marital / cohabiting status (ii) number of sexual partners, (iii) partners having other partners, (iv) curable sexually transmitted infection (STIs) and (v) HSV-2.

medRxiv preprint doi: https://doi.org/10.1101/2021.09.28.21264246; this version posted September 29, 2021. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY 4.0 International license .

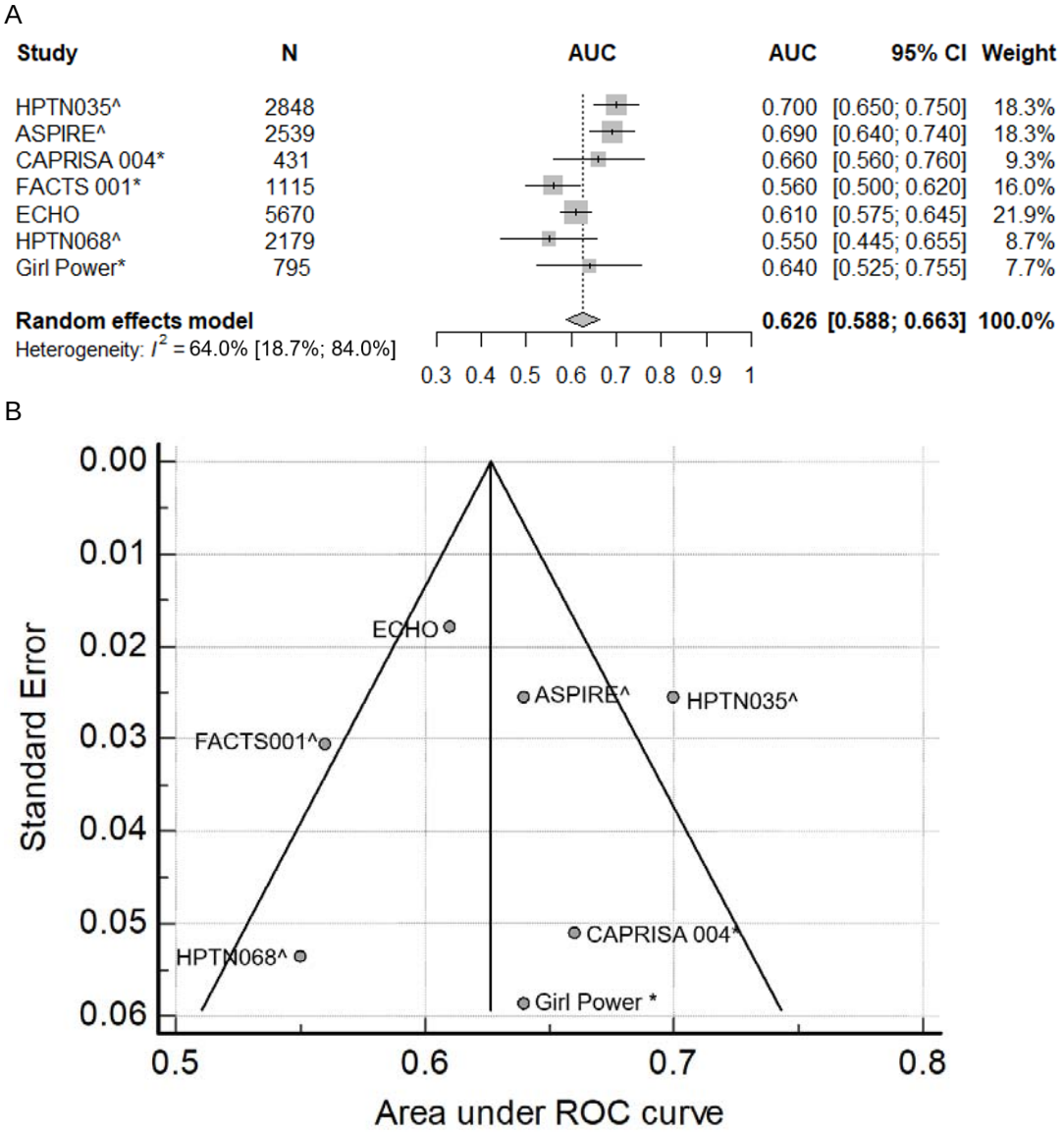

Figure 4. Forest plot (A) and funnel plot (B) for the area-under-curve of the receiver operating characteristic curves (AUC-ROCs) from external validation studies for the VOICE score. Studies with ^ did not collect all the predictors intended by the model (Details are in Table 3). Those with * used self-reported STIs history, syndromic management or self-reported symptoms in place of laboratory diagnosed STIs status at baseline as intended by the original VOICE score (Details are in Table 3 and Table S2).

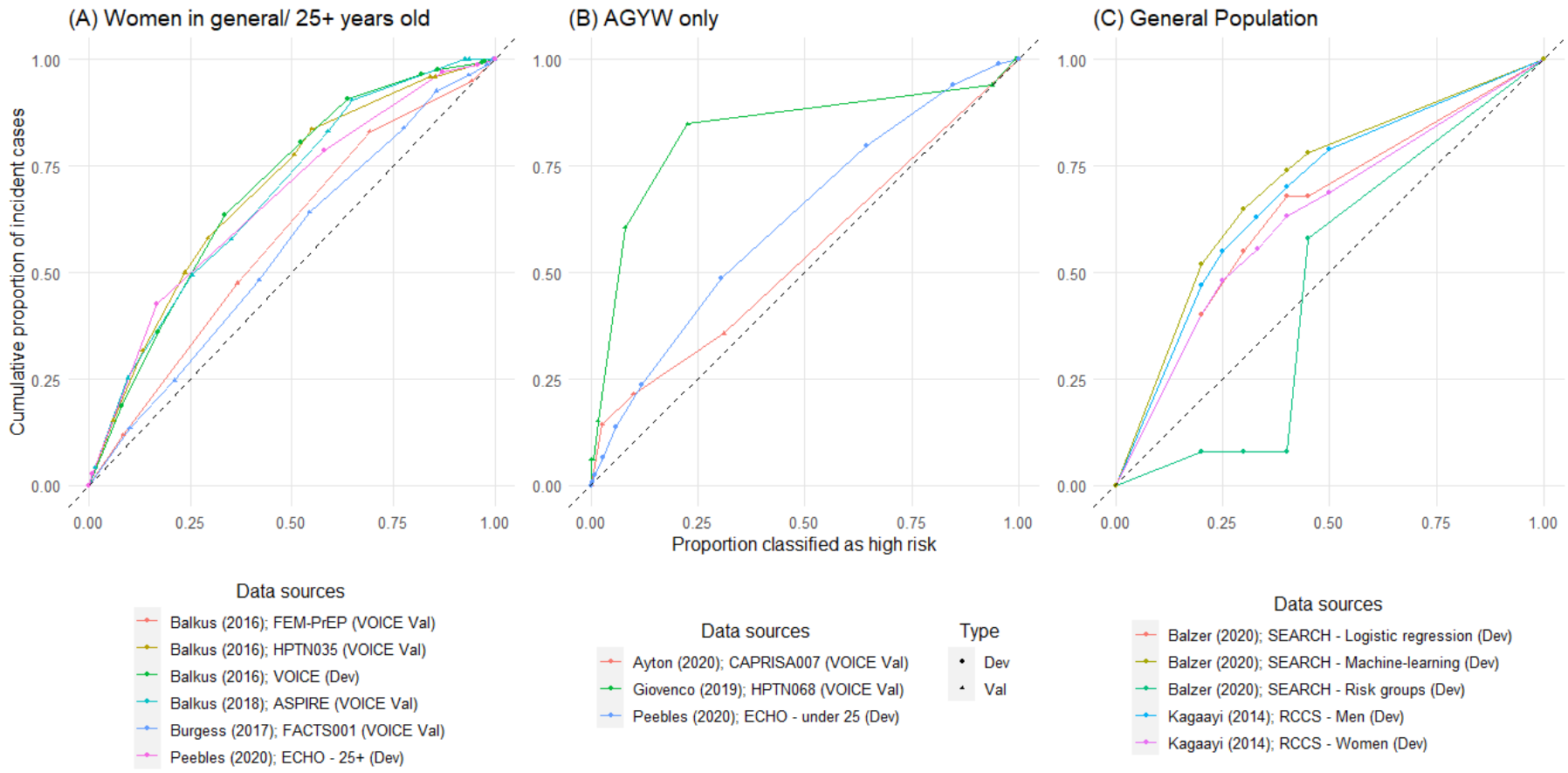

Figure 5. Percentage of individuals identified as high risk among incident cases versus proportion classified as high risk for: (i) women enrolled in clinical trials (A), (ii) AGYW (B), and (iii) the general population (C). When the highest score was used as the threshold, few or none were classified as high risk and they took up a small fraction of all HIV incident cases (indicated by the origin). When the lowest score was used, all were classified as high risk and all incident cases were among them. Dev: Development set; Val: Validation set.

medRxiv preprint doi: https://doi.org/10.1101/2021.09.28.21264246; this version posted September 29, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license.