UNAIDS 2021

REFERENCE

PrEP target-setting for key and high-priority populations

Estimating the number at risk

Contents

Acknowledgements

Three online consultations were held to solicit inputs from a number of technical experts. Written comments were solicited from additional people. The guide was also piloted in Thailand and Viet Nam with participants from the respective national AIDS programmes, whom we thank.

Monica Alonso Gonzalez

Rachel Baggaley

Stefan Baral

Trista Bingham

Carlos Cáceres

Katie Callahan

Rosalind Coleman

Shona Dalal

Sinead Delany-Moretlwe

Gaston Djomand

Megan Dunbar

Ian Fellows

Elizabeth Gardiner

Peter Ghys

Peter Godfrey-Faussett

Beatriz Grinsztejn

M. Itzel Loya-Montiel

Jerry Owen Jacobson

Joshua Kimani

Yuliang Liu

Laura Edith Martínez Gómez

Andrea Mejia

Maeve B. Mello

Pragna Patel

Navindra Persaud

Andrew Phillips

Carel Pretorius

Giovanni Ravasi

Jason Reed

Andrew Scheibe

Ray Shiraishi

Dawn Smith

Andre R. Verani

Peter Vickerman

Heather Watts

Joint United Nations Programme on HIV/AIDS (UNAIDS) staff

Abbreviations

BBS	biobehavioural surveillance
CI	confidence interval
DHS	demographic and health survey
HIV	human immunodeficiency virus
PHIA	population-based HIV impact assessment
PrEP	pre-exposure prophylaxis
RDS	respondent-driven sampling
TLS	time–location sampling
UNAIDS	Joint United Nations Programme on HIV/AIDS
WHO	World Health Organization

Introduction

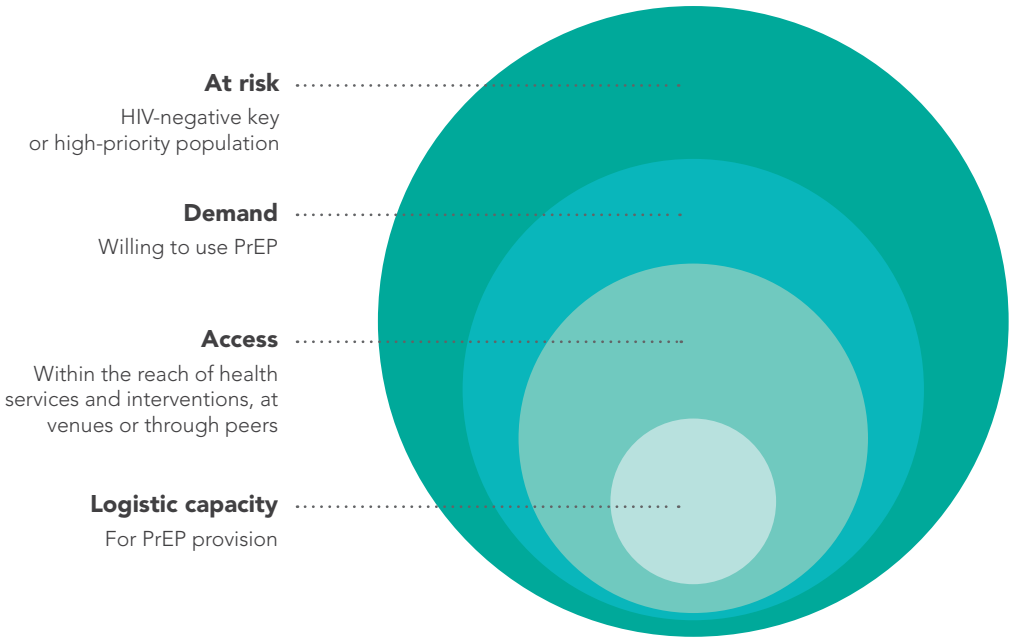
Estimating number at risk as a first level of target-setting

Oral pre-exposure prophylaxis (PrEP) is recommended by the World Health Organization (WHO) as an additional prevention option for people at substantial risk for HIV infection as part of combination HIV prevention. Providing access to PrEP is one of the five prevention pillars to meet global targets to reduce the number of new HIV infections by 2020 *(1)*. A key challenge when planning to implement PrEP is setting targets at national and subnational levels: how many high-risk individuals should a country or district aim to reach?

This document provides a method to estimate the number of individuals at substantial risk within specific key and high-priority populations (Figure 1). Estimating the total number at risk is a first step to gauging the need for HIV prevention, including PrEP, and is therefore a high-level target to work towards.

Final targets for implementation must address other considerations, such as demand, affordability, cost-effectiveness, service delivery capability, human rights challenges and political context.

Figure 1.
First level of target-setting: population members at substantial risk



Local or national PrEP targets may be used to suggest the extent of demand creation required to reach the population at substantial risk, as well as the scale-up of clinical services offering PrEP.

This guide is designed specifically for key populations, which are at elevated HIV risk in most countries globally *(2–5)*: men who have sex with men, transgender women, sex workers, and people who inject drugs. For the purposes of carrying out the estimates, male sex workers can be considered a subgroup of men who have sex with men.

The guide also addresses the high-priority population of adolescent girls and young women, who are at increased risk in several high-prevalence epidemics in southern and eastern Africa *(6)*.

The guide does not address targets for serodiscordant couples or other populations for which PrEP may be indicated. Similar approaches for other populations may be used, but this guide does not specifically address them or relevant data sources to inform those targets.

Many individuals in these groups may not be at substantial risk, because they engage in limited levels of risk behaviour or adhere to other effective preventive practices, or because of contextual factors that limit their risk, such as low HIV prevalence or high treatment coverage of people living with HIV. It is important to have a clear understanding of what part of the key or high-priority population of interest is at sufficient risk to make PrEP a useful addition to the prevention options available.

The method described here consists of six main steps:

1 Define the geographical area and population for target-setting.

2 Select an initial population size estimate for the key or high-priority population of interest.

3 If more than a year old, project the population size estimate to the desired year of PrEP implementation.

4 Narrow the population size estimate to the part of the population that is estimated to be HIV-negative.

5 If the population size estimate reflects only a part of the larger key or high-priority population, expand it (if the data permit).

6 Narrow the estimate to the part of the population that is at substantial HIV risk.

Step-by-step guidance and tools are provided for each step, with special emphasis on defining risk and estimating the proportion at risk.

WHO defines substantial risk as an HIV incidence of at least 3% in the absence of PrEP. Individuals at substantial risk may exist in locations where the overall incidence is lower. Thresholds for offering PrEP may vary depending on a variety of considerations, including resources, feasibility and demand.

Offering PrEP where the HIV incidence is greater than 3% is expected to save costs in many situations, but PrEP may still be cost-effective at lower incidence thresholds. For the purposes of estimates, however, it is often not possible to determine what part of the population would meet the definition of substantial risk using available data. Local cost-effectiveness studies may suggest a different incidence threshold based on local epidemiology, costs, and alternative prevention strategies already in place. For more details, see *WHO implementation tool for pre-exposure prophylaxis (PrEP) of HIV infection (7)*.

This guide provides tools for two approaches to define substantial risk:

- Exposures approach: substantial risk is defined as engaging in a minimum number of risky acts with a minimum number of high-risk partners. These minima are calculated by a simple mathematical model with reference to a specific HIV incidence threshold. The model takes into account the local epidemic context.
- Risk factors approach: all members of the key or high-priority population who are characterized by selected evidence-based risk factors are considered to be at substantial risk.

Other approaches may be possible, depending on the data available and the epidemic setting. This guide is meant to be informative and recognizes there is no one-size-fits-all approach.

Intended audience

This guide is designed for national programme planners to work with their surveillance, monitoring and evaluation specialists, other programme implementers and community members as some of the key steps require analysis of survey data or calculations that draw on estimates from surveys. It will also be useful for programme managers and partners who wish to determine the potential need for PrEP as part of planning for PrEP implementation or expansion.

Scope of method and estimates

How does this tool relate to other available tools?

A description of other available tools useful for PrEP programming is available *(8)*.

This tool uniquely presents an approach that uses population-based data to estimate the size of a population at risk for acquiring HIV that would benefit from PrEP. This provides a community-level target of people who should be reached and offered PrEP rather than a target primarily built from people who are already using other prevention services.

How does this method relate to cost-effectiveness?

This guide does not address economic analysis, but the estimated number at risk could serve as an input to economic analysis. If an economic analysis of PrEP is already available for the local setting, it could help to determine an appropriate incidence threshold for use in the exposures approach to defining risk.

How does this method relate to criteria for offering PrEP?

This guide is not intended for clinical use. Although it deals with defining substantial risk for the purpose of estimating the number of people at risk, the guide is not meant to be used to screen individuals in care settings.

This raises the question of whether the criteria used to define risk for the purposes of these estimates should be the same as the criteria used at the services level to screen for PrEP eligibility.

WHO recommends that PrEP is offered to all individuals who request it, and that local clinical guidelines should be in place to assess when an individual's risk may warrant recommending PrEP. Harmonizing definitions of risk between the estimates and clinical guidelines is ideal but may not always be possible due to data limitations.

For instance, the local surveys required to produce the estimates may not contain all the data needed to assess any clinical criteria that may be in place. Conversely, it may not be feasible for service providers in some settings to collect the same detailed behavioural information that may be available in surveys to develop a risk definition.

How does this method relate to procurement?

The aim of this guide is to estimate the total number of people at risk of acquiring HIV and thus to gauge the size of the population that could potentially benefit from PrEP. Due to other considerations (e.g. demand, cost-effectiveness, affordability), a PrEP programme may initially aim to reach only a part of the population at risk, in which case those smaller targets should be used for procurement purposes. This tool does not provide targets that can be use for procurement but rather targets for the number who should be reached. The PrEP-IT tool provides tables in which procurement targets can be estimated *(9)*.

A separate WHO guidance document on strategic planning for PrEP discusses how, where and to whom to offer PrEP services from an implementation perspective *(7)*.

How does this method relate to scaling up PrEP?

A phased introduction of PrEP, starting small and scaling up, with periodic assessments to improve quality, is a strong approach for building a robust programme.

The approaches described here could be used to support scale-up by estimating the size of particular subgroups at substantial risk (e.g. key or high-priority population members with sexually transmitted infections). In addition, it is advisable to calculate the total number at risk to gauge total need.

When planning for scale-up, care is needed to avoid stigmatizing any group of individuals, or the programme, by promoting PrEP specifically to a particular subgroup. PrEP should be available on demand to all people who think they will benefit. In some countries, promoting PrEP to sex workers made other people less likely to access it.

Is an estimate needed even if most people who seek PrEP are at high risk?

An estimate is needed. Many early-adopter providers have found that individuals who present for PrEP tend to have high-risk behaviours *(10)*. This is beneficial as it makes PrEP more cost-effective. However, there are more people who are at substantial risk who will not seek PrEP. For the purposes of target-setting, it is still important to estimate the total number of people at risk: there may be many other people at high risk who are not being reached, which may warrant demand-creation activities to increase the number of people accessing PrEP. A large gap will indicate a problem with the programme reach or the estimated target.

Should estimates be based on clients accessing services?

Several countries have approached estimates by using data from existing services, asking "How many of those who access services appear eligible for PrEP?" This is useful for planning, but it leaves out individuals at substantial risk who do not currently access services. Anecdotal evidence suggests that PrEP availability may draw people to services. The process described here is intended to gauge the greater need, which may suggest expanding the reach of existing services.

Will this tool help measure the impact of a PrEP programme?

No: the impact of PrEP uptake reaching targets can be modelled using GOALS and other packages.

Does this tool provide national targets?

No: the tool develops targets for geographic areas for which relevant data exist. Extrapolations of those targets to other areas is not part of this package.

Overview of method

Steps 1–4 should lead to an up-to-date estimate of the number of members of the key or high-priority population of interest who are HIV-negative when the study was conducted.

The method begins by selecting an initial population size estimate. Choosing an appropriate initial population size estimate is the most important step, as it greatly influences the resulting numbers at risk.

Because PrEP targets HIV-negative individuals, the population size estimate should be subset to people in the population who are HIV-negative. This requires an estimate of HIV prevalence in the key or high-priority population.

If data for the population size estimate were collected more than a year ago, the population size estimate should be projected forward to the desired year of PrEP implementation to account for growth in the general population over time. This can be achieved by applying census projections or population growth rates.

Applying a subgroup inflation factor is needed only if the population size estimate assessed the size of a particular subgroup rather than the whole key or high-priority population of interest. For example, population size estimates are often limited to individuals who frequent identifiable venues; or an estimate of the number of adolescent girls and young women could be limited to the number of girls and young women who attend schools. In this case, the population size estimate should be adjusted so it also reflects population members who are not part of the subgroup. Approaches and data sources to inflate the population size estimate are detailed in this step.

Step 5 narrows the estimate to population members who are at a level of risk great enough to make PrEP a cost-effective prevention option (substantial risk). Central to this guide are tools to define substantial risk in the local context and to estimate the proportion of the key or high-priority population at risk.

It is important to note these risk criteria are for the purposes of population-level estimates of the total number of people at risk. They may not be the same as any clinical assessment criteria in place to offer PrEP by service providers. They may be different due to differences in the types of data available for estimation versus the types of data available in service settings, or because of policies that lead to offering PrEP to a narrower group than the total number of people at substantial risk.

Two approaches are offered here to define risk. One approach is based on setting a threshold level of incidence and using a mathematical model (see the accompanying spreadsheet tools) to determine the minimum levels of risk behaviour required to reach that threshold.

The other approach is based on using risk factors (markers). Once risk criteria are defined, existing survey data are used to estimate the proportion of the key or high-priority population that meets the risk definition. If available surveys tested for HIV, then the risk proportion should be estimated from the subsample of HIV-negative participants.

Multiplying the updated population size estimate by the proportion at risk completes the calculation. Spreadsheet tools are provided to carry out both approaches to define risk and to document the estimates.

The final step is to review and refine the estimate, which involves comparing it with other information about the number of high-risk population members (e.g. programme data) to ensure it is a plausible number.

How the method was developed

The method was piloted drawing on available surveillance data from a range of cities in different key population groups. These sites were primarily in Central and South America and eastern and southern Africa. These trial exercises helped to clarify the practical issues that arise when applying the methods and how they can be resolved.

Sensitivity analysis was conducted to determine the impact of each step and the impact of reasonable variations in how each step is carried out. This examined choice of data sources, drawing on surveys using different sampling methods, the impact of weighting and pooling estimates, and other variations. The choice of initial population size estimate and the risk definition were the steps that tended to make the most impact.

A previous mathematical modelling framework for estimating incidence (modes of transmission model) was adapted to develop the Minimum Behaviours Calculator, used in the exposures approach to define risk criteria *(11)*. Spreadsheet tools were developed to organize and document the overall calculation.

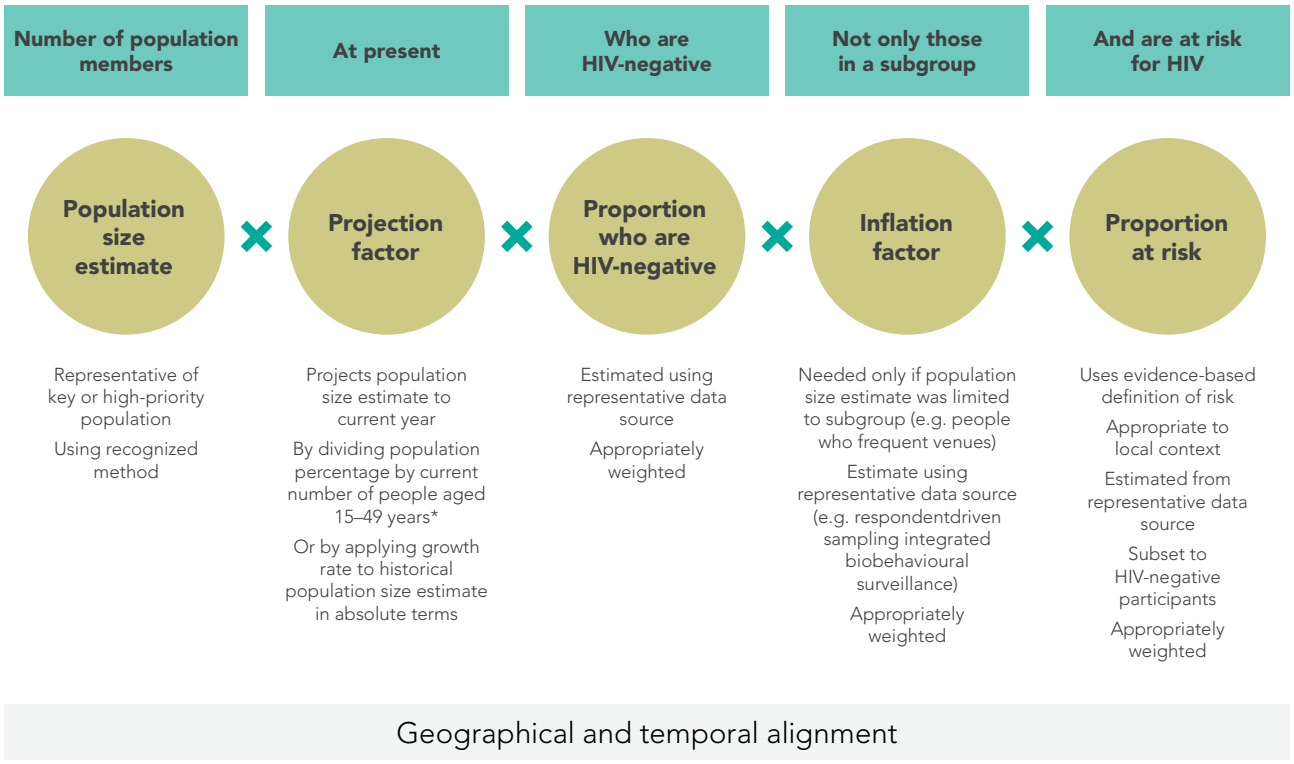
To provide an evidence base for the risk factors approach to define risk, literature reviews were conducted to identify risk factors for HIV that have consistent evidence in recent studies among men who have sex with men, transgender women, female sex workers, people who inject drugs, and adolescent girls and young women in low- and middle-income countries.

In addition, an analysis of cohort data from men who have sex with men participating in a sentinel surveillance programme in Central America was undertaken to illustrate how to use local data to identify risk factors. While this provides a step-by-step example of the analysis process, the risk factors identified did not predict HIV status well enough to recommend a universal risk definition for men who have sex with men. Therefore, bespoke risk analyses are preferred.

These analyses are described in the accompanying technical materials annex.

A global panel of experts participated in two virtual consultations, reviewed drafts of this guideline and provided input.

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



*People aged 15–49 years are the default for population size estimates.

Tools to carry out estimates

Table 1 lists the spreadsheet tools available to carry out the approaches described in this guide. The accompanying documents on target-setting and using the tools provide step-by-step examples.

Table 1.
Spreadsheet tools accompanying this guide

Tool	Description	Step in guide
PrEP estimates	Organizes and documents overall calculation of number of population members at risk	All steps
Project population size estimate	Projects population size estimate forward to present year	Step 3
Risk proportion	Organizes and documents calculation of risk proportion	Step 6B
Minimum Behaviours Calculator	Calculates number of exposures (risky partners and acts) required to meet specified incidence threshold, based on a force of infection model, as part of exposures approach to define substantial risk	Step 6B

How to use this guide

Establish a technical working group

Carrying out the estimates in the context of a technical working group can make them more accurate, improve their acceptability to stakeholders and affected communities, and ensure relevant data and perspectives are taken into account. A working group already in place (e.g. for surveillance, monitoring and evaluation, strategic information or estimates) may be well-suited.

Members of the working group should include:

- Local experts of the key or high-priority populations, including members of these populations.
- Individuals knowledgeable about existing data and estimates on these populations.
- Individuals knowledgeable about PrEP, HIV prevention and treatment services.

Gather data

Available data should be used to develop the estimates. Guidance is provided on how to move ahead in the absence of key data. The types of data that will be useful are:

- HIV prevalence and behavioural surveys and population size estimates for the populations of interest.
- Estimates of HIV prevalence, sexually transmitted infection prevalence, and antiretroviral therapy coverage in partner populations (see Table 4).
- Census estimates and projections.

Review the method and examples

Review the following sections before beginning in earnest:

- Overall method and each step.
- Flowcharts for key decisions for each step.
- Examples for men who have sex with men can be found in the accompanying documents on target-setting and using the tools, illustrating both approaches to defining risk. Review the examples for other populations of interest.

Step 1: Define the geographical area and population for target-setting

The entire calculation should be consistent with regard to the population being targeted in terms of the behavioural characteristics that define the population of interest, the age range and the geographical area.

The first step is to define the population and geographical area for which targets are needed. In analysis related to key populations, the following often need to be clarified:

- Will estimates be produced for men who have sex with men who frequent venues, or for the larger population of men who have sex with men?
- Will estimates be produced for sex workers who are venue-based or non-venue-based, or for both?
- Will estimates be produced for transgender women who engage in sex work, or for all transgender women?
- Will estimates be produced for adolescent girls and young women who engage in transactional sex, or for the larger population of adolescent girls and young women?

In this step, the entire key or high-priority population should be considered as data sources allow. Later steps will narrow the estimate to people at greater risk. One exception is age group, as surveys and other available data are often limited to adults aged over 15 or 18 years.

Similarly, the geographical area should be clearly defined. Common points to resolve often include the following:

- Are targets desired for the urban key or high-priority population, or will they include individuals in semi-urban and rural areas?
- Will targets be defined for key or high-priority population members who reside in the area or who frequent the area but may not live there?
- Will targets be at the national, subnational, district or other level?

The main data sources used in the calculation, such as surveys and size estimates, should be reviewed to determine how populations were defined and what part of the population of interest they represent. Defining the population and area is a balance between what is required for planning and the data available.

Census data are often needed as part of the calculations, so it is helpful to define targets for designated administrative areas.

Box 1: Aligning definitions and data sources for men who have sex with men in Guatemala City

Here we examine available data and alignment issues to estimate the number of men who have sex with men at substantial HIV risk in Guatemala City, as an initial target for a hypothetical start of PrEP interventions in 2017.

Data sources

- Initial population size estimate: the number of men who have sex with men was estimated in Guatemala City in 2009 by two different methods (capture–recapture and mapping and enumeration). Both aimed to estimate the number of men aged 18 years and over who frequented gay-identified venues. Both counted men who have sex with men and did not include transgender women.

- Estimating HIV prevalence and risk proportion: the most recent and representative survey of men who have sex with men available was the 2016 biobehavioural surveillance (BBS) using respondent-driven sampling (RDS). The survey was conducted at two sites (Coatepeque, Guatemala City) and sampled men who have sex with men and transgender women as a single population. At the Guatemala City site, there were 525 men who have sex with men and 189 transgender women participants. Results from the BBS can be used to estimate HIV prevalence because it included HIV testing. The behavioural data and syphilis test results were selected to estimate the proportion of men who have sex with men at risk. This may change as PrEP use increases in a given population over time: increased PrEP use leading to reduced condom use is being linked to increased prevalence of syphilis and other sexually transmitted infections.

- Subgroup inflation factor: although the 2009 population size estimates reflect men who have sex with men at venues, targets for PrEP may aim beyond venues, given increasing use of the internet and apps to meet sex partners. An adjustment factor is best taken from a representative survey that includes venue-going and non-venue-going men who have sex with men, such as an RDS BBS. The 2016 BBS questionnaire did not include questions on venue attendance, but a 2010 behavioural surveillance survey in Guatemala City that used RDS did include such a question.

- Population projections: the 2009 population size estimates correspond to 0.47–0.76% (enumeration) and 1.1% (capture–recapture) of the total number of males aged 15–49 years according to the population size estimate report. We can use these population percentages to obtain a more recent number of men who have sex with men by multiplying them by the number of males aged 15–49 years in Guatemala City as of 2015, the latest year available. To project from 2015 to 2017, annual growth rates for 2015 and 2016 for urban areas in Guatemala are available from the World Bank.

Considerations for alignment

- Geographical: all data sources (population size estimate, 2010 BBS, 2016 BBS) were designed to be representative of Guatemala City. The population size estimate examined venues in the city boundaries. BBS eligibility criteria required that participants lived, worked or studied in the city.

- Temporal: if planning for PrEP in 2017, the 2009 population size estimates can be projected forward. The 2010 venue inflation factor is a good source for adjusting the 2009 population size estimate as they are within a year of each other. Proportions based on the 2016 BBS are recent enough to be relevant.

- Age: all data sources were limited to age 18 years and over.

- Key population definition: the 2010 and 2016 BBS used eligibility criteria that limited participants to males who had anal sex with other males in the past 12 months. The 2009 population size estimates did not explicitly define men who have sex with men inclusion criteria. Therefore, some men who have sex with men counted by the population size estimate may not have had anal sex in the past 12 months, but there are no data to suggest how often this occurred. This is a potential source of misalignment that should be kept in mind when refining the final estimates. Also, because the BBS included men who have sex with men and transgender women, all analyses for the purposes of men who have sex with men estimates should exclude the transgender women participants.

Step 2: Select an initial population size estimate

Choice of the initial population size estimate greatly impacts the resulting number of individuals at risk. Often countries have more than one population size estimate available to choose from. Select an estimate:

- Conducted using a recognized population size estimate method:
- For key populations, this is generally a method for hard-to-reach populations.
- For adolescent girls and young women, this is generally based on the census or a general population survey.
- Representative of the population of interest (i.e. probability-based).
- Representative of the desired geographical area—this may be a city, a region or the national level, depending on the level of target-setting.
- As recent as possible.
- Accepted by local stakeholders.

Similar criteria were used by a recent global review of the availability and quality of population size estimates *(12)*.

For key populations, if local population size estimates are unavailable or seen as unreliable, a regional benchmark can be used. The median, range and interquartile range of published population size estimates in each region are available in the annual updated Joint United Nations Programme on HIV/AIDS (UNAIDS) Quick Start Guide for Spectrum as population percentages (e.g. percentage of males aged 15–49 years who form a part of specific key populations) *(13)*.

For example, a population size estimate could be developed by applying the median of published population percentages for the region to the census number of males or females aged 15–49 years in the area of interest.

Uncertainty bounds should be set to reflect a plausible range. Without local data, the bounds should be set with input from local stakeholders, but they should be at least ±20%.

Box 2: Selecting initial population size estimate for men who have sex with men in Guatemala City

The example in Box 1 looked at data sources available for Guatemala City. Two population size estimates were available, including one based on mapping and enumeration and another based on capture–recapture. Both population size estimates reflect men who have sex with men who frequent gay-identified venues. Both population size estimates used recognized methods.

The enumeration method summed the number of men who have sex with men found at each venue at a given time, without accounting for patterns over time, mobility across venues, or double-counting. In contrast, capture–recapture used a theoretical framework to estimate the number of all men who have sex with men who frequent venues over a given period (e.g. a number of weeks or months) and included an explicit method to avoid double-counting; in theory, this is more representative and was therefore selected to support the target setting exercise.

Box 3: Developing initial population size estimate in the absence of local data for men who have sex with men in Managua

- In the absence of a local population size estimate for Nicaragua, to develop a population size estimate for men who have sex with men in Managua we apply the median of population size estimates 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 population size estimates 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, but the number of males aged 15–49 years in Managua is unavailable.

- We assume the age distribution for the general male population is the same 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 of 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 population size estimate of 4035 [3228–4842] men who have sex with men in Managua in 2017.

Resources

- Regional ranges of published population size estimates *(12, 13)*.
- WHO/UNAIDS guidance on conducting population size estimates *(14)*.
- WHO/UNAIDS technical brief on population size estimates of men who have sex with men *(15)*.
- World Bank population growth rates.

Figure 3.
Selecting initial population size estimate

Step 3: Project the size estimate to the present year

The absolute number of individuals that form a part of the key or high-priority population of interest is likely to change over time along with patterns of growth in the general population. Population size estimates conducted more than a year before planned implementation of PrEP should be updated to account for population growth. Ideally, the size estimate will be based on the current year's census projection.

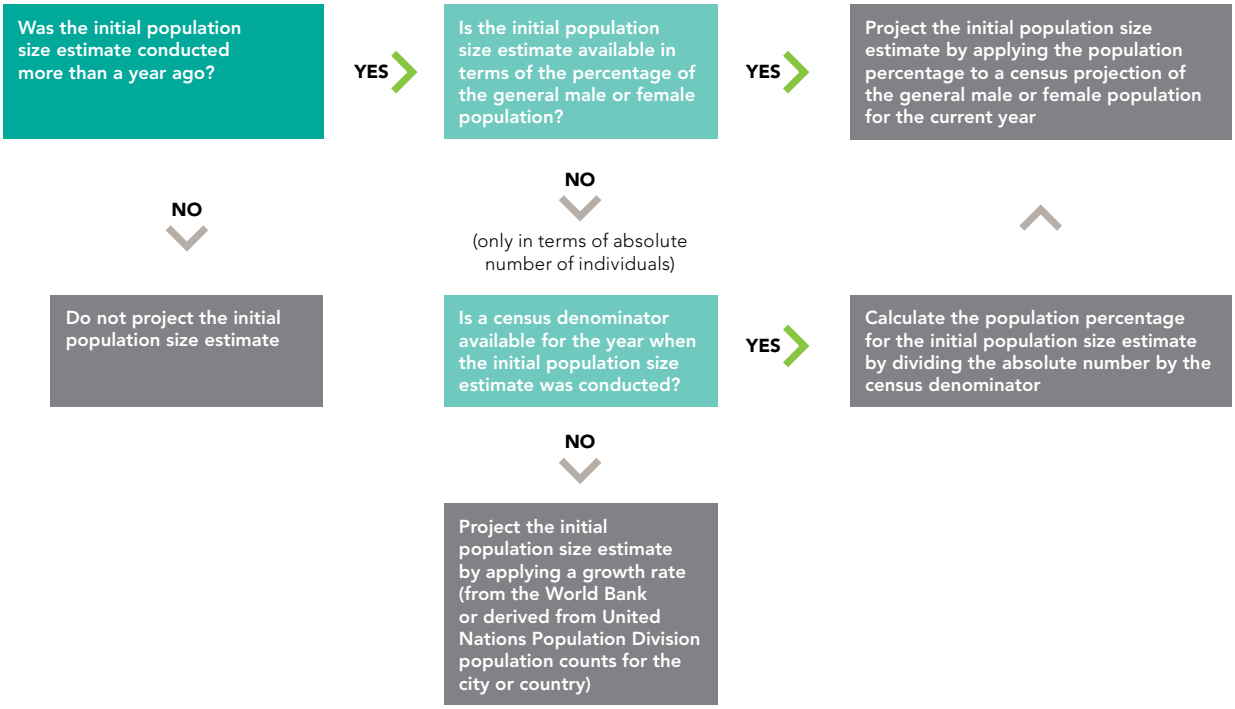
Three projection approaches are:

- If the population size estimate is available as a population percentage (e.g. males who inject drugs as 0.02% of all males aged 15–49 years), then this percentage should be multiplied by the census projection of the number of males aged 15–49 years in the target year.
- If the population percentage corresponding to the population size estimate is not available from the population size estimate study report, it can be calculated by dividing the population size estimate (e.g. 1500 males who inject drugs) by the census estimate of the number of males (or females, depending on the key population) aged 15–49 years from the same year as the population size estimate study. Age ranges of size estimates may vary from the standard 15–49 years, and other variable estimates should be harmonized to a standard age group to be decided during this target-setting activity.
- If the population percentage cannot be calculated from local census data because a census estimate or projection is not available by age and sex for the specific area where the population size estimate study was conducted, we can directly apply a growth rate to the number of key populations (e.g. 1500 males who inject drugs with 2% annual growth over a year would be 1500 × 1.02 = 1530). For growth over two or more years, the annual growth rates should be multiplied together. Average annual growth rates can be obtained from the World Bank or calculated from the total numbers of males (or females) aged 15–49 years at the start and end of the desired period. Use an urban growth rate if key populations are generally limited to urban areas.

Key points

- The older the initial population size estimate, the more important to project it to the current year.
- Online data from the United Nations Population Division *(16)* and the World Bank *(17)* can be helpful to determine population percentages and growth rates.
- When calculating the population percentage, be sure to use census data for the same year that the population size estimate was conducted.
- If the population of interest is mostly found in urban areas, use census figures for the urban population.

Figure 4.
Deciding whether and how to project the population size estimate



Box 4: Growth in number of people who inject drugs in South Africa

In 2015 a national programmatic mapping size estimation study estimated a range of 41 374–44 135 males and 31 489–34 402 females who inject drugs in South Africa. The population size estimate report was unable to present these numbers as population percentages for the same year of the estimate because census figures for 2015 were unavailable.

We use age and sex breakdowns from the United Nations Population Division online database to calculate a growth rate for 2015–2017 and use that growth rate to bring the population size estimate up to date:

	Females	Males
Aged 15–49 years in 2015	15 022 803	15 146 969
Aged 15–49 years in 2017	15 376 608	15 544 332

Based on the population counts, the number of females aged 15–49 years in the general population grew by 2.4% [(15 376 608 – 15 022 803)/15 022 803].

The male growth rate was 2.6% [(15 544 332 – 15 146 696)/15 146 969].

To project the 2015 population size estimates to 2017, we multiply the population size estimate range by the respective growth rate.

For females, we multiply the population size estimate (31 489–34 402) by 1.024, reflecting a growth rate of 2.4% as calculated above, which yields an estimated 32 231–35 212 females who inject drugs in 2017.

For males, 1.026 × (41 374–44 135) yields an estimated 42 459–45 293 males who inject drugs in 2017.

Box 5: Applying growth rates to update estimated number of men who have sex with men in Guatemala City

Guatemala City estimated the number of men who have sex with men to be 997–1601 using data collected by mapping and enumeration at venues in 2009. If PrEP were planned for 2017, we need to estimate population growth of men who have sex with men between 2009 and 2017.

According to the World Bank, annual urban population growth rates for Guatemala were as follows:

Year	2009	2010	2011	2012	2013	2014	2015	2016
Annual growth rate	3.082	3.065	3.046	3.026	2.997	2.969	2.932	2.896
Growth rate multiplier	1.03082	1.03065	1.03046	1.03026	1.02997	1.02969	1.02932	1.02896

We calculate the rate over the period 2009–2016 by multiplying the annual rates:

$1.0382 \times 1.03065 \times 1.03046 \times 1.03026 \times 1.02997 \times 1.02969 \times 1.02932 \times 1.02896 = 1.267.$

Then multiply the 2009 population size estimate by this period growth rate: the lower bound is $997 \times 1.267 = 1263$ and the upper bound is $1601 \times 1.267 = 2028$.

This leads to an estimated 1263–2028 men who have sex with men in Guatemala City in 2017.

Resources

- United Nations Population Division *(16)*.
- World Bank annual growth rates (overall and urban) *(17)*. Search for the country name and one of the following indicators, as appropriate: population growth, urban population growth, rural population growth.

Step 4: Multiply by the proportion HIV-negative

Population size estimates generally include both HIV-negative and HIV-positive members of the key or other high-priority population. Because PrEP is appropriate only for people who are HIV-negative, the population size estimate should be reduced to reflect only people who are HIV-negative. To do this, multiply the population size estimate by the proportion of the key or high-priority population that is estimated to be negative for HIV.

For example, if HIV prevalence among female sex workers in the city of interest is 4% (95% CI 2.5–5.5%), then the proportion that is HIV-negative is 96% (95% CI 94.5–97.5%).

Although this is straightforward, there are important points to keep in mind:

- Ensure the prevalence estimate is specific to the key or high-priority population in the geographical area of interest. The HIV prevalence estimate must be adjusted appropriately for the survey design. For example, when estimated from an RDS, time–location sampling (TLS) or demographic and health survey (DHS), the estimate should be weighted.
- Use prevalence estimates from representative surveys (e.g. RDS, TLS or DHS) rather than programme data, which can yield different prevalences due to programme use patterns.

If no local prevalence estimate is available, use a regional average or median from a recent UNAIDS global report, a national rate, a borrowed rate from a neighbouring country or published reviews of prevalence among the population of interest.

Resources

- UNAIDS data on HIV prevalence for 2017 *(18)*.
- Published regional estimates of HIV prevalence *(3–5, 19)*.

Box 6: Calculating proportion of HIV-negative men who have sex with men in Managua with BBS (preferred) versus programme data

A report from the 2016 BBS that sampled men who have sex with men and transgender women in Managua as a single population estimated HIV prevalence at 12.3% (95% CI 9.1–15.6%).

To determine HIV prevalence specifically for men who have sex with men, we use the software RDS-Analyst (RDS-2 estimator) and subset the estimate to men who have sex with men using the subset field (transwoman==0), where the transgender woman variable is from a survey item on gender identity.

This yields an estimated HIV prevalence of 12.9% (95% CI 5.7–20.1%), equivalent to an HIV-negative proportion of 0.87 (95% CI 0.80–0.94).

By contrast, among men who have sex with men consulting for the first time during a 2016 sentinel surveillance and prevention programme in Managua, HIV prevalence was 4.0% (95% CI 2.2–5.7%), equivalent to an HIV-negative proportion of 0.96 (95% CI 0.94–0.98).

Using the BBS estimate is preferred as it is more representative of the population of men who have sex with men in Managua.

Step 5: Multiply by an inflation factor if the population size estimate is limited to a subgroup

Population size estimates are often limited to the part of the population that is easier to reach. For key populations, this is often key population members at identifiable venues, such as gay bars, known sex work locations or shooting galleries.

Population size estimates limited to subgroups may be considerably smaller than those that assess the larger key or high-priority population. If plans for PrEP implementation are not limited to the same subgroup, it is important to increase the population size estimate accordingly using an inflation factor.

The inflation factor should be based on evidence of the proportion of population members who belong to the subgroup (e.g. who frequent venues). For key populations, often an inflator can be derived from RDS surveys.

Two examples are shown in Table 2, where the estimates have been adjusted for the RDS survey design. For adolescent girls and young women, population-level data may be needed, such as from a census or general population survey.

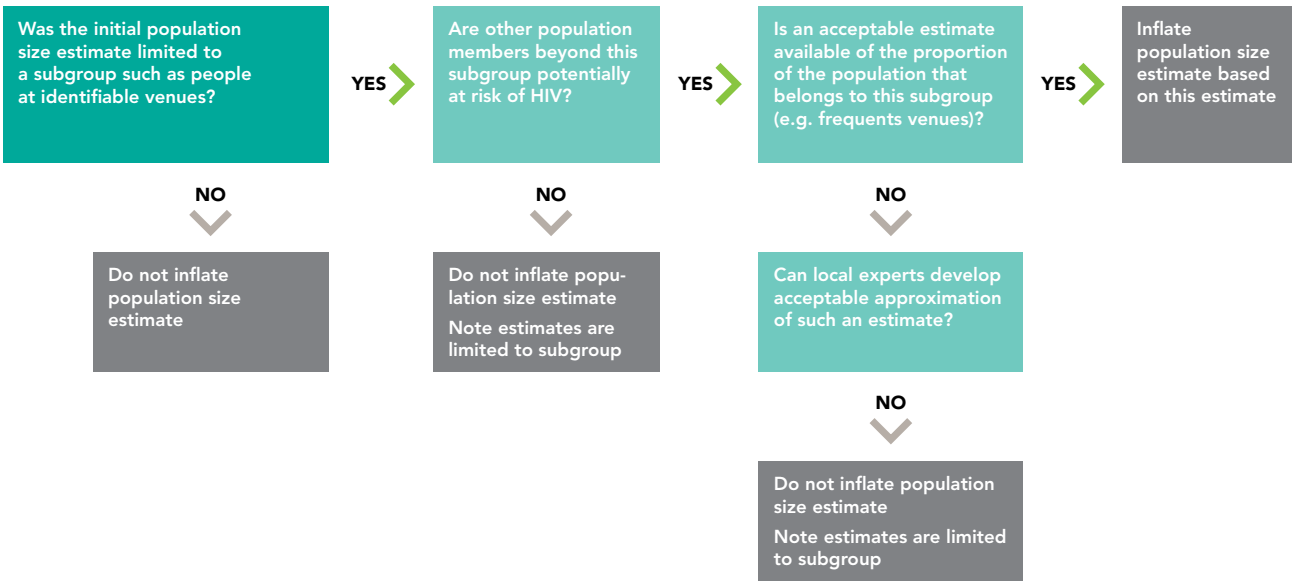
If data to develop an inflation factor are unavailable, consider asking local experts from the key or high-priority population for an informed estimate of the proportion of population members that belong to the subgroup. It is best to develop a range to ensure the uncertainty is captured in the resulting estimates.

If it appears there is no reliable data source to develop an inflation factor, the estimated number at risk may need to be limited to the subgroup reflected by the population size estimate.

Table 2.
Inflation factors for venue-based population size estimates

Tool	Question for inflator	Estimated proportion of Yes response (95% CI)	Venue-based inflation factor (1/estimate)
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, behavioural surveillance survey	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)

Figure 5.
Deciding whether and how to inflate population size estimate



Key points

- The inflation factor is only needed when the population size estimate was limited to a subgroup (e.g. individuals at venues).
- The inflation factor inflates the population size estimate so that population members who are not part of the subgroup are counted.
- Estimate the inflation factor using a representative data source (usually an RDS BBS).
- Subset the data to reflect the desired age group, geographical area and population of interest.
- Apply statistical weighting and adjustments as appropriate.
- If local data are not available, consider developing an expert consensus of the proportion of the population of interest believed to attend venues. Alternatively, work with the data available and acknowledge as a limitation that the estimated number of people at risk reflects only people accessible at venues.

Step 6: Multiply by the proportion at risk

Choose a representative data source

The proportion of the key or high-priority population at risk should be estimated from a representative data source, such as a probability survey (e.g. BBS using RDS, or DHS for adolescent girls and young women). If plans for PrEP centre on venue-going key populations, a TLS survey may be more appropriate. It is generally necessary to conduct new analyses of the survey dataset to produce the estimates.

Programme data are generally not a good source for estimating risk proportions, because attendees may have a very different risk profile than the larger key or high-priority population.

Define substantial risk of HIV infection

We describe two approaches to develop an operational definition of substantial risk and to estimate the proportion of the key or high-priority population that satisfies those criteria: the exposures approach and the risk factor approach. These approaches are not exhaustive, and there may be other strategies more appropriate to the local context.

WHO recommends that PrEP is considered as an additional prevention option for subgroups with at least 3% HIV incidence in the absence of PrEP. Cost-effectiveness analysis may suggest a different threshold based on local conditions. Where incidence estimates are not available, determining the number of people who face a 3% incidence is not easy.[1]

It is important to recognize that risk is context-dependent. Risk depends not only on a person's behaviours but also on HIV and sexually transmitted infection prevalence and treatment coverage among potential partners. The exposures approach described here accounts for differences in the epidemic context.

Both approaches require analysis of a behavioural survey so the dataset must be available. The approaches are generally equivalent in strength, relying on quality data availability. The limitations of each approach will be similar to those of the available data.

1 Even where incidence estimates are available, it is important to exclude population members who have little or no risk behaviour. For guidance on assessing incidence for planning purposes, see the WHO PrEP implementation tool (*7*).

Box 7: Defining substantial risk of HIV infection among individuals

Exposures approach

The exposures approach is based on the number of potential exposures to HIV in a given year. Survey participants are considered to be at substantial risk if they report a number of partners with risks of HIV exposure and potential exposures per partner (e.g. condomless sex, unsafe injections) that exceed specific thresholds. These minimum thresholds are themselves based on a threshold level of HIV incidence, which can be set as desired: substantial risk can be defined as the WHO-recommended 3% incidence, or another level can be determined to be more appropriate to the local setting.

The minimum numbers of partners and acts are calculated by a force of infection model. This model considers the local context of risk, including HIV prevalence, sexually transmitted infection prevalence, and antiretroviral therapy coverage in the partner population.

The threshold levels of risky partners and acts can be calculated for three different profiles of survey participants to improve accuracy, if data are available: people with a known HIV-positive partner; people with sexually transmitted infections; and people without sexually transmitted infections or of unknown sexually transmitted infection status.

The risk proportion is estimated as the proportion of the population that meets the minimum exposure thresholds.

Risk factors approach

The risk factors approach defines as at risk all individuals characterized by a risk factor or combination of risk factors shown to be associated with increased risk of HIV infection. This approach may be useful to support plans to implement PrEP in a particular subgroup—for example, as part of scale-up. There are important considerations related to the sensitivity and specificity of risk factors for purposes of prediction, and potential to produce or reinforce stigma.

Figure 6.
Selecting approach to define substantial risk

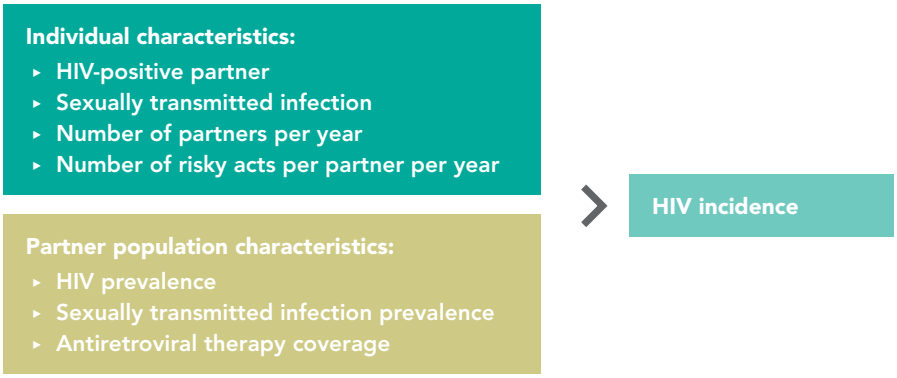
Exposures approach

The conceptual framework for the exposures approach is shown in Figure 7. HIV incidence for a susceptible individual is a function of the individual's risk behaviours, sexually transmitted infection status, and levels of HIV, sexually transmitted infections and antiretroviral therapy coverage among potential partners.

In this context, potential partners are not specific individuals mentioned in the survey but the larger partner population—for example, male partners of female sex workers are all male sex clients; male partners of adolescent girls and young women are all males in the general population; male partners of men who have sex with men are all men who have sex with men and potentially transgender women, and so on.

The mathematical model is described in the technical materials annex. A calculator to determine the minimum levels of risk behaviour is included in the annex on using the tools.

Figure 7.
Framework for HIV acquisition in exposures approach for defining risk criteria



Four steps are needed to define risk under the exposures approach (Figure 8):

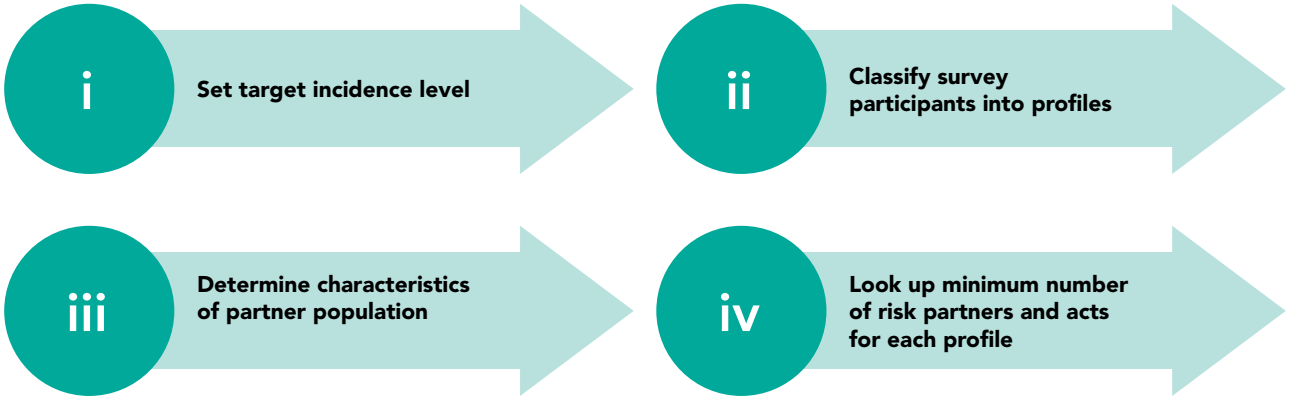
1 Decide on the threshold level of HIV incidence that will define substantial risk.

2 Determine which of the three survey participant profiles can be used, based on whether the survey contains information on sexually transmitted infection and HIV status of partners.

3 Determine levels of HIV and sexually transmitted infection prevalence and antiretroviral therapy coverage among the partner population.

4 Enter this information into the Minimum Behaviours Calculator to determine the number of partners with substantial risk, and the number of unprotected sex or injection acts per partner, required to meet the target incidence for each profile.

Figure 8.
Exposures approach to define substantial risk



Step i: Set target incidence level

The higher the incidence target, the greater the number of acts or partners at risk required for survey participants to be classified as substantial risk.

Enter the incidence target in the spreadsheet tools Minimum Behaviours Calculator.

Step ii: Classify survey participants into profiles

Survey participants are classified into one of three profiles based on presence of sexually transmitted infections and having a partner living with HIV. These factors influence the probability of HIV transmission during condomless sex and unsafe injection. Consequently, the profiles differ in the number of partners potentially engaged in risky behaviours and the number of unprotected acts per partner required to meet the target incidence level.

The profiles are as follows:

- Profile 1: participants with a recent partner living with HIV. Classify survey participants into this profile if the survey asked about the HIV status of partners for participants who said one or more partners was living with HIV or unknown status.
- Profile 2: participants with sexually transmitted infections and no known partners living with HIV. Classify participants into this profile if they have a positive sexually transmitted infection test result or have reported recent (past year) sexually transmitted infection symptoms. In this profile, all risky sexual acts are assumed to occur in the presence of sexually transmitted infection. No assumption is made about the HIV status of partners. The HIV status of partners is determined randomly, based on the HIV prevalence in the partner population.
- Profile 3: participants without sexually transmitted infections or of unknown sexually transmitted infection status, and with no known partners living with HIV. Classify participants into this if their sexually transmitted infection status is negative or in surveys where sexually transmitted infection status was not assessed, either by testing or by questionnaire items on recent symptoms. In this profile, the presence of sexually transmitted infections during risky sexual acts is a function of the prevalence of sexually transmitted infections in the partner population. No assumption is made about the HIV status of partners. The HIV status of partners is determined randomly, based on the HIV prevalence in the partner population.

Table 3 shows which profiles to use according to whether data for sexually transmitted infection and HIV status of partners are available. The Risk Proportion worksheet determines which profiles can be used.

Figure 9.
Classifying survey participants into profiles as part of exposures approach

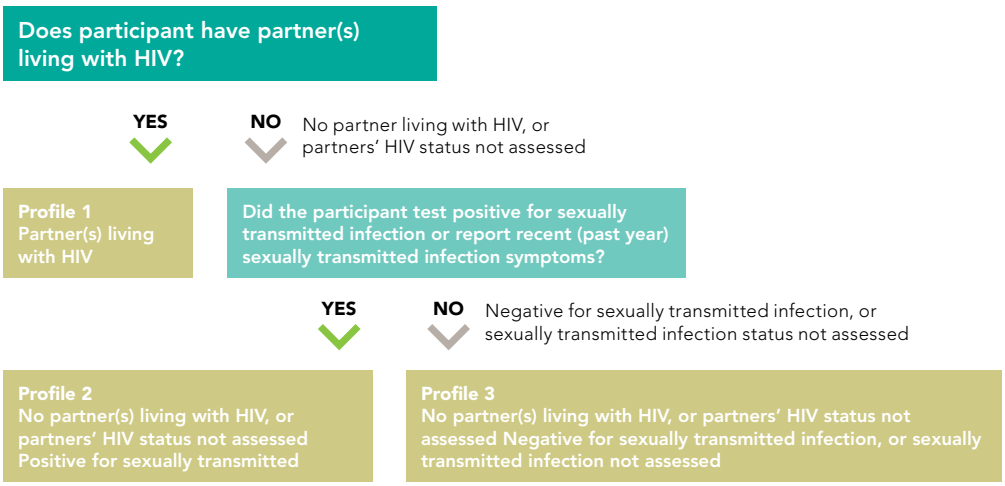


Table 3.
Selecting profiles based on available data

Were participants asked whether their partner(s) were living with HIV?	Were sexually transmitted infection tests conducted, or were participants asked about recent sexually transmitted infection symptoms?	Profiles		
		Men who have sex with men Transgender women Adolescent girls and young women	Female sex workers [a]	People who inject drugs [b]
Yes	Yes	1, 2, 3	2, 3	1, 3
Yes	No	1, 3	3	1, 3
No	No	2, 3	2, 3	3

a For female sex workers, having a partner living with HIV is not taken into account as it is assumed that risk derives primarily from sex clients, and female sex workers are unlikely to be aware of their clients' HIV status.

b For people who inject drugs, sexually transmitted infection status is not taken into account.

Step iii: Determine characteristics of partner population

Characteristics of the partner population influence the probability of HIV acquisition, including HIV prevalence, sexually transmitted infection prevalence, and antiretroviral therapy coverage among potential partners with HIV. The partner population refers to the pool of potential partners (Table 4).

Potential data sources for the partner characteristics are listed in Table 5.

These estimates should be entered into the Minimum Risk Behaviours Calculator worksheet.

Table 4.
Partner population definitions

Key or high-priority population	Partner population
Men who have sex with men	Men who have sex with menMen who have sex with men and transgender women if data on combined population are available
Transgender women	Men who have sex with men Men who have sex with men and transgender women if data on combined population are available
People who inject drugs	People who inject drugs [a]
Female sex workers	Male sex clients If unavailable, males aged 15–49 years in general population
Adolescent girls and young women	Males aged 15–49 years in general population (AIS/PHIA may help refine age range)

a For people who inject drugs, the model considers transmission from injection partners but not from sexual partners.

Step iv: Look up minimum numbers of partners and unprotected acts for each profile

Once the target incidence and estimates for the partner population have been determined, the Minimum Behaviours Calculator can be used to produce a table of the minimum number of risky partners, and number of unprotected acts per partner, for each profile.

These thresholds may be used to determine which participants in the survey(s) used to provide data are at substantial risk.

Table 6 defines the partners and acts to count.

Table 5.
Potential sources of data for estimating characteristics of partner population

Estimate for partner population	Data source
HIV prevalence	HIV prevalence survey in partner population For female sex workers, if there is no estimate of HIV prevalence among male sex clients, an alternative is to use the results from a population-based survey such as the AIDS Indicator Survey or Public Health Impact Assessments, or Spectrum estimate of HIV prevalence in the general male population
Sexually transmitted infection prevalence	Sexually transmitted infection prevalence in partner population If survey assessed few sexually transmitted infections in the population, consider estimating proportion of individuals who had either a positive sexually transmitted infection test result or recent (past 12 months) sexually transmitted infection symptoms if assessed
Antiretroviral therapy coverage	Antiretroviral therapy coverage estimate should reflect coverage of people living with HIV specifically in the partner population, including people living with HIV who are aware and unaware of their HIV-positive status If the partner population is the same as the key population (e.g. men who have sex with men, transgender women, people who inject drugs), antiretroviral therapy coverage could be calculated as: If this cannot be calculated because the number of individuals in the partner population on antiretroviral therapy is unknown, then antiretroviral therapy coverage in the general male population may be used, which could be calculated as: Total number of males on antiretroviral therapy/Spectrum estimate of total number of males living with HIV

Table 6.
Partners and risky acts counted in exposures approach

Key or high-priority population	How to measure number of partners	How to measure number of risky acts per partner
Men who have sex with men, transgender women	Total number of male partners with whom participant had unprotected anal intercourse	Unprotected anal intercourse with male or transgender woman partner
Female sex workers	Total number of male sex clients with whom participant had condomless sex	Condomless sex with male client
Adolescent girls and young women	Total number of male partners with whom participant had condomless sex	Condomless sex with male partner
People who inject drugs	Total number of other people with whom participant shared a needle, syringe, container, solution, cooker or other injection equipment If possible, include only receptive sharing	Unsafe injection
Reference period	All measures should refer to a recent reference period (e.g. 1, 3, 6 or 12 months) Counts should be annualized (converted to number in past 12 months) before using tables or tools	

Table 7.
Applying exposures approach with incomplete data

Data not collected	How to proceed
HIV status of partner(s)	Exposures approach can be used without using Profile 1
Sexually transmitted infection test results	Consider using self-report of recent sexually transmitted infection symptoms (past 12 months) instead of sexually transmitted infection test results If neither sexually transmitted infection results or symptoms are available, use Profile 3
Number of risky acts per partner per year	Exposures approach cannot be used without data on number of risky acts per partner Consider using risk factors approach instead
Number of risky partners per year	Exposures approach cannot be used without data on the number of risky partners Consider using risk factors approach instead
Sexually transmitted infection prevalence	If sexually transmitted infection testing was not conducted by survey, consider using a sexually transmitted infection prevalence estimate from a similar, neighbouring context
Antiretroviral therapy coverage	If estimate of antiretroviral coverage in target population is not available, consider assuming the same level of antiretroviral therapy coverage in general population It may be desirable to reduce this figure if the target population has more limited access to or uptake of antiretroviral therapy

Box 8: Annualizing numbers of partners with increased risk and acts per partner

In the exposures approach to define risk, numbers of risky partners and risky acts per partner reported by survey participants are compared with the minimum levels produced by the Minimum Behaviours Calculator. These minimum numbers are displayed in terms of the past 12 months.

Survey responses that refer to a period other than the past 12 months should be annualized before comparing them with the minimum numbers produced by the Minimum Behaviours Calculator.

One way to annualize is to multiply by where m is the reference period used by the survey in months.

For survey responses in terms of the past 6 months, annualize by multiplying both the number of risky partners and the number of risky acts per partner by or, which is 1.4.

For survey responses in terms of the past 30 days, annualize by multiplying by or, which is 3.5.

We do not simply multiply by 2 to annualize from the past 6 months to the past 12 months. This would lead to an inconsistent rate of total number of exposure acts over time. To understand this, imagine a survey used a reference period of the past 6 months. If we attempt to annualize by doubling the number of partners (P) and doubling the number of acts per partner (A), this implies the total number of acts during a year is $2P \times 2A = 4PA$, which is 4 times the number of acts in the past 6 months, which is not correct. If we multiply by instead of 2, then the total number of acts scales correctly: $P \times A = 2PA$.

This is only one way to annualize. It scales the number of partners and acts by the same amount. It does not, however, take into account the duration of partnerships. Depending on the information available, there may be better ways to annualize.

Risk factors approach

In the risk factors approach, the definition of substantial risk is based on one or more specific HIV risk factors. The data that inform this approach will come from biobehavioural surveys. For example, a risk definition for people who inject drugs might be:

People who inject drugs at substantial risk = Syphilis infection; or five or more drug injections using a shared needle, syringe or other injection equipment in the past week.

It may be desirable to define multiple levels of risk for planning purposes. For example:

- High risk: syphilis infection; or five or more drug injections using a shared needle, syringe or other injection equipment in the past week.
- Medium risk: five or more drug injections using a shared needle, syringe or other injection equipment in the past month.
- Low risk: none of the above.

These examples are based on two important risk factors for people who inject drugs: frequency of unsafe injection and having a sexually transmitted infection.

The choice of risk factors is flexible. It should be based on the best evidence available regarding local transmission patterns. Only risk factors that can be measured using surveys of the target population should be considered.

The risk factors approach may be useful when the data required for the exposures approach are not available. It may also be useful to support plans to implement PrEP in a specific subgroup.

Criteria for selecting risk factors

The following are important considerations when selecting risk factors:

- Empirically supported: draw on risk factors that are well-supported by empirical evidence, ideally from the local context.
- Proximal rather than distal: proximal risk factors directly and causally affect the probability of acquiring HIV, such as having a sexually transmitted infection or having a greater number of condomless sex or unsafe injection partners, in particular partners of unknown or positive HIV status who are not virally suppressed. Distal risk factors increase risk indirectly, such as having limited knowledge of HIV or engaging in sex work. Proximal risk factors are preferable when defining risk, because of their direct link to HIV. Distal risk factors can also contribute if there is evidence that they are highly predictive of acquiring HIV in the local context.
- Incident rather than prevalent: draw on risk factors that have been shown to be linked to incident (future) HIV infection. These are more likely to be causally related to HIV acquisition. Risk factors for prevalent (current) infection may be more subject to biases and more likely to reflect correlation rather than causation. In practice, incident and prevalent risk factors can be very different and even opposite *(20)*.
- Proven predictive power: even strong statistical associations may not translate to accurate prediction of future risk. Consider whether a risk factor is actually a good predictor of acquiring HIV. Does it have high sensitivity and specificity to predict HIV outcomes? See the section on cohort analysis in the technical materials annex for an example of highly (statistically) significant variables that nonetheless performed poorly at predicting HIV acquisition.

- Lower bar for higher-burden settings: an individual's risk varies by context. There is greater risk in areas where HIV and sexually transmitted infection prevalence among potential partners are higher and levels of antiretroviral therapy coverage are lower. When using risk factors to develop a risk definition, be aware of the context. Lower levels of risk factors should qualify as substantial risk in areas where HIV and sexually transmitted infections are more prevalent and antiretroviral therapy coverage is lower. Higher levels of risk factors should be required to qualify in areas where HIV and sexually transmitted infections are less prevalent and treatment coverage is greater.
- Avoid stigma: labelling an identifiable subgroup of the population as being at risk can increase stigma of that group and stigmatize a prevention programme. For example, early promotion of PrEP to male sex workers in some settings dampened interest in PrEP by other men who have sex with men. When selecting risk factors for target-setting, especially distal factors, avoid definitions that could generate or reinforce stigma *(21)*.

One way to identify risk factors is to consult the published literature to identify variables that appear to meet the criteria. A series of literature reviews were carried out to inform this guideline and offer a point of reference. Findings are summarized in the technical materials annex. Variables are listed that were found to be statistically associated with HIV among men who have sex with men, transgender women, female sex workers, people who inject drugs, and adolescent girls and young women in low- and middle-income countries, by studies that met predefined eligibility criteria.

The findings are organized by variables associated with incident versus prevalent HIV, by proximal versus distal risk factors, by the consistency of findings across studies, and by the strength of the estimated effect size.

Listed separately are risk factors identified by previous published systematic reviews.

Presence of a risk factor among the literature review findings does not guarantee it will predict risk well in every context. At the same time, a variable may be an excellent predictor even if it does not appear among the findings. Be sure to select variables considered to be locally relevant.

See Examples 2, 4, 5 and 6 in the annex on using the tools for a step-by-step illustration of using the risk factors approach to carry out the estimates.

Which risk factors have the strongest evidence?

The proximal risk factors in Box 9 are most directly linked to HIV transmission. A risk definition for PrEP target-setting should always include at least one of these risk factors. It is advisable to prioritize a greater frequency of these risk behaviours as this increases exposure.

Box 9: Proximal risk factors for key populations and adolescent girls and young women

- Men who have sex with men and transgender women: condomless anal sex, especially receptive anal sex, with a male or transgender woman partner of unknown or positive HIV status.
- Female sex workers: condomless anal or vaginal sex with a male sex client of unknown or positive HIV status.
- People who inject drugs: receptive sharing of needles, syringes or other injecting equipment or drug solutions when injecting drugs with a partner of unknown or positive HIV status.
- Adolescent girls and young women: condomless anal or vaginal sex with a male partner of unknown or positive HIV status.
- Having a sexually transmitted infection such as syphilis, herpes, chlamydia, gonorrhoea or bacterial vaginosis is a risk factor for sexual transmission of HIV in all groups.

Greater prevalence of HIV or other sexually transmitted infections in a specific geographical area may be considered as part of a risk definition as it increases the likelihood of HIV exposure and transmission.

Other behaviours and legal and social conditions can be important determinants of HIV infection in specific contexts. Some examples include engaging in sex work and non-injection drug use.

When choosing to include distal risk factors in a risk definition for target-setting, it may be helpful to review evidence from other settings. For this purpose, a review of risk factors is included in the technical materials annex.

The review was specifically for men who have sex with men, transgender women, female sex workers, people who inject drugs, and adolescent girls and young women in low- and middle-income countries. Findings include evidence of risk factors identified by studies conducted over the past 5 or 10 years (depending on the population). These include measures of effect size and whether findings were consistent across studies.

These review findings are intended only as a point of reference. There are important limitations. For example, the factors identified are a function of what researchers chose to examine over the time period and differences in methods, such as variables controlled for in regression models.

Notably, recent evidence is more limited for people who inject drugs and adolescent girls and young women.

Other factors may be relevant in specific contexts.

Box 10: Distal risk factors identified by recent research in low- and middle-income countries

Men who have sex with men

- Lower HIV knowledge.
- Sex work or transactional sex.
- Forced to have sex.
- Cohabitating or stable partner.
- Discrimination related to men who have sex with men.

Transgender women

- Sex work.

Female sex workers

- Use of drugs.
- Lower education.
- Forced to have sex.
- Lower price charged for sexual services.
- Number of children, pregnancies or abortions.
- Type of workplace or modality of sex work.

These risk factors are consistently identified by two or more recent studies. See the technical materials annex for other variables that did not meet the consistency definition and other details.

Identifying locally relevant risk factors through risk factor analysis

Risk factor analysis conducted using local data is more likely to lead to a risk definition relevant in the local context. Carrying out risk factor analysis can help determine the importance of distal risk factors.

Proximal risk factors (e.g. sexually transmitted infections, condomless sex, unsafe injections) are unlikely to be any different in the local context, but risk factor analysis can help to determine their effect size and usefulness as predictors in the local setting.

Advantages and limitations of conducting risk factor analysis with local data to develop a risk definition are listed in Table 8.

Table 8.
Advantages and limitations of identifying risk factors with a local survey

Advantages	Limitations
Local data may be more relevant to the local context and more up to date than published evidence	Variables that can be examined are limited to information collected by the survey
Local data may be of higher quality than published evidence, depending on methods used	Statistical power might be insufficient to identify risk factors, depending on sample size and local transmission patterns
Conducting risk factor analysis may provide an opportunity to examine variables not examined elsewhere	If the local survey is not representative of the target population for PrEP, risk factors identified may not be as relevant as published evidence
Evidence may be more acceptable to stakeholders	

Data sources for risk factor analysis

Data used to conduct risk factor analysis should meet basic quality criteria, including the following:

- The data were collected following a protocol based on sound methodology and approved by an ethical review board.
- The protocol includes a clear definition of the target population, which was strictly enforced during data collection through explicit eligibility criteria and screening procedures.
- The survey instrument was designed with input from the target population and was pilot-tested before data collection.
- All field staff underwent training. Appropriate supervision and data quality assurance mechanisms were in place.
- Laboratory procedures, including collection, storage and testing of specimens, were designed and carried out by appropriately trained laboratory specialists.
- The sample size provides sufficient statistical power to detect differences in HIV infection between subgroups.
- The dataset is well-documented, so all items and responses can be linked to the data collection instruments.
- All data elements required to apply standard statistical adjustments (e.g. survey weights, clustering adjustments to standard errors) are available in the dataset.

Recognizing bias

Risk factor analysis can be subject to several sources of bias (systematic error) that may cause variables to appear predictive of HIV when they are not, and vice versa.

When selecting a data source, consider the potential for the following:

- Confounding: this is often called "omitted variable bias" or "third variable bias". It occurs due to failure to control for another cause of the outcome that is correlated with the hypothesized risk factor. Developing a conceptual framework that describes how the hypothesized risk factor and other variables are expected to lead to HIV risk can help identify ways to reduce confounding.
- Temporal ambiguity: temporal relationships can cause bias in cross-sectional surveys. For example, some analyses have found a positive association between HIV and condom use *(22)*; one explanation is that survey participants aware of their HIV infection may use condoms more than others to protect partners. In this case, a potential confounder would be knowledge of HIV infection, which influenced the participant to use a condom.
- Social desirability bias: survey participants, particularly in face-to-face interviews, may underreport risk behaviours that are seen as socially unacceptable, with the result that they are less likely to be identified as risk factors.
- Selection bias: non-random survey recruitment can lead to greater or reduced representation of a source of exposure. For example, studies of transgender women limited to transgender women sex workers cannot identify sex work as a risk factor; studies of female sex workers limited to entertainment venues cannot identify selling sex on the street as a risk factor; and studies of adolescent girls and young women limited to students cannot identify being out of school as a risk factor. As a more subtle example, in a peer-referral study of men who have sex with men, men who have sex with men and who do not use drugs may fail to recruit men who have sex with men who do use drugs, or vice versa.
- Recall bias: risk behaviours may be under- or overreported due to difficulties remembering when they occurred. This is more likely when survey questions use longer reference periods (e.g. past 12 months versus past 30 days).
- Censoring: when participants in a cohort study with a specific risk factor or outcome are more likely to drop out than others, estimates of association may be biased and overall statistical power reduced.

More general discussions of bias are available *(23–25)*.

Methods for risk factor analysis

There are many statistical approaches to carrying out risk factor analysis. The most common approach is to use data from a cross-sectional survey that links a questionnaire with HIV test results, for example from a BBS, DHS or AIDS indicator survey. An advantage is that questionnaires in such surveys are often rich in variables to examine as possible risk factors. A disadvantage is that cross-sectional data reflect a snapshot in time and are more vulnerable to temporal ambiguity or confounding.

These sources of bias can be reduced by the following:

- Exclude participants who knew they were living with HIV at the time of the survey, since their behaviours may have changed as a result of their diagnosis.
- If the sample size permits, limit the analysis to the subsample of participants who underwent HIV testing recently (e.g. past year) and received a negative test result. In these individuals, any positive test results will be more linked to recent behaviours and other risk characteristics. Limiting analysis to individuals who get tested more frequently may introduce other biases, so this should be conducted as a sensitivity analysis.
- Often a stronger approach (compared with using cross-sectional data) is to use longitudinal data from study cohorts or participants in services or programmes over time. Longitudinal analysis limits temporal ambiguity but is still subject to other sources of confounding (given the lack of a randomized comparison group). Longitudinal analysis is subject to censoring, so it is important to apply appropriate statistical techniques such as survival analysis. Screening tools are often developed using survival analysis *(26–29)*. Consult a practical guide to conducting survival analysis with Stata software (StataCorp, College Station, TX, USA) *(30)*.

Risk factor analysis should evaluate the predictive performance of identified predictors. Common measures include sensitivity, specificity, the area under the receiver–operator curve, and Harrell's concordance statistic. In large samples, the dataset can be randomly divided into derivation and validation subsamples to derive the estimates and test their predictive performance, respectively *(28)*.

It is important to apply appropriate statistical adjustments, such as sampling weights, depending on the survey design. After analysis, it is important to evaluate whether regression models meet theoretical assumptions by examining diagnostics. This typically includes assessing model fit, assessing whether the functional form is adequate, and detecting influential observations. In Cox proportional hazards models, the proportional hazards assumption should be evaluated.

Developing risk definition based on risk factor analysis

Once a regression model has been fit to the data, the findings can be used to develop a risk definition, which can be used to estimate the proportion at risk. In broad terms, there are two ways this can be done.

One option is to define risk based on one or more of the risk factors identified. This is probably the most realistic option if the data used to conduct the risk factor analysis are different from the data that will be used to estimate the risk proportion. For example, when using programme or cohort data for risk factor analysis, it may make sense to use more representative data, such as a BBS or DHS survey, to estimate the proportion at risk. An example is a recent study applying French national PrEP criteria to a representative survey of men who have sex with men in Ireland to estimate the proportion of Irish men who have sex with men at risk *(31)*.

If the same data are used to conduct risk factor analysis and to estimate the risk proportion, then the complete regression equation (or risk score) can be used to determine a more precise risk level of each participant. For example, the proportion of the population at low, medium and high risk could be estimated by classifying participants according to cut-off points along the risk score.

See the technical materials annex for two examples of risk factor analysis: using services data to identify risk factors for incident HIV among men who have sex with men; and using nationally representative population-based HIV impact assessment (PHIA) surveys to identify risk factors for HIV prevalence among adolescent girls and young women.

Developing risk definition for adolescent girls and young women

Developing risk criteria for the purposes of PrEP targeting is more challenging for adolescent girls and young women than for key populations. While many members of a key population are by definition engaged in an activity that increases HIV risk, the population of adolescent girls and young women includes all adolescent girls and young women, many of whom may not be engaged in any risk behaviour or subject to a context of vulnerability.

Even in areas where as a group adolescent girls and young women have elevated HIV prevalence, many or even most individual adolescent girls and young women may in fact face minimal risk.

For PrEP targeting, it is critical to adopt risk criteria that are highly specific, so that targets do not include large segments of the population that would not benefit from PrEP. In this respect, it is important to carefully consider the local HIV context when defining risk, especially subnational differences in HIV burden. This is because adolescent girls and young women, as a demographic group defined by age and sex, reside nearly everywhere—in urban and rural areas, and in high-prevalence and low-prevalence areas.

Adolescent girls and young women who reside and find male sexual partners in areas with relatively low HIV burden are unlikely to be at high risk because their partners are unlikely to be living with HIV.

In sum, a risk definition for adolescent girls and young women should consider:

- The geographical context of HIV, specifically levels of HIV prevalence among males (the pool of potential partners for adolescent girls and young women).
- Risk behaviours and vulnerability factors that are closely linked to HIV acquisition.

Evidence to inform risk definitions for adolescent girls and young women

There is currently limited research to support a particular set of HIV risk criteria for adolescent girls and young women in high-prevalence settings. One resource is the risk score developed by Pintye et al. to identify females at antenatal clinics who might benefit from PrEP *(28)*. This score is based on five predictors: lifetime number of sex partners, having a male partner of unknown HIV status, syphilis, bacterial vaginosis, and vaginal candidiasis. The index had a sensitivity of 64% in the study and its specificity was not reported. It is unclear whether it would accurately predict HIV among adolescent girls and young women outside antenatal care settings.

A second resource is the technical materials annex, which lists HIV risk factors reported in the literature through 2018 specifically among young women. The available evidence, although from a small number of studies, points to sexually transmitted infections, number of lifetime and recent sex partners, and HIV status of partners as important risk factors. As in Pintye's index, these are proximal risk factors linked closely to HIV acquisition.

PHIA surveys from Malawi and Zambia were analysed for this guide to identify risk factors nationally representative of adolescent girls and young women. Several risk factors were strongly associated with HIV, but none of the resulting risk models predicted HIV status well: sensitivity was 70% and specificity 52%, even when analysis was limited to high-prevalence areas. This subpar performance may be due to the fact that PHIA surveys are cross-sectional and thus reflect prevalent HIV. Consequently, it was not possible to recommend a specific risk definition for adolescent girls and young women based on the PHIA analysis. See the technical materials annex for details of the methods and findings.

Looking ahead, cohort studies of adolescent girls and young women in areas of elevated HIV burden will be needed to develop more effective risk criteria for target-setting.

Box 11: Risk factors for HIV among adolescent girls and young women identified in recent literature

- Sexually transmitted infection.
- Large age gap (five or more years) with older sexual partners.
- Partners of unknown or positive HIV status.
- Condomless sex with such partners.
- Greater number of lifetime or recent sexual partners.
- Non-cohabitating partners.
- Residing in an area with elevated HIV prevalence among males.

Data sources to estimate risk proportion for adolescent girls and young women

Ideally the risk proportion should be estimated from a survey that is representative of adolescent girls and young women and includes data on sexual risk behaviours and sexually transmitted infections.

PHIA surveys are one such data source. These are currently available from 14 countries in sub-Saharan Africa.[2] PHIA can be used to estimate levels of risk behaviour, HIV prevalence, and often sexually transmitted infection prevalence among males and females aged 15 years and over. Data can be broken down into subnational areas.

Considerations when estimating a risk proportion from a PHIA survey include the following:

- Consider developing estimates of HIV prevalence among males in different parts of the country (zones, provinces, districts) using the male survey data. Then HIV prevalence in the local area can be built into the risk definition for adolescent girls and young women.
- Many survey participants are not sexually experienced, so consider classifying them as not at risk.

2 See https://phia.icap.columbia.edu/.

- Test results or question items on self-reported sexually transmitted infection can assess sexually transmitted infections.
- Question items on the last three sexual partners can be used to determine the age gap with older male partners.
- Other items on the last three partners that may be useful to define risk include knowledge of HIV status, condom use, and whether the partner lives away from the household and has other wives or live-in partners.
- See the PHIA Data Use Manual for technical details on how to *(32)*:
- Apply survey weights so that estimates are representative.
- Combine the main survey data (adult dataset) with the HIV and sexually transmitted infection testing data (biological dataset) to limit the risk proportion to HIV-negative adolescent girls and young women.

Box 12: Estimating risk proportions and targets for adolescent girls and young women in Zambia using PHIA survey

We used the adult and biological datasets from the Zambia 2016 PHIA. We determined the HIV prevalence among adult males ranged from 3.4% to 13.7% across Zambia's 10 provinces. Male HIV prevalence was 5% or less in 3 provinces and over 5% in 7 provinces.

As an example of a risk definition for adolescent girls and young women, we used the following:

- High risk:
- Active syphilis test result or self-reported sexually transmitted infection diagnosis in the past year.
- Resides in a province with over 5% HIV prevalence and had 2 or more male sex partners in the last year.
- Condomless sex in the past year with a male partner who is of unknown or positive HIV status or is five or more years older.
- Medium risk:
- Resides in a province with 5% or less HIV prevalence and had 2 or more male sex partners in the last year.
- Condomless sex in the past year with a male partner who is of unknown or positive HIV status or is five or more years older.

We limited the adult dataset to female survey participants aged 15–24 years (adolescent girls and young women) with a valid HIV test result and merged in HIV and sexually transmitted infection test results from the biological dataset.

Following the PHIA Data Use Manual's indications for applying jackknife survey weights (32), we found that nationally, 3.22% (95% CI 2.61–3.97%) of adolescent girls and young women were HIV-negative and met our high-risk definition; 0.28% (95% CI 0.16–0.49%) of adolescent girls and young women were HIV-negative, met our medium-risk definition and did not meet the high-risk definition.

According to the World Population Prospects, we found that the projected number of females aged 15–24 years residing in Zambia in 2021 is 2 003 000. Multiplying the risk proportions with this census projection led to a projected 64 497 (95% CI 52 278–79 519) adolescent girls and young women at high risk and 5 608 (95% CI 3 205–9 815) adolescent girls and young women at medium risk in Zambia in 2021. See the annex on using the tools for details.

This example includes some arbitrary decisions (e.g. 5% prevalence cut-off for provinces). Other risk definitions are possible and would best be developed by a national technical working group.

Classifying risk level of survey participants

This step is needed only when the risk proportion will be estimated by conducting additional analysis of data from a survey. Construct variables corresponding to the risk level of each participant based on the risk criteria that have been defined in the previous steps. The variables should generally be dichotomous: equal to 1 if the participant is in the risk level, otherwise 0.

Generally these indicator variables are best constructed by writing a program in statistical software or a spreadsheet formula so the analysis is well-documented and can be repeated or modified if needed.

As part of this process, it may be useful to review the recommendations for reviewing survey data in Annex 1.

Box 13: Constructing risk-level indicators

We assume risk criteria for people who inject drugs have been defined as follows:

- High-risk: syphilis infection, or five or more drug injections using a shared needle in the past week.
- Medium risk: five or more drug injections using a shared needle in the past month.
- Low risk: none of the above.

We ensure risk variables are defined as dichotomous 0/1 indicators. We assume variable names as follows:

- syphilis = syphilis infection.
- inj5_week = five or more injections using a shared needle in the past week.
- inj5_month = five or more drug injections using a shared needle in the past month.

Using these 0/1 risk factor variables, we define indicator variables corresponding to each risk level. The syntax will depend on the software used, but it should read approximately as:

- high_risk = 1 if syphilis = 1 OR inj5_week = 1, otherwise 0.
- med_risk = 1 if high_risk = 0 AND inj5_month = 1, otherwise 0.
- low_risk = high_risk = 0 AND med_risk = 0, otherwise 0.

Dichotomous risk variables have the advantage that the mean is the same as the proportion at the risk level. However, proportions at risk can also be estimated by defining one categorical risk variable with the three levels. Be sure to check missing values are handled appropriately so that participants for whom the risk level cannot be determined are excluded from the analysis.

Estimating proportion at risk among HIV-negative participants

After the dichotomous risk level variables are defined, we estimate the population proportion of each risk level. Be sure to subset the estimate to HIV-negative participants and to the intended key population group (e.g. male or female people who inject drugs, men who have sex with men or transgender women, adolescent girls and young women aged 15–24 years). We apply weighting and statistical adjustments appropriate to the study design.

Refer to the software instructions to subset the data appropriately. For example, in RDS-Analyst, subset the data using the subset field of the frequency estimates dialogue.

If using Stata survey procedures (svy), do not to use if. Instead, use the subpop option (e.g. svy, subpop(transwoman==0): tab risk_indicator, ci).

Uncertainty intervals

It is important to communicate the uncertainty of the estimated number of individuals at risk. There is uncertainty in the estimate because the inputs to the calculation below are estimates:

- Initial population size estimate, projected as needed.
- The proportion that are HIV-negative.
- The venue inflation factor, if included.
- The proportion at risk.

In the spreadsheet tools that accompany this guide, an uncertainty range may be specified for each of these estimates. The uncertainty range for the estimated number at risk is then calculated by the delta method.

The delta method assumes the four input estimates are based on large samples (asymptotic assumption) and their uncertainty ranges are 95% confidence intervals. The latter assumption may not be met when consensus ranges are used.

The calculation assumes no covariance among the inputs, since such information is seldom available.

For these reasons, the final uncertainty range should be considered an approximation rather than a proper confidence interval.

Key points

- Estimate the population proportion at each risk level.
- Subset to HIV-negative participants.
- Subset to the desired age, geographical and key population group.
- Apply statistical weighting and adjustments as appropriate.
- Communicate the uncertainty of the estimate with an uncertainty interval.

Step 7: Refine the targets

Comparing the target with programme data can help to assess its face validity.

The most useful comparisons are with programmes that have high coverage of the same target population and that perform well at reaching high-risk population members.

The estimated number at risk is likely to be significantly larger than the number of beneficiaries of programmes with low coverage of high-risk individuals.

Data from programmes that primarily serve people living with HIV cannot provide a point of comparison for PrEP targets.

Other factors relevant to setting targets for PrEP, apart from the level of risk, may be relevant to planning but are beyond the scope of this document. They include:

- Demand: willingness to use PrEP.
- Feasibility of reaching high-risk members of the target population.
- Other service limitations.

References

1 Prevention gap report 2016. Geneva: Joint United Nations Programme on HIV/AIDS; 2016 (http://www.unaids.org/sites/default/files/media_asset/2016-prevention-gap-report_en.pdf).

2 Poteat T, Scheim A, Xavier J, Reisner S, Baral S. Global epidemiology of HIV infection and related syndemics affecting transgender people. J Acquir Immune Defic Syndr. 2016;72:S210–S219.

3 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.

4 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.

5 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.

6 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.

7 WHO implementation tool for pre-exposure prophylaxis (PrEP) of HIV infection: module 9—strategic planning. Geneva: World Health Organization; 2017 (https://www.who.int/hiv/pub/prep/prep-implementation-tool/en/).

8 A user's guide to PrEP tools. New York: AVAC (https://www.prepwatch.org/users-guide-prep-tools-use/).

9 PrEP-it user guide version 1.0. Washington, DC: United States President's Plan for Emergency AIDS Relief;2019 (https://www.prepwatch.org/wp-content/uploads/2019/11/PrEP-it_UserGuide_sept24_2019.pdf).

10 Grant RM, Glidden DV. HIV moments and pre-exposure prophylaxis. Lancet. 2016;387:1507–1508.

11 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.

12 Sabin K, Zhao J, Garcia Calleja JM, Sheng Y, Arias Garcia S, Reinisch A, et al. Availability and quality of size estimations of female sex workers, men who have sex with men, people who inject drugs and transgender women in low- and middle-income countries. PLoS One. 2016;11:e0155150.

13 National HIV estimates file. Geneva: Joint United Nations Programme on HIV/AIDS (https://www.unaids.org/en/dataanalysis/datatools/spectrum-epp).

14 Guidelines on estimating the size of populations most at risk to HIV. Geneva: Joint United Nations Programme on HIV/AIDS and World Health Organization; 2010 (https://www.unaids.org/en/resources/documents/2011/2011_Estimating_Populations).

15 Recommended population size estimates of men who have sex with men. Geneva: World Health Organization and Joint United Nations Programme on HIV/AIDS; 2020 (https://apps.who.int/iris/rest/bitstreams/1321427/retrieve).

16 World population prospects. New York: United Nations Department of Economic and Social Affairs (https://population.un.org/wpp/).

17 Data. Washington,DC: World Bank (https://data.worldbank.org/indicator/SP.POP.TOTL).

18 UNAIDS data 2017. Geneva: Joint United Nations Programme on HIV/AIDS; 2017 (http://www.unaids.org/en/resources/documents/2017/2017_data_book).

19 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.

20 Tang W, Babu GR, Li J, Zhang Y, Fu G, Huan X, et al. The difference between HIV and syphilis prevalence and incidence cases: results from a cohort study in Nanjing, China, 2008–2010. Int J STD AIDS. 2015;26:648–655.

21 Haire BG. Preexposure prophylaxis-related stigma: strategies to improve uptake and adherence—a narrative review. HIV/AIDS. 2015;7:241–249.

22 Johnston LG, Vaillant TC, Dolores Y, Vales HM. HIV, hepatitis B/C and syphilis prevalence and risk behaviors among gay, transsexuals and men who have sex with men, Dominican Republic. Int J STD AIDS. 2013;24:313–321.

23 Blumenthal UJ, Fleisher JM, Esrey SA, Peasey A. Epidemiology: a tool for the assessment of risk. In: Fewtrell L, Bartram J, editors. Water quality: guidelines, standards and health. Geneva: World Health Organization; 2001 (https://apps.who.int/iris/handle/10665/42442).

24 Delgado-Rodríguez M, Llorca J. Bias. J Epidemiol Community Health. 2004;58:635–641.

25 Siegel D, Golden E, Washington AE, Morse SA, Fullilove MT, Catania JA, et al. Prevalence and correlates of herpes simplex infections: the population-based AIDS in multiethnic neighborhoods study. Jama. 1992;268:1702–1708.

26 Smith DK, Pals SL, Herbst JH, Shinde S, Carey JW. Development of a clinical screening index predictive of incident HIV infection among men who have sex with men in the United States. J Acquir Immune Defic Syndr. 2012;60:421–427.

27 Smith DK, Pan Y, Rose CE, Pals SL, Mehta SH, Kirk GD, et al. A brief screening tool to assess the risk of contracting HIV infection among active injection drug users. J Addict Med. 2015;9:226–232.

28 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.

29 Jin F, Crawford J, Prestage GP, Zablotska I, Imrie J, Kippax SC, et al. Unprotected anal intercourse, risk reduction behaviours, and subsequent HIV infection in a cohort of homosexual men. Aids. 2009;23:243–252.

30 Cleves M, Gould WW, Marchenko YV. An introduction to survival analysis using STATA. College Station, TX: Stata Press; 2010 (https://www.stata.com/bookstore/survival-analysis-stata-introduction/).

31 Lochlainn LN, O'Donnell K, Hurley C, Lyons F, Igoe D. Using data from a behavioural survey of men who have sex with men (MSM) to estimate the number likely to present for HIV pre-exposure prophylaxis (PrEP) in Ireland, 2017. Eurosurveillance. 2017;22.

32 PHIA data use manual: reference guide for using data from the population-based HIV impact assessments. New York: PHIA Project; 2019 (https://dms-filesystem.s3.amazonaws.com/uploads/170b96fe126534c7ec04076907e6f7ef/01_PHIA%20Data%20Use%20Manual.pdf).

Annex 1:
Recommendations for reviewing survey data

Before producing the estimates, it is important to review the data, including:

- Question items and variables that will be used to construct specific risk measures.
- Question items and variables used to define the target key population group (e.g. age, sex, sexual orientation, sex work venue).
- Weights and other inputs needed for statistical adjustments, depending on the study design.

Review of RDS data

Data preparation for the RDS surveys should follow standard guidelines:

- Verify numbering of coupon codes.
- Identify repeated participant or coupon codes.
- Identify subjects whose recruiter is not in the dataset.
- Coordinate with study teams to resolve these inconsistencies. Attempt to apply reasonable corrections to the data to allow analysis to move forward.
- Prepare the network size variable:
- Identify reported network sizes that are inconsistent with the number of peers observed. Network size logically must be at least the number of peers who are in the dataset plus one to represent the participant's recruiter. Where inconsistent, the network size should be replaced with this minimum value or imputed at the sample mean or median
- Identify missing network sizes and impute similarly.
- Identify respondents whose answers to the series of network size questions increased where it logically should have decreased or remained constant, and impute similarly.
- Examine diagnostics of the risk factor variables to be used for the analysis to identify any clear problems of failure to converge or indications of bias: homophily, the convergence plot, or the bottleneck plot. Consider removing or redefining variables where problems are significant.

Review of TLS data

Preparation of data for TLS surveys depends on the sampling design. Examine whether sampling weights are available, and unique codes that identify sampling events or venues, which are often needed for clustering adjustments to correct the confidence intervals.

Resources

Gile KJ, Johnston LG, Salganik MJ. Diagnostics for respondent-driven sampling. J R Stat Soc Ser A Stat Soc. 2015;178:241–269.

Karon JM, Wejnert C. Statistical methods for the analysis of time–location sampling data. J Urban Health. 2012;89:565–586.

Raymond HF, Ick T, Grasso M, Vaudrey J, McFarland W. Resource guide : time location sampling (TLS)—choosing a sampling method. San Francisco, CA: San Francisco Department of Public Health; 2007 (https://globalhealthsciences.ucsf.edu/sites/globalhealthsciences.ucsf.edu/files/tls-res-guide-2nd-edition.pdf).

Introduction to HIV/AIDS and sexually transmitted infection surveillance: module 4—introduction to respondent-driven sampling. Cairo: World Health Organization Regional Office for the Eastern Mediterranean (http://applications.emro.who.int/dsaf/EMRPUB_2013_EN_1539.pdf).

Annex 2:
Recommendations for constructing measures needed to define risk

Once the risk definitions have been developed, review the survey questionnaire and determine the best way to construct variables that are needed to classify level of risk. While the details will depend on how questions have been asked and how the data are arranged, be aware of the following common issues that tend to arise when constructing risk variables:

Examine levels of non-response and absence or incompleteness of key variables

High levels of non-response may indicate bias. When levels exceed 5% or 10%, it is important to investigate further to determine whether the data for the item in question are usable, or whether it would be better to use a different item.

Check skip patterns

Behavioural surveys often contain instructions to skip out of a series of questions or entire sections of the survey that are not relevant to a given participant.

When reviewing the dataset, skipped items may be coded as non-response or using a missing code (e.g. 999). For the purposes of obtaining accurate estimates of the risk proportions, it is often critical to recode skipped responses.

For example, imagine constructing a risk variable about having two or more recent casual sex partners in the past six months. Respondents who skip out of questions on numbers of recent casual sex partners because they responded "no" to an initial question about whether they had any casual partners may initially appear as missing or as non-response in the data.

However, how many casual partners did these respondents actually have? Certainly, they had 0 and should not be excluded from analysis because the item is skipped and coded as missing. Excluding such a participant would bias the risk proportion upward.

The same issue arises with questions on drug use. Consider constructing a risk variable about injection frequency in the past week. Participants who skip out of the frequency questions because they responded "no" to any injection in the past month should be counted as not having the risk factor (coded as 0) rather than counted as non-response or missing and mistakenly excluded from the analysis.

It is important to carefully review the questionnaire for skip patterns, examine how they were coded in the data, and recode any risk variables accordingly.

Verify eligibility criteria

Be sure to exclude participants who clearly did not complete with survey eligibility criteria. For example, surveys of sex workers are often limited to females who report having received money in exchange for sex in the past 12 or 6 months. If survey responses indicate this is not the case, it is important to review how eligibility criteria were verified during the survey. If there is a chance that eligibility criteria were not verified exhaustively, such participants should not be included in the analysis.

Check inconsistent responses

Some survey respondents may provide responses that, when examined together, simply cannot be true (they are internally inconsistent). For example, if participants report having 10 total sex partners in the past 6 months and later report having 15 receptive sex partners in the past 6 months, this is inconsistent and may indicate problems with understanding the items or lack of interest in providing accurate responses.

In either case, a decision should be made about whether the responses should be excluded from the analysis. As much as possible, examine consistency across responses, recognizing that an exhaustive check of the data is probably not possible given the time available.

Use short recall periods and direct measures of interest

If there is a choice about which survey items to use to construct a risk factor variable, and all else is equal, it is preferable to use direct measures of the risk behaviour of interest.

For example, if the desired risk variable is number of unprotected anal sex partners in the past six months, it is better to use a question that directly asks about the number of partners with whom unprotected anal sex occurred, as opposed to building the measure from several different items about unprotected anal sex with different partner types (e.g. stable, casual, commercial, non-commercial).

If there is a choice between items that assessed behaviours over different time spans (e.g. 30 days, 6 months, 12 months), shorter time spans are preferred as they are less vulnerable to recall bias.

Improve quality of future surveys to assess risk factors

Issues of non-response, correct application of eligibility criteria, and internal consistency can all be improved by strengthening data quality monitoring during survey training and implementation. If a survey will be designed to estimate the number of key populations at risk, consider taking special measures to strengthen training and quality monitoring around the specific items needed to construct the estimates.

Annex 3:
Should estimates be pooled across survey sites?

When estimating the proportion at risk as a part of Step 6, in some cases it may be attractive to pool the estimates across survey sites. Pooling data is a way to increase the precision of the estimates when sample size is limited at specific sites where targets are needed.

Pooled estimates are a kind of weighted average. The advantage is that they will have smaller confidence intervals than site-specific estimates. This strategy works best when sites are similar to each other. Caution is needed when estimates differ greatly across sites, in which case the pooled average may not be meaningful to any site. Pooled estimates may also serve to develop a national estimate of the number at risk.

Pooled estimates from RDS studies can be obtained using the aggregate estimates procedure in RDS Analyst software. This requires a population size estimate for each site. If unavailable, an approach taken by published studies has been to enter the size of the general (male or female) population aged 15–49 years as an approximation *(1–3)*.

To pool estimates from TLS studies, ensure the venue or event identifiers used for the clustering adjustment are unique across sites. For DHS studies, see weighting instructions provided with the survey dataset.

Regardless of survey design, consider weighting the sites relative to population size so that sites with a larger key population are given more weight. Consult a statistician as needed.

Box A3.1: Pooling risk proportions for men who have sex with men across four Colombian cities

To estimate the number of men who have sex with men at risk in four Colombian cities (Bogotá, Cali, Cartagena, Cucuta), a BBS that used RDS could be used to estimate the proportion of men who have sex with men who were HIV-negative as well as the proportions of men who have sex with men at low-, medium- and high-risk levels.

Sample sizes in the BBS ranged from 286 to 444 men who have sex with men across the sites. The size of the general population varied more widely, from approximately 170 000 to 2.1 million males aged 15–49 years in 2017.

Population size estimates for men who have sex with men ranged from 3.4% to 4.7% based on an earlier study.

Pooling was done using RDS Analyst, weighting by the relative population size estimates.

Key risk variables that might be considered for defining high-risk criteria also varied. Syphilis prevalence ranged from 1% to 3% (2% when pooled). History of sexually transmitted infection varied from 3% to 6% (4% pooled). Behavioural risk factors varied more widely.

A risk variable representing unprotected anal intercourse in the past year and multiple partners in the past year ranged from 24% to 39% (31% pooled). High perceived risk ranged more widely from 8% to 40%, so the pooled estimate of 17% does not adequately represent any of the four sites.

With a risk definition (hypothetical) that includes syphilis, unprotected anal intercourse and multiple partners, applying the risk proportion calculated using the pooled data leads to an estimated number of men who have sex with men at high risk that is from −13% to 14% of the estimates obtained when using site-specific proportions. In this scenario, pooling would not be recommended due to these differences, and given that each site has sufficient data to support site-specific estimates.

References

1 Goodenow C, Szalacha LA, Robin LE, Westheimer K. Dimensions of sexual orientation and HIV-related risk among adolescent females: evidence from a statewide survey. Am J Public Health 2008;98:1051–1058.

2 Szwarcwald CL, De Souza Júnior PRB, Damacena GN, Junior AB, Kendall C. Analysis of data collected by RDS among sex workers in 10 Brazilian cities, 2009: estimation of the prevalence of HIV, variance, and design effect. J Acquir Immune Defic Syndr. 2011;57:S129-35.

3 Mendoza MLR, Jacobson JO, Morales-Miranda S, Alarcón CÁS, Núñez RL. High HIV burden in men who have sex with men across Colombia's largest cities: findings from an integrated biological and behavioral surveillance study. PLoS One. 2015;10.

Annex 4:
Additional tools and materials to support PrEP target-setting

The following documents accompany this guide:

Spreadsheet tools	Support carrying out the methods described in this guide and documenting estimates There are separate spreadsheets for men who have sex with men, transgender women, female sex workers, people who inject drugs, and adolescent girls and young women
PrEP Target-setting for Key and High-priority Populations: Using the Tools	Provides step-by-step examples illustrating how to use the spreadsheet tools
PrEP Target-setting for Key and High-priority Populations: Technical Materials	Provides additional detail on the methods used to develop the guidance and other useful resources: • Mathematical model to determine minimum levels of risk behaviour to reach threshold level of HIV incidence • Literature reviews of risk factors for HIV among men who have sex with men, transgender women, female sex workers, people who inject drugs, and adolescent girls and young women • Example of cohort analysis to identify risk factors among men who have sex with men • Analysis using PHIA surveys to identify risk factors among adolescent girls and young women • Examples of target calculations using real data from diverse settings • Sensitivity analysis illustrating how choice of data source and methods at each step can influence targets

UNAIDS/JCxxxxE

UNAIDS
Joint United Nations
Programme on HIV/AIDS

20 Avenue Appia
1211 Geneva 27
Switzerland

+41 22 791 3666

unaids.org