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**:
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.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Variable Label</th>
<th>Variable Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADM0_CODE</td>
<td>GADM Admin 0 Code</td>
<td>GADM code identifying the administrative unit</td>
</tr>
<tr>
<td>ADM0_NAME</td>
<td>Admin 0 Name</td>
<td>Zero level administrative unit (Country) name as found in the GADM shapefile</td>
</tr>
<tr>
<td>ADM1_CODE</td>
<td>GADM Admin 1 Code</td>
<td>GADM code identifying the administrative unit (Only in the admin1 files)</td>
</tr>
<tr>
<td>ADM1_NAME</td>
<td>Admin 1 Name</td>
<td>First level administrative unit name as found in the GADM shapefile</td>
</tr>
<tr>
<td>ADM2_CODE</td>
<td>GADM Admin 2 Code</td>
<td>GADM code identifying the administrative unit (Only in the admin2 files)</td>
</tr>
<tr>
<td>ADM2_NAME</td>
<td>Admin 2 Name</td>
<td>Second level administrative unit name as found in the GADM shapefile (Only in the admin2 files)</td>
</tr>
<tr>
<td>year</td>
<td>Year</td>
<td>Time period of estimate. Possible values: years in the range 2000-2017</td>
</tr>
<tr>
<td>age_group_id</td>
<td>Age Group ID</td>
<td>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</td>
</tr>
<tr>
<td>age_group_name</td>
<td>Age Group Name</td>
<td>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</td>
</tr>
<tr>
<td>sex_id</td>
<td>Sex ID</td>
<td>Unique numeric identifier for the sex generated and stored in an IHME database of data dimensions. Possible values: 1,2,3</td>
</tr>
<tr>
<td>sex</td>
<td>Sex</td>
<td>Sex estimated: Possible values: Males, Females, Both</td>
</tr>
<tr>
<td>measure</td>
<td>Measure</td>
<td>The measure (indicator) estimated. Possible values:</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- HIV prevalence</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- People living with HIV</td>
</tr>
<tr>
<td>metric</td>
<td>Metric</td>
<td>Metric/unit of measure for the estimate. Possible values: Counts, Percent</td>
</tr>
<tr>
<td>Variable</td>
<td>Variable Label</td>
<td>Variable Definition</td>
</tr>
<tr>
<td>----------</td>
<td>----------------</td>
<td>---------------------</td>
</tr>
<tr>
<td>mean</td>
<td>Mean</td>
<td>Mean posterior population-weighted estimate for the administrative unit</td>
</tr>
<tr>
<td>lower</td>
<td>Lower Confidence Interval</td>
<td>2.5% population-weighted posterior quantile estimate for the administrative unit</td>
</tr>
<tr>
<td>upper</td>
<td>Upper Confidence Interval</td>
<td>97.5% population-weighted posterior quantile estimate for the administrative unit</td>
</tr>
</tbody>
</table>

**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
www.healthdata.org

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