



Institute for Health Metrics and Evaluation

GBD Sources Tool Guide

This document is intended to serve as a basic guide for using the Global Burden of Disease (GBD) Sources Tool. The tool lets you explore records of the raw data sources used to generate GBD 2021 estimates and research findings. You can filter the input sources by GBD component, geography, and cause, risk, covariate, or impairment. After you have made your selection, you can view and access catalog entries for input sources used by GBD through the Global Health Data Exchange (GHDx).

You can download your results set as a CSV file of citations. You can also download results as a CSV file (or files) of source metadata to see more information about how the sources were used in the analysis of the GBD. This download contains 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.

Contents

GBD Sources Tool Guide	1
Contents.....	1
Accessing the Tool	2
Section 1: Forming a Query	2
Section 2: Global Health Data Exchange (GHDx) Results.....	2
Section 3: Downloading Citations.....	2
Section 4: Downloading Source Metadata	3
Section 5: Downloading Smaller Source Metadata Results Sets	5
Section 6: Downloading Larger Source Metadata Results Sets.....	6
Section 7: CSV content.....	7
Section 8: Troubleshooting.....	8
Combining multiple CSV files	8
Opening CSV files with special characters in Excel.....	8
Additional Information	9
Contact information.....	9

Accessing the Tool

Access the tool through any Internet browser at <http://ghdx.healthdata.org/gbd-2021/sources>.

Section 1: Forming a Query

Begin by selecting either a GBD component or a geography from the **Components** or **Locations** drop-down menu. The tool does not allow downloads for “ALL GBD” components *and* “Global.”

Options for GBD components are:

- **ALL GBD**
 - **Mortality and Population**
 - Mortality
 - Fertility
 - Population
 - **Causes of Death and Disability**
 - Causes of Death
 - Nonfatal Health Outcomes
 - **Risk Factors**
 - Relative Risks
 - Exposure and TMREL
 - **Covariates**
 - **Etiologies**

Location options are all locations in the standard GBD reporting hierarchy, including GBD super regions, regions, countries, territories, and subnational locations.

Depending on the GBD component you select, you are given options to refine your query by cause, impairment, risk, or outcome. You can refine by location at any time.

Click the **search** button to generate a set of results.

Section 2: Global Health Data Exchange (GHDx) Results

After you submit a query, you will see an alphabetical list of the input sources used for the part of GBD you specified. Clicking on a title of any of these citations brings you to the GHDx record for the source, which contains metadata describing the source, including geography, time period, data type, citation, and topics covered by the dataset.

Whenever possible, the GHDx record includes a link to where the data can be freely downloaded, registered for, or requested from the provider. For more about the GHDx, visit [About the GHDx](#).

Section 3: Downloading Citations

To get a simple list of unique citations for the input sources generated by your query, click the **download citations CSV** button. This will export a CSV file of only citations.

Section 4: Downloading Source Metadata

Once the list of input sources for your query has been retrieved, you have the option to download a full set of metadata about the sources. This will include not just citations but a large array of other GATHER-recommended metadata that illustrate how sources were used in GBD analysis. In a source metadata set, a single source may be represented by multiple rows in a CSV download: one for each location, age group, sex, etc. for which the source was used.


Click on the **download source metadata CSV** button to begin the download process for your results set.

Note: If you plan to download the full results of your query, we recommend that you make the narrowest selection possible. Because the download represents the use of input data at essentially the data point level, many result sets can easily be several thousand rows – or even *millions* of rows.

If you make no selections with the filters, the tool will prompt you to narrow your criteria.

[Home](#) > [IHME Data](#)

Global Burden of Disease Study 2021 (GBD 2021) Sources Tool

 Please narrow down your search criteria.

The Global Burden of Disease Study 2021 (GBD 2021) synthesizes a large number of input sources to estimate mortality, causes of death and illness, and risk factors.

The Data Input Sources Tool lets you explore GBD 2021 input sources and retrieve relevant metadata.

Use the selection boxes below to see these input sources by GBD component, geography, cause, risk, and more. After you have made your selection, you can view and access GHDx catalog records for input sources used by GBD.

Download these input sources as a CSV file to see more information about how they were used in the analysis for the Global Burden of Disease. This CSV 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.

For detailed information on the tool and the contents of the CSV files, refer to the [GBD 2021 Sources Tool User Guide](#).

Components

ALL GBD


Locations

Global

search

If your selection is not sufficiently narrow, the tool may further prompt you to make an additional choice.

Global Burden of Disease Study 2021 (GBD 2021) Sources Tool

 Please select a cause or impairment.

The Global Burden of Disease Study 2021 (GBD 2021) synthesizes a large number of input sources to estimate mortality, causes of death and illness, and risk factors.

The Data Input Sources Tool lets you explore GBD 2021 input sources and retrieve relevant metadata.

Use the selection boxes below to see these input sources by GBD component, geography, cause, risk, and more. After you have made your selection, you can view and access GHDx catalog records for input sources used by GBD.

Download these input sources as a CSV file to see more information about how they were used in the analysis for the Global Burden of Disease. This CSV 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.

For detailed information on the tool and the contents of the CSV files, refer to the [GBD 2021 Sources Tool User Guide](#).

Components

Nonfatal Health Outcomes

Causes

All causes

Impairments

Impairments

Locations

Nigeria

search

When your selection is sufficiently narrowed, you will receive no further prompts and can proceed with the source metadata CSV download.

For the source metadata download, your selection will still fall into one of two categories:

- A “smaller” results set
- A “larger” results set

Here is an example of a “smaller” results set:

Results

Results for cause: *Malaria*, location: *Nigeria*, component: *Nonfatal Health Outcomes*.

Total Citations: 76

Total Source Metadata Rows: 2098

download source metadata CSV

download citations CSV

You can know if your selection has produced a smaller results set or a larger results set by hovering your cursor over the **download source metadata CSV** button.

A smaller results set example:

Results

Results for cause: *Malaria*, location: *Nigeria*, component: *Nonfatal Health Outcomes*.

Total Citations: 76

Total Source Metadata Rows: 2098

[download source metadata CSV](#) [download citations CSV](#)

Citations Download a smaller set of results in the browser.

1 2 [next >](#) [last »](#)

A larger results set example:

Results

Results for component: *Nonfatal Health Outcomes*, cause: *HIV/AIDS and sexually transmitted infections*, location: *Global*.

Total Citations: 1459

Total Source Metadata Rows: 291748

[download source metadata CSV](#) [download citations CSV](#)

Citations Request a larger result set. We will email you when it is ready.

1 2 3 4 5 6 7 8 9 ... [next >](#) [last »](#)

Again, click on the **download source metadata CSV** button when you are ready to begin the download process.

Section 5: Downloading Smaller Source Metadata Results Sets

For smaller result sets, you will be taken to a download page. While the CSV is prepared, the page will display a message of “Task state: working.”

Global Burden of Disease Study 2021 (GBD 2021) Sources Tool

A link to your download will be available below when ready.

Task state: working

[Back to my filter](#)

[Back to the Data Sources Tool main page](#)

For smaller results sets, CSV preparation can take from less than a minute to several minutes – or sometimes even longer. (The amount of time depends primarily on the numbers of rows the file will include.) During this time, the download page will remain active as your download is prepared. When it is ready, the page will have a message of “Task state: success” and a link to your download. Click the link to download your source metadata CSV file.

Global Burden of Disease Study 2021 (GBD 2021) Sources Tool

A link to your download will be available below when ready.

Task state: success

• [Download #1](#)

 ←
[Back to my filter](#)
[Back to the Data Sources Tool main page](#)

Note: If the time the download page spends in a task state of “working” begins to seem quite long, try refreshing your browser; every so often the page will get stuck in the “working” stage, and a page refresh can result in a completed download link.

Section 6: Downloading Larger Source Metadata Results Sets

As mentioned, your selections can produce very large result sets. In these cases, when you click on the **download source metadata CSV** button, the tool will not take you to a download page. Instead, you will go to a page where you are required to provide an email address and have the option to choose how many rows you want per file. The maximum number of rows for a single file is 1 million. The number you enter determines the number of files your download will include, not the number of total results in that download.

This is a large sources results set and may take some time to process, so we will notify you via email when your CSV download is ready. You will receive an email about this request, and a secondary email with a download link when your CSV is ready to download.

Email *

Please indicate the maximum number of rows per file. 1000000 is the limit. *

A large request may take up to several hours to complete. You will receive a confirmation email with a link to a page where you may check the status of your request. When the process is complete, that page will contain the file, or files, download link(s). You will receive a second email confirming that your download is ready once the process completes.

Section 7: CSV content

While the GHDx results contain metadata on the dataset itself, the CSV generated from your query includes additional information about the way the dataset was used in producing GBD estimates. Below is a description of what the columns in the CSV represent.

Column Header	Description
Citation	Suggested citation for the input source
Component	GBD Component
Subcomponent	Additional information on what an input was used for in analysis (e.g., severity splits, risk exposure, etc.)
Location	Geographic location for which the input source was used
Cause	Cause, etiology, or impairment for which the input source was used (where relevant)
Risk	Risk for which the input source was used (where relevant)
Covariate	Covariate for which the input source was used (where relevant)
Sustainable Development Goal	Sustainable Development Goal (SDG) for which the input source was used (where relevant)
Publication status	Indication if the input source is unpublished or forthcoming (if blank, the input source is considered published)
Provider	Provider of the data (provided whenever possible except for scientific literature)
Provider URL	Web address where you can find or inquire about the input source
GHDx URL	Web address for the GHDx record of the input source
Secondary GHDx URL	Web address for the GHDx record of the granular input source (where relevant) Note: In some cases, a large input source is cataloged at a granular level to increase transparency on what parts of the source were used. This is generally signified by a citation that says “as it appears in.” The secondary GHDx URL points you to catalog records for the more granular input source. For example, the WHO Mortality Database represents many country years of vital registration data, and each country year is documented by a separate GHDx catalog record.
Data Collection Method	Information on the mechanism of data collection
Year Start	Starting year of the data derived from the input source
Year End	Ending year of the data derived from the input source
Sex	Sex of the population of the data derived from the input source
Age Start	Numerical value of starting age of the population of the data derived from the input source
Age End	Numerical value of ending age of the population of the data derived from the input source
Age Type	Unit of age values in previous columns (e.g., years, months, days)
Representativeness	Relevant quality of the population in the input source (e.g., nationally representative, representative of urban areas only, etc.)
Urbanicity Type	Urbanicity value of the population in the input source (e.g., urban, rural, etc.)

Column Header	Description
Population Representativeness Covariates	Other qualities of the population in the input source (where relevant and collected)
Sample Size	Sample size of data derived from input source
Sample Size Unit	Unit of sample size
Standard Error	Standard error of data derived from input source

Section 8: Troubleshooting

This section includes workarounds for some of the known downloading issues with the GBD Data Input Source Tool.

Combining multiple CSV files

Sometimes, queries return data in multiple CSV files. Each file contains a segment of the data and has identical columns and a header row. Note that you cannot have more than 1 million rows in a CSV file. Combining results over one million rows requires use of another program, such as Stata.

To combine multiple CSV files in Excel:

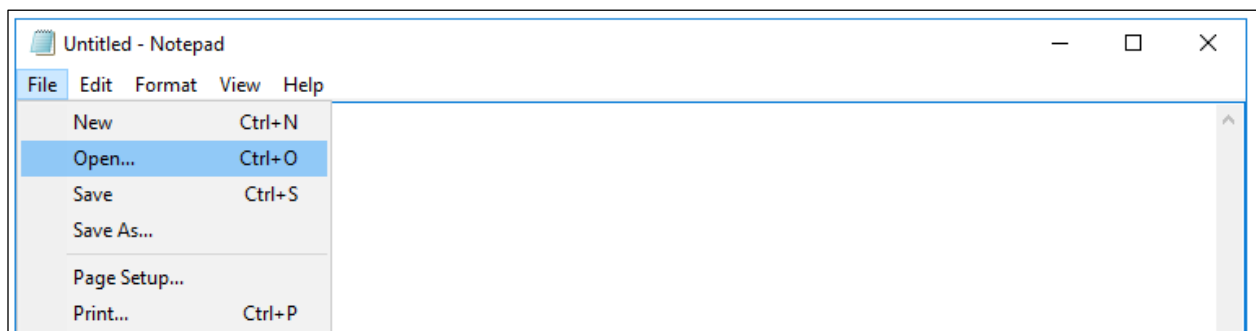
1. Open the first file in the set you want to combine in Excel.
2. Open the next file in the set.
3. Select all cells below the first line in this next file.
Note: A simple way to do this is to select cell A2 (the first and leftmost data-containing cell) and press the following keys: **Ctrl** + **Shift** + **↓** + **→**.
4. Copy the selected cells.
5. Paste the copied cells into the first and leftmost blank cell in the first file.
6. Repeat steps 2-5 until the data from all files has been copied into the first one.

Opening CSV files with special characters in Excel

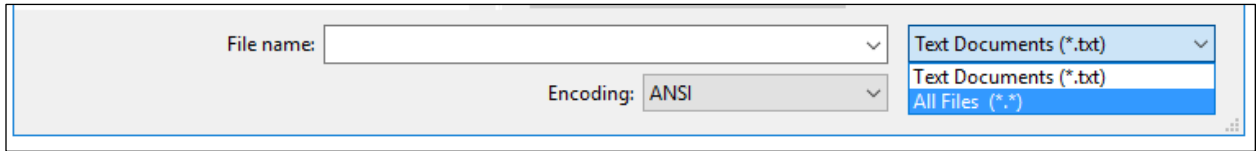
Sometimes, queries result in files that contain special characters, such as place names with accents like “México.” These characters may not display properly when the CSV file is opened in Excel.

To solve this problem:

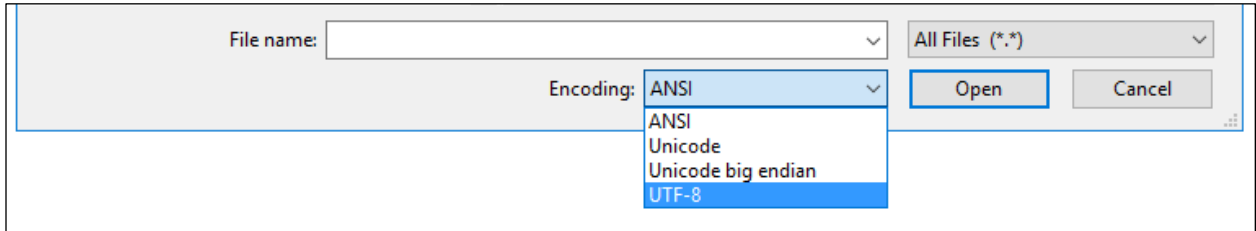
1. Open Notepad.
2. Click File > Open (Ctrl + O).



3. Browse to and open the exported CSV, making sure to select "All Files" in the bottom right.



4. In the Encoding drop-down menu at the bottom, change from ANSI to UTF-8.



5. Save the file. When you reopen it in Excel, the special characters should render correctly.

Additional Information

Contact information

For questions not answered by this guide, please contact IHME:

Address:

Institute for Health Metrics and Evaluation
Population Health Building/Hans Rosling Center
3980 15th Ave. NE, Seattle, WA 98195 USA
UW Campus Box #351615

Telephone: +1-206-897-2800

Fax: +1-206-897-2899

Email: data@healthdata.org

Website: www.healthdata.org