UNAIDS 2021

REFERENCE

PrEP target-setting for key and high-priority populations

Technical materials

Contents

Abbreviations

ALPM	area-level HIV prevalence among adult males
AUC	area under the curve
BBS	biobehavioural surveillance
BIC	Bayesian information criterion
CI	confidence interval
HIV	human immunodeficiency virus infection
HR	hazard ratio
IQR	interquartile range
MOT	modes of transmission
OR	odds ratio
PHIA	population-based HIV impact assessment
PrEP	pre-exposure prophylaxis
PSE	population size estimate
RDS	respondent-driven sampling
RPR	rapid plasma regain
SS	successive sampling
TLS	time–location sampling
VDRL	Venereal Disease Research Laboratory
VICITS	Estrategia de Vigilancia Centinela de las Infecciones de Transmisión Sexual
WHO	World Health Organization

Minimum risk behaviours model

The guidance describes approaches to developing a definition of substantial risk of HIV based on setting an incidence threshold (e.g. 3%, as recommended by WHO) and determining the minimum number of risk partners and risky exposure acts required to reach that threshold.

The tools provided include spreadsheets that calculate these minimum numbers of acts. This section describes the mathematical model implemented to develop these minimum numbers.

The model included was adapted from the modes of transmission (MOT) framework *(1)*. MOT is based on a binomial force of infection equation that calculates the number of new infections over a year in different population subgroups as a function of levels of risk behaviour. We are interested in the reverse: calculating risk behaviours needed to meet a predefined incidence target.

In addition, the force of infection equation was modified to take into account antiretroviral therapy coverage in the partner population, which reduces the force of infection. Antiretroviral therapy coverage was not incorporated in the initial MOT model as first published, but it is incorporated in more recent MOT spreadsheets.

Modes of transmission model

The original MOT incidence formula is *(1)*:

$$I = U\left[1 - \{pS[1-\beta']^{\alpha(1-v)} + p(1-S)[1-\beta]^{\alpha(1-v)} + (1-p)\}^{n}\right]$$

where:

I = incidence of HIV in the key population.

U = number of uninfected individuals susceptible to being infected.

p = HIV prevalence in the partner population.

S = sexually transmitted infection prevalence in the target or partner population.

β = probability of transmission of HIV during a single contact in the absence of a sexually transmitted infection.

β′ = probability of transmission of HIV during a single contact in the presence of a sexually transmitted infection.

V = proportion of acts currently protected.

α = number of contacts per partner.

n = number of partners.

In the MOT formula, the probability of transmission is conditioned on the probability that a partner randomly drawn from the population is living with HIV, and whether this partner has a sexually transmitted infection. The output of the model is the number of new infections over the period (e.g. one year).

Adapted model

Below, the formula is adapted for the purposes of calculating minimum levels of behaviour. There are three main changes:

- To account for antiretroviral therapy, the probability of transmission is conditioned additionally on the coverage level of antiretroviral therapy in the partner population (A).
- The formula is simplified to consider only risky (unprotected) acts of exposure.
- The left-hand side is recast as the probability of HIV acquisition during the period. Because probability over the course of a year is the same as incidence, this change allows us to relate risk behaviours (numbers of risky partners and risky exposure acts over the course of a year) to a target incidence level, expressed as a percentage (e.g. 3%).

$$P(HIV) = 1 - \left\{ \begin{matrix} pS[A(1-rsa)^{\gamma} + (1-A)(1-rs)^{\gamma}] + \\ p(1-S)[A(1-ra)^{\gamma} + (1-A)(1-r)^{\gamma}] + \\ (1-p) \end{matrix} \right\}^{n}$$

where:

P(HIV) = probability a susceptible individual acquires HIV during the period (one year).

p = HIV prevalence in the partner population.

S = sexually transmitted infection prevalence in the partner population.

r = probability of HIV transmission during a single risky (unprotected) act of exposure in the absence of sexually transmitted infection and antiretroviral therapy.

s = multiplier on transmission probability in the presence of sexually transmitted infection.

a = multiplier on transmission probability in the presence of antiretroviral therapy.

γ = number of risky acts of exposure per partner per year.

n = number of partners per year.

Circumcision is not included in the model because we are not modelling heterosexual males. As in the MOT and Goals models, we assume men who have sex with men, transgender women and females receive no protective benefit from male circumcision of male partners.

The model is limited to risky (unprotected) acts and does not account for additional probability of infection from protected sexual acts, for example due to incorrect use of condoms or defective condoms. In Goals, condom effectiveness is assumed to be 80%.

The transmission probabilities (r, s, a) are as in the most recent *(2012)* MOT spreadsheets available online (Table 1).

For people who inject drugs, the model accounts for HIV risk due to injection but not due to sexual relationships (the sexually transmitted infection multiplier s is set to 1.0).

Table 1.
Model parameters: HIV transmission probability per risky exposure act

Population	Base probability (r)	Sexually transmitted infection multiplier (s)	Antiretroviral therapy multiplier (a)
Men who have sex with men	3 384 160		24%
Transgender women	0.01	4.8	0.90
Female sex workers	897 644		6%
Adolescent girls and young women	0.001	4.8	0.96
People who inject drugs	0.01	1.0	0.80

Using force of infection formula to calculate minimum numbers of risk partners and risky acts

The spreadsheet tools include a Minimum Behaviors Calculator, which calculates P(HIV) over a reasonable range of numbers of partners and acts. For each number of partners, the spreadsheet identifies the number of acts that just reaches the target incidence level. The spreadsheet lists all combinations of minimum numbers of partners and acts.

These calculations are carried out for three risk scenarios, which vary the values of HIV and sexually transmitted infection prevalence (p and S):

- Participants with partners living with HIV (where p in the formula is set to 1).
- Other participants with sexually transmitted infections (S is set to 1).
- Other participants without sexually transmitted infection (or sexually transmitted infection status unknown) (the formula as written, where p and S are the prevalence parameters).

Interpretation

It is important to interpret the minimum numbers produced by the model at the population level because they rely on averages. The numbers are intended for the purposes of developing population-level estimates and do not necessarily reflect the minimum levels of risk behaviour for an individual person.

Review of evidence of risk factors for HIV

Overview of search methods

We reviewed the published literature (plus grey literature for transgender women) to identify recent evidence of variables associated with the risk of HIV infection among men who have sex with men, transgender women, female sex workers, people who inject drugs, and adolescent girls and young women in middle- and low-income countries.

We initially sought to identify robust evidence of consistency of risk factors identified across studies without necessarily being exhaustive. Due to lack of consistency with respect to known risk factors (e.g. unprotected anal intercourse) across the studies initially reviewed, we conducted exhaustive systematic reviews for all populations.

Methods differed by population in accordance with the volume of evidence available for each group and time limitations. PubMed searches required terms for the respective population, HIV and analysis of associations.1 Lessons learned during the first reviews (men who have sex with men and female sex workers) led to improved search terms for remaining groups. Search terms for key populations were adapted from previous reviews *(2–5)*.

For transgender women, we also reviewed the AIDSHub database and UNAIDS reports because evidence from publications was limited.

Study selection criteria differed by population. More strict criteria were applied where more studies were available.

Low- and middle-income countries were identified using the World Bank Atlas method: gross national income per capita of US$ 12 235 or less in 2016.

Titles were reviewed for evidence of not meeting study selection criteria. Abstracts of remaining search results were reviewed. Full text documents were obtained and reviewed when records appeared to meet study selection criteria or when eligibility was unclear.

Data extraction

Data on all estimates of associations with HIV infection were entered into a spreadsheet using a standardized format. Data were entered for each variable that could be evaluated for association with HIV infection. When a paper reported associations with both incident and prevalent infection, only the former were recorded. For categorical

1 HIV search terms for men who have sex with men, female sex workers and transgender women were "HIV" or "AIDS" (MeSH) or "HIV" (in title or abstract). For people who inject drugs and for adolescent girls and young women, they were "HIV" or "human immunodeficiency virus" and "infection", "prevalence" or "incidence" (in title). Search terms for analysis of associations for men who have sex with men, female sex workers and transgender women were "logistic regression," "regression analysis" or "longitudinal studies" (MeSH) or "risk factor*," "correlate*" or "associate*" (in title or abstract). For people who inject drugs and for adolescent girls and young women, "risk factor" or "risk factors" was added to the MeSH terms and "regress*" or "predictor*" to the title or abstract terms.

variables, a record was entered for each level analysed. Data included the name of the variable, reference condition for the comparison, reference time period (e.g. past 12 months) and estimates.

In the first reviews (men who have sex with men, female sex workers, transgender women), we recorded the estimate of effect size (e.g. odds ratio, OR; hazard ratio, HR) and confidence interval (CI) only when the association was significant ($P \leq 5\%$) as we did not initially aim to conduct meta-analysis. For people who inject drugs and for adolescent girls and young women, we also recorded the *P*-value of all associations, regardless of statistically significance.

When a paper presented associations from multiple subsamples (e.g. by study site or by males and females who inject drugs), we recorded the subgroup for each association. When a paper did not report effect size estimates but presented HIV prevalence by subgroup, we recorded frequencies, prevalence and results of any bivariate testing (e.g. Chi-squared test) for each subgroup.

We reviewed papers to determine whether they reported on data from the same study. We excluded papers that reported on associations with prevalent HIV infection if a different paper included in the review reported on associations with incident infection. In other cases when multiple papers reported on the same data, we identified variables examined by both papers and eliminated redundant associations by retaining those for which methods appeared most robust.

Statistical analysis

Associations were classified into a risk factor category. Categories were developed iteratively to accommodate the types of association that emerged over the course of the review.

Risk factor categories were defined to reflect a direction of the association (e.g. condomless sex versus condom use; younger versus older age at first sex) to best align with the evidence. Each association was coded to reflect a positive, negative or non-significant relationship between HIV and the category, based on the estimated effect size and reference group of the association.

For example, in the risk factor category "condomless sex with a male partner", an OR of 1.5 for the variable "always versus never used condoms in the past 3 months" would be coded as "negative" in direction, while a significant OR of 1.5 for the variable "did not use condoms at last sex" would be coded as "positive" in direction. If the ORs were less than 1.0 in these examples, the labels would be reversed.

When studies reported HIV prevalence by subgroup, we calculated ORs based on reported frequencies, estimated confidence intervals for the OR using the delta method,2 and calculated *P*-values using unpooled Z tests.

Consistency of associations

We developed measures of the consistency and strength of the associations for each risk factor category. Associations were considered consistent if all were in the same direction. Consistency was not evaluated when there were fewer than two associations. Each subgroup analysis (e.g. by site) from a given study could contribute associations.

2 SE[ln(OR)] = sqrt(1/*a* + 1/*b* + 1/*c* + 1/*d*) where *a*–*d* are the cell counts used to calculate the OR.
The CI on the log scale is then ±1.96 × SE[ln(OR)] and the CI for the OR is obtained by exponentiating these limits.

Meta-analysis of effect size

The strength of the available evidence was assessed by fixed effects meta-analysis, separately for associations with incident and prevalent HIV infection. The combined effect size was calculated as the average of the reported estimates, weighted by the inversed variance of each estimate. Variance was rarely reported by studies and so was approximated from the 95% CI as (upper limit − lower limit)/2 × 1.96.

Before calculating the variances and weighted average, if necessary effect sizes and CI limits were reversed (by taking the reciprocal) to ensure the direction of the estimate was consistent with the direction of the category. The 95% CI for the combined estimate was derived by calculating the variance as the inverse of the sum of the weights *(6)*.

The meta-analysis for people who inject drugs and for adolescent girls and young women includes both statistically significant and non-significant estimates. The meta-analysis is partial for men who have sex with men, transgender women and female sex workers as it includes only statistically significant estimates because data for non-significant estimates were not extracted during the review.

If a study reported on multiple estimates in a category, we first retained only statistically significant estimates. If there were still multiple estimates, we retained the one with the greatest value of effect size/variance.

Findings from previous reviews

PubMed searches identified a number of published reviews of factors associated with HIV. We present variables that authors of these past reviews concluded were associated with HIV infection. We did not include previous reviews that that drew primarily on evidence from high-income countries or that did not specify the location of the underlying evidence. No other study selection criteria were applied to the previous reviews.

Findings

The variables identified by the review and meta-analysis estimates are intended to be informative, as a point of reference for countries to consider when carrying out the risk factors approach to define risk.

Findings are presented by population in three tables:

- Variables with any significant association reported with incident HIV infection.
- Other variables (i.e. no association with incident HIV) with any significant association reported with prevalent HIV infection.
- Variables found to be associated with HIV by previous systematic reviews.

Men who have sex with men

Search strategy:

- Databases searched: PubMed.
- Publication dates: 2012 to 4 September 2017.

- Population search terms: men who have sex with men, males who have sex with males, homosexual*, bisexual* (in title or abstract).

Study selection criteria:

- Data collected from low- or middle-income country.
- Some part of data collection took place during 2008–2017.
- Sample size ≥ 250 participants.
- Included HIV laboratory test and behavioural data.
- Participants aged ≥ 15 years.
- Eligibility criteria referenced specific behaviour and time period (e.g. sex with a male in past 12 months). Studies defining eligibility as "ever" engaging in the behaviour were excluded.
- < 10% of participants identified as transgender women (if reported).
- Did not aim to recruit male sex workers exclusively.
- If cross-sectional, sampling was by respondent-driven sampling (RDS) or time–location sampling (TLS).
- Reported estimate of association between HIV infection and at least one variable, or reported HIV prevalence by subgroup so that associations could be calculated.
- Estimates weighted or otherwise adjusted for study design.

Search results:

- 2378 papers returned by PubMed query.
- 23 papers meeting study selection criteria and providing associations from unique studies.
- 7 incidence/cohort studies and 16 prevalence/cross-sectional studies.

Meta-analysis:

- Limited to statistically significant associations ($P \leq 0.05$), which may bias estimated pooled effect sizes upwards.
- Tables list all risk factor categories examined.

In the following tables:

- N = number of included associations. Papers could contribute up to one association per risk factor category per subgroup (e.g. study sites).
- Combined effect: pooled effect size (OR, HR) and 95% CI.
- Range of effects: minimum and maximum effect size across studies.
- Consistent relationship: all associations were in the same direction (positive or negative). This was not evaluated if $N < 1$.

Table 2.
Variables significantly associated with incident HIV among men who have sex with men[a]

Variable	Significant association with incident HIV				Significant association with prevalent HIV			
	N	Pooled estimate	Range of significant estimates	Consistent relation-ship?	N	Pooled estimate	Range of significant estimates	Consistent relation-ship?
Sexually transmitted infection (laboratory-confirmed)	7	2.27 (1.56–2.99)	1.48–17.7	Yes	5	3.86 (2.7–5.01)	3–4.93	Yes
Number of male partners	3	3.16 (1.67–4.65)	2.52–5.07	Yes	2	2.17 (0.97–3.36)	2.09–2.25	Yes
Anal sex	2	3.39 (0.95–5.83)	2.86–9.16	Yes	0	–	–	–
Unprotected anal intercourse	2	3.01 (1.22–4.8)	2.48–6.47	Yes	3	0.78 (0.22–1.33)	0.59–5.1	No
Condomless sex with female partners	1	9.78 (1.19–80.24)	–	–	0	–	–	–
Condomless sex	1	4.84 (1.78–13.19)	–	–	2	0.99 (0.52–1.45)	0.4–1.25	No
Receptive or versatile role in anal sex	1	1.67 (1.24–2.25)	–	–	3	2.74 (1.58–3.9)	2.33–7.2	Yes
HIV knowledge	2	0.26 (0–0.78)	0.11–0.5	Yes	1	2.1 (1.1–4.3)	–	–
Race or ethnicity[b]	1	5.7 (1.5–21.5)	–	–	1	8.3 (2.4–29.1)	–	–
Venues where partners are met[b]	1	3.61 (1.03–12.47)	–	–	3	3.36 (1.91–4.81)	3–8.98	–
Sex partners' locations of residence	1	3.75 (1.52–9.26)	–	–	0	–	–	–
Sex partners' ages	1	3.4 (1.11–10.39)	–	–	0	–	–	–
Lower educational attainment	1	2.12 (1.12–4.03)	–	–	3	0.72 (0.28–1.16)	0.34–1.72	No
Married/previously married to woman	1	1.99 (1.01–3.93)	–	–	3	0.52 (0.08–0.95)	0.21–3.33	No
Drug use	1	1.99 (1.36–2.91)	–	–	1	0.15 (0.04–0.65)	–	–
Younger age at first male sex	1	0.41 (0.21–0.8)	–	–	3	2.82 (1.75–3.9)	1.82–4.3	Yes
Exposure to HIV prevention	1	0.38 (0.17–0.83)	–	–	2	2.7 (1.15–4.26)	2.6–2.82	Yes
Tested for HIV previously	1	0.3 (0.16–0.56)	–	–	2	5.18 (2.78–7.57)	4.7–5.9	Yes
Meets partners online	1	0.06 (0–0.81)	–	–	1	3.1 (1.2–7.6)	–	–

[a] Based on review of studies published in 2012–2017 in low- and middle-income countries. Prevalent associations are included for comparison.

[b] Specific categories and comparisons vary by study.

Table 3.
Variables significantly associated with prevalent HIV among men who have sex with men[a]

Variable[b]	Significant association with prevalent HIV			
	N	Pooled estimate	Range of significant estimates	Consistent relationship?
Proximal				
Sexually transmitted infection symptoms/syndrome	3	3.27 (1.69–4.85)	3.06–3.4	Yes
Receptive anal sex (or number of partners)	3	2.17 (1.4–2.94)	1.7–12	Yes
Sexually transmitted infection history	2	2.74 (1.58–3.9)	2.6–4.3	Yes
Insertive anal sex (or number of partners)	2	0.64 (0.25–1.04)	0.5–0.71	Yes
Partner of unknown HIV status	1	7.9 (1–60.9)	–	–
Condom breakage/slippage	1	3.6 (1.5–8.6)	–	–
Insertive role	1	0.24 (0.14–0.42)	–	–
Injecting drug use	1	0.08 (0.01–0.6)	–	–
Unprotected anal intercourse receptive	2	0.61 (0–1.34)	0.36–2.7	No
Distal				
Sex work or transactional sex	3	2.86 (1.53–4.18)	2.17–4.9	Yes
Discrimination related to men who have sex with men	2	6.97 (3.83–10.11)	5.38–12.7	Yes
Forced sex	2	1.95 (0.85–3.05)	1.6–3.11	Yes
Cohabitating or stable partner	2	1.44 (0.83–2.05)	1.4–1.48	Yes
Religion[c]	2	1.34 (0.73–1.95)	1.23–3.9	–
Study site or area or residence[c]	1	2.9 (1.3–6.5)	–	–
History of prison	1	4.37 (1.38–13.84)	–	–
Unable to access condoms	1	2.8 (1.3–6.2)	–	–
Openness about sexual orientation	1	2.5 (1.96–3.23)	–	–
Higher socioeconomic status	1	1.37 (1.15–1.61)	–	–
Higher income	1	0.78 (0.64–0.95)	–	–
Internalized homophobia	1	0.4 (0.22–0.71)	–	–
Had casual partner (or number of partners)	1	0.26 (0.08–0.85)	–	–
Paid for sex	1	0.03 (0.1–0.3)	–	–
Female partners (or number of partners)	6	0.4 (0.1–0.7)	0.1–2.7	No
Alcohol use (frequent or problematic)	3	0.88 (0.18–1.59)	0.4–4.72	No
Lubricant use	3	0.67 (0–1.37)	0.12–2.89	No

[a] Based on review of studies published in 2012–2017 in low- and middle-income countries.

[b] Does not include variables shown in Table 2.

[c] Specific categories and comparisons vary by study.

Table 4.
Risk factors for HIV among men who have sex with men identified by previous reviews published during 2012–2017

Variable	Reference	Time period reviewed	Geographical area of studies reviewed
Proximal			
HSV-2	(7)	2003–2017	Africa, Americas, Europe, Southeast Asia, Western Pacific
Syphilis	(8)	Past 30 years	Asia (Bangladesh, Cambodia, China, India, Indonesia, Lao People's Democratic Republic, Malaysia, Myanmar, Philippines, Singapore, Taiwan, Thailand, Viet Nam)
Syphilis	(9)	2005–2014	China
Syphilis	(10)	1991–2014	China
Multiple male sex partners	(8)	Past 30 years	Asia (Bangladesh, Cambodia, China, India, Indonesia, Lao People's Democratic Republic, Malaysia, Myanmar, Philippines, Singapore, Taiwan, Thailand, Viet Nam)
Multiple male sex partners	(9)	2005–2014	China
Multiple male sex partners	(10)	1991–2014	China
Receptive anal sex or receptive role in anal sex	(8)	Past 30 years	Asia (Bangladesh, Cambodia, China, India, Indonesia, Lao People's Democratic Republic, Malaysia, Myanmar, Philippines, Singapore, Taiwan, Thailand, Viet Nam)
Receptive anal sex or receptive role in anal sex	(9)	2005–2014	China
Unprotected anal intercourse	(9)	2005–2014	China
Unprotected anal intercourse	(10)	1991–2014	China
Sex with females	(11)	2000–2014	China
Transactional sex	(12)	2004–2013	Southeast Asia (Indonesia, Lao People's Democratic Republic, Thailand, Viet Nam), South Asia (India, Nepal), East Asia (China), Latin America (Argentina, Ecuador, El Salvador, Peru), sub-Saharan Africa (Kenya, Senegal, South Africa, Uganda), North America (United States of America), Middle East (Israel)
Uncircumcised	(9)	2005–2014	China
Distal			
Age < 25 years	(9)	2005–2014	China
Homosexual identification	(9)	2005–2014	China
Lower educational attainment	(9)	2005–2014	China
Migration and mobility	(10)	1991–2014	China

HSV, herpes simplex virus.

Transgender women

Search strategy:

- Databases searched: PubMed, AIDSHub, reports provided by UNAIDS.
- Publication dates: 2008–2017.
- Population search terms: transgender persons (MeSH) or transgender, travesty, koti, hijra, MTF, male to female transgender, transsexual*, transvest*, mahuvahine, mahu, waria, katoey, cross dresser, bantut, nadleehi, berdachel, xanith (in title or abstract).

Study selection criteria:

- Data collected from low- or middle-income country.
- Some part of data collection took place during 2008–2017.
- Sample size ≥ 250 participants.
- Included HIV laboratory test and behavioural data.
- Participants aged ≥ 15 years.
- Eligibility criteria referenced specific behaviour and time period (e.g. sex with male in past 12 months). Studies defining eligibility as "ever" engaging in the behaviour were excluded.
- Did not aim to recruit transgender women sex workers exclusively.
- If cross-sectional, sampling was by RDS or was venue-based (including mapping stage and with venue selection random or by census).
- Reported estimate of association between HIV infection and at least one variable, or reported HIV prevalence by subgroup so that associations could be calculated.

Search results:

- 513 documents returned by databases.
- 6 documents meeting study selection criteria and providing associations from unique studies.
- 0 incidence/cohort studies and 6 prevalence/cross-sectional studies.

Meta-analysis:

- Limited to statistically significant associations ($P \leq 0.05$), which may bias estimated pooled effect sizes upwards.
- Tables list all risk factor categories examined.

In the following tables:

- N = number of included associations. Papers could contribute up to one association per risk factor category per subgroup (e.g. study sites).
- Combined effect: pooled effect size (OR, HR) and 95% CI.
- Range of effects: minimum and maximum effect size across studies.
- Consistent relationship: all associations were in the same direction (positive or negative). This was not evaluated if $N < 1$.

Table 5.
Variables significantly associated with prevalent HIV among transgender women[a]

Variable	Significant association with prevalent HIV			
	N	Pooled estimate	Range of significant estimates	Consistent relationship?
Proximal				
Number of male partners or sex acts	2	2.0 (0.81–3.19)	1.72–13	Yes
Sexually transmitted infection symptoms/syndrome	2	1.86 (0.92–2.8)	1.58–2.4	Yes
Sexually transmitted infection (laboratory-confirmed)	2	1.8 (0.84–2.77)	1.56–2.7	Yes
Unprotected anal intercourse	1	3.84 (1.58–9.33)	–	–
Injection drug use	1	3.25 (1.37–7.71)	–	–
Anal sex	1	3.2 (1.17–8.96)	–	–
Distal				
Sold sex	2	1.66 (0.46–2.86)	1.5–30.7	Yes
Study site or area of residence[b]	1	3.32 (1.36–8.09)	–	–
Occupationb	1	2.9 (1.2–7.01)	–	–
Strong transgender identity	1	34.1 (5.8–220.2)	–	–
Race (Brazil)	1	22.8 (2.9–178.9)	–	–
Lower income	1	6.15 (1.05–36)	–	–
Younger age at first male sex	1	5.1 (0.08–0.47)	–	–
Years of sex work	1	4.44 (1.02–19.28)	–	–
Takes gender-affirming hormones	1	4.4 (1.2–17.3)	–	–
Drug use	1	4.4 (1.4–14.1)	–	–
Perceived HIV risk	1	4.1 (1.55–10.92)	–	–
Low self-esteem	1	3.25 (1.35–7.85)	–	–
Drug use during or after sex	1	2.9 (1.09–7.73)	–	–
Urban residence	1	2.7 (1.1–6.5)	–	–
Has sex with partners at hotels or lodges	1	2.4 (1.89–3.17)	–	–
Physical violence related to transgender women	1	2.35 (1.09–4.53)	–	–
Dresses as woman all the time	1	2.1 (1.2–3.8)	–	–

Variable	Significant association with prevalent HIV			
	N	Pooled estimate	Range of significant estimates	Consistent relationship?
Meets partners online	1	1.9 (1.2–3.2)	–	–
Has cohabitating or stable partner	1	0.61 (0.41–0.89)	–	–
Alcohol (frequency, binge drinking)	1	0.6 (0.36–0.93)	–	–
Lower educational attainment	3	0.82 (0.15–1.49)	0.2–2.62	No

[a] Based on review of studies published in 2008–2017 in low- and middle-income countries.

[b] Specific categories and comparisons vary by study.

Table 6.
Risk factors for HIV among transgender women identified by previous reviews published during 2008–2017

Variable	Reference	Time period reviewed	Geographical area of studies reviewed
Sex work	(13)	1980–2007	14 countries in 5 continents

Female sex workers

Search strategy:

- Databases searched: PubMed.
- Publication dates: 2012 to 4 September 2017.
- Population search terms: sex worker* (MeSH) or sex work*, female sex worker*, commercial sex worker* (in title or abstract).

Study selection criteria:

- Data collected from low- or middle-income country.
- Some part of data collection took place during 2008–2017.
- Sample size ≥ 250 participants.
- Included HIV laboratory test and behavioural data.
- Participants aged ≥ 15 years.

- Eligibility criteria referenced a specific behaviour and time period (e.g. sold sex in past 12 months). Studies defining eligibility as "ever" engaging in the behaviour were excluded.
- If cross-sectional, sampling was by RDS or was venue-based (including a mapping stage and with venue selection random or by census).
- Reported estimate of association between HIV infection and at least one variable, or reported HIV prevalence by subgroup so that associations could be calculated.

Search results

- 794 papers returned by PubMed.
- 23 papers meeting study selection criteria and providing associations from unique studies.
- 2 incidence/cohort studies and 21 prevalence/cross-sectional studies.

Meta-analysis

- Limited to statistically significant associations ($P \leq 0.05$), which may bias estimated pooled effect sizes upwards.
- Tables list all risk factor categories examined.

In the following tables:

- N = number of included associations. Papers could contribute up to one association per risk factor category per subgroup (e.g. study sites).
- Combined effect: pooled effect size (OR, HR) and 95% CI.
- Range of effects: minimum and maximum effect size across studies.
- Consistent relationship: all associations were in the same direction (positive or negative). This was not evaluated if $N < 1$.

Table 7.
Variables significantly associated with incident HIV among female sex workers[a]

Variable	Significant association with incident HIV				Significant association with prevalent HIV			
	N	Pooled estimate	Range of significant estimates	Consistent relation-ship?	N	Pooled estimate	Range of significant estimates	Consistent relation-ship?
Proximal								
Sexually transmitted infection (laboratory-confirmed)	1	7.19 (1.68–30.77)	–	–	4	2.9 (1.86–3.93)	2.19–9.63	Yes
Number of clients or encounters	1	4.9 (1.81–13.13)	–	–	3	1.92 (1.2–2.63)	1.8–3.05	Yes
Injection drug (or heroin) use	1	3.7 (1.11–12.35)	–	–	8	4.81 (3.92–5.69)	2.8–22.05	Yes
Condomless sex with clients	1	2.9 (1.03–8.34)	–	–	4	1.15 (0.73–1.58)	0.79–3.33	No
Distal								
Drug use	1	6.7 (2.25–19.93)	–	–	3	2.34 (1.66–3.03)	1.44–6.0	Yes
Lives with family (versus apartment or brothel)	1	3.8 (1.34–10.69)	–	–	0	–	–	–
Alcohol use or frequency	1	0.3 (0.12–0.77)	–	–	0	–	–	–

[a] Based on review of studies published in 2012–2017 in low- and middle-income countries. Prevalent associations are included for comparison.

Table 8.
Variables significantly associated with prevalent HIV among female sex workers[a]

Variable[b]	Significant association with prevalent HIV			
	N	Pooled estimate	Range of significant estimates	Consistent relationship?
Proximal				
Sexually transmitted infection symptoms/syndrome	4	1.83 (1.27–2.38)	1.58–3.04	Yes
Condomless sex (type of partner unspecified)	2	0.85 (0.28–1.43)	0.71–3.36	No
Condomless sex with non-commercial partner	2	0.81 (0.02–1.6)	0.5–3.99	No
Distal				
Lower educational attainment or illiteracy	9	1.88 (1.53–2.23)	1.35–3.72	Yes
Workplace/modality[c]	8	1.69 (1.32–2.07)	1.34–9.95	–
Study site or area or residence[c]	5	1.43 (0.93–1.92)	1–11.11	–

Variable[b]	Significant association with prevalent HIV			
	N	Pooled estimate	Range of significant estimates	Consistent relationship?
Forced sex	3	1.91 (1.21–2.61)	1.74–3.77	Yes
Lower price charged per sex	3	1.58 (1.03–2.13)	1.4–1.79	Yes
Tested for HIV or sexually transmitted infection	3	0.43 (0.11–0.75)	0.18–0.61	Yes
Number of children, pregnancies or abortions	2	4.15 (2.29–6.01)	3.06–13	Yes
Sex work is main source of income or only job	2	1.34 (0.75–1.93)	1.22–1.68	Yes
Recruiter has HIV (RDS)	1	4.6 (2.3–9.19)	–	–
Exposure to HIV prevention or harm reduction	1	3.25 (2.36–4.48)	–	–
Perceived HIV risk	1	2.65 (2.11–3.33)	–	–
Physical violence	1	2.52 (1.41–4.51)	–	–
Condom as contraception	1	2.5 (1.7–3.5)	–	–
Sexually transmitted infection history or treatment	1	2.3 (1.2–4.1)	–	–
Condom use negotiable with clients	1	2.21 (1.24–3.94)	–	–
Alcohol during sex work	1	1.62 (1.04–2.53)	–	–
Urban concentration in district	1	1.55 (1.05–2.29)	–	–
Sex partner injects drugs	1	1.45 (1.07–1.97)	–	–
Mobility or migration	1	0.62 (0.4–0.96)	–	–
Purchased condoms	1	0.5 (0.3–0.83)	–	–
Hormonal contraception	1	0.5 (0.1–0.9)	–	–
Condom requirement met in district	1	0.5 (0.26–0.99)	–	–
Requests condom use of non-commercial partner	1	0.3 (0.15–0.6)	–	–
HIV knowledge	3	1.04 (0.57–1.5)	0.7–1.51	No
Years of sex work or younger age of initiation	8	1.01 (0.83–1.18)	0.25–2.11	No

[a] Based on review of studies published in 2012–2017 in low- and middle-income countries.

[b] Does not include variables shown in Table 7.

[c] Specific categories and comparisons vary by study.

Table 9.
Risk factors for HIV among female sex workers identified by previous reviews published during 2012–2017

Variable	Reference	Time period reviewed	Geographical area of studies reviewed
Proximal			
HSV-2	(7)	2003–2017	Africa, Americas, Europe, Southeast Asia, Western Pacific
Sexually transmitted infections in general	(14)	2000–2010	Sub-Saharan Africa
Trichomoniasis	(10)	1991–2014	China
Anal sex with clients	(14)	2000–2010	Sub-Saharan Africa
Condomless sex with clients	(14)	2000–2010	Sub-Saharan Africa
Multiple concurrent partnerships	(14)	2000–2010	Sub-Saharan Africa
Injection drug use	(10)	1991–2014	China
Intravaginal cleansing with soap	(15)	1997–2015	Uganda
Distal			
Occupational context (poverty, violence, criminalization, high mobility, hazardous alcohol use)	(14)	2000–2010	Sub-Saharan Africa
Sex work is sole income source	(15)	1997–2015	Uganda
Sex work location: hair salons, massage parlours, small hotels, streets	(10)	1991–2014	China
Sex work location: streets	(15)	1997–2015	Uganda
Shorter duration practising sex work	(14)	2000–2010	Sub-Saharan Africa
Alcohol consumption	(15)	1997–2015	Uganda
Age > 25 years	(15)	1997–2015	Uganda
Widowed	(15)	1997–2015	Uganda
Lower educational attainment	(15)	1997–2015	Uganda
Migration and mobility	(10)	1991–2014	China
Incarceration	(10)	1991–2014	China
Not knowing own HIV status	(15)	1997–2015	Uganda

People who inject drugs

Search strategy:

- Databases searched: PubMed.
- Publication dates: 2008–2017.
- Population search terms: inject* (in title).

Study selection criteria:

- Data collected from low- or middle-income country.
- Some part of data collection took place during 2008–2017.
- Sample size ≥ 250 participants.
- Included HIV laboratory test and behavioural data.
- Participants aged ≥ 15 years.
- Eligibility criteria referenced a specific behaviour and time period (e.g. injection in past 12 months). Studies defining eligibility as "ever" engaging in the behaviour were excluded.
- If cross-sectional, sampling was by RDS, venue-based (including mapping stage and with venue selection random or by census), or any recruitment of clients of harm reduction services or drug treatment facilities.
- Reported an estimate of the association between HIV infection and at least one variable, or reported HIV prevalence by subgroup so that associations could be calculated.

Search results:

- 183 papers returned by PubMed.
- 19 papers meeting study selection criteria and providing associations from unique studies.
- 5 incidence/cohort studies and 14 prevalence/cross-sectional studies.

Meta-analysis:

- Included both statistically significant and non-significant associations.
- Tables list only risk factor categories with a statistically significant pooled meta-analysis estimate.

In the following tables:

- N = number of included associations. Papers could contribute up to one association per risk factor category per subgroup (e.g. study sites).
- Combined effect: pooled effect size (OR, HR) and 95% CI.
- Range of effects: minimum and maximum effect size across studies.
- Consistent relationship: all associations were in the same direction (positive or negative). This was not evaluated if N < 1.

Table 10.
Variables with a significant pooled association with incident HIV among people who inject drugs[a]

Variable	Association with incident HIV				Association with prevalent HIV			
	N	Pooled estimate	Range of estimates	Consistent relation-ship?	N	Pooled estimate	Range of estimates	Consistent relation-ship?
Proximal								
Number of injections sharing needles or paraphernalia	1	2.53 (0.99–6.44)	–	–	1	1.56 (1.04–2.33)	–	–
Religion (India and Pakistan)[b]	1	1.7 (1.4–2.7)	–	–	0	–	–	–
Study site or area of residence[b]	1	1.61 (1.06–2.44)	–	–	1	15.2 (1.5–145.2)	–	–
Number of drug dealers used	1	8.46 (2.25–31.8)	–	–	0	–	–	–
Hepatitis B vaccination	1	3.56 (1.29–9.79)	–	–	0	–	–	–
Lives alone, not with family or rents	1	2.7 (1.05–6.67)	–	–	1	1.11 (0.56–2.5)	–	–
Unstable housing or homeless	1	1.7 (1.2–2.5)	–	–	1	1.47 (1–2.15)	–	–
Alcohol use or frequency	1	0.3 (0.12–0.77)	–	–	0	–	–	–

[a] Based on review of studies published in 2008–2017 in low- and middle-income countries. Prevalent associations are included for comparison.

[b] Specific categories and comparisons vary by study.

Table 11.
Variables with a significant pooled association with prevalent HIV among people who inject drugs[a]

Variable[b]	Association with prevalent HIV			
	N	Pooled estimate	Range of estimates	Consistent relationship?
Proximal				
Hepatitis C (laboratory-confirmed)	5	1.53 (1.08–1.98)	1.29–15.7	Yes
Sexually transmitted infection symptoms	1	8.8 (1.4–12.6)	–	–
Injected with a person living with HIV	1	1.86 (1.64–2.11)	–	–
Cleans needles or syringes before reuse	1	0.3 (0.1–0.6)	–	–
Distal				
Location of injection (hidden, at dealer's home, geographical area)[c]	2	4.43 (2.14–6.72)	3.41–5.1	–
Problematic relationship with family	1	1.57 (1.18–2.08)	–	–
Overdose	1	1.54 (1.02–2.32)	–	–

Variable[b]	Association with prevalent HIV			
	N	Pooled estimate	Range of estimates	Consistent relationship?
Years of drug use	1	1.25 (1.13–1.39)	–	–
Injected with self-used syringe	1	1.17 (1.07–1.28)	–	–
Number of drugs injected	1	1.17 (1.05–1.3)	–	–
Has access to syringes	1	0.5 (0.2–0.8)	–	–

[a] Based on review of studies published in 2008–2017 in low- and middle-income countries.

[b] Does not include variables shown in Table 10.

[c] Specific categories and comparisons vary by study.

Adolescent girls and young women

Search strategy:

- Databases searched: PubMed.
- Publication dates: 2008–2017.
- Population search terms: women, girl*, female*, adolescent*, youth* (in title).
- In addition, citations from Dellar RC et al. *(16)* were reviewed.

Study selection criteria:

- Data collected from a low- or middle-income country.
- Some part of data collection took place during 2008–2017.
- Included HIV laboratory test and behavioural data.
- Participants aged 15–24 years (age range broadened given data limitations[3]).
- Did not aim to recruit sex workers exclusively.
- If cross-sectional, probability sampling was used.
- Reported estimate of association between HIV infection and at least one variable, or reported HIV prevalence by subgroup so that associations could be calculated.

Search results:

- 843 papers returned by search.
- 8 papers meeting study selection criteria and providing associations from unique studies.
- 6 incidence/cohort studies and 2 prevalence/cross-sectional studies.

3 Few studies were found in this age range, so we broadened the criteria to include participants aged at most 30 years, or with a mean or median age of 24 years or younger if the age range could not be determined.

Meta-analysis:

- Included both statistically significant and non-significant associations.
- Tables list only risk factor categories with a statistically significant pooled meta-analysis estimate.

In the following tables:

- *N* = number of included associations. Papers could contribute up to one association per risk factor category per subgroup (e.g. study sites).
- Combined effect: pooled effect size (OR, HR) and 95% CI.
- Range of effects: minimum and maximum effect size across studies.
- Consistent relationship: all associations were in the same direction (positive or negative). This was not evaluated if $N < 1$.

Due to the limited evidence available in this population, and the more flexible criteria applied regarding age, the specific studies included in the review are listed below.

Table 12.
Studies included in the adolescent girls and young women review

Reference	Study population	Age of participants (years)	Data collection period	Country	Design
(17, 18)	Pregnant and postpartum women	19–27 IQR (median 22)	2011–2013	Kenya	Antenatal clinic cohort
(19)	Community cohort	15–24	1999–2008	Uganda	Cohort
(20)	Pregnant women	15–24 (mean 20)	2009	Brazil	Cross-sectional
(21)	Young women residing in Kasumi	16–32 (82% aged 15–24)	2007–2010	Kenya	Cross-sectional
(22)	Pregnant women and partners at large hospital	Median 25	2015–2016	Malawi	Case-control
(23)	Females	15–24	1998–2013	Zimbabwe	Cohort
(24)	Females	15–30	2003–2012	South Africa	Cohort

IQR, interquartile range.

Table 13.
Variables with significant pooled association with incident HIV among adolescent girls and young women[a]

Variable	Association with incident HIV				Association with prevalent HIV			
	N	Pooled estimate	Range of estimates	Consistent relationship?	N	Pooled estimate	Range of estimates	Consistent relationship?
Proximal								
Number of recent sex partners	3	2.09 (1.13–3.06)	1.82–2.27	Yes	2	1.13 (0–2.43)	0.68–4.95	No
Partner living with HIV or newly diagnosed with HIV	1	126.4 (33.8–472.2)	–	–	0	–	–	–
Partner of unknown HIV status	1	10.75 (3.13–36.94)	–	–	0	–	–	–
Sexually transmitted infection history or treatment	1	3.48 (1.31–9.27)	–	–	2	2.71 (0.41–5.01)	1.38–42.5	Yes
Number of lifetime sex partners	1	1.14 (0.99–1.33)	–	–	2	2.19 (0.54–3.83)	2–2.52	Yes
Unstable housing or homeless	1	1.7 (1.2–2.5)	–	–	1	1.47 (1–2.15)	–	–
Distal								
Yeast infection	1	2.78 (1.17–6.63)	–	–	0	–	–	–
Husband migrates or partners from outside community	1	1.77 (1.23–2.55)	–	–	0	–	–	–
Trading village versus rural	1	1.48 (1.04–2.11)	–	–	0	–	–	–
Shorter duration of current marriage	1	1.06 (1–1.13)	–	–	0	–	–	–

[a] Based on review of studies published in 2008–2017 in low- and middle-income countries. Prevalent associations are included for comparison.

Table 14.
Variables with significant pooled association with prevalent HIV among adolescent girls and young women[a,b]

Variable[b]	Association with prevalent HIV			
	N	Pooled estimate	Range of estimates	Consistent relationship?
Distal				
Number of lifetime marriages	1	9.1 (2.4–34.2)	–	–
Study site or area or residence[c]	1	2 (1.07–3.73)	–	–
Couples HIV testing and counselling	1	0.2 (0.04–0.9)	–	–

[a] Based on review of studies published in 2008–2017 in low- and middle-income countries.

[b] Does not include variables shown in Table 13.

[c] Specific categories and comparisons vary by study.

Table 15.
Risk factors for HIV among adolescent girls and young women identified by previous reviews published during 2008–2017

Variable	Reference	Time period reviewed	Geographical area of studies reviewed
Proximal			
Uncircumcised male partner	(25)	To September 2009	Kenya, Rwanda, Uganda, United Republic of Tanzania, Zimbabwe, 14 sites in eastern and southern Africa
Distal			
Intimate partner violence	(26)	To May 2013	United States of America (8 studies), South Africa (4 studies), eastern Africa (10 studies), India (3 studies), Brazil (1 study), multiple low-income countries (2 studies)
Experience of orphanhood	(27)	1980 to June 2009	Primarily sub-Saharan Africa

Limitations

Variables identified by the reviews are limited by the time period of literature examined:

- Men who have sex with men, female sex workers: past 5 years.
- Transgender women, people who inject drugs, adolescent girls and young women: past 10 years.

For adolescent girls and young women, results should be interpreted as suggestive because the included studies were among a broader age range than 15–24 years.

For men who have sex with men, female sex workers, and transgender women, the review was limited to statistically significant associations, which may have led to overestimation of the combined effect sizes.

The estimated CIs for the pooled effects are likely to be conservative (too wide) in all populations. CIs are derived from the meta-analysis weights, which are calculated from standard errors of the reported effect sizes. In most cases, these standard errors had to be approximated based on the reported CIs. Consequently, many of the non-significant combined effects may in fact be significant.

Evidence for known risk factors such as unprotected anal intercourse and injection drug use was mixed. This is probably due to most evidence coming from cross-sectional rather than more robust cohort studies.

Consistency could not be evaluated for several risk factor categories where evidence was limited to a single study.

Example of cohort analysis to identify risk factors among men who have sex with men

Objectives

This section shows how to conduct analysis with cohort data to identify risk factors associated with incident HIV infection among men who have sex with men. The risk factor findings are used to construct risk definitions that could be used as part of the risk factors approach to defining substantial risk.

Although these specific data did not lead to a risk definition with an acceptable level of predictive performance, the section is included here to illustrate the analysis process.

Data source

Data for the cohort analysis were all visit records from individuals presenting to Estrategia de Vigilancia Centinela de las Infecciones de Transmisión Sexual (VICITS) clinics in Guatemala from 2014 to 2017 and in Nicaragua from 2011 to 2017 (Table 16).

During these periods, clinics were located in Guatemala City (three clinics) and three other cities in Guatemala, and in Managua (three clinics) and Chinandega in Nicaragua.

VICITS is a sentinel surveillance strategy with a combined prevention component, including diagnosis and treatment of other sexually transmitted infections. VICITS clinics provide services to men who have sex with men, female sex workers, and transgender women throughout Central America, implemented and financed by ministries of health, the Centers for Disease Control and Prevention, and the United States President's Emergency Plan for AIDS Relief.

We draw on data from clinics in Guatemala and Nicaragua.

On their first visit, people receive testing for HIV and sexually transmitted infections, a medical examination and syndromic management. Follow-up HIV tests are recommended every six months or every three months for people who believe they are at risk.

At the first visit and then annually, a standardized interview on demographics, risk behaviours, history of sexually transmitted infection, and use of prevention is carried out by HIV counsellors or medical staff.

In addition to follow-up visits, clients can present at any time for testing for sexually transmitted infections.

For the time period available for this analysis, biological tests in the data were limited to HIV and syphilis. Syphilis testing was by Venereal Disease Research Laboratory (VDRL) or rapid plasma regain (RPR) tests.

Behavioural data collected include:

- Sexual orientation and transgender identity.
- Sex work in the past 12 months.
- Months or years engaged in sex work.
- Number of male sex clients in the past week.
- Whether sex work is the main income source.
- Educational attainment.
- Ethnic group.
- Knowledge of HIV status and testing history.
- Sexually transmitted infection diagnoses in the past 12 months.
- Sexually transmitted infection signs and symptoms in the past 12 months.
- Frequency of alcohol consumption in the past 30 days.
- Drug use (lifetime and past 30 days).
- Condom use at last anal sex with a male partner.
- Condom use at last sex with a stable partner, a casual partner and a sex client.
- Frequency of condom use in the past 30 days by partner type.

Records for individuals were linked across visits and across clinics using a code with characters representing initials, date of birth, department and city of birth, and a sequential number. The sequential number allows the code to differentiate between clients with the same demographic information.

Data preparation and measures

The first visit record during the period for which data were available was the starting point for defining baseline characteristics. Because the first record of the period was not necessarily the person's first consultation, some records did not have data on all variables of interest.

We drew on subsequent records to complete the baseline data, within a defined time period:

- Within six months of the first visit for sexually transmitted infection and syndromic sexually transmitted infection results.
- Within six months for risk behaviours.
- No time limit for sexual orientation and gender identity.

If the HIV test result was missing from the first record and recorded as HIV-negative on a subsequent visit, we assigned HIV-negative at baseline.

Where there were multiple records for the same person on the same date, we compared data items; where they differed, we drew data from the record with the latest time stamp.

Data were reviewed for consistency of patterns over time. Sex, date of birth and transgender woman identity were consistent across records.

A total of 67 people (61 men who have sex with men, 1 transgender woman, 5 female sex workers) had a positive HIV test result followed by a negative HIV test result; these individuals were excluded from the analysis.

We excluded people with no HIV result at any consultation and HIV-negative people with no follow-up during the period.

Transgender woman status was defined as male sex and self-identifying as transsexual, transgender or transvestite. All other participants of male sex were classified as men who have sex with men. All females were classified as female sex workers.

Receptive sex in the past 30 days, a key risk factor based on findings from the literature review, was missing for 19% (*N* = 191) of men who have sex with men HIV-negative at baseline in Nicaragua and 5% (*N* = 148) in Guatemala. To avoid disregarding these records in their entirety, we inferred the receptive sex variable using multiple imputation by chain equations. This is a stochastic technique that led to 20 different datasets with different probabilistic realizations of the receptive variable. Increasing the number of imputed datasets to 50 did not alter the findings.

Statistical analysis

Analysis was limited to VICITS clients who were HIV-negative at first visit during the time period. For the purposes of survival analysis, we defined analysis time as the difference in visit dates between the first observed visit and the first positive HIV result, or, if none, to the last HIV test result, expressed in years.

We summarized the number of HIV seroconversions and estimated HIV incidence rates per 100 person-years by city, defined as the city of the first visit. We examined Kaplan–Meier survivor curves and tested for differences in the survivor function by city and country using log-rank and Wilcoxon tests to guide decisions regarding pooling data across sites for risk factor analysis. When survivor curves appeared to differ across sites, we tested whether differences persisted after adjusting for variables to be included in the analysis.

The remainder of the analysis was limited to men who have sex with men due to limited sample size and number of seroconversions among transgender women and female sex workers.

We examined the prevalence of each of the behavioural and sexually transmitted infection predictor variables at baseline across all sites and separately for the two sites with the largest sample (Guatemala City, Managua) for binomial variables. We examined medians and IQRs for discrete variables (age, number of sex clients).

Predictive models

We used Cox proportional hazards regression to build the predictive multivariable models. In the first stage, bivariate associations were estimated for Guatemala City, Managua and pooling all sites. Bivariate analysis was conducted on the data before imputation. To determine the best specification for discrete variables, we examined Lowess curves versus HIV and constructed splines with knots fit to any inflection points we observed visually.

We also examined categorical groups at these points and at quartiles. We continued to consider any specifications that were associated with the HIV outcome ($P \leq 0.10$) and improved Harrell's concordance statistic.

As bivariate findings for Guatemala City and Managua appeared similar in direction and magnitude, and due to limited power in Managua, models were constructed using the data pooled across all Guatemala and Nicaragua sites. Survivor curves were not significantly different across sites overall ($P = 0.071$ log-rank, $P = 0.582$ Wilcoxon) or between Guatemala City and Managua specifically ($P = 0.660$ and $P = 0.677$).

Following bivariate analysis, we estimated an initial model including variables with bivariate association of $P \leq 0.10$ and then tested remaining variables in turn, in order of strongest unadjusted hazards ratio, in case they became significant conditional on other covariates.

We tested 2-way interactions between the main effects and retained those significant at $P \leq 0.05$. Interactions with clinics were significant, but we did not retain them since they did not substantially change the other coefficient estimates and because we aimed to build a model applicable in other settings. Alternative specifications were compared based on the concordance statistic.

We evaluated the sensitivity and specificity of different thresholds of the resulting model's risk score (regression equation) to predict HIV acquisition within two years of the first visit. We examined potential thresholds of the risk score corresponding to 1% increments between 1% and 10% predicted probabilities of acquiring HIV by 2 years.

The sensitivity and specificity calculations excluded people who were censored (without a HIV test result at two or more years). As an additional measure of predictive performance, we calculated the area under the receiver–operator curve (AUC) from logistic regression of the same regression equation.

Given the imputation, we had to calculate and average some of the above statistics on each of the imputed datasets ($N = 20$), including regression estimates, concordance statistics, predicted probabilities, sensitivity, specificity and AUC. Analysis was conducted in Stata v15.1.

Incidence of HIV at VICITS sites

Among 3722 men who have sex with men who were HIV-negative at first visit, average follow-up was 1.4 years and there were 167 seroconversions (126 in Guatemala City, 41 Managua). HIV incidence was 3.3/100 person-years overall and was not statistically different across sites.

Among 265 transgender women who were HIV-negative at first visit, average follow-up was 1.8 years and there were 13 seroconversions (11 in Managua). HIV incidence was 2.8/100 person-years and was not statistically different from incidence among men who have sex with men.

Among 1814 female sex workers who were HIV-negative at first visit, average follow-up was 1.4 years. HIV incidence was considerably lower, at 0.2/100 person-years.

There were 2 or more years of follow-up for 983 (26%) men who have sex with men, 81 (31%) transgender women and 507 (28%) female sex workers.

The survivor curves in Figure 1 suggest a higher HIV incidence among men who have sex with men in Guatemala City compared with Managua. Log-rank and Wilcoxon tests ($P = 0.660$ and $P = 0.677$), however, did not indicate a difference between the survival curves. There was weak evidence of differences across all sites ($P = 0.071$ and $P = 0.582$). Figure 1 suggests a precipitous drop in the proportion of transgender women who are still HIV-negative at around 5 years; this appears to be due to the small sample size (only 15 transgender women were followed for 5 years or more).

Table 16.
HIV incidence among men who have sex with men attending VICITS clinics

Clinic location	N	Person–years	HIV seroconversions	Incidence rate (/100 person–years) (95% CI)
Guatemala (2014–2017)				
Guatemala City	2491	3479	126	3.6 (3.0–4.3)
Xela and Cuatpeque	257	299	7	2.3 (1.1–4.9)
Escuintla and Tecún Umán	78	64	2	3.1 (0.8–12.5)
Puerto Barrios and Flores	7	11	1	9.2 (1.3–65.0)
Nicaragua (2011–2017)				
Managua	862	1447	41	2.8 (2.1–3.8)
Chinandega	27	1.5	0	0
All sites	3722	5301	177	3.3 (2.9–3.9)

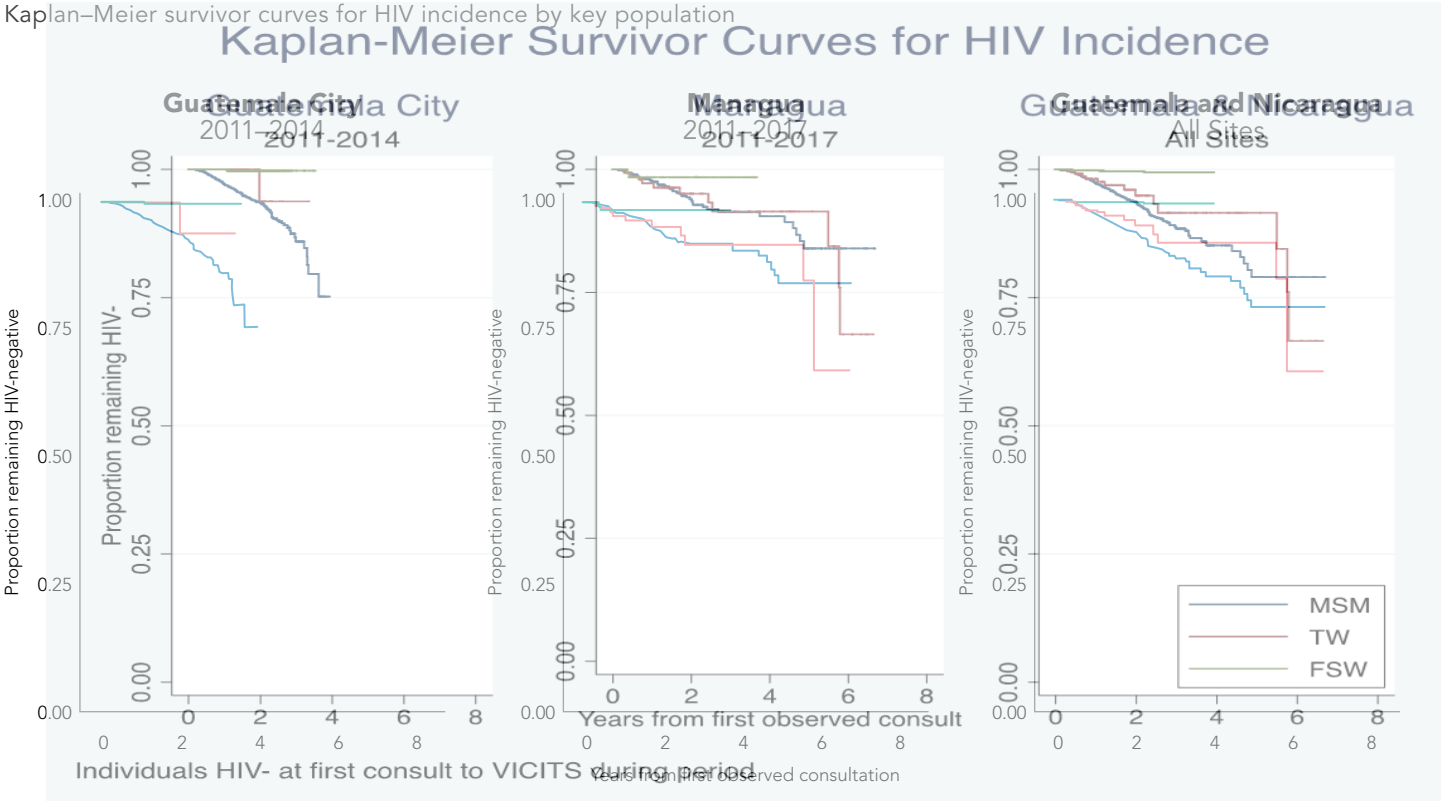
Table 17.
HIV incidence among transgender women attending VICITS clinics

Clinic location	N	Person–years	HIV seroconversions	Incidence rate (/100 person–years) (95% CI)
Guatemala (2014–2017)				
Guatemala City	54	77	1	1.3 (0.2–9.2)
Xela and Cuatepeque	6	7	0	0
Escuintla and Tecún Umán	32	33	1	3.0 (0.4–21.3)
Puerto Barrios and Flores	1	3	0	0
Nicaragua (2011–2017)				
Managua	158	347	11	3.2 (1.8–5.7)
Chinandega	14	1	0	0
All sites	265	468	13	2.8 (1.6–4.8)

Table 17.
HIV incidence among female sex workers attending VICITS clinics

Clinic location	N	Person–years	HIV seroconversions	Incidence rate (/100 person–years) (95% CI)
Guatemala (2014–2017)				
Guatemala City	767	1000	1	0.1 (0.0–0.7)
Xela and Cuatepeque	575	1046	1	0.1 (0.0–0.7)
Escuintla and Tecún Umán	132	110	1	0.9 (0.1–6.4)
Puerto Barrios and Flores	60	49	0	0
Nicaragua (2011–2017)				
Managua	178	241	2	0.8 (0.2–3.3)
Chinandega	102	170	0	0
All sites	1814	2616	5	0.2 (0.1–0.5)

Figure 1.
Kaplan–Meier survivor curves for HIV incidence by key population



Individuals HIV-negative at first VICITS clinic consultation during period

- Men who have sex with men
- Transgender women
- Female sex workers

Prevalence of risk variables and bivariate associations with HIV

About two-thirds of men who have sex with men in the sample engaged in any sexual intercourse or anal sex, 35% had sex with a casual partner, and 38% had unprotected intercourse (oral or anal) with a male partner over the past 30 days (Table 19).

The prevalence of unprotected anal intercourse at last anal sex was 49% in Guatemala and 81% in Nicaragua.

About 10% of men who have sex with men used drugs and 1% injected drugs in the past 30 days.

About 7% said they engaged in sex work in the past 12 months.

Eight per cent tested positive for syphilis, 6% met criteria for active syphilis, and 12% were diagnosed at a VICITS clinic with a sexually transmitted infection syndrome within the baseline period. Levels of sexually transmitted infection were higher in Guatemala City than in Managua ($P < 0.001$ by chi-squared tests).

In bivariate analysis, there were associations of incident HIV with syphilis, active syphilis, sexually transmitted infection syndrome, and receptive anal sex in the past 30 days, and a marginal association with having a recent casual partner (Table 20).

Variables for self-reported history of sexually transmitted infection in the past year, sex work in the past year, and drug use in the past 30 days did not reach significance despite positive estimated associations.

The relationship with age differed across sites. HIV was strongly associated with younger age in Guatemala City and with older age in the pooled data.

Multivariable models

The best multivariate model obtained included syphilis, sexually transmitted infection syndrome, receptive anal sex, age, and interaction between age and Nicaragua (Model 1 in Table 21). Adjusted ORs for the three behavioural terms were between 1.5 and 2.2 and were not statistically different from each other.

We fit a second model removing syndromic sexually transmitted infection, as it was marginally significant and data to assess syndromic sexually transmitted infection are generally not available in biobehavioural surveillance (BBS).

Predictive power of both models was poor. Model 1 reached a concordance of 60.8% and an AUC of 60.4% (where 50% indicates no predictive value). The estimated sensitivity and specificity at the risk score thresholds examined are shown in Table 22.

For Model 1, we considered medium- and high-risk thresholds set at predicted probabilities of ≥ 5% (sensitivity 98%, specificity 14%) and 9% (sensitivity 60%, specificity 55%), respectively. For Model 1, sensitivity drops more rapidly, so we considered thresholds instead at 5% (77% sensitivity, 27% specificity) and 7% (61% sensitivity, 45% specificity).

To simplify application of the risk score, we divided by the maximum coefficient, multiplied by 100 and rounded (Table 23).[4]

The interaction with age leads to different risk definitions in the two countries. In Guatemala, men who have sex with men would be considered as high risk by Model 1

4 Coefficients were simplified by dividing by the maximum coefficient, multiplying by 100 and rounding. Thresholds are at predicted probabilities 5% and 9% for Model 1 and at 5% and 7% for Model 2 in Table 23.

if they are aged 25 years or younger, or if aged 34 years or younger with a syndromic sexually transmitted infection, or aged 37 years or younger and report recent receptive anal sex.

In Nicaragua, Model 1 suggests risk increases with age and would classify as high risk age 55 years or over in the presence of any of the behavioural risk factors.

Predictions should not be considered beyond the 19–55 years age range in the data.

Investigating the age effect in Nicaragua further, we find it disappears after we limit the model to men who have sex with men aged under 50 years. This is because HIV prevalence was 14% among 171 subjects aged 50 years or over in Nicaragua and 0% among 68 subjects aged 50 years or over in Guatemala, leading to the interaction.

If we limit the model to age 49 years or younger, the interaction is non-significant ($P = 0.529$) and the main effect of a negative association between age and HIV persists with a similar coefficient (−0.04), as in Model 1. This suggests that among HIV-negative men who have sex with men aged under 50 years in both countries, HIV risk declines with age.

Table 19.
Characteristics of HIV-negative men who have sex with men at first observed consultation in VICITS clinics

Variable	Guatemala City, 2011–2014			Managua, 2011–2017			Guatemala and Nicaragua (all sites)		
	N	n	%	N	n	%	N	n	%
Syphilis	2366	239	10.1	852	44	5.2	3568	298	8.4
Active syphilis	2365	171	7.2	852	29	3.4	3567	207	5.8
Sexually transmitted infection syndrome	2491	380	15.3	862	29	3.4	3722	437	11.7
Sexually transmitted infection history (12 months)	2491	278	11.2	862	20	2.3	3722	327	8.8
Any intercourse (30 days)	2491	1795	72.1	862	390	45.2	3722	2517	67.6
Anal sex (30 days)	2491	1658	66.6	862	383	44.4	3722	2359	63.4
Receptive anal sex (30 days)	2370	1050	44.3	696	148	21.3	3416	1409	41.2
Casual partner (30 days)	2491	884	35.5	862	230	26.7	3722	1299	34.9
Condomless sex in past 30 days[a] with									
Stable male partner	2491	518	20.8	862	135	15.7	3722	753	20.2
Casual male partner	2491	471	18.9	862	179	20.8	3722	738	19.8
Sex client	2491	24	1.0	862	83	9.6	3722	114	3.1
Any of above	2491	935	37.5	862	287	33.3	3722	1394	37.5
Unprotected anal intercourse at last anal sex[b]	2491	1229	49.3	862	698	81.0	3722	2080	55.9
Condomless sex at last sex in past 30 days with									
Stable male partner	2491	415	16.7	862	99	11.5	3722	588	15.8
Casual male partner	2491	333	13.4	862	101	11.7	3722	495	13.3
Sex client	2491	19	0.8	862	55	6.4	3722	79	2.1
Any of above	2491	709	28.5	862	203	23.5	3722	1040	27.9
Sex work (12 months)	2491	131	5.3	862	90	10.4	3722	244	6.6
Drug use (30 days)	2491	319	12.8	862	42	4.9	3722	381	10.2
Injection drug use (30 days)	2491	33	1.3	862	8	0.9	3722	43	1.2
Number of sex clients (30 days), median (IQR)	2488	0	0-0	859	0	0-0	3714	0	0-0
Age, median (IQR)	2491	28	24-33	862	27	23-34	3722	27	24-33

[a] Sometimes/always versus always used a condom.

[b] No reference period.

Table 20.
Bivariate associations with incident HIV among men who have sex with men attending VICITS

Variable	Guatemala City, 2011–2014		Managua, 2011–2017		Guatemala and Nicaragua (all sites)	
	HIV (%)	HR (95% CI)	HIV (%)	HR (95% CI)	HIV (%)	HR (95% CI)
Syphilis	7.9	3.1 (1.9–5.1)[a]	4.5	1.7 (0.4–6.9)	7.4	2.9 (1.8–4.6)[a]
Active syphilis	6.4	2.5 (1.4–4.7)[a]	0.0	0	5.8	2.2 (1.2–4.0)[a]
Sexually transmitted infection syndrome	7.1	2.3 (1.5–3.5)[a]	3.4	2.4 (0.3–17.6)	6.4	2.2 (1.4–3.3)[a]
Sexually transmitted infection history (12 months)	5.8	1.3 (0.8–2.2)	5.0	2.8 (0.4–20.7)	5.5	1.3 (0.8–2.2)
Any intercourse (30 days)	4.6	1.1 (0.7–1.5)	3.8	1.8 (0.9–3.6)	4.3	1.3 (0.9–1.8)
Anal sex (30 days)	4.7	1.1 (0.8–1.6)	3.7	1.6 (0.8–3.2)	4.3	1.3 (0.9–1.8)
Receptive anal sex (30 days)	5.9	2.0 (1.4–2.9)[a]	3.4	2.8 (1.0–7.8)	5.3	2.2 (1.5–3.0)[a]
Casual partner (30 days)	5.1	1.3 (0.9–1.9)	3.5	1.9 (0.9–4.4)	4.5	1.4 (1.0–2.0)[b]
Condomless sex in past 30 days[c] with						
Stable male partner	3.3	0.7 (0.4–1.1)	3.7	1.2 (0.5–3.2)	3.5	0.8 (0.5–1.3)
Casual male partner	5.1	1.3 (0.9–2.1)	1.7	0.8 (0.2–2.7)	3.8	1.2 (0.8–1.8)
Sex client	12.5	2.8 (0.9–8.8)	3.6	1.1 (0.3–3.5)	6.1	1.7 (0.8–3.8)
Any of above	4.2	0.9 (0.6–1.4)	2.8	1.1 (0.5–2.4)	3.7	1.0 (0.7–1.4)
Unprotected anal intercourse at last anal sex[d]	5.5	1.0 (0.7–1.4)	5.2	0.8 (0.3–2.1)	5.1	0.9 (0.7–1.3)
Condomless sex at last sex in past 30 days with						
Stable male partner	3.6	0.7 (0.4–1.3)	5.1	1.7 (0.6–4.3)	3.9	0.9 (0.6–1.5)
Casual male partner	5.1	1.3 (0.8–2.2)	2.0	1.0 (0.2–4.4)	4.0	1.3 (0.8–2.0)
Sex client	10.5	2.6 (0.6–10.4)	3.6	1.0 (0.2–4.1)	6.3	1.6 (0.6–4.2)
Any of above	4.4	1.0 (0.7–1.5)	3.4	1.3 (0.6–3.0)	4.1	1.1 (0.8–1.6)
Sex work (12 months)	6.1	1.3 (0.7–2.8)	4.4	1.4 (0.5–3.9)	5.3	1.3 (0.8–2.4)
Drug use (30 days)	4.1	0.9 (0.5–1.6)	2.4	0.9 (0.1–6.9)	3.7	0.8 (0.5–1.5)
Injection drug use (30 days)	9.1	2.0 (0.6–6.4)	12.5	5.7 (0.8–42.2)	9.3	2.4 (0.9–6.4)
Number of sex clients (30 days)	–	0.9 (0.7–1.1)	–	1.0 (1.0–1.1)	–	1.0 (0.9–1.1)
Age	–	0.9 (0.9–1.0)[a]	–	1.0 (1.0–1.1)	–	1.0 (0.9–1.0)[a]

[a] Strong statistical association: $P \leq 0.01$.

[b] $P \leq 0.05$.

[c] Sometimes/always versus always used a condom.

[d] No reference period.

Table 21.
Multivariable models of incident HIV among men who have sex with men attending VICITS clinics (all sites, *N* = 3948)

Variable	Model 1		Model 2	
	AHR (95% CI)	P	AHR (95% CI)	P
Syphilis	2.17 (1.28–3.7)	0.004	2.75 (1.78–4.4)	<0.001
Sexually transmitted infection syndrome	1.53 (0.99–2.62)	0.084	–	–
Receptive anal sex (30 days)	1.81 (1.02–2.92)	< 0.001	1.83 (0.99–2.8)	0.057
Age (centred)	0.95 (0.92–0.98)	0.003	0.95 (0.92–0.98)	0.001
Age (centred) × Nicaragua	1.06 (1.01–1.11)	0.018	1.06 (1.01–1.12)	0.015

AHR, adjusted hazard ratio.

Table 22.
Predictive performance of three models at various thresholds of predicted probability of HIV at two years in men who have sex with men attending VICITS clinics in Guatemala and Nicaragua

Variable	Model 1		Model 2	
	AHR (95% CI)	P	AHR (95% CI)	P
Predicted P(HIV) at 2 years				
≥ 5%	98	14	77	27
≥ 6%	82	26	61	39
≥ 7%	79	36	61	45
≥ 8%	66	42	26	65
≥ 9%	60	55	15	78
≥ 10%	19	70	10	94
Harrell's concordance		0.608		0.628
Area under curve		0.604		0.625

Table 23.
Simplified risk scores for VICITS Models 1 and 2

Variable	Model 1		Model 2	
	Coefficient[a]	Simplified	Coefficient	Simplified
Syphilis	0.777	100	1.011	100
Receptive anal sex (30 days)	0.428	76	–	60
Sexually transmitted infection syndrome	0.594	55	0.602	–
Centred age (age – 29 years) interaction with:				
Guatemala	−0.046	−6	−0.048	−5
Nicaragua	0.013	2	0.014	1
Thresholds[b]				
Medium risk	−0.068	−9	0.128	13
High risk	0.539	69	0.471	47

[a] Coefficients were simplified by dividing by the maximum coefficient, multiplying by 100 and rounding.

[b] Thresholds are at predicted probabilities 5% and 9% for Model 1 and 5% and 7% for Model 2.

Limitations

Key limitations of the survival analysis included the following:

- There was a lack of power to build models for sites outside Guatemala City and for key populations other than men who have sex with men due to limited sample size and seroconversions.
- The risk definitions contain few variables, due to the limited behavioural variables in the available data.
- Known proximal risk factors, including unprotected anal intercourse and number of partners, could not be constructed from the data. The available condom use and injection drug use variables were not predictive of incident HIV and therefore could not be included in the models and risk definitions.
- The predictive performance of the models obtained was relatively poor.
- The behavioural data are self-reported and may be subject to recall bias and social desirability bias.

Using population-based HIV impact assessment surveys to identify risk factors for adolescent girls and young women

Objectives

Population-based HIV impact assessment (PHIA) surveys are representative of the general female (and male) population aged 15 years and over and include extensive information related to HIV knowledge, attitudes, risk behaviours, making them a unique resource to identify risk factors for prevalent HIV infection and to estimate the proportion of the population that meets specified HIV risk criteria.[5]

At the time of writing, PHIA surveys had been conducted or were ongoing in 14 countries in sub-Saharan Africa: Cameroon, Côte D'Ivoire, Eswatini, Ethiopia, Kenya, Lesotho, Malawi, Mozambique, Namibia, Rwanda, Uganda, the United Republic of Tanzania, Zambia and Zimbabwe.

This section illustrates using PHIA surveys to:

- Evaluate whether a HIV risk index developed by Pintye *et al. (17)* for use among pregnant and postpartum women can be constructed from PHIA data.
- Identify risk factors for prevalent HIV infection among adolescent girls and young women.
- Assess the predictive performance of multivariable regression models of prevalent HIV infection to identify combinations of risk factors that could be used as a definition of substantial risk for PrEP target-setting.

Data sources

PHIA surveys are cross-sectional, household-based, nationally representative surveys of adults and adolescents aged 15 years and over. Participation is voluntary and requires informed consent. Computer-assisted personal interviews are administered by trained staff using tablets. The interviews collect data on themes including household characteristics, sociodemographics, and HIV-related risk factors.

Blood specimens are collected for laboratory assessment of HIV and, in some surveys, other sexually transmitted infections. Participants who test positive for HIV or other sexually transmitted infections are referred to care and treatment services.

This analysis draws on adult and biological datasets from the Malawi *(2015–2016)* and Zambia *(2016)* PHIA surveys.

Can Pintye's HIV risk index be assessed with PHIA?

An HIV risk index to assess the need for PrEP among females accessing antenatal services was developed by Pintye *et al. (17)* based on data from a prospective study

5 For more information on PHIA and to request access to survey datasets, see https://phia.icap.columbia.edu/.

of incident HIV among 1304 pregnant and postpartum women in western Kenya conducted between 2011 and 2014. The risk score is calculated from five measures often assessed in antenatal clinics:

- Having a male partner of unknown HIV status.
- Number of lifetime sexual partners.
- Syphilis infection.
- Bacterial vaginosis.
- Vaginal candidiasis.

The authors also developed a simplified risk index that excluded bacterial vaginosis and vaginal candidiasis.

The complete index had a sensitivity of 64% (correctly predicted HIV acquisition of 64% of the women who eventually acquired HIV), and the simplified index had a sensitivity of 54%. The specificity of the two indices was not reported.

Although originally developed for antenatal settings, Pintye's risk score might also be used to assess HIV risk among adolescent girls and young women if it could be measured by a representative survey such as PHIA.

Review of PHIA data for Malawi *(2015–2016)* and Zambia *(2016)* found that two of the measures (bacterial vaginosis, vaginal candidiasis) were not available in either survey. Two measures (number of lifetime sexual partners, syphilis test results) are available in Zambia but not Malawi. One measure (male partner's HIV status) is available in both PHIA surveys.

Currently PHIA cannot be used to assess Pintye's complete risk score. Some PHIA surveys, such as Zambia *(2016)*, can be used to assess the simplified score, but the sensitivity to detect incident HIV was relatively low (54%) in the Kenya cohort.

Future PHIA surveys would be able to assess the complete index by collecting data on lifetime sexual partners, syphilis, bacterial vaginosis and vaginal candidiasis.

Data preparation and measures

Records for individuals were linked across the adult and biological datasets following guidance in the PHIA Data Use Manual *(28)*.

Analysis was limited to females aged 15–24 years with a blood test result for HIV who reported ever having had sexual intercourse. Previous sexual experience was ascertained from the question item on age at first sex (refusal code 96).

Among the 5555 sexually experienced female participants aged 15–24 years with a blood test result in the Malawi and Zambia surveys, 377 tested positive for HIV. Of these, 184 (48.8%) reported being aware of their HIV-positive status and were excluded from the analysis, since awareness of living with HIV may have led to changes in behaviour and biased the analysis.

Candidate risk factors were derived from the literature and limited to those that could be assessed with the Malawi and Zambia PHIAs. Related factors were also considered. Predictors were limited to those available in both surveys, as we sought to combine the two datasets to improve statistical power. Candidate predictors are listed in Table 24.

Area-level HIV prevalence among adult males (ALPM) was considered as a candidate risk factor and was estimated at the zone (Malawi) and province (Zambia) levels. The ALPM estimates were weighted following PHIA guidance (described later).

Skip patterns were reviewed and candidate predictors were recoded as needed. For example, predictors related to sexual activity with a male partner of positive or unknown HIV status were coded 0 (no such activity) if none of the past three partners was a male of positive or unknown HIV status.

Similarly, predictors related to sexual activity with partners who had other wives were coded 0 if the respondent did not have a male spouse or live-in partner, or if that partner did not have other wives.

Predictors based on multiple question items were coded as incomplete (missing) and excluded from the analysis if any one of the required items was missing due to non-response, unless it was clear from the non-missing items that the respondent satisfied the respective risk factor.

Overall, the candidate risk variables had low levels of non-response and were complete for 96.5–100% of respondents. There were lower levels of completeness for experience of forced sex (69.2%), engaging in sex work (82% complete in the Malawi data), and having an uncircumcised partner in the past year (92.7%). Sex work and forced sex were therefore not evaluated.

Variants of the age gap variable (number of years separating the age of the respondent from the ages of their last three sexual partners in the past 12 months) were also examined. These variables included:

- Maximum age difference with last three partners.
- Sum of age differences across last three partners.
- Natural logarithm and square root transforms of the above.

All age gap predictors were coded 0 for respondents who were the same age as or older than the respective partner.

Table 24.
Candidate risk variables available in PHIA data

Candidate risk factor	Construction from PHIA data
Age	
Number of sexual partners in past 12 months	(Question item phrased similarly)
Male sexual partner of positive or unknown HIV status	Based on items on last 3 sexual partners HIV status was considered unknown if response to item on partner's HIV status was not "told me" or "tested together"
HIV-positive male sexual partner	As above
Male partner > 5 or > 10 years older	Based on items on last 3 sexual partners
No condom used at last sex with male partner	For Malawi, based on items on last 3 sexual partners For Zambia, based on general item on condom use at last sex These were the most compatible items across the 2 surveys
No condom used at last sex with male partner of positive or unknown HIV status	Based on items on last 3 sexual partners
No condom used at last sex with male partner living with HIV-positive	Based on items on last 3 sexual partners
Number of last 3 sex partners who were male, of positive or unknown HIV status, and with whom no condom was used at last sex	Based on items on last 3 sexual partners
Has married or live-in male partner who was staying elsewhere at the time of the survey interview	Based on items related to ever being married, current marital status, and whether primary partner was living elsewhere
Has married or live-in male partner who has other wives or lives with other women	Based on items related to ever being married, current marital status, and whether husband has other wives
No condom used at last sex with married or live-in partner who has other wives or live-in partners	Based on above and whether condom was used at last sex with married or live-in partner
Uncircumcised male partner in past 12 months	Based on items on last 3 sexual partners
Has non-cohabitating male partner	Based on items on last 3 sexual partners
Non-monogamous with a non-cohabitating male partner in past 12 months	If number of partners in past 12 months was ≥ 2 and any of past 3 sexual partners was non-cohabitating male
Engaged in transactional sex in past 12 months	If respondent was with any of the last 3 male partners for material support or help
Ever forced to have sex	(Question item phrased similarly)
Forced to have sex in past 12 months	(Question item phrased similarly)
Out of school	Of school age, not enrolled in school, and not completed secondary school[a]
Sexually transmitted infection diagnosis in past 12 months	Told by doctor, clinical officer or nurse in past 12 months that respondent had sexually transmitted disease
Sexually transmitted infection symptoms in past 12 months	Based on item on having experienced "abnormal vaginal discharge or pelvic pain" The Zambia questionnaire asked additionally about experience of a vaginal ulcer or sore, but this item was not included as it was not available in the Malawi data
Area-level HIV prevalence among adult males	Weighted estimates calculated from PHIA data at the zone (Malawi) and province (Zambia) levels

[a] For Malawi, aged ≤ 17 years; for Zambia, aged ≤ 18 years; in both countries, had not completed grade 12 (last grade of secondary school).

Statistical analysis

Overall approach

We aimed to develop multivariable models to identify independent risk factors and assess their ability to predict HIV infection based on the AUC measure, sensitivity and specificity. Goodness of fit of the models was also assessed based on the Bayesian information criterion (BIC).

Initially, models predicting HIV as a function of behavioural measures were fit to the Malawi PHIA (as the derivation sample) and evaluated on the Zambia PHIA (as the validation sample). The predictive performance of the best models resulting from this approach was poor, however, with an AUC of 0.71, but with no risk score threshold yielding both sensitivity and specificity of at least 70%.

To improve performance, we tested ALPM as a potential predictor, and interactions between ALPM and other predictors. Additionally, the derivation sample was redefined as a 75% random sample of the combined Malawi and Zambia PHIA data, instead of Malawi data alone, to improve statistical power and variability. The remaining 25% of survey participants comprised the validation sample.

The remainder of this section describes this approach.

Derivation and validation samples

The Malawi and Zambia PHIA adult and biological data were combined into a single dataset. Participants not meeting the eligibility criteria for analysis (sexually experienced females aged 15–24 years with a blood test result for HIV and who were not already HIV-positive and aware of their HIV-positive status) were removed from the dataset.

A simple random sample of 75% of eligible HIV-positive participants and, separately, of 75% of eligible HIV-negative participants, was drawn and designated as the derivation sample. Remaining eligible participants comprised the validation sample.

The derivation sample contained participants residing in all 7 zones in Malawi and all 10 provinces in Zambia, with 114–383 participants per zone or province.

Additional characteristics of the samples are shown in Table 25.

Table 25.
Characteristics of derivation and validation samples

Characteristic	Derivation sample	Validation sample
Number of participants	4029 (75.0%)	1342 (25.0%)
HIV-negative	3884 (75.0%)	1294 (25.0%)
HIV-positive	145 (75.1%)	48 (24.9%)
From Malawi PHIA	1881 (52.5%)	1699 (47.5%)
From Zambia PHIA	2148 (51.6%)	2017 (48.4%)

Bivariate models

Given the importance of HIV prevalence in the local setting as a determinant of individual-level HIV risk, we required each candidate predictor to be independently associated with HIV infection after controlling for ALPM to be considered for inclusion in the multivariable model. ALPM was strongly associated with individual HIV status in the derivation sample (OR = 1.16, $P < 0.001$). We assessed the bivariate association of each candidate predictor with HIV infection by fitting a logistic regression model controlling for ALPM, with ALPM specified as a continuous variable.

Other specifications of ALPM were examined, including categorical (with cut-off points identified by visual examination of Lowess curves versus HIV) and transformations including square root, natural logarithm, quadratic, cube and other polynomials. None, however, improved statistical significance and goodness of fit, as assessed by the AUC in bivariate models.

We determined whether the effect size of the predictor varied significantly with ALPM by estimating a second logistic regression model, which included interaction between the predictor and ALPM. Candidate predictors and interactions that were marginally significant ($P \leq 0.20$) in either of the models were considered for inclusion in the multivariable model.

We determined the best categorical specification of discrete variables (age and number of partners) by examining Lowess curves versus HIV. We then evaluated any promising cut-off points using a bivariate logistic model (as before, controlling for ALPM). This led to consideration of age both as continuous and 20 years and older, and number of partners as 1 or more and 2 or more partners, in the multivariable model.

Finally, we assessed the bivariate association of fixed effects for zones and provinces with HIV, for consideration as possible controls in addition to ALPM.

Multivariable models

We developed a multivariable model by backward selection. The model initially included ALPM and all predictors and interactions with ALPM that had a bivariate association with HIV of $P \leq 0.20$. The sample for multivariable analysis was restricted to participants with valid responses to all such predictors. The forced sex predictor was not considered in multivariable analysis due to high non-response of 30.8% in the derivation sample.

Before backward selection, we examined variance inflation factors to identify excessive multicollinearity among candidate predictors (variance inflation factor ≥ 10). There was multicollinearity between the number of recent partners and the number of recent condomless sex partners, as well as among different specifications of the same predictor (age gap variations). In these cases, we included the predictor that yielded the greatest AUC and BIC.

Predictors with $P < 0.10$ in the initial multivariable model were successively removed. After completing backward selection, we tested two-way interactions between the main effects. Any interaction with a bivariate association of $P \leq 0.05$ (as before, controlling for ALPM) was added to the multivariable model. Interactions were retained if they remained significant at $P \leq 0.05$ in the model and improved the AUC and BIC in the unweighted version of the model.[6]

6 BIC is not available for weighted regression models in Stata.

With the predictors finalized, we again checked whether any of the alternative specifications of ALPM improved the AUC and BIC.

Assessing predictive performance

We evaluated the sensitivity and specificity of the model to predict HIV status of participants in the derivation sample at different thresholds of the model's risk score (regression equation). Risk score thresholds were evaluated in 1% increments between 1% and 4% predicted probability of HIV, as this was the range in which sensitivity dropped below 70%.

Initially we aimed to test the best-performing model on the validation sample. None of the models performed well enough (sensitivity and specificity ≥70%) to warrant validation, however.

Models stratified by subnational HIV prevalence

In addition to the model-building approach described above, in which ALPM was incorporated into the model as a predictor and interactions, we developed 2 additional models for participants residing in zones (Malawi) or provinces (Zambia) of lower prevalence (ALPM < 6%) and greater prevalence (ALPM ≥ 6%). These models were developed through the same process, from bivariate and multivariable analysis to assessment of sensitivity and specificity. We chose the 6% threshold because it appeared as the most notable inflection point on the Lowess curve of ALPM versus HIV.

Statistical weighting

We developed jackknife replicate weights adjusted for multicountry analysis as recommended by the PHIA Data Use Manual *(28)*. We verified these weights correctly reproduced the male and female HIV prevalence estimates published in the Malawi and Zambia PHIA survey reports. All bivariate and multivariable logistic regression models incorporated the weights using commands for analysis of complex surveys (e.g. *svy* commands) in Stata. AUC and BIC scores were based on unweighted models. Analysis was conducted in Stata v15.1.

Risk factor findings

Bivariate findings

Several of the candidate risk factors were significantly associated with HIV infection in bivariate analysis. The strongest risk factors in the larger sample that was not stratified by ALPM (Table 26) were:

- Condomless sex with multiple partners of positive or unknown HIV status (OR = 8.3), or number of such partners (OR = 2.4).
- Sexually transmitted infection diagnosis (OR = 3.7) or symptoms (OR = 2.2).
- Multiple partners (OR = 2.6), or having multiple partners including a non-cohabitating partner (OR = 2.8).
- Partner of unknown status (OR = 2.0), or HIV-positive or unknown status (OR = 2.1).

- Age gap with partners of more than 5 years (OR = 1.6), or other specifications (OR = 1.1–1.3).

Marginally significant risk factors ($P \leq 0.2$) were having a married or live-in partner who was staying elsewhere at the time of the survey, having a partner living with HIV, respondent's age 20 years or over, forced sex (lifetime and past year), and transactional sex.

Unexpectedly, being out of school was negatively associated with HIV (OR = 0.06, $P = 0.09$).

Three variables had significant or marginally significant interactions with ALPM, suggesting their effect size varied with ALPM:

- Sexually transmitted infection symptoms (interaction OR = 0.81, $P = 0.09$).
- Being out of school (interaction OR = 1.3, $P = 0.04$),
- Having a married or live-in partner who was staying away at the time of the survey (interaction OR = 1.3, $P = 0.05$).

These associations were all from models that controlled for ALPM.

Among adolescent girls and young women in high-prevalence areas (Table 27), bivariate findings were similar, but effect sizes were generally greater and additional predictors became significant ($P \leq 0.05$), including:

- Condomless sex with a partner of positive or unknown HIV status (OR = 2.2).
- Having a non-cohabitating partner (OR = 1.9).
- Being out of school (OR = 1.9).

Also in the high-prevalence sample, condomless sex with a married or live-in partner who had other wives or live-in partners (OR = 0.28, $P < 0.001$) and forced sex (OR = 0.53, $P = 0.123$) were negatively associated with HIV.

Among adolescent girls and young women in low-prevalence areas (Table 28), two risk factors were identified:

- Age ≥ 20 years (OR = 4.4) and age specified as a continuous variable (OR = 1.3).
- Condomless sex with a partner of positive or unknown HIV status (OR = 2.2) and number of such partners (OR = 2.38).

Marginally significant were having a married or live-in partner who was staying away at the time of the survey and condomless sex with such a partner.

Several of the associations could not be estimated in the low-prevalence sample because there were too few HIV-positive survey participants: multiple partners; one or more HIV-positive partners; non-monogamous and having a non-cohabitating partner; out of school; and sexually transmitted infection diagnosis.

Table 26.
Bivariate associations with HIV among adolescent girls and young women in all areas (N = 4029)

Candidate risk factor	N	n	HIV (%)	OR	95% CI	P
Age (continuous)	4029	4029	3.6	1.04	0.95–1.12	0.412
Age ≥ 20 years[a]	4029	2651	3.9	2.99	0.86–10.42	0.097
Number of partners (continuous) in past 12 months	3910	3276	3.6	1.31	0.97–1.78	0.088
Number of partners in past 12 months ≥ 2	3910	198	8.1	2.63	1.54–4.49	0.002
HIV-positive partner among last 3 partners in past 12 months	4025	20	10.0	4.46	0.46–42.95	0.207
Partner of unknown HIV status among last 3 partners in past 12 months	4026	1372	5.4	2.00	1.42–2.81	0.001
Partner of HIV-positive or unknown status among last 3 partners in past 12 months	4025	1391	5.5	2.11	1.48–3.02	0.000
Maximum age gap among last 3 partners in past 12 months						
≥ 5 years	4029	1442	4.6	1.55	1.08–2.22	0.024
≥ 10 years	4029	358	5.0	1.29	0.77–2.18	0.347
Continuous	3910	–	–	1.05	1.02–1.08	0.003
Square root	3910	–	–	1.26	1.07–1.47	0.009
Sum of age gap over last 3 partners in past 12 months	4029	–	–	1.05	1.02–1.09	0.002
No condom at last sex among last 3 partners in past 12 months						
With any partner	4029	2515	3.4	0.91	0.64–1.30	0.616
With ≥ 1 HIV-positive/unknown status partner	3910	1027	6.0	2.34	1.66–3.32	0.000
With ≥ 2 HIV-positive/unknown status partners	3897	35	22.9	8.33	3.51–19.80	0.000
Number of HIV-positive/unknown status partners with whom no condom was used at last sex	3897	–	–	2.42	1.76–3.34	0.000
Married/live-in partner staying elsewhere at time of survey[a]	3988	248	4.0	7.66	1.06–55.16	0.054
Married/live-in partner with other wives or lives with other women	3971	136	2.9	0.93	0.28–3.05	0.901
No condom at last sex with this partner	3987	108	2.8	0.96	0.21–4.35	0.953
Uncircumcised partner in past 12 months	3734	1853	2.9	0.85	0.58–1.26	0.433
Non-cohabitating partner	3910	1528	4.8	1.64	1.15–2.32	0.011
Non-monogamous with a non-cohabitating partner in past 12 months	3910	189	8.5	2.76	1.61–4.72	0.001
Ever forced to have sex	2788	334	2.7	0.51	0.24–1.07	0.088

Candidate risk factor	N	n	HIV (%)	OR	95% CI	P
Forced to have sex in past 12 months	2788	67	7.5	1.93	0.73–5.11	0.199
Transactional sex in past 12 months	3887	563	4.6	1.36	0.86–2.17	0.201
Out of school[a]	4029	416	4.3	0.06	0.00–1.31	0.087
Sexually transmitted infection diagnosis in past 12 months	3917	69	13.0	3.71	1.50–9.16	0.009
Sexually transmitted infection symptoms in past 12 months[a]	3897	267	7.5	2.24	1.32–3.81	0.006

[a] Estimates from model controlling for interaction with ALPM as this improved significance of the predictor.

Table 27.
Bivariate associations with HIV among adolescent girls and young women in high-prevalence areas (N = 2963)

Candidate risk factor	N	n	HIV (%)	OR	95% CI	P
Age (continuous)	2963	2963	4.3	1.01	0.93–1.11	0.760
Age ≥ 20 years	2963	1967	4.5	1.10	0.74–1.64	0.650
Number of partners (continuous) in past 12 months	2875	2394	4.3	1.36	1.00–1.85	0.058
Number of partners ≥ 2 in past 12 months	2875	161	9.9	3.00	1.74–5.15	0.001
HIV-positive partner among last 3 partners in past 12 months	2961	15	13.3	4.71	0.45–49.85	0.209
Partner of unknown HIV status among last 3 partners in past 12 months	2961	1015	6.5	2.01	1.40–2.88	0.001
Partner of HIV-positive or unknown status among last 3 partners in past 12 months	2961	1030	6.6	2.14	1.47–3.12	0.001
Maximum age gap among last 3 partners in past 12 months						
≥ 5 years	2963	1121	5.4	1.48	1.01–2.16	0.055
≥ 10 years	2963	274	6.2	1.31	0.76–2.25	0.340
Continuous	2875	–	–	1.05	1.02–1.09	0.008
Square root	2875	–	–	1.24	1.05–1.47	0.018
Sum of age gap over last 3 partners among last 3 partners in past 12 months	2963	–	–	1.06	1.02–1.09	0.002
No condom at last sex among last 3 partners in past 12 months						
With any partner	2963	1802	4.1	0.90	0.62–1.30	0.582
With ≥ 1 HIV-positive/unknown partner	2875	752	7.2	2.19	1.52–3.16	0.000
With ≥ 2 HIV-positive/unknown partners	2864	28	28.6	10.17	4.20–24.61	0.000

Candidate risk factor	N	n	HIV (%)	OR	95% CI	P
Number of HIV-positive/unknown partners with whom no condom was used at last sex among last 3 partners in past 12 months	2864	–	–	2.40	1.69–3.43	0.000
Married/live-in partner staying elsewhere at time of survey	2929	193	3.6	0.73	0.32–1.62	0.444
Married/live-in partner with other wives or lives with other women	2914	90	2.2	0.43	0.08–2.19	0.316
No condom at last sex with this partner	2927	70	1.4	0.28	0.20–0.39	0.000
Uncircumcised partner in last 12 months	2736	1281	3.6	0.79	0.53–1.19	0.274
Non-cohabitating partner	2875	1162	5.9	1.90	1.31–2.75	0.002
Non-monogamous with non-cohabitating partner in last 12 months	2875	154	10.4	3.15	1.83–5.43	0.000
Ever forced to have sex	2012	264	3.0	0.53	0.24–1.16	0.123
Forced to have sex in last 12 months	2012	47	8.5	1.92	0.67–5.45	0.234
Transactional sex in last 12 months	2859	431	5.8	1.51	0.94–2.43	0.099
Out of school	2963	296	6.1	1.87	1.08–3.26	0.036
Sexually transmitted infection diagnosis in last 12 months	2879	60	15.0	3.75	1.51–9.30	0.009
Sexually transmitted infection symptoms in last 12 months	2862	218	8.7	2.44	1.41–4.24	0.004

Table 28.
Bivariate associations with HIV among adolescent girls and young women in low-prevalence areas (N = 1066)

Candidate risk factor	N	n	HIV (%)	OR	95% CI	P
Age (continuous)	1066	–	–	1.27	1.06–1.51	0.013
Age ≥ 20 years	1066	684	2.3	4.44	1.23–15.95	0.031
Number of partners (continuous) in past 12 months	1035	–	–	0.73	0.29–1.84	0.512
Number of partners ≥ 2 in past 12 months	998	–	–	–	–	–
HIV-positive partner among last 3 partners in past 12 months	1059	5	0.0	–	–	–
Partner of unknown HIV status among last 3 partners in past 12 months	1065	357	2.2	1.64	0.59–4.57	0.353
Partner of HIV-positive/unknown status among last 3 partners in past 12 months	1064	361	2.2	1.61	0.58–4.47	0.371

Candidate risk factor	N	n	HIV (%)	OR	95% CI	P
Maximum age gap among last 3 partners in past 12 months						
≥ 5 years	1066	321	1.9	1.52	0.50–4.63	0.466
≥ 10 years	1066	84	1.2	0.78	0.44–1.41	0.421
Continuous	1035	–	–	1.03	0.96–1.10	0.413
Square root	1035	–	–	1.18	0.77–1.83	0.452
Sum of age gap over last 3 partners among last 3 partners in past 12 months	1066	–	–	1.01	0.92–1.10	0.872
No condom at last sex among last 3 partners in past 12 months						
With any partner	1066	713	1.7	1.05	0.36–3.04	0.926
With ≥ 1 HIV-positive/unknown partner	1035	275	2.9	3.27	1.12–9.55	0.040
With ≥ 2 HIV-positive/unknown partners	1026	7	0.0	–	–	–
Number of HIV-positive/unknown partners with whom no condom was used at last sex among last 3 partners in past 12 months	1033	–	–	2.38	1.18–4.80	0.023
Married/live-in partner staying elsewhere at time of survey	1059	55	5.5	2.83	0.55–14.68	0.227
Married/live-in partner with other wives or lives with other women	1057	46	4.3	5.01	0.58–43.51	0.157
No condom at last sex with this partner	1060	38	5.3	6.14	0.67–56.33	0.121
Uncircumcised partner in past 12 months	998	572	1.2	0.72	0.25–2.10	0.550
Non-cohabitating partner	1035	366	1.1	0.40	0.11–1.55	0.197
Non-monogamous with non-cohabitating partner in past 12 months	1000	35	0.0	–	–	–
Ever forced to have sex	776	70	1.4	0.43	0.10–1.94	0.285
Forced to have sex in past 12 months	776	20	5.0	1.77	0.36–8.84	0.491
Transactional sex in past 12 months	1028	132	0.8	0.72	0.39–1.32	0.293
Out of school	946	120	0.0	–	–	–
Sexually transmitted infection diagnosis in past 12 months	1029	9	0.0	–	–	–
Sexually transmitted infection symptoms in past 12 months	1035	49	2.0	0.73	0.16–3.22	0.678

Multivariable findings

The models obtained differed by ALPM (Tables 6–8). In all three models, condomless sex with a partner of positive or unknown HIV status, or multiple such partners, was among the strongest risk factors identified.

A past-year diagnosis of sexually transmitted infection and greater age gap with older partners (whether expressed as the maximum gap or the sum of gaps with the last three partners) were strong predictors in the larger sample and for adolescent girls and young women in high-prevalence areas.

Surprisingly, the number of partners was negatively associated with HIV in the larger sample and in high-prevalence areas.

Having a married or live-in partner who was staying elsewhere at the time of the survey was an independent risk factor in the larger sample, but this had diminishing importance as ALPM increased owing to an interaction term, so much so that it became negatively associated with HIV in the high-prevalence model.

In contrast, being out of school was a risk factor in the larger sample and its effect size increased with ALPM (again, owing to an interaction). Being out of school did not, however, emerge as a risk factor in the high-prevalence areas model.

An additional independent risk factor identified was having a non-cohabitating partner in the high-prevalence model. For adolescent girls and young women in low-prevalence areas, the only independent risk factor identified in addition to condomless sex with partners of positive or unknown status was older age.

Table 29.
Multivariable model of HIV in adolescent girls and young women in all areas (N = 3832)

Candidate risk factor	Adjusted OR	95% CI	P
No condom at last sex with ≥ 2 HIV-positive/unknown partners among last 3 partners in past 12 months	8.00	2.49–25.67	0.001
Married/live-in partner staying elsewhere at time of survey	7.90	0.96–65.12	0.055
Sexually transmitted infection diagnosis in past 12 months	4.20	1.66–10.61	0.004
ALPM (natural logarithm)	3.03	1.71–5.36	0.001
Partner of HIV-positive/unknown status among last 3 partners in past 12 months	2.34	1.46–3.76	0.001
Sum of age gap over last 3 partners among last 3 partners in past 12 months	1.06	1.02–1.09	0.003
Number of partners (continuous) in past 12 months	0.54	0.31–0.92	0.024
Out of school[a]	0.02	0.00–0.65	0.029
Interactions with ALPM[a]			
ALPM × out of school	1.47	1.09–1.97	0.012
ALPM × married/live-in partner staying elsewhere at time of survey	0.81	0.65–1.01	0.060

[a] Estimates from model controlling for interaction with ALPM as this improved significance of the predictor.

Table 30.
Multivariable model of HIV in adolescent girls and young women in high-prevalence areas (N = 2799)

Candidate risk factor	Adjusted OR	95% CI	P
Sexually transmitted infection diagnosis in past 12 months	4.56	1.82–11.43	0.002
Number of HIV-positive/unknown status partners with whom no condom was used at last sex among last 3 partners in past 12 months	2.50	1.65–3.80	< 0.001
Non-cohabitating partner	1.89	1.27–2.81	0.003
Maximum age gap (continuous) among last 3 partners in past 12 months	1.05	1.01–1.09	0.019
Number of partners (continuous) in past 12 months	0.62	0.34–1.12	0.109
No condom at last sex with married/live-in partner with other wives or lives with other women	0.24	0.15–0.37	< 0.001

Table 31.
Multivariable model for HIV in adolescent girls and young women in low-prevalence areas (N = 1025)

Candidate risk factor	Adjusted OR	95% CI	P
No condom at last sex with ≥ 1 HIV-positive/unknown status partner among last 3 partners in past 12 months	3.39	1.04–11.07	0.043
Age (continuous)	1.22	0.98–1.53	0.074

Predictive performance

The predictive power of all three models was poor (Table 32). The AUC was similar across models (0.69–0.70). The best cut-off points of the risk score were chosen where sensitivity was 70% or greater. Even at this modest level of sensitivity, specificity ranged from 52% in the high-prevalence areas model to 70% in the low-prevalence areas model. Since the risk models did not predict well in the derivation sample, they were not evaluated on the validation sample.

Table 32.
Predictive performance of best multivariable models

Measure	All areas model	High-prevalence areas model	Low-prevalence areas model
AUC	0.70	0.69	0.70
Sensitivity	70.1%	71.3%	73.3%
Specificity	56.6%	51.9%	69.9%

Limitations

Important limitations of this analysis include the following:

- PHIA surveys are cross-sectional, so risk factors identified reflect prevalent rather than incident HIV. This provides weaker evidence for causation and may have led to some of the unexpected findings, such as inverse relationships between number of partners, particular partner types and HIV.
- The predictive performance of the models was poor, which may be due to the cross-sectional nature of the data.
- Pintye's risk index could not be assessed since key inputs to the index (bacterial vaginosis, vaginal candidiasis, lifetime number of sexual partners, syphilis) were not available in one or both surveys.

- Although data were weighted as recommended by PHIA guidance, it was not possible to adjust standard errors for clustering of participants within zones or provinces. Consequently, the estimated effect size of ALPM may include other unobserved differences across zones and province regarding the likelihood of HIV infection.
- The behavioural and sexually transmitted infection diagnosis measures were self-reported and may be subject to recall bias and social desirability bias.

Conclusion

PHIA surveys are a valuable resource for estimating the proportion of adolescent girls and young women at risk since they are nationally representative, are available in many high-burden countries, and gather extensive information on risk behaviours and sexually transmitted infection.

This analysis identified several risk factors strongly associated with prevalent HIV in Malawi and Zambia in terms of statistical significance and large estimated effect sizes.

The low sensitivity and specificity of the best risk models obtained, however, suggests these risk factors will incorrectly classify about 30% of adolescent girls and young women living with HIV and about 40–50% of HIV-negative adolescent girls and young women. Importantly, the low level of specificity of these indicators may led to substantially overestimated PrEP targets.

A key limitation of PHIA data for risk factor analysis is that they are cross-sectional. Cohort studies of adolescent girls and young women in high-burden settings are needed to develop a risk definition linked to incident HIV that provides a more acceptable level of predictive performance.

Exploring the estimation method in practice

Objectives

This section aims to:

- Apply the cohort-derived risk definition to a range of sites to identify issues that arise in practice when using local surveillance data.
- Develop and apply additional risk definitions informed by the literature review.
- Explore how reasonable variations in risk definition may influence the estimated number of men who have sex with men at risk.
- Identify patterns across the selected sites that offer insight regarding how to simplify the risk definition without loss of accuracy in the resulting estimates.

Additional risk definitions

We developed additional risk definitions informed by the literature review to allow us to explore how changes in the risk definition affect the estimates in practice, focusing on variables that tend to be available in BBS surveys (Table 33).

The VICITS model for Guatemala from the previous section, which excludes syndromic sexually transmitted infection, is listed as Risk Definition A.

Risk Definition B begins from this model and replaces age with multiple partners and unprotected anal intercourse, as it is likely that the age effect in the VICITS models is capturing unobserved risk behaviours. We selected multiple partners and unprotected anal intercourse as the risk factors as they are directly linked to transmission risk.

Risk Definition C is a reduced model, including only syphilis, multiple partners and unprotected anal intercourse. It serves as a point of reference for examining the effect of including additional variables consistently identified as risk factors in past studies.

Self-reported sexually transmitted infection history or symptoms and high perceived risk of HIV are added in Risk Definitions D and E, respectively.

Having a current partner with HIV is examined in Risk Definition F, as discordant couples are a natural target for PrEP.

Apart from the VICITS model, which has medium- and high-risk thresholds linked to predicted probabilities of acquiring HIV, the remaining definitions are intended to classify individuals as at risk or not at risk.

Table 33.
Risk definitions explored for men who have sex with men

Risk definition	Risk classification criteria
A: VICITS Guatemala model	100 x syphilis + 60 x (receptive anal sex) −5 x (age − 29) ≥ 13 (medium risk) ≥ 47 (high risk)
B: partners and unprotected anal intercourse in place of age	Syphilis Or receptive anal sex Or > 1 male partner and unprotected anal intercourse
C: partners and unprotected anal intercourse reduced definition	Syphilis Or > 1 male partner and unprotected anal intercourse
D: sexually transmitted infection history or symptoms	Syphilis Or > 1 male partner and unprotected anal intercourse Or sexually transmitted infection history/symptoms
E: perceived risk	Syphilis Or > 1 male partner and unprotected anal intercourse Or high perceived HIV risk
F: partner living with HIV	Syphilis Or > 1 male partner and unprotected anal intercourse Or partner living with HIV

BBS and population size estimate data sources

Estimates were developed using BBS from one city in eastern Africa and four cities in Latin America. All surveys used RDS and tested for HIV and syphilis infections. Sample sizes were over 300 men who have sex with men. HIV prevalence ranged from 9% to 22% (Table 34).

Two of the BBS surveys included men who have sex with men and transgender women. Others were limited to men who have sex with men by eligibility criteria.

Population size estimates available for these sites included methods designed to estimate the number of all men who have sex with men (the multiplier method using RDS in conjunction with distribution of a unique object) and venue-based methods (e.g. programmatic mapping size estimate, mapping and enumeration, capture–recapture by distributing unique objects at venues) (Table 35).

Data preparation for the RDS surveys included:

- Verifying numbering of coupon codes, identifying repeated participant or coupon codes, identifying subjects whose recruiter was not in the dataset, and coordinating with study teams to resolve inconsistencies
- Preparing the network size variable:
    - Network sizes reported at less than the number of peers observed in the data (participant's recruits plus 1 for the participant's own recruiter were replaced with this value).
    - Network size was imputed at the mean where missing and where the series of network size items increased where it logically should have decreased or stayed constant.

Table 34.
BBS survey data used to construct estimates for men who have sex with men[a]

Country	City	Year	Sampling	N	HIV prevalence (95% CI)
Colombia	Bogotá	2016	RDS	415	22.4% (13.0–31.8%)
	Cali	2016	RDS	444	17.3% (9.3–25.2%)
	Cucuta	2016	RDS	307	10.9% (6.2–15.7%)
	Cartagena	2016	RDS	285	9.6% (1.8–17.5%)
Nicaragua[b]	Managua	2016	RDS	520	12.9% (5.7–20.1%)
Country A[b]	City A	2011–2012	RDS	459	11.4% (5.7–17.2%)
Ecuador	Guayaquil	2017	RDS	454	11.2% (6.6–15.7%)
Guatemala[b]	Guatemala City	2016	RDS	512	8.2% (3.2–13.2%)

[a] Estimates adjusted with RDS-2.

[b] *N* and prevalence estimates exclude transgender women.

Table 35.
Population size estimates used to construct estimates for men who have sex with men

Country	City	Year	Method	Size estimate	Estimate relative to males aged 15–49 years	Males aged 15–49 years in 2017	UNAIDS reference range for 2017[a]
Colombia	Bogotá	2010	Multiplier (RDS–unique object)	92 593 (43 855–141 330)	4.3%	2 139 512	31 879 (12 623–117 031)
	Cali	2010	Multiplier (RDS–unique object)	22 727 (9949–35 506)	3.9%	636 796	9488 (3757–34 833)
	Cartagena	2010	Multiplier (RDS–unique object)	8095 (4359–11 831)	3.4%	259 740	3870 (1532–14 208)
	Cucuta	2010	Multiplier (RDS–unique object)	7365 (4747–9983)	4.7%	170 481	2540 (1006–9325)
Ecuador	Guayaquil	2015	Programmatic mapping size estimate	13 416 (no data)	2.4%	573 952	8552 (3386–31 395)
Guatemala	Guatemala City	2009	Venue-based capture–recapture (unique object)	4999 (4634–5366)	–	–	–
		2009	Enumeration	1299 (997–1601)	–	–	–
Nicaragua	Managua	–	–	No data	–	270 826	4035 (3228–4842)
Country A	City A	2017	3-source capture -recapture (unique object)	10 807 (7371–14 244)	3.5%	305 951	3916 (490–9790)

[a] Calculated by applying the median (IQR) percentage of males aged 15–49 years estimated to be men who have sex with men in the respective region: 1.28% (0.16–3.20%) for eastern and southern Africa, and 1.49% (0.59–5.47%) for Latin America.

Available risk factor measures

We reviewed the BBS questionnaires and data to construct measures as similar as possible to the risk factors identified by the literature review. The definitions vary across the datasets (Table 36). A number of practical issues were identified:

- Some questionnaires lacked direct measures of unprotected anal intercourse or total number of recent partners and had to be constructed from items across types of partners (e.g. stable, casual, commercial), requiring assumptions that partner types are mutually exclusive and exhaustive.
- For unprotected anal intercourse, questions on anal sex and condomless sex with a particular type had to be combined, requiring the assumption that both anal sex and condomless sex had occurred with the same person.
- There were high levels (over 40%) of missingness on some items, leading to concern about bias for purposes of estimating proportion at risk.
- Constructing the variables requires attention to skip patterns to ensure the proportion is calculated over all men who have sex with men and does not exclude individuals who skipped out of the item. For example, respondents who report no casual partners are not asked about condom use with casual partners; in the data, the condom use item needs to be changed from missing to 0, so the proportion correctly counts the person as not having engaged in the risk behaviour. The same issue applies to most of the measures constructed.
- Where the sexually transmitted infection history or symptoms item included responses about which sexually transmitted infection had been diagnosed, we did not count responses of syphilis, because syphilis infection was also part of the risk definitions.

Table 36.
Definitions of risk factors across BBS in men who have sex with men

Risk factor	City A	Guatemala	Managua	Guayaquil	Colombia
Sexually transmitted infection history/ symptoms	Sexually transmitted infection sign or symptom in past 6 months	Sexually transmitted infection other than syphilis in past 12 months	Sexually transmitted infection in past 12 months	Sexually transmitted infection sign or symptom in past 12 months	Diagnosed with gonorrhoea, chlamydia, herpes, hepatitis B or genital warts in past 12 months (excludes syphilis)
Receptive anal sex	≥ 1 receptive anal sex act with male partner in past 6 months	Receptive or versatile anal sex was most common sex practice with male or transgender women stable, casual or paid partners or clients in past 12 months	As for Guatemala	Receptive or receptive and insertive in anal sex with male or transgender women partners	As for Guayaquil
Number of male partners	Number of steady, casual, paid and paying male partners in past 6 months (includes clients)	Number of male and transgender women sex partners in past 6 months (includes clients)	As for Guatemala	Number of male and transgender sex partners in past 12 months	As for Guayaquil
Unprotected anal intercourse	For last stable, casual or paid partners or clients: Number of rounds of sex with a condom in past 6 months less than number of total rounds of sex in past 6 months ≥ 50% missingness on condom use items	Any of following true for any of past 3 partners: Male Anal sex most common type of sex in past 12 months Sometimes or never used condom in past 12 months	Sometimes or never used condom in anal sex with stable, casual or paid partners or clients in past 30 days	As for Guatemala	As for Guatemala
Partner living with HIV	Stable or casual partner in past 6 months has HIV 40% missing	No data	No data	Most recent stable or sex partner in past 12 months living with HIV	As for Guayaquil
High perceived risk of HIV	How likely to get HIV or afraid of HIV 43–46% missing	No data	No data	73% missing	High perceived risk of acquiring HIV

Analysis to construct the estimates

The number of men who have sex with men at risk was calculated following the proposed method by multiplying together the following:

- Initial PSE.
- Proportion of men who have sex with men who are HIV-negative.
- Growth in general urban population since PSE was conducted.
- Inflation factor if PSE reflects only those men who have sex with men who frequent venues.
- Proportion of men who have sex with men at risk, according to the respective risk definition.
- Initial PSEs.

For Guatemala City, we used the capture–recapture (unique object) PSE instead of the enumeration estimate. Enumeration is limited to individuals present at venues at a particular moment in time, whereas capture–recapture attempts to capture the larger venue-going population.

In Managua, where no local PSE is available, we used the median of published PSEs in Latin America (from the UNAIDS Spectrum Quick Start guide), with bounds at ±20% of the median.

Projecting PSEs forward

We carried out the estimates assuming the desired year of PrEP implementation would be 2017. Because many of the PSEs were conducted earlier (Guatemala in 2009, Colombia in 2010, Guayaquil in 2015), the PSEs needed to be updated to account for the fact that the population of men who have sex with men and the general population of each city are growing.

For example, there were an estimated 4999 men who have sex with men in Guayaquil by capture–recapture in 2009. From 2009 to 2017, the general urban population of Guatemala grew by 26.7%, so by 2017 there were likely more men who have sex with men. There are two ways of projecting forward the PSE:

- If the population percentage is known, such as 4.3% of males aged 18–49 years, it can be multiplied by the census projection of the number of males aged 18–49 years in 2017.
- If we are beginning from an absolute number of men who have sex with men, it can be multiplied by the growth rate for the respective period.

We took the latter approach because the available PSEs were in absolute terms. We obtained annual average rates of growth for the urban population in each country from the World Bank for each year between the PSE and 2017. We multiplied the annual growth rates to obtain an overall growth rate for the period (Table 37).

Table 37.
Growth rate multipliers to project PSEs in absolute terms to 2017

City	PSE year	National urban growth rate from PSE year to 2017	Growth rate multiplier
City A	2017	0	1.0
Guayaquil	2015	3.7%	1.037
Bogotá	2010	9.9%	1.099
Cali	2010	9.9%	1.099
Cucuta	2010	9.9%	1.099
Cartagena	2010	9.9%	1.099
Guatemala City	2009	26.7%	1.267
Managua	No PSE	–	–

Box 1. Developing initial PSE in absence of local data: men who have sex with men in Managua, Nicaragua

To develop a PSE for men who have sex with men in Managua, in the absence of a local PSE, we apply the median of PSEs published in Latin America of 1.49% of males aged 15–49 years. As bounds, we use ±20%, because the published median is not far from other PSEs obtained in large cities elsewhere in Central America.

To determine the denominator, Nicaraguan census projections indicate 334 967 males aged 15 years and over in Managua in 2016. The number for the 15–49 years age group specifically in Managua is unavailable.

We assume the same age distribution for the general male population as in nearby Guatemala City in 2016, where 79.5% of males aged 15 years and over were aged 15–49 years. Applied to Managua, this yields 334 967 × 0.795 = 266 299 males aged 15–49 years in 2016.

To project this 2016 figure to 2017, we apply the average annual urban population growth rate for Nicaragua in 2016, 1.7%, obtained from the World Bank. This yields 266 299 × 1.017 = 270 826 males aged 15–49 years in Managua in 2017.

With 270 826 as our denominator, we multiply by the published median population percentage of 1.49% and ±20% bounds to obtain our PSE estimate of 4035 [3228–4842] men who have sex with men in Managua in 2017.

Venue-inflation factors

Two of the initial PSEs included in the risk calculations collected data at venues and thus reflect men who have sex with men who frequent venues. To inflate the estimate to also reflect men who have sex with men who do not frequent venues, we developed inflation factors derived from RDS surveys.

For Guayaquil, we used the BBS question item, "Do you frequent public sites for meeting or socializing with gay, homosexual, bisexual or male sex workers and/or transgender women?"

For Guatemala City, we used an item from the 2010 BBS: "In the past 12 months, have you gone to sites for meeting or hooking up with male partners, like bars, dance clubs or parks?"

The population proportion of these items was estimated similar to other variables: among HIV-negative men who have sex with men using RDS-2, yielding 0.333 (95% CI 0.269–0.397) for Guayaquil and 0.613 (95% CI 0.613-0.787) for Guatemala City.

We calculated the inflation multiplier as the reciprocal of the estimate (e.g. 1/0.333), yielding a factor of 1.43 (95% CI 1.27–1.63) for Guatemala City and 3.00 (95% CI 2.52–3.71) for Guayaquil (Table 38). These factors differ by a factor of 2, suggesting that in Guatemala City in 2010, men who have sex with men were twice as likely to frequent venues identified by men who have sex with men than in Guayaquil in 2017.

Table 38.
Inflation factors for venue-based-based PSE

City	Question item used for inflator	Estimated proportion of "yes" response (95% CI)	Inflation factor (1/estimate) (95% CI)
Guayaquil, 2017 BBS	Do you frequent public sites for meeting or socializing with gay, homosexual, bisexual or male sex workers or transgender women?	0.333 (0.269–0.397)	3.00 (2.52–3.71)
Guatemala City, 2010 BBS	In the past 12 months, have you gone to sites for meeting or hooking up with male partners, such as bars, dance clubs or parks?	0.700 (0.613–0.787)	1.43 (1.27–1.63)

Proportion at risk

Dichotomous risk indicator variables were constructed representing each of the risk definitions in Table 33, coded as 1 if the study participant satisfied the condition (e.g. for definition 3, if the participant had syphilis or reported both multiple partners and unprotected anal intercourse) and otherwise coded as 0.

For Risk Definition 1, participants' risk scores were calculated and used to construct 0/1 indicators of low, medium and high risk, respectively, based on the medium- and high-risk thresholds in Table 33.

We estimated the population prevalence of the risk indicator variables by applying the RDS-2 estimator in RDS Analyst software, subset to HIV-negative participants. HIV-negative status was determined by the HIV test result rather than self-report.

Estimates for BBS that recruited both men who have sex with men and transgender women were subset additionally to participants who did not identify as transwomen.

Proportion HIV-negative

This was estimated from the BBS data using the RDS-2 estimator subset to men who have sex with men in the case of surveys that included transgender women.

Prevalence of risk variables in BBS data

The prevalence of risk factors was calculated similarly. Among HIV-negative men who have sex with men, prevalence of syphilis (RPR/VDRL) varied from 1% to 9%. Self-report of sexually transmitted infection history or symptoms was higher at all sites, varying from 6% to 60% (Table 39).

Half or more of HIV-negative men who have sex with men engaged in receptive anal sex at all sites except Colombia, where the proportion was a third.

Having multiple partners was more prevalent than unprotected anal intercourse at all sites except Guatemala. Compared with either behaviour alone, the prevalence of both unprotected anal intercourse and multiple partners was reduced by a factor of about two to three at most sites.

The Colombian sites were distinct in that rates of sexually transmitted infection history and symptoms were lower and closer to the levels of syphilis infection (Table 40).

The perceived HIV risk variable was only available in the Colombian BBS. In Cartagena prevalence of perceived risk was similar to the prevalence of the risk behaviours examined, whereas at the other three sites it was much lower. The estimates pooled across the Colombian cities are a weighted average. The biggest impact of pooling is seen in the perceived risk variable, where the pooled prevalence is 17% compared with 40% in Cartagena and 6–10% in the other cities.

The prevalence of having a partner living with HIV was 14% in City A, which has a generalized epidemic, compared with 1–3% across the Colombia sites, which are concentrated epidemics.

Median age was similar across surveys, at 22–24 years.

Estimated men who have sex with men at risk and sensitivity to risk definition

Applying the risk definition from the VICITS Guatemala model leads to classifying about half of HIV-negative men who have sex with men as at risk and about 80% as high or medium risk at all sites except Bogotá. In Bogotá, 31% would be classified as high risk and 72% as at high or medium risk.

These risk proportions lead to estimates of 1100 and 1900 HIV-negative men who have sex with men at high-risk in City A, Guatemala and Managua, 20 700 in Guayaquil and 24 400 in Bogotá (Table 41).

Relative to the VICITS model, Risk Definition B (positive syphilis test result or multiple partners plus unprotected anal intercourse) decreases the risk proportions by 22–67% (Table 41), except in Bogotá, where the proportion is increased. Thus, at most sites, accounting for multiple partners and unprotected anal intercourse, combined, does not replace the effects of age and receptive sex in the VICITS Model.

Relative to Risk Definition B, adding receptive sex increases the proportion considerably, from 62% to 260%. This increase is greatest where the difference in prevalence between receptive anal sex and multiple partners and unprotected anal intercourse (combined) is greatest (e.g. Managua).

Adding sexually transmitted infection history or symptoms to the risk definition increases the proportion at risk at all sites. The increase is more modest in Colombia, where levels of reported sexually transmitted infection were lowest.

Adding high perceived risk to the risk definition increases the risk proportion from 15% to 51%.

Adding having a partner living with HIV to the definition increases the proportion most in City A (18%), led to a 5–9% increase at the other Colombian sites, and had no effect in Cartagena, where the prevalence of having a partner living with HIV was lowest, at 1%.

Table 39.
Prevalence of risk factors among HIV-negative men who have sex with men from BBS surveys[a]

Risk factor	City A (%, 95% CI)	Guatemala (%, 95% CI)	Managua (%, 95% CI)	Guayaquil (%, 95% CI)	Bogotá (%, 95% CI)
Syphilis	1 (0–4)	9 (1–17)	5 (0–10)	6 (3–10)	1 (0–5)
Sexually transmitted infection history or symptoms	58 (49–68)	39 (11–67)	10 (4–15)	15 (10–20)	6 (0–16)
Receptive anal sex	49 (38–60)	45 (34–56)	48 (37–60)	60 (53–68)	32 (19–45)
Number of male partners > 1	75 (66–84)	43 (32–54)	34 (23–46)	70 (62–77)	79 (71–87)
Unprotected anal intercourse	40 (28–51)	58 (47–69)	27 (17–37)	58 (51–65)	52 (39–66)
Number of male partners > 1 and unprotected anal intercourse	33 (22–44)	27 (17–37)	11 (3–18)	41 (33–48)	39 (26–53)
Partner living with HIV	13 (4–22)	Not available	Not available	4 (2–7)	3 (0–11)
Age, median [IQR]	23 [20–26]	24 [22–29]	24 [21–31]	22 [20–28]	24 [21–29]

[a] RDS-2 estimates.

Table 40.
Prevalence of risk factors among HIV-negative men who have sex with men, Colombia, 2016 BBS[a]

Risk factor	Bogotá	Cali	Cartagena	Cúcuta	Pooled
Syphilis	1 (0–5)	3 (0–6)	1 (0–5)	1 (0–4)	2 (0–3)
Sexually transmitted infection history or symptoms	6 (0–16)	4 (0–10)	4 (0–11)	3 (0–7)	4 (1–8)
Receptive anal sex	32 (19–45)	37 (26–48)	43 (31–56)	41 (31–50)	38 (32–44)
Number of male partners > 1	79 (71–87)	70 (62–79)	66 (54–79)	75 (66–84)	72 (67–77)
Unprotected anal intercourse	52 (39–66)	29 (17–40)	52 (39–65)	37 (26–47)	42 (36–48)
Number of male partners > 1 and unprotected anal intercourse	39 (26–53)	24 (13–35)	36 (24–49)	28 (18–37)	31 (25–38)
High perceived HIV risk	8 (0–17)	10 (3–18)	40 (28–53)	6 (1–11)	17 (12–22)
Partner living with HIV	3 (0–11)	3 (0–9)	1 (0–6)	2 (0–5)	2 (0–5)
Age, median [IQR]	24 [21–29]	23 [20–29]	22 [20–25]	22 [20–25]	23 [20–27]

[a] RDS-2 estimates.

Table 41.
Estimated men who have sex with men at risk based on VICITS Guatemala risk definition

Location	Men who have sex with men at risk (%)			Men who have sex with men at risk (n)		
	High risk	Medium risk	Low risk	High risk	Medium risk	Low risk
City A	55 (44–67)	32 (22–42)	12 (5–19)	1900 (200–6200)	1100 (100–3900)	400 (0–1800)
Guatemala	51 (39–62)	30 (20–40)	19 (10–29)	4200 (2500–6700)	2500 (1300–4300)	1600 (700–3100)
Managua	46 (35–58)	31 (20–41)	23 (13–33)	1600 (900–2600)	1100 (500–1900)	800 (300–1500)
Guayaquil	58 (51–65)	25 (18–31)	17 (12–23)	21 500 (14 900–31 600)	9100 (5200–15 200)	6500 (3500–11 100)
Bogotá	31 (18–44)	41 (27–54)	28 (16–41)	24 400 (5800–59 600)	32 200 (900–73 100)	22 400 (5300–5500)

Table 42.
Increase in risk proportion when incorporating additional risk factors identified by literature review[a]

Location	A: VICITS model	B: syphilis Or multiple partners and unprotected anal intercourse	Versus A[b]	When also considering as at risk							
				C: receptive anal sex	Versus B[b]	D: sexually transmitted infection history or symptoms	Versus B[b]	E: high perceived risk	Versus B[b]	F: partner living with HIV	Versus B[b]
City A	55 (44–67)	34 (22–46)	–(38%)	70 (58–81)	(106%)	73 (62–84)	(115%)	–	–	40 (20–60)	(18%)
Guatemala	51 (39–62)	34 (23–45)	–(33%)	65 (54–75)	(91%)	72 (42–103)	(112%)	–	–	–	–
Managua	46 (35–58)	15 (6–24)	–(67%)	54 (43–66)	(260%)	23 (13–32)	(53%)	–	–	–	–
Guayaquil	58 (51–65)	45 (38–53)	–(22%)	75 (68–83)	(67%)	53 (45–60)	(18%)	–	–	49 (41–57)	(9%)
Bogotá	31 (18–44)	39 (25–53)	(26%)	63 (51–76)	(62%)	41 (27–55)	(5%)	45 (31–60)	(15%)	41 (27–56)	(5%)
Cali	43 (31–54)	25 (14–37)	–(42%)	53 (41–65)	(112%)	27 (15–39)	(8%)	32 (20–43)	(28%)	27 (16–39)	(8%)
Cartagena	49 (35–63)	37 (24–50)	–(24%)	70 (57–83)	(89%)	38 (25–50)	(3%)	56 (42–70)	(51%)	37 (24–50)	(0%)
Cucuta	51 (40–61)	27 (18–37)	–(47%)	57 (47–68)	(111%)	30 (19–40)	(11%)	31 (21–42)	(15%)	29 (19–40)	(7%)

[a] See Table 33 for complete risk definitions.

[b] Percentage difference relative to indicated column.

Sensitivity of estimates to data sources and methods

Beginning from the proposed method to estimate the number of key populations at high HIV risk (Figure 2), we conducted sensitivity analysis to explore how much difference each step in the procedure matters in practice, across available sites, relative to not carrying out the step or doing it differently.

We identified as best practices those steps that made a difference of 20% or more at any site.

This analysis examined sensitivity of the estimated number of high-risk men who have sex with men and did not examine medium- or low-risk thresholds. The reference scenario or base case was the proposed method including all recommended steps. For most of the comparisons, we adopted Risk Definition B in Table 33 as the base case.

Details of the methods and the results of each comparison are described below.

Using BBS versus programme data to estimate proportions at risk and HIV-negative

The method recommends using a representative survey, such as RDS or TLS BBS, as the data source for estimating proportions among HIV-negative men who have sex with men.

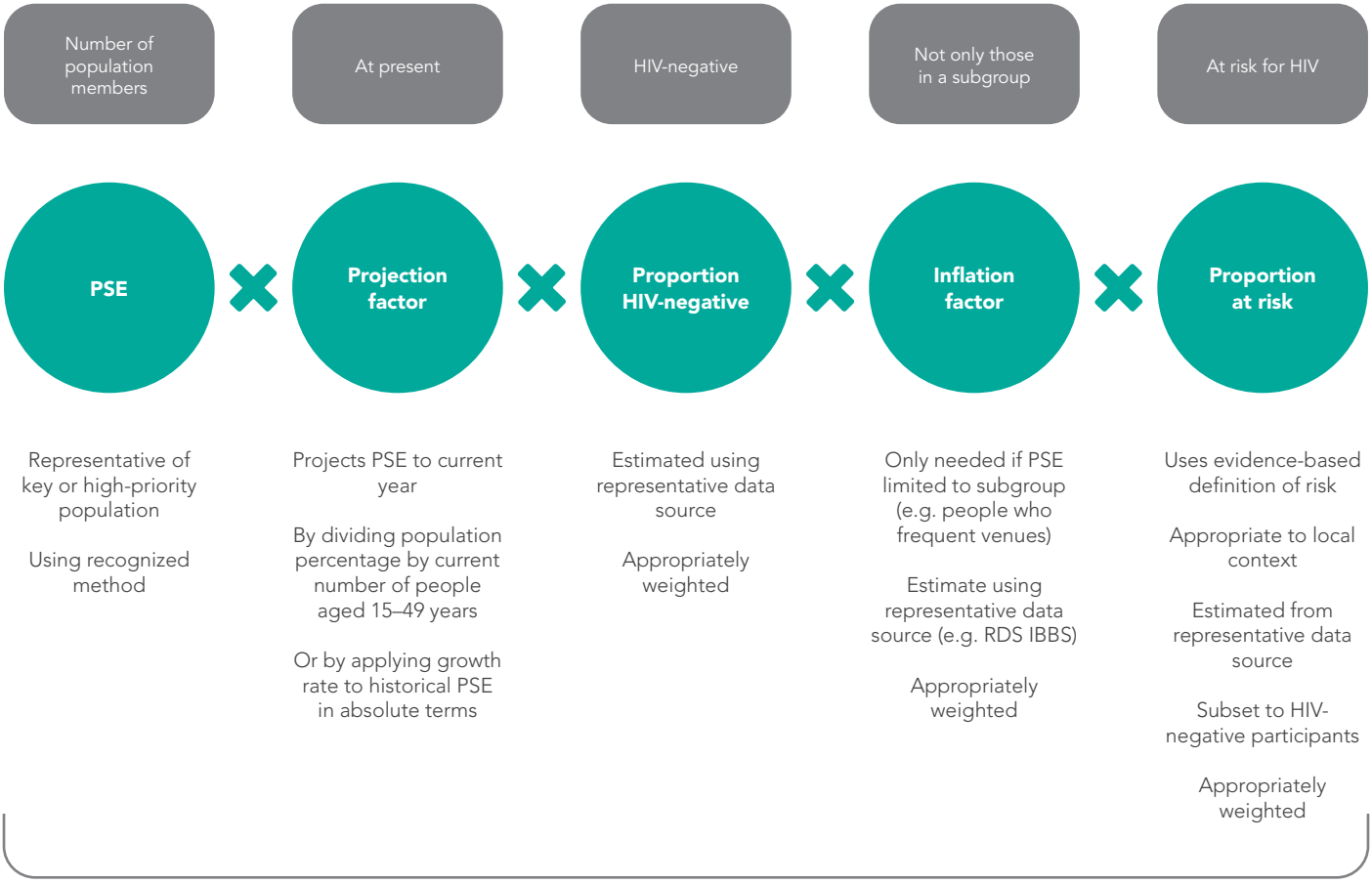
Countries could instead turn to programme data as a source of estimates. We compared the estimated number of men who have sex with men risk using estimates from BBS with estimates based on men who have sex with men programme attendees at first visit to VICITS clinics during 2016, the same year as the Guatemala and Nicaragua BBS. For this comparison, we use the VICITS Guatemala risk definition as the others cannot be estimated from VICITS data.

For Guatemala City, using BBS, the base case, leads to estimates of 1093 men who have sex with men, compared with 1063 when using VICITS data for the risk proportion, 1167 when using VICITS for the HIV-negative proportion, and 1136 when using VICITS for both proportions, changes equivalent to −3% to 7% relative to the base case (Table 45).

For Managua, using VICITS data leads to larger changes: a 35% change due to the risk proportion, 10% due to differences in the HIV-negative proportion, and 48% due to both.

This suggests that in Managua, HIV-negative VICITS clients had a higher-risk profile and a lower HIV prevalence relative to the population reached by BBS. These differences are likely to be due to programme use patterns.

Figure 2.
Method for estimating number of key or high-priority population members at risk for HIV



Using RDS versus TLS BBS to estimate risk proportion

We draw on data from a comparative study of RDS versus TLS as BBS sampling methods *(29)*. The two BBS were conducted concurrently in Guatemala City in 2010 among men who have sex with men and transgender women, collected behavioural data using a standardized instrument, and were closely coordinated.

Because the BBS did not test for HIV, we were not able to estimate proportions subset to HIV-negative participants. Instead, we subset the analysis to participants who did not report a previous HIV diagnosis. Analysis was also subset to men who have sex with men participants.

Similarly, the PSE could not be subset to HIV-negative men who have sex with men using HIV prevalence estimates from either BBS. Therefore, we used a more recent 2016 HIV prevalence for purposes of this comparison. Thus, this comparison is limited to examining differences in the number at risk stemming from differences in the estimated risk proportions.

RDS estimates were adjusted using RDS-2. TLS estimates were weighted to account for the multistage design and standard errors adjusted for clustering by recruitment event, as in the published analyses *(29)*.

As data for syphilis were not available, the risk definition used was based on the number of the following risk factors present:

- High perceived risk.
- Self-reported sexually transmitted infection symptoms in past year.
- Unprotected anal intercourse in past six months.
- Multiple partners.

Because the number of sex partners is available only for non-commercial partners, multiple partners was defined as any of the following:

- Selling sex in past 30 days.
- Buying sex in past year.
- Two or more non-commercial anal sex partners in past six months.

Thresholds were more than two variables for high risk and one variable for medium risk.

Table 44 presents the derivation of the estimated number of HIV-negative men who have sex with men in Guatemala City in 2010. The TLS survey estimated a greater prevalence of multiple partnerships than RDS (72% versus 46%, unpooled Z-test $P = 0.008$). The other risk factors examined were not statistically different. This led to a high-risk proportion of 46% for TLS and 37% for RDS, although not statistically different. This results in 23% more high-risk men who have sex with men being estimated when using the TLS survey (799 versus 652). TLS resulted in 4% fewer men who have sex with men at medium risk being estimated.

Weighted versus unweighted estimates when drawing on RDS surveys

We compared adjusting the risk proportions and HIV-negative proportions using RDS-2, Gile's successive sampling (SS) estimator and no statistical adjustments. Differences from Gile's SS versus RDS-2 were less than 1%, and differences when using unweighted estimates ranged from −3% to 4% across the sites examined (Table 46).

The limited difference when using SS is expected as the chief correction made by Gile's SS, the finite population correction, is less relevant to small samples of men who have sex with men in large cities.

Subsetting risk proportion and PSE to HIV-negative individuals

The method recommends subsetting BBS data to HIV-negative participants when estimating risk proportions and multiplying the PSE by the proportion of men who have sex with men who are HIV-negative. Here we examine omitting these steps.

Failing to exclude HIV-negative survey participants from the risk proportion calculation led to differences in expected numbers of men who have sex with men of −2% to 7% across

sites. There was no consistent direction in the impact. In three sites, including HIV-positive participants in the proportion led to higher risk; at two sites, it led to lower risk.

Not multiplying the initial base by the HIV-negative proportion had a larger impact, increasing the estimated number at risk from 9% to 29% across sites. With neither adjustment, the increase ranges from 11% to 23%.

Using pooled versus site-specific proportions

When the sample size available in BBS studies is limited, countries may consider pooling their data across study sites to estimate risk proportions and the proportion HIV-negative. We examine the effect of pooling across four sites included in the Colombian 2016 BBS. The base case uses estimates calculated from the data for each respective city (site-specific estimates). For the pooled scenario, we estimated proportions using RDS-2 in RDS-A on the combined dataset, after ensuring participant and coupon codes were unique to each city.

Using pooled proportions led to increases and decreases in the estimates, depending on where each site was with respect to the average. Changes ranged from −13% to 14% relative to using site-specific estimates for the risk proportion, and from 6% to 9% when pooling the proportion HIV-negative. Combined changes were −18% to 5% (Table 47).

Projecting PSE to the present year

When an available PSE was conducted before the anticipated year of PrEP implementation, the method recommends projecting the PSE forward, either by applying a growth rate to the PSE in absolute terms or by dividing the PSE population percentage by the number of males aged 18–49 years in the planned implementation year.

Here, for sites with a PSE before 2017, we examine not projecting the PSE forward.

Results in Table 49 suggest a 21% decrease in the estimates when failing to project the PSE in Guatemala City (2009 PSE), 9% in Bogota (2010 PSE) and 4% in Guayaquil (2015 PSE).

Summary of sensitivity analysis findings

We identified five aspects of the proposed estimation procedure that, when omitted or carried out differently, led to differences of 20% or more in the estimates:

- Choice of risk profile.
- Using programme data instead of BBS to estimate risk proportion.
- Not subsetting the PSE to HIV-negative individuals.
- Not projecting the PSE to the present year (Table 50).
- Varying the risk profile (this led to the largest observed changes).

Limitations

The sensitivity analysis could yield different results when examining additional sites, particularly outside concentrated settings or in other regions. The sensitivity analysis does not reflect all possible methodological choices and alternatives. Other important

considerations include the impact of how each risk factor is measured (e.g. differences in reference time periods of risk behaviours; looking at more than two, or more than three, partners instead of more than one).

We conducted one-way sensitivity analysis and in most of the comparisons did not examine the impact of altering multiple steps at once, which could have led to greater changes in the estimates.

We were unable to explore how much the risk definitions could be simplified without sacrificing the accuracy of the estimates because of the lack of a gold standard, due to the poor performance of the cohort-derived model.

The accuracy of the estimates and comparisons depends on the accuracy and reliability of the underlying PSEs and BBS data.

Table 43.
How using BBS versus programme data to estimate proportions impacts estimated number at risk[a,b]

		Guatemala City, 2016		Managua, 2016	
			Change versus base case		Change versus base case
Using BBS (base case)	Est	1093		1627	
	LB	547		899	
	UB	1984		2638	
Using programme data (VICITS) to estimate					
Proportion at high risk	Est	1063	–(3%)	2189	(35%)
	LB	638	(17%)	1414	(57%)
	UB	1690	–(15%)	3166	(20%)
Proportion HIV-negative	Est	1167	(7%)	1795	(10%)
	LB	614	(12%)	1061	(18%)
	UB	2023	(2%)	2735	(4%)
Both	Est	1136	(4%)	2415	(48%)
	LB	716	(31%)	1670	(86%)
	UB	1723	–(13%)	3283	(24%)

Est, point estimate; LB, lower bound of 95% CI; UB, upper bound of 95% CI.

[a] All estimates use VICITS Guatemala risk definition and incorporate all recommended steps, apart from what is varied in the first column.

[b] Programme data based on VICITS attendees who were HIV-negative at first visit during 2016.

Table 44.
Calculation of number of HIV-negative men who have sex with men in Guatemala City, 2010

	Estimate	Lower bound of 95%	Upper bound of 95%
PSE, by capture–recapture, 2009	4999	4634	5366
Growth rate, 2009–2010[a]	1.031	1.031	1.031
Proportion of men who have sex with men who are HIV-negative[b]	0.918	0.868	0.968
Inflation factor for venue-based PSE			
Proportion of men who have sex with men who went to a venue to meet partners in past year[c]	0.705	0.611	0.799
Inflation multiplier (= 1/above)	1.42	1.64	1.25
Estimated number of HIV-negative men who have sex with men in 2010 (PSE × growth × %HIV-negative × inflation factor)	6708	6783	6701
Age, median [IQR]	23 [20–26]	22 [20–28]	24 [21–29]

[a] World bank estimate of growth in urban population of Guatemala.

[b] Estimated from 2016 RDS BBS.

[c] Estimated from 2016 RDS BBS subset to HIV-negative participants.

Table 45.
How using RDS versus TLS surveys to estimate risk proportion impacts estimated number at risk, Guatemala City, 2010

Risk factor	RDS BBS, 2010 (N = 377)			TLS BBS, 2010 (N = 486)			P[a]
	Estimate	Lower bound	Upper bound	Estimate	Lower bound	Upper bound	
High perceived risk of HIV	35.4%	19.8%	51.1%	30.0%	21.2%	40.6%	0.565
Sexually transmitted infection symptoms in past 12 months	9.0%	0.9%	17.1%	10.1%	6.8%	14.6%	0.816
Unprotected anal intercourse in past 6 months	27.0%	10.3%	43.8%	26.8%	18.5%	37.1%	0.981
Multiple partners indicator (any 1 of following):	58.9%	42.4%	75.3%	75.4%	63.1%	84.6%	0.099
Sold sex in past 30 days	24.4%	10.1%	38.7%	18.6%	9.7%	32.8%	0.538
Paid for sex in past 12 months	7.8%	0.0%	18.5%	11.9%	7.0%	19.5%	0.474
≥ 2 non-commercial anal sex partners in past 6 months	45.9%	28.9%	62.9%	72.1%	62.2%	80.2%	0.008
Proportion at risk (number of above risk factors)							
Low risk (0)	18.6%	5.1%	32.1%	12.1%	6.5%	21.2%	0.403
Medium risk (1)	44.0%	28.4%	59.6%	42.1%	33.6%	51.1%	0.835
High risk (≥ 2)	37.4%	21.4%	53.4%	45.8%	38.8%	53.1%	0.345
Estimated number of HIV-negative men who have sex with men (from Table 44)[b]	1743	1459	1999	1743	1459	1999	
Estimated number of HIV-negative men who have sex with men in 2010 at:							
Low risk	324	75	641	210	95	424	–(35%)[c]
Medium risk	767	415	1191	734	491	1021	–(4%)[c]
High risk	652	313	1067	799	566	1061	(23%)[c]

[a] *P*-values from unpooled 2-sample Z-tests.

[b] Adjusted estimates using RDS-2 (RDS) and sampling weights and clustering by sampling event (TLS), respectively.

[c] Percentage difference relative to RDS. Comparison does not reflect potential RDS-TLS differences in the estimated proportion HIV-negative, as these did not test for HIV.

Table 46.
Increase in risk proportion when incorporating additional risk factors identified by literature review[a]

Estimator used to weight data		City A	% change versus base case	Guatemala	% change versus base case	Managua	% change versus base case	Guayaquil	% change versus base case	Bogotá	% change versus base case
RDS-2 (base case)	Est	2414		5362		1911		27 926		50 065	
	LB	235		3514		1108		20 100		16 626	
	UB	7506		8036		3003		39 809		10 2987	
Adjustment of proportion at risk											
Gile's SS	Est	2412	(0%)	5367	(0%)	1911	(0%)	27 941	(0%)	50 065	(0%)
	LB	251	(7%)	3724	(6%)	1215	(10%)	20 356	(1%)	18 155	(9%)
	UB	7133	−(5%)	7704	−(4%)	2817	−(6%)	39 421	−(1%)	96 709	−(6%)
Unweighted	Est	2342	−(3%)	5578	(4%)	1920	(1%)	28 086	(1%)	48 541	−(3%)
	LB	253	(8%)	4072	(16%)	1289	(16%)	20 982	(4%)	18 482	(11%)
	UB	6718	−(11%)	7669	−(5%)	2707	−(10%)	38 785	−(3%)	90 156	−(12%)
Adjustment of proportion HIV-negative											
Gile's SS	Est	2413	(0%)	5361	(0%)	1910	(0%)	27 917	(0%)	50 065	(0%)
	LB	239	(2%)	3524	(0%)	1140	(3%)	20 242	(1%)	17 139	(3%)
	UB	7376	−(2%)	8014	(0%)	2927	−(3%)	39 521	−(1%)	100 497	−(2%)
Unweighted	Est	2402	−(1%)	5327	−(1%)	1873	−(2%)	27 770	−(1%)	52 229	(4%)
	LB	242	(3%)	3594	(2%)	1142	(3%)	20 366	(1%)	18 807	(13%)
	UB	7245	−(3%)	7772	−(3%)	2816	−(6%)	38 906	−(2%)	100 350	−(3%)

Est, point estimate; LB, lower bound of 95% CI; UB, upper bound of 95% CI.

[a] All estimates use Model 2 risk definition and incorporate all recommended steps, apart from what is varied in the first column.

Table 47.
How subsetting risk proportion and PSE to HIV-negative individuals impacts estimated number at risk[a]

		City A	% change versus base case	Guatemala	% change versus base case	Managua	% change versus base case	Guayaquil	% change versus base case	Bogotá	% change versus base case
Risk proportion and PSE subset to HIV-negative men who have sex with men (base case)	Est	2414		5362		1911		27 926		50 065	
	LB	235		3514		1108		20 100		16 626	
	UB	7506		8036		3003		39 809		10 2987	
Risk proportion not subset	Est	2366	–(2%)	5469	(2%)	2050	(7%)	28 469	(2%)	47 787	–(5%)
	LB	235	(0%)	3638	(4%)	1241	(12%)	20 764	(3%)	15 517	–(7%)
	UB	7263	–(3%)	8106	(1%)	3128	(4%)	40 136	(1%)	99 754	–(3%)
PSE not subset	Est	2726	(13%)	5842	(9%)	2194	(15%)	31 438	(13%)	64 512	(29%)
	LB	284	(21%)	4050	(15%)	1387	(25%)	23 857	(19%)	24 364	(47%)
	UB	7959	(6%)	8301	(3%)	3185	(6%)	42 622	(7%)	118 420	(15%)
Neither subset	Est	2672	(11%)	5959	(11%)	2354	(23%)	32 050	(15%)	61 578	(23%)
	LB	283	(21%)	4193	(19%)	1554	(40%)	24 645	(23%)	22 738	(37%)
	UB	7701	(3%)	8374	(4%)	3317	(10%)	42 973	(8%)	114 703	(11%)

Est, point estimate; LB, lower bound of 95% CI; UB, upper bound of 95% CI.

[a] All estimates use Model 2 risk definition and incorporate all recommended steps, apart from what is varied in the first column.

Table 48.
How using pooled instead of site-specific proportions impacts estimated number at risk: men who have sex with men, Colombia, 2016[a]

		Bogotá	% change versus base case	Cali	% change versus base case	Cartagena	% change versus base case	Cucuta	%change versus base case
Site-specific estimates (base case)	Est	50 065		11 021		5634		4145	
	LB	16 626		3375		2257		2058	
	UB	102 987		23 141		10 605		7020	
Estimates pooled across four cities									
Of % at high risk	Est	481 84	−(4%)	12 608	(14%)	4905	−(13%)	4400	(6%)
	LB	18 055	(9%)	4489	(33%)	2169	−(4%)	2416	(17%)
	UB	90 688	−(12%)	23 753	(3%)	8575	−(19%)	6912	−(2%)
Of % HIV-negative	Est	54 612	(9%)	11 277	(2%)	5278	−(6%)	3938	−(5%)
	LB	19 623	(18%)	3634	(8%)	2204	−(2%)	1965	−(5%)
	UB	105 115	(2%)	22 655	−(2%)	9583	−(10%)	6641	−(5%)
Of both	Est	52 560	(5%)	12 901	(17%)	4595	−(18%)	4181	(1%)
	LB	21 309	(28%)	4834	(43%)	2118	−(6%)	2307	(12%)
	UB	92 562	−(10%)	23 254	(0%)	7749	−(27%)	6538	−(7%)

Est, point estimate; LB, lower bound of 95% CI; UB, upper bound of 95% CI.

[a] All estimates use Model 2 risk definition and incorporate all recommended steps, apart from what is varied in the first column.

Table 49.
How projecting initial PSE to present year impacts estimated number at risk[a,b]

		Guatemala, 2009 PSE	% change versus base case	Bogotá, 2010 PSE	% change versus base case	Guayaquil, 2015 PSE	% change versus base case
Projected to 2017 (base case)	Est	5362		50 065		27 926	
	LB	3514		16 626		20 100	
	UB	8036		102 987		39 809	
Not projected: historical PSE used	Est	4232	−(21%)	45 555	−(9%)	26 929	−(4%)
	LB	2773	−(21%)	15129	−(9%)	19 383	−(4%)
	UB	6343	−(21%)	93710	−(9%)	38 388	−(4%)

Est, point estimate; LB, lower bound of 95% CI; UB, upper bound of 95% CI.

[a] Analysis limited to sites where most recent PSE was conducted before 2017.

[b] All estimates use Model 2 risk definition and incorporate all recommended steps, apart from what is varied in the first column.

Table 50.
Summary of changes in estimated number of men who have sex with men at risk as function of methods[a]

Steps and alternatives	Range of relative difference across sites	Sites with ≥ 20% difference
Risk profile definitions	4999	5366
Risk Definition B relative to VICITS Guatemala model (A)	−67% to 26%	8/8
Risk Definitions C–F relative to B	0% to 260%	8/8
Using programme data (versus BBS) for estimates of:	75 (66–84)	79 (71–87)
Proportion at high risk	−3% to 35%	1/2
Proportion HIV-negative	7% to 10%	0/2
Both	4% to 48%	1/2
Using TLS behavioural survey to estimate risk proportions (versus RDS)	23%	1/1
Statistical adjustment of RDS BBS (versus RDS-2 estimator)		
Proportion at risk		
Gile's SS	0	0/5
Unweighted	−3% to 4%	0/5
Proportion HIV-negative		
Gile's SS	0	0/5
Unweighted	−2% to 4%	0/5
Not subsetting inputs to HIV-negative men who have sex with men (versus subsetting)		
Proportion at risk	−5% to 7%	0/5
PSE	9% to 29%	1/5
Both	11% to 23%	2/5
Pooling estimates of inputs across sites (versus site-specific estimates)		
Proportion at risk	−13% to 14%	0/4
Proportion HIV-negative	−6% to 9%	0/4
Both	−18% to 17%	0/4
Not projecting PSE to present year (versus projecting)	−21% to −4%	1/3[b]

[a] Analysis limited to point estimates, not bounds.

[b] Among sites with PSE conducted in a before the simulated year of PrEP implementation.

References

1. Gouws E, White PJ, Stover J, Brown T. Short term estimates of adult HIV incidence by mode of transmission: Kenya and Thailand as examples. Sex Transm Infect. 2006;82:iii51–iii55.

2. Beyrer C, Baral SD, Van Griensven F, Goodreau SM, Chariyalertsak S, Wirtz AL, et al. Global epidemiology of HIV infection in men who have sex with men. Lancet. 2012;380:367–377.

3. Baral SD, Poteat T, Strömdahl S, Wirtz AL, Guadamuz TE, Beyrer C. Worldwide burden of HIV in transgender women: a systematic review and meta-analysis. Lancet Infect Dis. 2013;13:214–222.

4. Mathers BM, Degenhardt L, Phillips B, Wiessing L, Hickman M, Strathdee SA, et al. Global epidemiology of injecting drug use and HIV among people who inject drugs: a systematic review. Lancet. 2008;372:1733–1745.

5. Baral S, Beyrer C, Muessig K, Poteat T, Wirtz AL, Decker MR, et al. Burden of HIV among female sex workers in low-income and middle-income countries: a systematic review and meta-analysis. Lancet Infect Dis. 2012;12:538–549.

6. Borenstein M, Hedges LV, Higgins JPT, Rothstein HR. A basic introduction to fixed-effect and random-effects models for meta-analysis. Res Synth Methods. 2010;1(2):97–111.

7. Looker KJ, Garnett GP. A systematic review of the epidemiology and interaction of herpes simplex virus types 1 and 2. Sex Transm Infect. 2005;81:103–107.

8. Kainne Dokubo E, Kim AA, Le L V, Nadol PJ, Prybylski D, Wolfe MI. HIV incidence in Asia: a review of available data and assessment of the epidemic. AIDS Rev. 2013;15:67–76.

9. Zhang W, Xu J-J, Zou H, Zhang J, Wang N, Shang H. HIV incidence and associated risk factors in men who have sex with men in mainland China: an updated systematic review and meta-analysis. Sex Health. 2016;13:373.

10. Zhao Y, Luo T, Tucker JD, Wong WCW. Risk factors of HIV and other sexually transmitted infections in China: a systematic review of reviews. PLoS One. 2015;10:1–15.

11. Wang H-Y, Xu J-J, Zou H-C, Reilly KH, Zhang CM, Yun K, et al. Sexual risk behaviors and HIV infection among men who have sex with men and women in China: evidence from a systematic review and meta-analysis. Biomed Res Int. 2015;2015:1–12.

12. Oldenburg CE, Perez-Brumer AG, Reisner SL, Mimiaga MJ. Transactional sex and the HIV epidemic among men who have sex with men (MSM): results from a systematic review and meta-analysis. AIDS Behav. 2015;19:2177–2183.

13. Operario D, Soma T, Underhill K. Sex work and HIV status among transgender women: systematic review and meta-analysis. J Acquir Immune Defic Syndr. 2008;48:97–103.

14. Scorgie F, Chersich MF, Ntaganira I, Gerbase A, Lule F, Lo YR. Socio-demographic characteristics and behavioral risk factors of female sex workers in sub-Saharan Africa: a systematic review. AIDS Behav. 2012;16:920–933.

15. Muldoon KA. A systematic review of the clinical and social epidemiological research among sex workers in Uganda. BMC Publ Health. 2015;15:1226.

16. Dellar RC, Dlamini S, Karim QA. Adolescent girls and young women: key populations for HIV epidemic control. J Int AIDS Soc. 2015;18:64–70.

17. Pintye J, Drake AL, Kinuthia J, Unger JA, Matemo D, Heffron R, et al. A risk assessment tool for identifying pregnant and postpartum women who may benefit from pre-exposure prophylaxis (PrEP). Clin Infect Dis. 2016;64:ciw850.

18. Kinuthia J, Drake AL, Matemo D, Richardson BA, Zeh C, Osborn L, et al. HIV acquisition during pregnancy and postpartum is associated with genital infections and partnership characteristics. Aids. 2015;29:2025–2033.

19. Edelstein ZR, Santelli JS, Helleringer S, Schuyler AC, Wei Y, Mathur S, et al. Factors associated with incident HIV infection versus prevalent infection among youth in Rakai, Uganda. J Epidemiol Glob Health. 2015;5:85–91.

20. Miranda AE, Pinto VM, McFarland W, Page K. HIV infection among young pregnant women in Brazil: prevalence and associated risk factors. AIDS Behav. 2014;18:S50–S52.

21. Gumbe A, McLellan-Lemal E, Gust DA, Pals SL, Gray KM, Ndivo R, et al. Correlates of prevalent HIV infection among adults and adolescents in the Kisumu incidence cohort study, Kisumu, Kenya. Int J STD AIDS. 2015;26:929–40.

22. Rosenberg NE, Graybill LA, Wesevich A, McGrath N, Golin CE, Maman S, et al. Individual, partner, and couple predictors of HIV infection among pregnant women in Malawi: a case–control study. AIDS Behav. 2018;22:1775–1786.

23. Schaefer R, Gregson S, Eaton JW, Mugurungi O, Rhead R, Takaruza A, et al. Age-disparate relationships and HIV incidence in adolescent girls and young women: evidence from Zimbabwe. Aids. 2017;31:1461–1470.

24. Harling G, Newell M-L, Tanser F, Kawachi I, Subramanian SV, Bärnighausen T. Do age-disparate relationships drive HIV incidence in young women? Evidence from a population cohort in rural KwaZulu-Natal, South Africa. J Acquir Immune Defic Syndr. 2014;66:443–451.

25. Weiss HA, Hankins CA, Dickson K. Male circumcision and risk of HIV infection in women: a systematic review and meta-analysis. Lancet Infect Dis. 2009;9:669–677.

26. Li Y, Marshall CM, Rees HC, Nunez A, Ezeanolue EE, Ehiris JE. Intimate partner violence and HIV infection among women: a systematic review and meta-analysis. J Int AIDS Soc. 2014;17(1).

27. Operario D, Underhill K, Chuong C, Cluver L. HIV infection and sexual risk behaviour among youth who have experienced orphanhood: systematic review and meta-analysis. J Int AIDS Soc. 2011;14:25.

28. PHIA data use manual. New York: PHIA; 2021 (https://dms-filesystem.s3.amazonaws.com/uploads/21352ae0cb0eca87931ff495ee92c0b2/01%20PHIA%20Data%20Use%20Manual.pdf Project site: https://phia.icap.columbia.edu/).

29. Paz-Bailey G, Miller W, Shiraishi RW, Jacobson JO, Abimbola TO, Chen SY. Reaching men who have sex with men: a comparison of respondent-driven sampling. AIDS Behav. 2013;17:3081–3090.

Annexes

Countries included in literature reviews

The following low- and middle-income countries were identified by the World Bank Atlas method (gross national income per capita of US$ 12 235 or less in 2016):

Afghanistan
Albania
Algeria
American Samoa[7]
Angola
Argentina
Armenia
Azerbaijan
Bangladesh
Belarus
Belize
Benin
Bhutan
Bolivia
Bosnia and Herzegovina
Botswana
Brazil
Bulgaria
Burkina Faso
Burundi
Cabo Verde
Cambodia
Cameroon
Central African Republic
Chad
China
Colombia
Comoros
Congo
Costa Rica
Côte d'Ivoire
Croatia
Cuba
Democratic People's Republic of Korea
Democratic Republic of the Congo
Djibouti
Dominica
Dominican Republic
Ecuador
Egypt
El Salvador
Equatorial Guinea
Eritrea
Ethiopia
Fiji
Gabon
Gambia
Georgia
Ghana
Grenada
Guatemala
Guinea
Guinea-Bissau
Guyana
Haiti
Honduras
India
Indonesia
Iran (Islamic Republic of)
Iraq
Jamaica
Jordan
Kazakhstan
Kenya
Kiribati
Kosovo
Kyrgyzstan
Lao People's Democratic Republic
Lebanon
Lesotho

7 The World Bank uses the term "country" interchangeably with "economy". This does not imply political independence but refers to any territory for which authorities report separate social or economic statistics.

Liberia

Libya

Madagascar

Malawi

Malaysia

Maldives

Mali

Marshall Islands

Mauritania

Mauritius

Mexico

Micronesia
(Federated States of)

Mongolia

Montenegro

Morocco

Mozambique

Myanmar

Namibia

Nauru

Nepal

Nicaragua

Niger

Nigeria

North Macedonia

Pakistan

Panama

Papua New Guinea

Paraguay

Peru

Philippines

Republic of Moldova

Romania

Russian Federation

Rwanda

Saint Lucia

Saint Vincent
and the Grenadines

Samoa

Sao Tomé and Principe

Senegal

Serbia

Sierra Leone

Solomon Islands

Somalia

South Africa

South Sudan

Sri Lanka

Sudan

Suriname

Swaziland

Syrian Arab Republic

Tajikistan

Thailand

Timor-Leste

Togo

Tonga

Tunisia

Turkey

Turkmenistan

Tuvalu

Uganda

Ukraine

United Republic
of Tanzania

Uzbekistan

Vanuatu

Venezuela
(Bolivarian Republic of)

Viet Nam

West Bank and Gaza

Yemen

Zambia

Zimbabwe

UNAIDS/JCxxxxE

UNAIDS
Joint United Nations
Programme on HIV/AIDS

20 Avenue Appia
1211 Geneva 27
Switzerland

+41 22 791 3666

unaids.org