



Institute for Health Metrics and Evaluation

Data Release Information Sheet

Data Summary

Dataset name: Onchocerciasis Prevalence Geopositioned Database

Date of release: May 22, 2019

Summary:

IHME researchers conducted a systematic review of published literature containing data on onchocerciasis-related infection and disability indicators from 1975-2017, encompassing the period of implementation for control and local elimination programs among the Americas, Africa and Yemen. Ultimately, geographic data, as well as relevant epidemiological metadata, were extracted from 259 peer-reviewed sources reporting prevalence of onchocerciasis. This dataset contains the following: a screening sheet detailing all studies reviewed; a database of onchocerciasis prevalence containing 14,043 unique location, diagnostic, age and sex-specific records marked for collapse where a single geo-position is shared between multiple records; and a conversion file that connects the names of diagnostics extracted and standardized during the review to diagnostic codes and diagnostic groups.

Relevant publications and visualizations:

Hill E, Hall J, Letourneau ID, Donkers K, Shirude S, Pigott DM, et al. A database of geopositioned onchocerciasis prevalence data. *Scientific Data*. 22 May 2019.

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). Onchocerciasis Prevalence Geopositioned Database. Seattle, United States: Institute for Health Metrics and Evaluation (IHME), 2019.

File Inventory

File Name	Description	Version Date
IHME_ONCHO_PREV_DATABASE_Y2019M05D22.CSV	Onchocerciasis Prevalence Geopositioned Database [CSV]	May 22, 2019
IHME_ONCHO_PREV_DATABASE_CODEBOOK_Y2019M05D22.CSV	Codebook [CSV]	May 22, 2019
IHME_ONCHO_PREV_DATABASE_DX_CONVERSION_Y2019M05D22.XLSX	Diagnostic conversion file: connects the names of diagnostics extracted and standardized during the review to diagnostic codes and diagnostic groups [XLSX]	May 22, 2019
IHME_ONCHO_PREV_DATABASE_SCREENING_SHEET_Y2019M05D22.XLSX	Systematic review screening sheet	May 22, 2019
IHME_ONCHO_PREV_DATABASE_INFO_SHEET_Y2019M05D22.PDF	Data Release Information Sheet	May 22, 2019

Database Variable Information

Variable	Variable Label	Variable Definition
NID	NID	Unique identifier for the catalog record for the study publication in the Global Health Data Exchange (GHDx)
SITE_MEMO	Site memo	A character string that details out the breakdown of the record's location
LOC_GROUP	Location Group	A unique identifying number that groups record rows that are georeferenced to the same location
LOC_UNIQ	Unique Location	A unique identifying number that groups record rows within a unique LOC_GROUP number that share the SITE_MEMO. This is done for collapse purposes when a location could not be individually georeferenced
LOC_SPEC	Location Specificity	A character string to identify within each unique LOC_GROUP and LOC_UNIQ combo to signify whether that row represents a portion of the location total by a Sex, Age, or Sex Age subset.
COUNTRY	Country Code	A unique character string used by IHME to identify country and or subnational location within a country
POINT	Point Georeference	Whether the record has been georeferenced to a point or a polygon. 1 = point; 0 = polygon

Variable	Variable Label	Variable Definition
LAT	Latitude	If POINT is 1, this is a decimal point value to represent the latitude of the point, otherwise this value will be NA
LONG	Longitude	If POINT is 1, this is a decimal point value to represent the longitude of the point, otherwise this value will be NA
POLY_REFERENCE	Polygon Reference	If POINT is 0, this is a character string to represent the name of the shapefile that contains the georeferenced shape for this record
POLY_ID_FIELD_NAME	Polygon ID Field Name	If POINT is 0, this is a character string to represent the name of the column in the shapefile that contains the unique identifier to connect the record to the polygon within the specified shapefile in POLY_REFERENCE
POLY_ID	Polygon ID	If POINT is 0, this is a unique identifying numeric number to reference within the shapefile and column specified in POLY_REFERENCE and POLY_ID_FIELD_NAME to the polygon this record has been georeferenced to
AGE_START	Starting Age	A number that represents the start of the age range tested for this record. If no age is provided, assumed value of 0
AGE_END	Ending Age	A number that represents the end of the age range tested for this record. If no age is provided, assumed value of 99
SEX	Sex	A character string to represent which sex was tested in this record. Possible Values: Both, Female, Male
YEAR_START	Start Year	Starting year of the data collected from the study. If no year provided, assumed start year to be 3 years before the publication date
MONTH_START	Start Month	If provided, a numeric value between 1 and 12 to represent the starting month, otherwise it is given the value NA
YEAR_END	End Year	Ending year of the data collected from the study. If no year provided, assumed end year to be 1 year before the publication date
MONTH_END	End Month	If provided, a numeric value between 1 and 12 to represent the ending month, otherwise it is given the value NA
DX_CODE	Diagnostic Code	A whole numeric value to represent the test for the presence or symptoms of onchocerciasis performed in this record. A table of specific diagnostic names to codes has been included as part of this upload.

Variable	Variable Label	Variable Definition
DX_GROUP	Diagnostic Group	A character string that represents what diagnostic grouping this record belongs to. In the table of specific diagnostic names to codes you can find how codes were grouped. Possible values: Prevalence - ss (Skin Snip), nod (Nodules), sero (Serology), otherPrev (Other Prevalence); Sequelae - skin (Skin Symptoms), eye (Eye Symptoms)
N	Sample Size	A numeric value representing the number of surveyed participants for this record
CASES	Cases	A whole numeric value representing the number of persons who tested positive for this record

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.