

---

## Plan Overview

*A Data Management Plan created using DMPonline*

**Title:** Key Population HIV Estimates in Sub-Saharan Africa

**Creator:** Oliver Stevens

**Principal Investigator:** Oliver Stevens, Jeffrey W. Eaton

**Data Manager:** Oliver Stevens

**Project Administrator:** Oliver Stevens, Jeffrey W. Eaton

**Affiliation:** Imperial College London

**Template:** Imperial College London Generic DMP

**ORCID ID:** 0000-0001-6842-9434

**ORCID ID:** 0000-0001-7728-728X

### Project abstract:

Key populations (KP), including female sex workers (FSW), men who have sex with men (MSM), people who inject drugs (PWID), and transgender women (TGW), are disproportionately vulnerable to HIV in sub-Saharan Africa (SSA). The Global AIDS Strategy 2021-2026 calls for equitable and equal access to HIV prevention and treatment programmes through the removal of structural barriers to reduce HIV incidence and end HIV/AIDS as a public health threat by 2030. Delivering appropriate HIV prevention and treatment programming for these populations, and monitoring attainment of an equitable HIV response, requires robust information on key population size, HIV prevalence, the treatment cascade, and new HIV infections.

We seek to consolidate key population size estimates (KPSE), HIV prevalence, and ART coverage data and create key population estimates to guide policymaking and programme decision making

**ID:** 118392

**Start date:** 01-04-2023

**End date:** 31-12-2026

**Last modified:** 28-10-2024

### Copyright information:

The above plan creator(s) have agreed that others may use as much of the text of this plan as they would like in their own plans, and customise it as necessary. You do not need to credit the creator(s) as the source of the language used, but using any of the plan's text does not

imply that the creator(s) endorse, or have any relationship to, your project or proposal

# Key Population HIV Estimates in Sub-Saharan Africa

---

## Administrative details

### Project Name/Title

Key Population HIV Estimates in Sub-Saharan Africa

### Principle Investigator/Researcher

Lead investigator: Oliver Stevens  
Principal Investigator: Jeffrey W. Eaton  
Co-investigators: Rebecca Anderson

### ORCID iD (if applicable)

Oliver Stevens: 0000-0001-6842-9434  
Jeffrey W. Eaton: 0000-0001-7728-728X  
Rebecca Anderson: 0009-0002-6127-3769

## Project Description

Key populations (KP), including female sex workers (FSW), men who have sex with men (MSM), people who inject drugs (PWID), and transgender women (TGW), are disproportionately vulnerable to HIV in sub-Saharan Africa (SSA). The Global AIDS Strategy 2021-2026 calls for equitable and equal access to HIV prevention and treatment programmes through the removal of structural barriers to reduce HIV incidence and end HIV/AIDS as a public health threat by 2030. Delivering appropriate HIV prevention and treatment programming for these populations, and monitoring attainment of an equitable HIV response, requires robust information on key population size, HIV prevalence, the treatment cascade, and new HIV infections.

We seek to consolidate key population size estimates (KPSE), HIV prevalence, and ART coverage data and create key population estimates to guide policymaking and programme decision making

## Data Collection

### What data will you create or collect?

Existing databases or systematic reviews collating key population HIV survey data have been created and maintained by several organisations including international organisations and academic institutions. This analysis will draw on five databases as the initial basis for sources:

- UNAIDS Key Population Atlas and UNAIDS Global AIDS Monitoring
- The Global Fund to Fight HIV/AIDS, TB, and Malaria key population database
- US Centers for Disease Control and Prevention (CDC) key population database
- [Degenhardt et al. \(2017\) systematic review](#)
- Stannah et al. [2019](#) and [2022](#) (in press) systematic reviews

Sources will additionally be identified through literature searches.

### Will you be reusing existing data (e.g. data sourced from a 3rd party data provider)?

- Yes (please give details)

All data are secondary data collated from the five databases above and literature searches.

### **What file formats will be used?**

All data will be recorded electronically with no physical copies created. Aggregate data will be extracted from survey reports using Microsoft Excel and saved as comma separated value files (.csv). Analysis will be conducted in R, with outputs saved as .csv and .rds object files.

### **Will you create any software or write any code to process or analyse data?**

- Yes (please give details)
- Code will be written to analyse aggregate data in R
- Individual-level data will be processed and harmonised to standard variable names across surveys using code in R
- Data will be analysed using code in R and in C++

## **Ethics and Legal Responsibilities**

### **Does your research involve human participants?**

- Yes

### **Have you applied for/obtained ethical approval?**

- Yes

Ethical approval obtained from Imperial College Research Ethics Committee #6412027

### **Have you obtained informed consent?**

- Informed consent is not needed (explain why)

This study involves only secondary analysis is of anonymised data. Therefore no further consent was required.

### **Will you be processing/collecting personal data?**

- No

### **Will you be processing/collecting special categories of personal data (please select all that apply)?**

- Sex life or sexual orientation
- Racial or ethnic origin
- Health data

### **Are there any IP or copyright restrictions which might influence your use or sharing of the data?**

- Yes (please give details)

A small number of the survey reports and most of the individual-level survey datasets accessed for this study are not publicly

available information. Data owners have provided these data for inclusion in this study, but retain ownership of the data and determination of further data sharing. We will provide contact information and any information about data request processes for data alongside publication of results of the study.

Any compilations of publicly available data and data arising from analytical or modeled results of this study will be made publicly available under open access licenses. Where possible, this will include relevant aggregations of individual-level survey data with permission from collaborating Data Owners.

## Data Storage and Security

### How much data do you expect to generate?

- < 1 TB

### How will you store and back-up your data during the project?

All data will be stored on the Imperial College London Sharepoint and an Imperial College London server at the South Kensington campus, ensuring continuous back-up. No local copies of the data will be stored.

Code will be backed up using GitHub.

### How will you manage access and security?

The Imperial College London Sharepoint and server are only accessible to members of the research team while logged into the Department of Infectious Disease Epidemiology's Virtual Private Network. Data will be read directly into analysis software from these locations. All data are anonymised and no personally identifying information will be stored, reducing data sensitivity.

## Data Documentation and Metadata

### How will the data be documented to ensure it can be understood?

Metadata will be recorded for each key population/study combination. Data elements will include:

- Number of survey respondents
- Year
- Recruitment methodology
- Ethics and IRB number
- Eligibility / inclusion criteria
- Name of data owner
- Contact of data owner
- Primary methods and results source (survey report / primary publication)
- Reuse permissions

Variable names and values will be recoded to a standardised set shared across all survey datasets. Original and recoded values will be stored in a spreadsheet alongside the data

### Will you be using any domain specific or widely used metadata standards to describe your data?

- No

## Data Preservation and Sharing

### What are your plans for long-term preservation and data sharing?

Publicly available aggregate data collated for this analysis will be published as supplementary material with all manuscripts arising from this analysis.

We are exploring the possibility of setting up a data repository for individual-level data, subject to the permissions of the original data owners, but this is not yet part of the study due to resource constraints. If this becomes possible during the course of the study, an amendment to the study protocol and data management plan describing this activity will be submitted to the Imperial College Research Ethics Committee for review.

### Will there be any restrictions on accessing the data?

- Yes (please give details)

Private aggregate and all individual-level data will not be shared publicly or upon request without the permission of the original data owners. Requests for private data will be directed towards data owners. Contact information for Data Owners and any instructions for data requests will be published as supplementary information to any manuscript arising from this study.

### How will potential users find out about your data?

Public aggregate data will be included as supplementary material in publications arising from the analysis.

A data access statement for private data will be included in manuscripts for users to raise requests with data owners for data access.

## Responsibilities and Resources

### Who is responsible for implementing this plan?

Along with the PI, Jeffrey W. Eaton, Oliver Stevens will have responsibility for study-wide data management, metadata creation, data security and quality assurance of data.

### Will you require any additional resources to deliver this plan?

No additional resources are required to deliver the key deliverables in this data management plan. The creation of a new data repository for key population data is not presently funded.