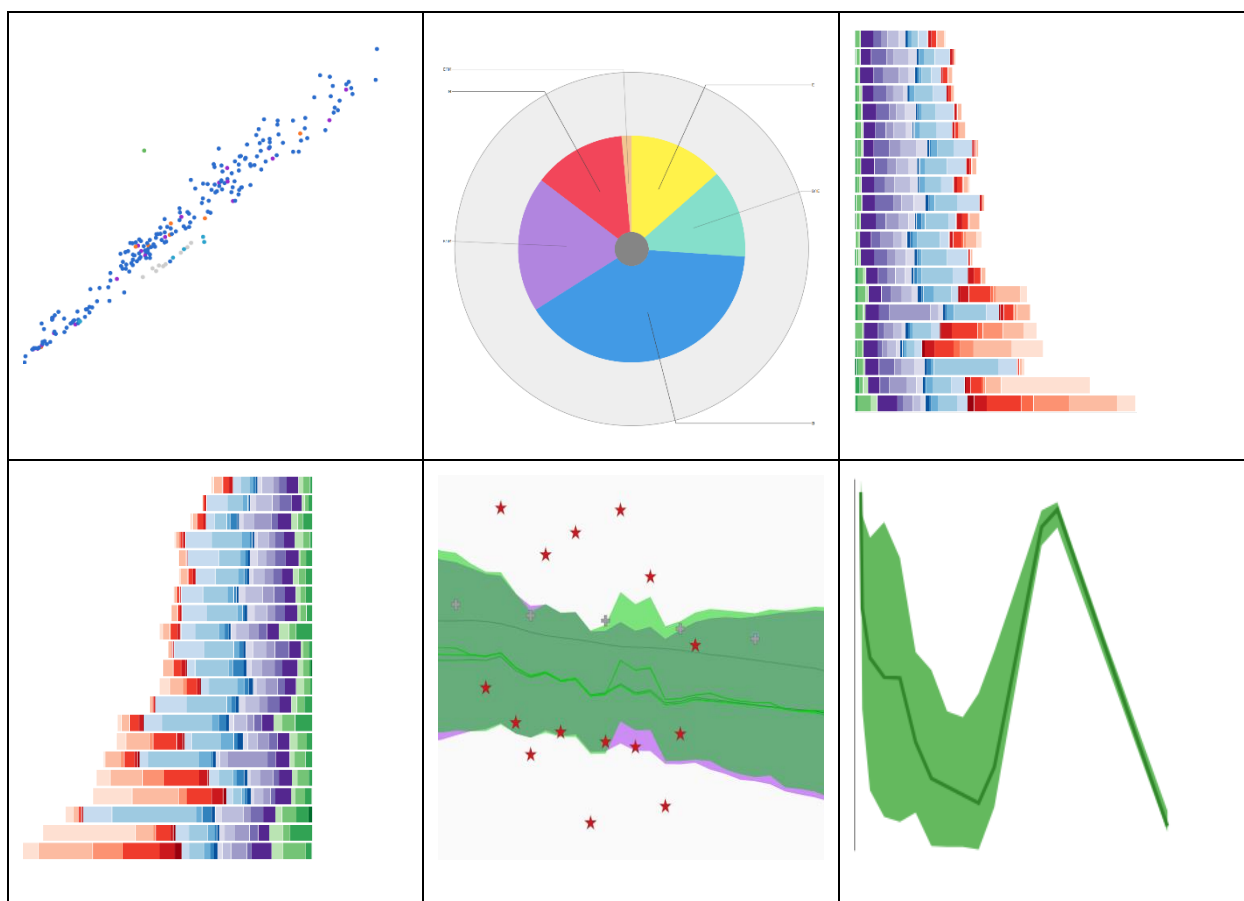Institute for Health Metrics and Evaluation

GBD 2019 Data and Tools Overview

October 2020

Introduction

This document is a basic guide to results from the Global Burden of Disease Study (GBD) and the suite of web-based tools used to disseminate these results. These tools allow students, researchers, policymakers, and other members of the public to access, view, and interact with GBD data outputs. They are all freely available for non-commercial use.

IHME maintains six online GBD query tools and data visualizations. These are the GBD Results Tool, GBD Compare, Mortality Visualization (MortViz), Causes of Death Visualization (CoDViz), Epi Visualization, (EpiViz), and GBD Data Input Sources Tool. With these, users can query, view, and download, in CSV format, information and data of the following types:

- Data input sources. These are lists of the raw data sources used to produce estimates for different components of the study, and relevant metadata about them
- Model input data. These are data points adjusted to meet GBD's format and quality requirements
- Estimates. These are final GBD results: the point estimates and 95% uncertainty intervals, where appropriate, for study indicators

While most GBD results are distributed through the tools listed above, certain results are published and made available for download as prepackaged files in the Global Health Data Exchange (GHDx), IHME's health and health-related data catalogue. For GBD 2019, child and adult mortality (5q0 and 45q0), life tables, population, fertility, and migration estimates are initially available only in the GHDx. Covariate data used in GBD and certain items of frequently requested documentation (disability weights, GBD cause-ICD code maps, relative risks, and more) are always available through GHDx records. Code used to produce GBD estimates is also hosted in the GHDx.

This guide offers the following:

- Brief descriptions of each GBD query tool, data visualization, the GHDx, and GBD code
- A glossary of key terms found in GBD and GBD results
- An overview of the indicators and data outputs users can expect to view and download from each tool and the GHDx
- Appendices files in Excel (XLSX) format containing:
  - GBD cause, risk/etiology/impairment/injury (REI), and location hierarchies
  - Lists of each cause and cause-risk, cause-impairment, and cause-injury pair, with the measures available for each
  - Age groups available for each context (e.g., cause, life expectancy, population, etc.) in GBD results
  - Metrics definitions for each measure
  - A list of models in EpiViz and the corresponding causes, risks, and impairments for each

Additional data resources for GBD 2019 are available at http://ghdx.healthdata.org/gbd-2019.

Information on the GBD study is available at http://www.healthdata.org/gbd

Contact Us

Do you have questions or comments about any of the GBD 2019 data or online tools?

Email us at data@healthdata.org

Call us at +1-206-897-2800

Fax us at +1-206-897-2899

Write us at

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

Website: www.healthdata.org

GHDx: www.ghdx.healthdata.org

Contents

Section 1: Tool Descriptions

1.1 GBD Results Tool

The GBD Results Tool allows the user to query and download in a CSV file the indicators from the most recently published annual GBD. Users can select a specific subset of results according to different dimensions including indicator, cause, risk, location, age, and year.
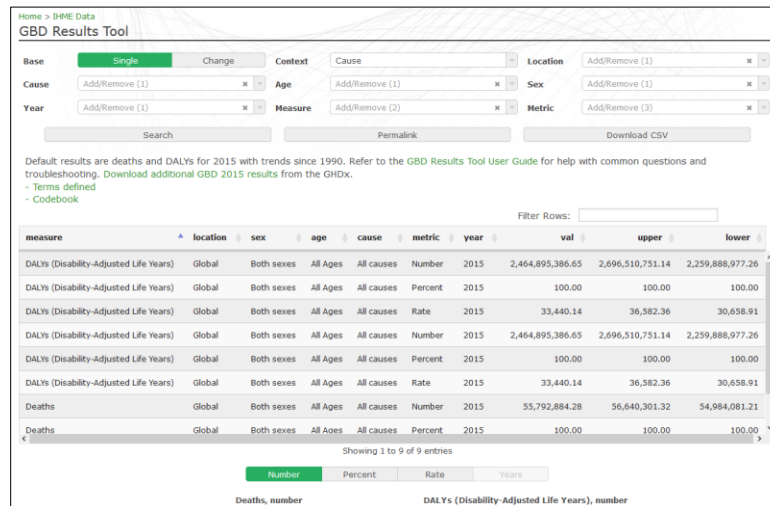

Figure 1. GBD Results Tool

The GBD Results Tool is found here: http://ghdx.healthdata.org/gbd-results-tool.

1.2 GBD Compare

GBD Compare is the GBD's most comprehensive visualization, displaying data from the widest range of GBD components. Users can use maps, plots, treemaps, arrow diagrams, and a dozen other charts to compare patterns and trends in causes and risks over time, to explore the health profile within a country by age and sex; to compare countries with one another; or to explore regional or global trends. Users can drill from a global view into country views including subnational details for some countries; can examine how disease patterns have changed over time; and can explore which causes of death and disability are increasing and which are decreasing.
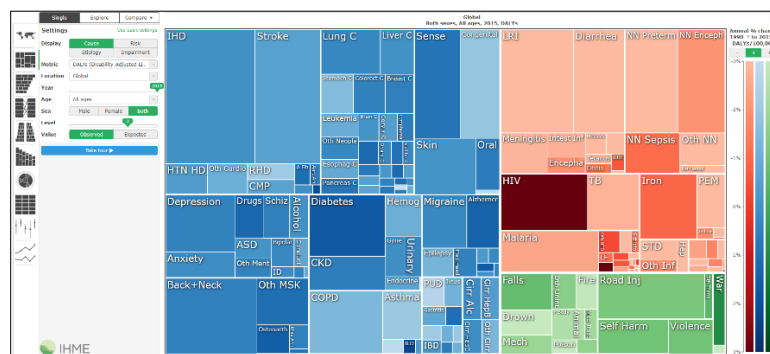

Figure 2. GBD Compare

GBD Compare is found here: http://www.healthdata.org/data-visualization/gbd-compare.

1.3 Mortality Visualization (MortViz)

The Mortality Visualization (MortViz) allows users to view data and indicators of all-cause mortality. Users are able to follow each analytic step from the beginning of the data preparation to the final results and each transformation along the way. The visualization provides views of both child mortality (defined as the probability of dying between birth and age 5) and adult mortality (defined as the probability of dying between age 15 and 60) for each of the locations for which indicators are produced for the annual global GBD enterprise.
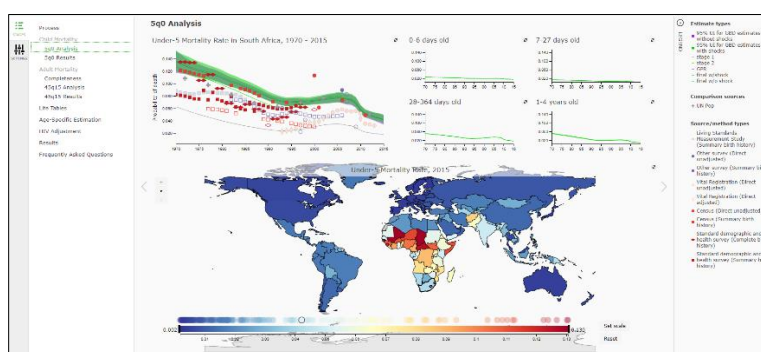

Figure 3. MortViz

MortViz is found here: http://www.healthdata.org/data-visualization/mortality-visualization.

1.4 Causes of Death Visualization (CoDViz)

The Causes of Death (CodViz) allows users to view data and estimates of death by cause, age, sex, and location over time. Deaths are represented as rates, cause fractions (i.e., the proportion of total deaths due to a cause), and numbers. Users have the ability to see the transformations of the input data to correct for different biases and to see the effects of garbage code redistribution. They also have access to the model parameters that were used to generate results for a given cause.
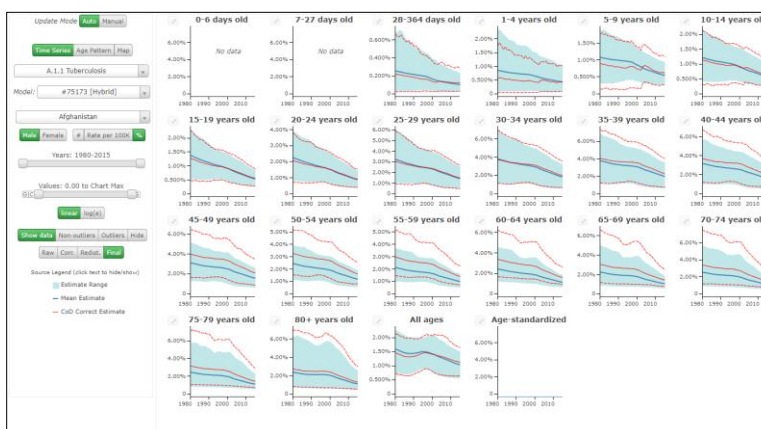

Figure 4. CoDViz

CodViz is found here: http://www.healthdata.org/data-visualization/causes-death-cod-visualization.

1.5 Epi Visualization (EpiViz)

The Epi Visualization (EpiViz) allows users to view data from the nonfatal health outcomes component of the GBD. It allows users to see the input data, data corrections, model settings and results of prevalence, incidence, remission, excess mortality and cause-specific mortality rates for each of the diseases and risk factors that have been analyzed in the GBD.
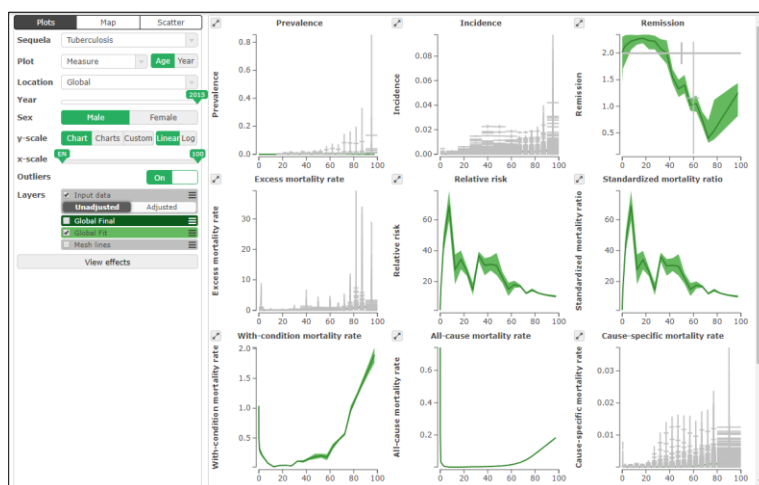

Figure 5. EpiViz

EpiViz is found here: http://www.healthdata.org/data-visualization/epi-viz.

1.6 Data Input Sources Tool

The Data Input Sources Tool allows users to view citations for sources of data by cause, location, and year for the most recently published set of annual results. For example, users can search for all sources associated with a particular cause and location and retrieve a set of citations used by GBD to produce those indicators.

The tool's CSV export function also provides the complete set of dimensions and metadata associated with that source.

Figure 6. GBD Data Input Sources Tool

The Data Input Sources Tool is found here: http://ghdx.healthdata.org/gbd-2019/data-input-sources.

1.7 GHDx

The Global Health Data Exchange (GHDx) is a catalogue of data sources related to health. It includes citations for all sources used in the most recent version of the annually published GBD. (It also includes sources that are not included in or relevant to the GBD, but are health-related and may be of use to other research.) Users can perform searches based upon keyword, location, data type, and year. Full citation information is provided for all sources. Wherever possible, links are provided to the data contained in each source. Where data holders prohibit direct access to the data and require further registration or request, links are provided to the relevant instructions.
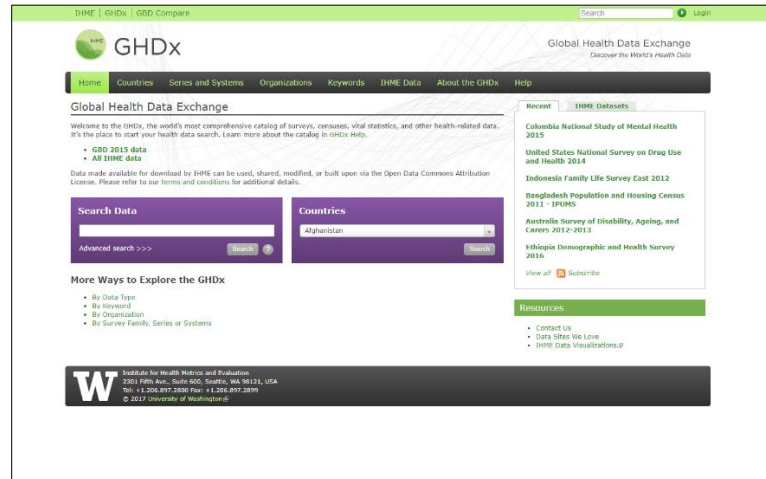

Figure 7. GHDx

The GHDx is found here: http://ghdx.healthdata.org/.

1.8 Code

Starting with the 2015 study, GBD publishes its analytic code in concordance with the GATHER guidelines.

Code is found here: http://ghdx.healthdata.org/gbd-2019/code.

Section 2: Term Definitions

Term	Definition
Age group	A population segment within a specified age range.
All-cause mortality rate	The number of deaths due to all conditions by the mid-year population.
Cause	A single disease or injury or an aggregation of diseases and injuries that causes death or disability.
Cause hierarchy	The classification of diseases and injuries. The causes in GBD are classified into 4 levels. At level 1, there are three large cause groupings: communicable, maternal and neonatal conditions and nutritional deficiencies (CMNN); non-communicable diseases (NCDs); and injuries. At level 2 there are 21 disease and injury categories. The finest level of detail in causes is provided at levels 3 and 4.
Cause-specific mortality Rate	The number of deaths due to cause divided by the mid-year population.
Continuous variable	A population characteristic that is measured on a continuous scale (e.g., the mean level of blood pressure or body mass index).
Covariate	Covariates are variables that have a positive or negative impact on disease or conditions in GBD. For countries that contain little data, covariates are important part of helping fill in the gaps of missing information. For example, since there is not very much information collected on North Korea, from North Korea's GDP we can extract data on how much food is being produced and use that to make estimates on the amount of food being consumed by individuals.
Deaths	Deaths occurring in a population during a certain time period.
Disability-adjusted life years (DALYs)	The sum of years lost due to premature death (YLLs) and years lived with disability (YLDs). DALYs are also defined as years of healthy life lost.
Etiology	The cause, set of causes, or manner of causation of a disease or condition. For example, diarrhea is a cause in the GBD. Diarrhea itself has many causes (like norovirus). Diarrhea is the cause; norovirus is the etiology.
Excess mortality rate	The number of excess deaths divided by the number of prevalent cases. It is equivalent to the cause-specific mortality rate divided by prevalence.
Expected value (life expectancy, deaths, YLLs, YLDs, DALYs)	The value of a specified measure (life expectancy, deaths, YLLs, YLDs, or DALYs) that is expected for a particular GBD location and year, given its socio-demographic development status as measured by SDI.
Fatal discontinuities	An increase in the death rate of more than 1 per million, resulting from conflict and terrorism, natural disasters, major transport accidents or epidemics.
Fertility	The actual level of reproduction of a population, based on the number of live births that occur. Fertility is often measured in terms of women of childbearing age, defined as 15-49 years. GBD estimates fertility for females ages 10-54.
Healthcare Access and Quality Index (HAQ)	A summary measure of personal health care access and quality for a given location. HAQ is based on risk-standardized mortality rates from causes that, in the presence of high-quality health care, should not result in death – also known as amenable mortality.

Term	Definition
Healthy life expectancy, or Health-adjusted life expectancy (HALE)	The number of years that a person at a given age can expect to live in good health, if the rates of all-cause mortality and all-cause disability in a specified year of interest would remain constant into the future.
Impairment	Consequences (or sequelae) of multiple underlying causes for which the main source of data pertain to the sum of these across all causes. GBD currently measures 9 impairments: vision loss, hearing loss, anemia, heart failure, epilepsy, infertility, developmental intellectual disability, pelvic inflammatory disease and Guillain-Barré syndrome.
Incidence	The number of new cases of a given cause during a given period in a specified population. In the results tools it is expressed as the number of new cases in a year divided by the mid-year population size. In the analytical tool EpiViz, it is defined as the number of new cases among those not yet affected by the cause in the population.
Injury n-codes (Injury by Nature)	Injuries are classified into two dimensions: the cause of injury (e.g. road injury, fall or interpersonal violence) and the nature of injury ('n-code') that determines the bodily consequences of the injury (e.g. fracture or head injury). Causes of death are classified by cause of injury; disability is determined by the nature of injury.
Life expectancy	The number of years a person is expected to live at a given age assuming he or she will experience the age-specific mortality rate observed in a given year throughout his or her lifetime. For GBD, the life expectancy associated with an age group (e.g., 50- to 54-year-olds) is life expectancy at the starting year of the age group.
Life expectancy (without fatal discontinuities or HIV)	Life expectancy when the impact of fatal discontinuities or the HIV epidemic is removed.
Life expectancy (without fatal discontinuities, with HIV)	Life expectancy as estimated without including the impact of fatal discontinuities but including the HIV epidemic.
Life expectancy decomposition (LE Decomp)	Changes in life expectancy over time attributed to the causes of death that resulted in the changes in life expectancy.
Life table	A table which shows, for a person at each age, what the probability is that they die before their next birthday. Life tables are used to measure mortality, survivorship, and the life expectancy of a population at varying ages.
Location	Includes country, non-sovereign region, principal administrative unit of a country (e.g., state, province), GBD region, or other custom administrative division, such as World Bank Income Level or WHO region.
Maternal mortality ratio (MMR)	The number of maternal deaths per 100,000 live births. GBD defines maternal deaths as any death of a woman while pregnant or within one year of termination of pregnancy, irrespective of the duration and site of the pregnancy, from any cause related to or aggravated by the pregnancy or its management but not from accidental or incidental causes. Ages included range from 10 to 54 years.

Term	Definition
Measure	The indicator for which estimates are produced.
Metric (Units)	The unit by which a measure is expressed. E.g., number, percent, rate, etc.
Other cause mortality rate	The number of all-cause deaths minus deaths due to a specific cause divided by the mid-year population.
Prevalence	The proportion of people in a population who are a case of a disease, injury or sequela. All results in GBD refer to point prevalence.
Probability of death	The probability that a person dies during an interval of two ages (e.g., between birth and age 5), if the rates of all-cause mortality in a specified year of interest would remain constant into the future.
Probability of death (without fatal discontinuities or HIV)	The estimated probability of death without the impact of fatal discontinuities or the HIV epidemic.
Probability of death (without fatal discontinuities, with HIV)	The estimated probability of death without the impact of fatal discontinuities but including the HIV epidemic.
Proportion	The number of cases with a certain characteristic in a population (e.g., the proportion of HIV that is due to sexual transmission or the proportion of households using solid fuels for cooking).
Relative risk	The ratio of the mortality rate in the diseased and the mortality rate in the non-diseased population.
Remission	The number of cases that resolve or are cured per person-year of follow-up.
Risk (Risk factor)	An attribute, behavior, exposure, or other factor which is causally associated with an increased (or decreased) probability of a disease or injury. If the probability decreased, the risk is a protective factor.
Sequela	The non-fatal consequence of a disease or injury. In GBD, mutually exclusive sequelae are defined for each disease and injury that in a parsimonious manner quantify the consequences that cause disability or may lead to disability in the future.
Sex	Male, female or both sexes combined.
Socio-demographic index (SDI)	A compound measure of income, average years of schooling and fertility for each GBD location and year which is used as a measure of socio-demographic development.
Standardized mortality ratio	The mortality rate in the diseased compared to the mortality rate in the entire population.
Summary exposure value (SEV)	A measure of a population's exposure to a risk factor that takes into account the extent of exposure by risk level and the severity of that risk's contribution to disease burden. SEV takes the value zero when no excess risk for a population exists and the value one when the total population is at the highest level of risk; we report SEV on a scale from 0% to 100% to emphasize that it is risk-weighted prevalence.

Term	Definition
Sustainable Development Goals (SDGs)	The United Nations established, in September 2015, the Sustainable Development Goals (SDGs), which specify 17 universal goals, 169 targets, and 232 indicators leading up to 2030. Member countries adopted this set of goals to end poverty, protect the planet, and ensure prosperity for all. The GBD study estimates progress made by countries for 41 health-related SDG indicators.
Uncertainty interval	A range of values that reflects the certainty of an estimate. In GBD, every estimate is calculated 1,000 times, each time sampling from distributions rather than point estimates for data inputs, data transformations and model choice. The 95th uncertainty interval is determined by the 25th and 975th value of the 1,000 values after ordering them from smallest to largest. Larger uncertainty intervals can result from limited data availability, small studies, and conflicting data, while smaller uncertainty intervals can result from extensive data availability, large studies, and data that are consistent across sources.
Value	The mean value of an estimate.
With-condition mortality rate	Total number of deaths among prevalent cases in a year of interest. It is equivalent to the sum of other cause mortality rate and cause-specific mortality rate.
Year	The period of 365 days (or 366 days in leap years) in the Gregorian calendar divided into 12 months beginning with January and ending with December.
Years lived with disability (YLDs)	Years lived with any short-term or long-term health loss weighted for severity by the disability weights.
Years of life lost (YLLs)	Years of life lost due to premature mortality. YLLs are the multiplication of deaths and a standard life expectancy at the age of death. The standard life expectancy is derived from a life table that contains the lowest observed mortality rate at each age that has been observed in any population greater than 5 million.

Section 3: Tool Outputs and Indicators Overview

GBD 2017 Tool		GBD Results Tool	GBD Compare	MortViz	CoDViz	EpiViz	GHDx	GBD Data Input Sources
GBD Component	Mortality	X	X	X			X	X
	Population						X	X
	Fertility						X	X
	Migration						X	
	Causes of death	X	X		X		X	X
	Nonfatal health Outcomes	X	X			X	X	X
	Risk factors	X	X				X	X
	Covariates						X	X
Output	Estimates	X	X	X	X	X	X	
	Model input data			X	X	X		
	Data input sources							X
Dimension	Age group	X	X	X	X	X	X	X
	Cause	X	X		X	X	X	X
	Impairment	X	X			X		X
	Injuries by nature	X	X					X
	Location	X	X	X	X	X	X	X
	Risk	X	X			X	X	X
	Sex	X	X	X	X	X	X	X
	Year	X	X	X	X	X	X	X
Measure / Indicator	Deaths	X	X	X	X			
	Disability-adjusted life years (DALYs)	X	X					
	Years lived with disability (YLDs)	X	X					
	Years of life lost (YLLs)	X	X					
	Prevalence	X	X			X		
	Incidence	X	X			X		
	Maternal mortality ratio (MMR)	X	X					
	Probability of death		X	X			X	
	Life expectancy	X	X	X			X	
	Health-adjusted life expectancy (HALE)	X	X				X	
	Summary exposure value (SEV)	X	X					
	Life expectancy decomposition		X					
	Expected value (life expectancy, deaths, YLLs, YLDs, DALYs)		X					

GBD 2017 Tool		GBD Results Tool	GBD Compare	MortViz	CoDViz	EpiViz	GHDx	GBD Data Input Sources
Measure / Indicator	Covariates						X	
	Population						X	
	Fertility						X	
	Universal healthcare (UHC) effective coverage index						X	
	Life expectancy (without fatal discontinuities or HIV)			X			X	
	Life expectancy (without fatal discontinuities, with HIV)			X			X	
	Probability of death (without fatal discontinuities or HIV)			X			X	
	Probability of death (without fatal discontinuities, with HIV)			X			X	
	Remission					X		
	Excess mortality rate					X		
	Standardized mortality ratio					X		
	With-condition mortality rate					X		
	All-cause mortality rate			X		X		
	Cause-specific mortality rate				X	X		
	Other cause mortality rate					X		
	Proportion					X		
	Continuous measure					X		

Appendices

Appendix 1: Cause, REI, and Locations Hierarchies [XLSX File]
File: IHME_GBD_2019_A1_HIERARCHIES_Y2020M10D15.XLSX

The **Cause Hierarchy, REI Hierarchy** (risk factor, etiology, impairment, and injury n-code), and **GBD 2019 Locations Hierarchy** sheets contain GBD 2019 reporting hierarchies with values in the order they appear in online tools, such as the GBD Results Tool and GBD Compare.

The **All Locations Hierarchies** sheet contains the GBD 2019 reporting hierarchy and a number of other hierarchies which will allow users of GBD 2019 results to aggregate results by location in various ways (by GBD regions, World Bank regions, OECD countries, African Union countries, etc.).

To filter for the child values of a given parent value in any hierarchy file (e.g., all countries in the GBD region "South Asia," or all causes under "Chronic respiratory diseases"), filter on the parent ID in the **Parent ID** column.

To view each hierarchy in its proper nested order (the order they appear in online tools), sort by the **Sort Order** column.

For the Cause Hierarchy, users can also filter by cause outline value (e.g., all causes in "A" or "A.1"), or for causes for which either only years of life lost (YLLs) or years lived with disability (YLDs) results were produced.

For the All Locations Hierarchies, filter first by **Location Set Version ID**, and then by Parent ID and/or Sort Order.

Locations and Levels: In the locations hierarchies files:

- Level 3 = country or territory
- Levels 4 and 5 = subnational units

Location Set Version IDs

Location Set ID	Location Set
543	GBD 2019 Reporting
587	African Union
592	Commonwealth
591	European Union
593	Four World Regions
588	G20
586	Nordic Region
589	OECD Countries
670	WHO region
596	World Bank Income Levels
595	World Bank Regions

Select Variable Definitions

Variable	Definition
Cause ID	Cause IDs for cause variables come from an IHME database that creates and stores unique numeric identifiers for each GBD cause.
Cause Outline	This outline represents the hierarchy of causes and the depth of each cause in the hierarchy by alpha-numeric code for the current round of GBD.
Level	This indicates a location value's level in a location hierarchy.
Location ID	Locations IDs for geographic variables come from an IHME database that creates and stores unique numeric identifiers for locations of various location types.
Location Set Version ID	IDs for GBD 2019 location sets (including the GBD locations and custom region locations) come from an IHME database that creates and stores unique numeric identifiers for location set versions.
Parent	This indicates the parent ID for a value's parent in the hierarchy. For example, all 31 Mexican states and 1 federal district (Mexico City) have a parent ID of 130, the location ID for the country of Mexico.
REI ID	REI IDs for risk, etiology, impairment, and injury n-code (injury by nature) variables come from an IHME database that creates and stores unique numeric identifiers for each risk, etiology, impairment, and injury.
REI Type	This indicates whether the value is a risk, etiology, impairment, or injury.
Sort Order	This can be used to sort a set of values in a particular order (not alphabetical, but the order in which the values appear in online tools, paper tables, etc.).
YLL Only	Indicates causes for which only mortality estimates (deaths, years of life lost (YLLs)) were produced.
YLD Only	Indicates causes for which only nonfatal estimates (years lived with disability (YLDs)) were produced.

Appendix 2: Results by Measure and Cause-Risk/Impairment/Injury Pairs [XLSX File]
File: IHME_GBD_2019_A2_RESULTS_BY_MEASURE_Y2020M10D15.XLSX

This Excel file contains a set of sheets that display the measures (deaths, DALYs, incidence, etc.) for which estimates are provided for each GBD cause, risk, impairment, and injury by nature.

Filter on cause, risk, etc., or a measure (deaths, YLDs, incidence, etc.), to see available results.

Additionally, the sheets display which causes each risk, impairment, and injury by nature are associated with. For example:

- In the "Risk" sheet, filter for "Ambient particulate matter pollution" in the "Risks" column and see the 18 causes associated with the Ambient particulate matter pollution risk

In the "Impairment" sheet, filter for "Blindness" in the "Impairments" column and see the 34 causes associated with the Blindness impairment

Appendix 3: Metrics by Measure Definitions [XLSX File]

File: IHME_GBD_2019_A3_MEASURE_METRIC_DEFINITIONS_Y2020M10D15.XLSX

This file (also included as a tab in the "Results by Measure" file) provides definitions for each metric, by measure, contained in the GBD Results Tool and GBD Compare.

Appendix 4: Contexts by Age [XLSX File]

File: IHME_GBD_2019_A4_CONTEXTS_BY_AGE_Y2020M10D15.XLSX

This Excel file provides the full set of age groups used, cumulatively, in results available from the GBD Results, and the context each age group appears in. Filter on contexts like Cause, Risk, Life expectancy, Population, etc. to the see the ages available for each.

Ages are largely the same in GBD Compare and the other GBD data visualizations, but similar tables for those tools are forthcoming.