Data Release Information Sheet

**Data Summary**

**Dataset name:** Low- and Middle-Income Country Under-5 Diarrhea Incidence, Prevalence, and Mortality Geospatial Estimates 2000-2017

**Date of release:** May 6, 2020

**Summary:**

Annual estimates were produced for the prevalence and incidence of diarrhea and diarrhea-related mortality among children younger than 5 years of age at the 5x5 km-level for 94 low- and middle-income countries (LMICs) between 2000-2017. These estimates were produced using geo-positioned data from 466 household surveys. Survey sources used include the Demographic and Health Survey (DHS) and UNICEF Multiple Indicator Cluster Survey (MICS) series, and other country-specific surveys.

This dataset includes the following:

- GeoTIFF raster files for pixel-level estimates of under-5 diarrhea prevalence, incidence, and diarrhea-related mortality
- CSV files of aggregated under-5 diarrhea prevalence, incidence, and diarrhea-related mortality for each country at zero, first and second administrative divisions
- Code files used to generate the estimates

**Get Data Files**

**Relevant publications and visualizations:**

- [Local Burden of Disease – Diarrhea](#)
**Acknowledgements**

**Contributing organizations:**
- Institute for Health Metrics and Evaluation (IHME)

**Funders:**
- Bill and Melinda Gates Foundation (BMGF)

**Suggested Citation:**

**Data Files Information**

**Data Update**
This documentation was updated on July 12, 2021, to include a new link in the Summary field above to the data files in IHME’s cloud storage space.

**CSV files of Aggregated Estimates of Childhood Diarrhea**
 Stored in files named `<MEASURE>_<LEVEL_OF_AGGREGATION>.CSV`
(Example: `IHME_LMIC_DIARRHEA_2000_2017_PREV_ADMIN_1_Y2020M02D29.CSV`)

- **Measure**: Incidence, Prevalence, Mortality (Diarrhea)
- **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) 2019 shapefiles, with adjustments made in some countries. Each row in each table is unique by administrative unit and year.

<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>Variable</td>
<td>Variable Label</td>
<td>Variable Definition</td>
</tr>
<tr>
<td>-----------------</td>
<td>-------------------------</td>
<td>---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</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: 1</td>
</tr>
<tr>
<td>age_group_name</td>
<td>Age Group Name</td>
<td>Age group estimated. Possible values: Under 5</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: 3</td>
</tr>
<tr>
<td>sex</td>
<td>Sex</td>
<td>Sex estimated: Possible values: Both</td>
</tr>
</tbody>
</table>
| measure         | Measure                 | The measure (indicator) estimated. Possible values:                                                                                     • Incidence  
                |                                                        • Mortality  
                |                                                        • Prevalence                                                                                                  |
| metric          | Metric                  | Metric/unit of measure for the estimate. Values: Counts                                                                                                  |
| mean            | Mean                    | Mean posterior population-weighted estimate for the administrative unit                                                                   |
| upper           | Upper Confidence Interval | 97.5% population-weighted posterior quantile estimate for the administrative unit                                                                |
| lower           | Lower Confidence Interval | 2.5% 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_LMIC_DIARRHEA_2000_2017_CODEBOOK_ADMIN_0_Y2020M01D14.CSV`

Variable names, labels, and value encoding for admin 1 files can be found in the machine-actionable codebook file `IHME_LMIC_DIARRHEA_2000_2017_CODEBOOK_ADMIN_1_Y2020M01D14.CSV`

Variable names, labels, and value encoding for admin 2 files can be found in the machine-actionable codebook file `IHME_LMIC_DIARRHEA_2000_2017_CODEBOOK_ADMIN_2_Y2020M01D14.CSV`

**GeoTIFF Raster Files for Pixel-level Estimates of Childhood Diarrhea**

Stored in files named `<MEASURE>_<METRIC>_<STAT>_<YEAR>.TIF`


- **Measure:** Incidence, Prevalence, Mortality (Diarrhea)
- **Metric:** Counts (CT) or Rates (RT)
- **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
- **Year:** From 2000 to 2017, 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). Rasters are stacked, with a layer for each year from 2000-2017. The first layer corresponds to 2000, the 18th layer corresponds to 2017.

**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)](http://www.healthdata.org/about/terms-and-conditions), a statement that promotes best practices in reporting health estimates.


**Additional Information**

**Terms and Conditions**

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