

UNAIDS 2022

---

# Key Population Data Review Guidance

---

4 January 2022, version 1.0

## Introduction

This manual provides guidance on reviewing key population surveillance data as part of creating national-level estimates of population size and new HIV infections for 2021 by four key population groups: female sex workers (FSW), men who have sex with men (MSM), people who inject drugs (PWID) and transgender people (TG). This manual is also available at [hivtools.unaids.org](http://hivtools.unaids.org)

It should be used alongside the Key Population Data Workbook which is available from within your country organisation on the [AIDS Data Repository](#). The Workbook has been pre-populated with existing surveillance data and previously created estimates compiled by:

- UNAIDS Key Population Atlas
- Global Fund for AIDS, TB, and Malaria
- US Centres for Disease Control and Prevention databases
- Goals, Optima, and Incidence Patterns Models

Key population data not included in the pre-populated data should also be added into the workbook, and guidance is provided below on where and how to search for it.

The key population workflow is organised into five sections and linked to the tabs of the workbook:

- 1) Compiling available data sources on key populations in your country;
- 2) Entering, cleaning, and de-duplicating data
- 3) Validating and reviewing data;
- 4) Forming consensus estimates at the national level; and
- 5) Creating estimates of new HIV infections

For any questions about this process please contact [KPestimates@unaids.org](mailto:KPestimates@unaids.org).

## Tab 2: Compiling available data sources on key populations

Key population information, including surveillance data and modelled estimates, are collected and stored by multiple stakeholders; a single database of all available key population data may not exist in your country. This section will:

- 1) Review common sources where data may be found on:
  - Key population size estimates;
  - HIV prevalence; and
  - ART coverage
- 2) Provide instructions on how to enter the data in the Key Population Data Workbook and remove duplicate data.

### Data sources

#### **National planning and programmatic sources**

- National Strategic HIV/STI Plans
- Programme data: The number of key population members engaging with programmes, HIV prevalence, and numbers on treatment

### Published reports

- Academic literature (e.g., [PubMed](#), [Google Scholar](#))
- Grey literature (e.g. BBS surveys, size estimation exercises, partner annual reports)
- Community organisations

## Tab 3: Entering and checking data

The Key Population Data Workbook has been pre-populated with estimates of population size, HIV prevalence, and ART coverage. However:

- 1) **This is unlikely to be a complete list of available data** and we encourage you to conduct an in-depth search through the sources listed above; and
- 2) **There may be pre-populated rows that are missing mandatory information in the *Data inputs* tab**

The pre-populated data has been checked against primary sources in cases where sources could be found. Rows with Yes in *Data checked* (**column B**) are complete and correct and do not need checking.

However, not all data could be sourced, and some of the pre-populated data may have No entered in **column B**. **Please try to source these data from your networks and colleagues and send any new data sources or reports to [KPestimates@unaids.org](mailto:KPestimates@unaids.org)**

To help you complete the workbook, some cells will be disabled depending on your dropdown menu selection. Accordingly, we recommend completing the **columns** left to right. The mandatory items in the **Data inputs** tab must be complete before moving to the **Validation** tab:

- Gender
- Age group
- Area name
- Province
- Year

If these indicators are not entered, the data will not carry through into the **Validation** and **Estimates** tabs:

After each row has been checked and completed, select Yes in *Data checked* (**column B**). This will help you keep track of which rows have been checked and is a **mandatory step** for data to appear in the Validation stage.

The table below provides information on each column in the **Data input** tab, the dropdown menu options, and notes on how to complete the column

Column name	Definition	Dropdown option	Notes
Country	Country name	-	
Surveillance type	How was the data created?	1) BBS 2) Size estimation exercise 3) Programme data 4) Modelled estimate 5) Academic/other study 6) National extrapolation	-
Indicator	Indicator	1) Population size estimate 2) HIV prevalence 3) ART coverage 4) Number on ART	If Number on ART is selected: <i>Surveillance Type</i> must be <i>Programme data</i>
Method	What estimation method was used?	If PSE: 1) 2S-CRC 2) 3S-CRC 3) Enumeration 4) Expert opinion/WOTC 5) Median / Delphi / Consensus 6) Network scale-up 7) Object/event multiplier 8) PLACE/Mapping 9) Regional benchmark/Literature 10) Service multiplier 11) SS-PSE  If HIV prevalence: 1) Self-report 2) Laboratory confirmed  If ART coverage: 1) Self-report 2) Laboratory confirmed 3) Viral load	5) If a median PSE is entered, please enter each method <b>and</b> median as separate entries (see table below)  3) If viral load suppression is available without an estimate of ART coverage, divide the VLS estimate by 0.9 as an approximation of ART coverage
Key population	Key population	1) FSW 2) MSM 3) PWID 4) TG	-
Gender*	Gender	1) Male 2) Female	For transgender populations, enter the sex to be used as a comparison with the general population, i.e., for trans women, select "Female."
Age group*	Age group	1) 15+ 2) 15-24 3) 15-49 4) 15-64 5) 20-24 6) 25-29 7) 25-49	If an exact match cannot be found, choose the age group range that is closest to the data

Column name	Definition	Dropdown option	Notes
Area name*	Area name (e.g. Districts, counties, or cities)	-	If several areas are reported in the surveillance data (e.g. Cape Town; Johannesburg; Pretoria), please find the original source data and enter separate rows for each area  If the area is not specific (e.g. "Urban areas" or "5 provinces") please find the original source data and enter the specific surveillance area names and area-specific estimates as separate data rows
Province*	Province in which area is found	List of country name and all first administrative level province names	If the area is at country level, enter the country as the province  For example: Area: Lilongwe    Province: Central Area: Ethiopia    Province: Ethiopia
Year*	Year	Year of data collection <b>not of publication</b>	-
Count lower bound	Count lower bound	-	Optional
Count estimate	Count estimate	-	Enter one or both of a count or proportion estimate for PSE
Count upper bound	Count upper bound	-	Optional
Proportion lower bound	Proportion lower bound	-	Optional
Proportion estimate	Population size estimate expressed relative to the age/sex matched to general population in of data collection year	-	Enter one or both of a count or proportion estimate for PSE
Proportion upper bound	Proportion upper bound	-	Optional
Sample size	Sample size	-	Not required for PSE HIV prevalence: The denominator tested ART coverage: The PLHIV denominator
Source	Source	-	e.g. "FSW BBS 2013 in Kigali"
Link	Web link	-	Link to or citation of original report or academic publication  Please download the source file and email it to <a href="mailto:KPEstimates@unaids.org">KPEstimates@unaids.org</a>

\* Mandatory fields that must be completed before moving to the **Validation** tab.

### Example for entering and disaggregating a median PSE

A median PSE for FSW in Harare in 2018 was 7500. In the PSE report, four methods were found that were used to create the median. Enter all the methods and the median in the **Data input** tab.

Method	KP	Area name	Year	Count estimate
Network scale-up	FSW	Harare	2018	2500
Two-source CRC	FSW	Harare	2018	2000
Wisdom of the crowds	FSW	Harare	2018	15000
Unique Object	FSW	Harare	2018	8000
Median	FSW	Harare	2018	7500

## Tab 4: Validating data

The next tab in the Workbook – **Validation** – will help you compare and assess the quality of your key population data. Key population surveillance exercises are conducted infrequently and often in selected subnational or urban areas using a range of methodologies. This means comparing and validating different surveillance exercises can be challenging. NOTE: This tab will be blank if you have not completed the mandatory fields in the **Data Input** tab.

### Check nationally representative data

If the area name is the same as your country name, a warning will be displayed, asking *Are the data nationally representative?* It is common for subnational data to be entered under the country name, rather than the subnational areas/cities where the data were actually collected. Please check each data item that was entered as “national data” in the first tab and edit the area in the **Data input** tab if it is not nationally representative.

### Record key population definition

Please record the specific key population definition used by each surveillance activity in **column Q**

- For example: “FSW were defined as women who use sex work as their primary source of income. Women who engage in infrequent transactional sex (less than five times a month) were excluded”

When comparing surveillance activities, it is important to consider if studies recruited from a specific subpopulation (e.g. venue vs street-based FSW), and if the different subpopulations are comparable? You may wish to exclude some data points if the key population definition is different to the majority of other data points.

### Record comments on methods used

Please record comments on methodological strengths and weaknesses and study quality in **column P**. These may include:

- Method quality
  - Population size methods are graded:
    - Non-empirical (Wisdom of the crowds, Delphi, expert opinion, regional benchmarking)

- Caution (PLACE/Programmatic mapping)
- Empirical (Multipliers, capture-recapture, SS-PSE, network scale up)

We recommend that:

- Non-empirical size estimates are **excluded** from final estimates
- PSE from programmatic mapping are **considered carefully**. Estimates may be biased by the availability of programmes or visibility of key population groups in those geographic areas

Note that data points derived using an empirical method can also occasionally be of poor quality. The method alone does not qualify a data point for inclusion or exclusion. **It is important that all non-empirically based data are still considered for inclusion in final estimates.**

- HIV prevalence and ART coverage methods are graded:
  - Caution (self-report)
  - Empirical (serological testing)
- Biases
  - If data were derived from household surveys, is the estimate likely to be an underestimate due to stigma and social desirability biases?
- Environmental and social factors
  - Were there any environmental factors that may have impacted on the surveillance estimate? For example, was the study conducted during:
    - A time when legal or criminal justice consequences would prevent key population members from coming forward?
    - Economic crisis (e.g. caused by COVID-19) which may increase the number of women engaging in sex work?
    - Changing patterns in drug markets which may increase/decrease the number of people who inject drugs?

Based on methodological or implementation limitations arising from the review, you may wish to exclude data points from your consensus estimates by selecting “Yes” in **column R**.

### Evaluate sample size

If the sample size is under 100, a warning will display: *This is a small sample size*. You may wish to exclude this data point from your consensus estimates by selecting “Yes” in **column R**. For HIV prevalence, we recommend excluding data points with a sample size <100. For ART coverage smaller sample sizes are typical and ART coverage estimates with sample size <100 may be retained.

### Comparing estimates with sub-Saharan Africa benchmark data

On the right-hand side of the **Validation** tab, you can compare your data with similar data across sub-Saharan Africa, in turn for each key population, using the dropdown menu. Three scatter plots are shown:

#### 1) Comparing population size estimates

Population size estimates expressed as a count cannot be compared among countries, due to varying country and total population size. Instead, absolute population size estimates which have been converted to proportions of the national or subnational adult population, which

allows for more meaningful comparison among countries. These proportions are calculated by dividing a key population size estimate by the total population size matched by area, age, sex, and year. UNAIDS and WHO recommend that population proportions for MSM under 1% should be reviewed, as they may be an under-estimate (UNAIDS & WHO, 2020).

The plot shows your country data in orange, and all other available population proportions across sub-Saharan Africa in grey.

- **Are your orange data similar:**
  - **To one another?** It may be unlikely that one city has a FSW proportion of 0.1%, and another in the same province or country has a FSW proportion of 5%
  - **To the rest of the sub-Saharan Africa (grey) data points?**

## 2) Comparing HIV prevalence estimates

HIV prevalence estimates in key populations cannot be directly compared among countries due to varying general population HIV prevalence. Instead, key population HIV prevalence estimates are expressed relative to the province/age/sex-matched overall population prevalence – the “HIV prevalence ratio”.

The workbook will calculate the prevalence ratio automatically, if the sex, age group, province and year were provided for the key population HIV prevalence data point (in tab **Data input**). The plot shows your country data in orange, and all other available prevalence ratios across sub-Saharan Africa in grey. Prevalence ratio is displayed on the y-axis, and general population prevalence on the x-axis. Compare your data with data from countries with a similar general population HIV prevalence.

- **Are there HIV prevalence ratios under 1?** We would expect KP HIV prevalence to be *greater* (not smaller) than the matched overall adult population HIV prevalence (blue dashed line). Therefore, scrutinize any data points that imply a prevalence ratio <1.
- **Are the orange HIV prevalence ratios similar to those from other SSA studies at comparable overall adult population HIV prevalence?**

## 3) Comparing ART coverage estimates

Similar to HIV prevalence, we cannot compare ART coverage in key populations among countries because 1) ART coverage varies between country and 2) due to rapid ART scale up, comparing ART coverage estimates from different years is not meaningful. Instead, key population ART coverage estimates are expressed relative to the province/age/sex/year-matched general population ART coverage – the “ART coverage ratio”.

The workbook will calculate the ART coverage ratio automatically, if the sex, age group, province and year were specified for the key population ART coverage data point (in tab **Data input**). The plot shows your country data in orange, and all other available prevalence ratios across sub-Saharan Africa in grey. ART coverage ratio is displayed on the y-axis, and general population prevalence on the x-axis. Compare your data with data from countries with a similar general population ART coverage.



Based on this validation check, if any of your data points do not appear to be in line with the range of SSA data points, you may wish to exclude them from your consensus estimates, by selecting “Yes” in **column S**.

## Tab 5: Creating consensus estimates at the national level

The tab – **Creating consensus estimates** – will assist with converting subnational estimates into a national-level estimate for PSE, HIV prevalence, and ART coverage for 2021, for each KP.

The tables for each indicator starting in row 20 will show data that:

- Have been **collected** in the last 10 years;
- Have no methodological or sample size concerns identified in the **Validation** tab;
- Have no validation concerns identified in the **Validation** tab.
- For a data point to meet these criteria, in the **Validation** tab you should confirm the compliance with those last 2 requirements by stating ‘No’, **columns R** and **S**. If you leave those empty (or click Yes on either or both), the data point will not carry forward into the **Creating consensus estimates** tab.

You do not need to use all validated data to create the consensus estimates and **it is important to balance data recency and data quality**. For example, data on MSM population size from 2015, 2017, and 2020, some options are:

- Approach 1: Only use data from 2020
- Approach 2: Use data from all years
- Approach 3: Discard data from 2015, and use a median of data from 2017 and 2020
- Approach 4: Discard data from 2015, and use a weighted average of data from 2017 and 2020, where the 2020 data has twice the weight of 2017 data ([weighted calculator available here](#)).

Please document your inclusion and weighting methods for the national population size, HIV prevalence, and ART coverage estimates carefully in **columns H, N, and X**.

### Population size estimates

- 1) If a nationally representative size estimate is available, enter it directly in **column G** or;
- 2) If urban subnational data are available:
  - Enter your average population proportion **for urban areas** in **column B**;
  - Enter the urban/rural ratio difference in **column C** (e.g enter 5 if you expect a 5x larger key population density in urban areas than rural areas)
  - **Column D** will estimate the implied national population size using:
    - 15-49 national Spectrum population by sex in 2021
    - The proportion of the national population living in urban areas (**entered in E6**). This has been pre-populated from the [UN Population Division World Urbanisation Prospects 2018](#).
    - The urban KP population proportion entered in **column B**.
  - If you accept the national population size estimate produced in **column F**, copy it in **column G** or;
- 3) If no local data are available, enter the modelled estimate generated by the workbook.

## HIV prevalence

In **column M** enter (in order of preference):

- 1) A nationally representative KP HIV prevalence estimate, or;
- 2) An average of subnational KP HIV prevalence estimate, or;
- 3) If no local prevalence data are available for FSW, MSM, or PWID, use the modelled estimate generated by the workbook. If no local data are available for TG, leave **column M** blank.

## ART coverage

In **column W** enter (in order of preference):

- 1) A nationally representative KP ART coverage estimate, or;
- 2) A median of subnational KP ART coverage estimate or;
- 3) If no local data are available on ART coverage, leave the cell(s) in **column W** blank.

Please detail your methods for creating national level estimates for PSE, HIV prevalence, and ART coverage, in **columns H, N, and X respectively**:

- 1) If a nationally representative estimate was already available in your surveillance data, which passed data quality review and validation against regional ranges, enter “*Nationally representative estimate*”
- 2) If you used the workbook’s modelled national level estimate for population size or HIV prevalence, enter “*Used automatically generated value*”
- 3) Or alternatively, please list all surveillance data used to create the national level estimate.

## Tab 6: Creating estimates of new infections

The final tab – ***Estimates of new infections*** – compares your consensus national KP population sizes and HIV prevalence estimates with those used in published epidemic and transmission models for your country and uses this to decide on a plausible transmission model-based estimate of new HIV infections. The transmission models used here may include Goals, the Incidence Patterns Model (IPM), Optima, and/or optionally another model fit for your country.

Estimates from Goals and the Incidence Patterns Model have been pre-populated for FSW and MSM, and some countries will have PWID estimates. Neither transmission model currently includes transgender populations.

Additional country-specific modelling exercises (e.g. Optima) that estimated new infections by key population may be available and can be entered in the empty table starting in cell B30.

If your consensus national estimates for population size and HIV prevalence are similar to those used in Goals, Optima, or another country-specific transmission model, we recommend you use the estimates of new infections from these as your estimates for 2021. Enter these in **column F**.

If your consensus national estimates for population size and HIV prevalence differ considerably from those used in the transmission model, please notify UNAIDS, who will support the creation of a new Goals model using your new consensus KP data and estimates as model inputs and benchmarks.

## Uploading to Spectrum

When complete, the workbook should be saved and uploaded to your final national Spectrum file. Close the KP workbook file and note the location of the file on your drive. Open your final Spectrum file. Under the AIM module, Program Statistics window, the final tab is labelled Key populations. Open the KP workbook Excel file by clicking on Open file.

You will then be able to visualise your key population infection estimates alongside estimates of new infections for the total population.



Save the Spectrum file. This will add your KP workbook excel file to the zipped Spectrum PJNZ file. Be sure to upload the updated Spectrum file to the AIDS Data Repository.



**UNAIDS**

20 Avenue Appia  
CH-1211 Geneva 27  
Switzerland

+41 22 791 3666

[unaids.org](http://unaids.org)