

Types of data sources used in the Global Burden of Disease (GBD) Study

This document describes the many secondary data sources used in the Global Burden of Disease (GBD) Study.

Data is critical to the GBD.

As a key principle, the GBD endeavors to use all available data in each of the source types outlined below and, where necessary, applies statistical and epidemiological methods to make the data comparable across different geographies and over time. Estimates are only as strong as the evidence base underlying them so inclusion of all possible data source types in the GBD is critical to the credibility of the study.

Documentation should accompany each data source.

Each data source should be accompanied by all documentation necessary for researchers to understand how to work with the data. Documentation may include a questionnaire, data collection form, codebook, data dictionary, and/or reports or word files that provide information on the methodology for data collection.

Microdata, tabulations, or reports?

The more detailed the data the better. The gold standard is individual-level microdata by detailed age, sex, and year (1980 to present) with GPS coordinates. In most cases, this level of detail is not available so the GBD applies methods to enable use of tabulated or report data by aggregate age groups, grouped years, and location names. IHME can provide information on how to generate GBD-friendly tabulations.

What about sub-populations?

Data should be representative of the total population within a given GBD geography (i.e. country or subnational unit). Data collected for only non-representative sub-populations (e.g. prisoners, refugees, specific occupational groups, etc.) are not able to be used.

The GBD always uses these types of data sources...

<i>Data source</i>	<i>Definition</i>	<i>Uses in the GBD</i>
Censuses	Collects data on every member of a population at a single point in time. Contains demographic, economic, and social data.	Population denominator, level of education attained, total fertility rate, GDP
Population registry	Some countries have adopted a population registry, thus removing the need for a census, which continuously collects data on the demographics of a population. It can also include births, deaths, migration, and educational attainment. <i>*Found in Austria, Denmark, Finland, Israel, Japan, The Netherlands, Norway, Sweden</i>	Population denominator
Vital Registration (VR) <i>*Also called civil registration or vital statistics</i>	System that continuously collects data on the “vital” events of every member of a population. Contains live births and deaths (including underlying cause of death).	Causes of death, live births, deaths, underlying cause of death <i>*Usable even if the full population is not covered, e.g. incomplete</i>



If national vital registration does not exist, the GBD uses...

Sample registration	Similar to a vital registration system but only covering a sample of the population <i>*Found in India, China, Indonesia, Bangladesh</i>	Causes of death
Demographic Surveillance Systems (DSS) <i>*Also called Health and Demographic Surveillance Systems (HDSS)</i>	Continuous demographic monitoring of a geographically defined population. Collects demographic data on births, migrations, pregnancies, and deaths (including cause of death). <i>*Found in many developing countries; INDEPTH is the largest network.</i>	Causes of death
Verbal Autopsies (VA)	An interview-based survey method of determining an individual's cause of death in populations without a complete vital registration system.	Causes of death

And the GBD always uses...

<i>Data source</i>	<i>Definition</i>	<i>Uses in the GBD</i>
Hospital data	Data from a public or private hospital system including inpatient administrative records, hospital discharge data, and outpatient clinic visits. <i>*Data must include International Classification of Diseases (ICD) diagnosis code(s)</i>	Disease incidence and prevalence (for morbidity only) <i>*Hospital data is not used for causes of death in the GBD approach.</i>
Health insurance claims data <i>*Also called claims data</i>	Data that describes patient-provider interactions. <i>*This source is new to the GBD and was added for the GBD 2016 round</i> <i>**In order to be used in the GBD, 1) public or private insurance must be widespread, 2) data must include ICD diagnosis code(s).</i>	Disease incidence and prevalence (for morbidity only)
Surveys <i>*Also called household surveys</i>	Surveys collect information for a defined population on health-related behaviors, and psychosocial well-being. Can include biomarker data <i>*E.g. DHS, MICS, LSMS, GATS, national surveys, etc.</i>	A broad range of indicators, from child mortality to diabetes prevalence to tobacco smoking behavior
Disease registries	Centralized databases for the collection of information on patients with a specific diagnosis, particularly for chronic conditions <i>*E.g. registries for cancer, congenital abnormalities, end-stage renal disease, etc.</i>	Incidence, prevalence, and mortality due to specific diseases or conditions
Morbidity notification data <i>*Also called epi surveillance or disease surveillance data</i>	A notifiable infectious disease or condition is one for which regular, frequent, and timely information regarding individual cases is considered necessary for the prevention and control of the disease or condition.	<i>*Although these data typically undercount, they can be useful to determine the geographic distribution</i>
Police records	Data describing type of offence. Useful for mortality related to injuries and homicides.	<i>*Although these data typically undercount, they can be used for mortality due to injuries, homicides</i>

<i>Published literature</i>	<p>Articles and papers containing health-related data that are published in peer-reviewed journals. Grey literature is unpublished or published in a non-commercial form and can also be useful.</p> <p><i>*It may be necessary to search local language publications as well as international literature (most of the latter should already be in the GBD databases if published in English).</i></p>	Disease incidence and prevalence (for morbidity only), as well as a variety of other indicators including disease duration and progression, etc.
------------------------------------	--	--

How to conduct a data inventory.

A data inventory documents the types of data available, where they are being held, and with whom at each institute you would need to speak with to acquire the data. When conducting a data inventory, it is important to be systematic and avoid choosing “favorite” data sources.

The Global Health Data Exchange (<http://ghdx.healthdata.org/>) is a good starting point for a data inventory as IHME has catalogued a number of sources already for each country.

Questions? Contact gbdsec@uw.edu for guidance on how to begin a data inventory or more information on the data types listed above.