



# Institute for Health Metrics and Evaluation

## Data Release Information Sheet

### ***Data Summary***

Dataset name: Sub-Saharan Africa HIV Prevalence Geospatial Estimates 2000-2018

Date of release: August 31, 2020

Summary:

Annual estimates were produced for HIV prevalence and the number of people living with HIV (PLHIV) at the 5x5 km-level for 43 countries in sub-Saharan Africa between 2000 and 2018. The estimates are by sex and for 5-year age groups in the 15-59 year range, as well as the aggregate age groups 15-49 years and 15-59 years. These estimates were produced using a geo-positioned dataset created from 95 household surveys as well as 10,351 site-years of sentinel surveillance reporting of Antenatal Care clinic attendees. Data for covariates associated with HIV prevalence were also incorporated from 294 surveys. Survey sources used include the Demographic and Health Survey (DHS), AIDS Indicator Survey (AIS), Multiple Indicator Cluster Survey (MICS), Core Welfare Indicators Questionnaire Survey (CWIQ), Population-based HIV Impact Assessment Survey (PHIA), and other country-specific surveys.

This dataset includes the following:

- GeoTIFF raster files for pixel-level estimates of HIV prevalence and PLHIV
- CSV files of aggregated HIV prevalence and PLHIV estimates for each country at zero, first, and second administrative divisions
- Code files used to generate the estimates

[Get Data Files](#)

### ***Acknowledgements***

Contributing organizations:

- Institute for Health Metrics and Evaluation (IHME)

Funders:

- Bill and Melinda Gates Foundation (BMGF)

Suggested Citation:

Institute for Health Metrics and Evaluation (IHME). Sub-Saharan Africa HIV Prevalence Geospatial Estimates 2000-2018. Seattle, United States of America: Institute for Health Metrics and Evaluation (IHME), 2020.

## Data Files Information

### CSV files of Aggregated Estimates of HIV Prevalence and PLHIV

Stored in files named <MEASURE/METRIC>\_<SEX>\_<LEVEL\_OF\_AGGREGATION>.CSV

(Example: IHME\_AFRICA\_MC\_HIV\_2000\_2017\_HIV\_COUNTS\_ADMIN\_1\_Y2020M08D024.CSV)

- **Measure/Metric:** HIV prevalence [Percent], People living with HIV [Counts]
- **Sex:** Males, Females, Both
- **Level of aggregation:** admin0, admin1, or admin2, corresponding to first and second administrative level areas as defined in the Database of Global Administrative Areas (GADM) 2020\_02\_20 shapefiles, with adjustments made in some countries. Each row in each table is unique by administrative unit, year, age group and sex.

Variable	Variable Label	Variable Definition
ADM0_CODE	GADM Admin 0 Code	GADM code identifying the administrative unit
ADM0_NAME	Admin 0 Name	Zero level administrative unit (Country) name as found in the GADM shapefile
ADM1_CODE	GADM Admin 1 Code	GADM code identifying the administrative unit (Only in the admin1 files)
ADM1_NAME	Admin 1 Name	First level administrative unit name as found in the GADM shapefile
ADM2_CODE	GADM Admin 2 Code	GADM code identifying the administrative unit (Only in the admin2 files)
ADM2_NAME	Admin 2 Name	Second level administrative unit name as found in the GADM shapefile (Only in the admin2 files)
year	Year	Time period of estimate. Possible values: years in the range 2000-2017
age_group_id	Age Group ID	Unique numeric identifier for the age group generated and stored in an IHME database of data dimensions. Possible values: age IDs in the range of 8-16,24,199
age_group_name	Age Group Name	Age group estimated. Possible values: 15-19,20-24,25-29,30-34,35-39,40-44,45-49,50-54,55-59,15-49,15-59 years
sex_id	Sex ID	Unique numeric identifier for the sex generated and stored in an IHME database of data dimensions. Possible values: 1,2,3
sex	Sex	Sex estimated: Possible values: Males, Females, Both
measure	Measure	The measure (indicator) estimated. Possible values: <ul style="list-style-type: none"> <li>• HIV prevalence</li> <li>• People living with HIV</li> </ul>
metric	Metric	Metric/unit of measure for the estimate. Values: Counts, Percent

Variable	Variable Label	Variable Definition
mean	Mean	Mean posterior population-weighted estimate for the administrative unit
lower	Lower Confidence Interval	2.5% population-weighted posterior quantile estimate for the administrative unit
upper	Upper Confidence Interval	97.5% population-weighted posterior quantile estimate for the administrative unit

**Codebooks**

Variable names, labels, and value encoding for admin 0 files can be found in the machine-actionable codebook file:

[IHME\\_AFRICA\\_HIV\\_PREVALENCE\\_2000\\_2018\\_CODEBOOK\\_ADMIN\\_0\\_Y2020M08D24.CSV](#)

Variable names, labels, and value encoding for admin 1 files can be found in the machine-actionable codebook file:

[IHME\\_AFRICA\\_HIV\\_PREVALENCE\\_2000\\_2018\\_CODEBOOK\\_ADMIN\\_1\\_Y2020M08D24.CSV](#)

Variable names, labels, and value encoding for admin 2 files can be found in the machine-actionable:

[IHME\\_AFRICA\\_HIV\\_PREVALENCE\\_2000\\_2018\\_CODEBOOK\\_ADMIN\\_2\\_Y2020M08D24.CSV](#)

**GeoTIFF Raster Files for Pixel-level Estimates of HIV Prevalence and PLHIV**

Stored in files named <MEASURE>\_<STAT>\_<AGE>\_<SEX>\_<YEAR>.TIF

(Example:

IHME\_AFRICA\_MC\_HIV\_2000\_2018\_HIV\_PREVALENCE\_MEAN\_2010\_8\_1\_Y2020M08D24.TIF)

- **Measure:** HIV prevalence [Percent], People living with HIV [Counts]
- **Stat:** mean, upper, or lower summary statistics from the predictive posterior distribution at each pixel. Lower and upper correspond to 2.5% and 97.5% quantiles
- **Age:** Values corresponding to the age group of the estimate
- **Sex:** Males, Females, Both
- **Year:** From 2000 to 2018, corresponding to the time period of the estimate

Note that rasters mask (i.e., have NA values) for lakes and areas with low population (10 people per 1km and classified as barren/sparsely vegetated).

## **Data Input Sources**

This file contains relevant metadata about the input sources as suggested in the [Guidelines for Accurate and Transparent Health Estimates Reporting \(GATHER\)](#), a statement that promotes best practices in reporting health estimates.

### **Household Survey Sources**

[IHME\\_AFRICA\\_HIV\\_2000\\_2018\\_DATA\\_INPUT\\_SOURCES\\_SURVEY\\_Y2020M08D24.XLSX](#)

### **Antenatal Care Clinic Sentinel Surveillance Sources**

[IHME\\_AFRICA\\_HIV\\_2000\\_2018\\_DATA\\_INPUT\\_SOURCES\\_ANC\\_Y2020M08D24.XLSX](#)

### **Covariate Sources**

[IHME\\_AFRICA\\_HIV\\_2000\\_2018\\_DATA\\_INPUT\\_SOURCES\\_COVARIATES\\_Y2020M08D24.XLSX](#)

## ***Additional Information***

### **Terms and Conditions**

<http://www.healthdata.org/about/terms-and-conditions>

### **Contact information**

To request further information about this dataset, please contact IHME:

Institute for Health Metrics and Evaluation

2301 Fifth Ave., Suite 600

Seattle, WA 98121

USA

Telephone: +1-206-897-2800

Fax: +1-206-897-2899

Email: [data@healthdata.org](mailto:data@healthdata.org)

[www.healthdata.org](http://www.healthdata.org)

These files may be updated periodically, so we appreciate hearing feedback or additional information about how these data are being used.