

Global, Regional and National Burden of Cardiovascular Diseases and Risk Factors in 204 countries and territories, 1990-2023

Appendix 3: Non-fatal estimation methods

Preamble

This appendix contains methodological details about the Global Burden of Disease non-fatal estimation process. Parts of this section have been reproduced from the supplementary appendix of “Non-fatal burden of 375 diseases and injuries, risk-attributable burden of 88 risk factors, and healthy life expectancy in 204 countries and territories, including 660 subnational locations, 1990–2023: a systematic analysis for the Global Burden of Disease Study 2023”.

Hay et al. Non-fatal burden of 375 diseases and injuries, risk-attributable burden of 88 risk factors, and healthy life expectancy in 204 countries and territories, including 660 subnational locations, 1990–2023: a systematic analysis for the Global Burden of Disease Study 2023. *Lancet* (In press).

Appendix 1: Non-fatal methods appendix to “Burden of 375 diseases and injuries, risk-attributable burden of 88 risk factors, and healthy life expectancy in 204 countries and territories, including 660 subnational locations, 1990 – 2023: a systematic analysis for the Global Burden of Disease Study 2023”

Preamble

This appendix provides further methodological detail for “Burden of 375 diseases and injuries, risk-attributable burden of 88 risk factors, and healthy life expectancy in 204 countries and territories, including 660 subnational locations, 1990 – 2023: a systematic analysis for the Global Burden of Disease Study 2023”. This study complies with the Guidelines for Accurate and Transparent Health Estimates Reporting (GATHER) recommendations.¹ It includes detailed tables and information on data in an effort to maximise transparency in our estimation processes and provide a comprehensive description of analytical steps. We intend this appendix to be a living document, to be updated with each iteration of the Global Burden of Disease Study.

Portions of this appendix have been reproduced or adapted from appendices for Lim et al 2012,² GBD 2015 Disease and Injury Incidence and Prevalence Collaborators,³ GBD 2016 Disease and Injury Incidence and Prevalence Collaborators,⁴ GBD 2017 Disease and Injury Incidence and Prevalence Collaborators,⁵ GBD 2019 Diseases and Injuries Collaborators,⁶ and GBD 2021 Diseases and Injuries Collaborators.⁷ References are provided for reproduced or adapted sections.

Table of Contents

Preamble	3
Table of Contents	4
List of methods appendix figures	5
Section 1: GBD Overview	7
Section 1.1: Global Burden of Diseases, Injuries, and Risk Factors Study 2023	7
Section 1.2: Geographical locations of the analysis	8
Section 1.3: Statement of GATHER compliance	8
Section 1.4: GBD cause hierarchy	8
Section 1.5: Data input sources overview	9
Section 1.6: Funding sources	9
Section 1.7: Abbreviations	10
Section 2: Non-fatal outcome estimation	11
Section 2.1: Data sources, identification, and extraction	11
Section 2.1.1: Systematic reviews	12
Section 2.1.2: Survey data preparation	13
Section 2.1.3: Disease registries	13
Section 2.1.4: Case notifications	13
Section 2.2: Clinical data sources and methods summary	13
Section 2.2.1: Mapping diagnoses to GBD diseases and injuries	14
Section 2.2.2 Individual-level clinical data sources with unique identifiers in GBD 2023	14
Section 2.2.3: Aggregate inpatient hospital admission sources used in GBD 2023	18
Section 2.2.4: Estimation of the inpatient utilization envelope	19
Section 2.2.5: Correction Factors	23
Section 2.2.6: Outpatient encounter data	29
Section 2.2.7: New data added for GBD 2023	30
Section 2.3 Data Adjustments	31
Section 2.3.1: Crosswalking	31
Section 2.3.4 Network Analysis	34
Section 2.3.5 Age sex splitting	35
Section 2.4: Spatiotemporal Gaussian process regression (ST-GPR) modelling	35
Section 2.4.1 Estimating mean functions	36

Section 2.4.1: Estimating error variance	38
Section 2.4.2: Estimating the covariance function	39
Section 2.4.3: Prediction using GPR	40
Section 2.4.4: Subnational scaling and aggregation	40
Section 2.5: MR-BRT meta-regression modelling	40
Section 2.5.1 MR-BRT Overview	40
Section 2.5.2 MR-BRT Formula.....	40
Section 2.5.3 MR-BRT Features	41
Section 2.6: DisMod-MR 2.1 estimation	42
Section 2.6.1: Estimation of sequelae and causes	42
Section 2.6.2: DisMod-MR 2.1 description	42
Section 2.6.3: DisMod-MR 2.1 likelihood estimation.....	47
Section 2.7: DisMod-AT estimation	48
Section 2.8: Impairment and underlying cause estimation	51
Section 2.8.1: Impairment squeeze.....	52
Section 2.9: Disability weights	54
Section 2.9.1: GBD 2010 Disability Weights Measurement Study	55
Section 2.9.2: GBD 2013 European disability weights measurement study	56
Section 2.10: Comorbidity correction (COMO)	57
Section 2.11: YLD computation, uncertainty, and residual YLDs.....	59
Section 2.11.1: Residual YLDs.....	59
Section 2.12: Birth prevalence.....	59
Section 3: SDI	60
Section 3.1: SDI definition	60
Section 3.2: Development of revised SDI indicator.....	60
Section 4: Estimation process for DALYs	62
Section 5: HALE.....	62
References.....	63
Section 6: Cardiovascular non-fatal methods.....	66

List of methods appendix figures

Appendix Figures

Figure S1: Number of countries and territories with non-fatal data by disease group for a given year by GBD region

Figure S2: Number of countries and territories with non-fatal data for a given year by GBD region

Figure S3: Processing steps for individual-level clinical data sources with unique identifiers

Figure S4: Merative MarketScan Research Databases Enrollee Population by Year

Figure S5: Medicare Part A Enrollee Population by Year

Figure S6: Medicare Part A+B Enrollee Population by Year from 5% Enhanced Sample

Figure S7: Poland Enrollee Population by Year

Figure S8: Processing steps for aggregate inpatient hospital admission sources

Figure S9: Overview process of estimation of hospital utilization envelope

Figure S10: Overview of Correction Factor process

Figure S11: Readmissions correction model in linear space for acute encephalitis

Figure S12: Non-primary correction model in linear space for acute encephalitis

Figure S13: Non-primary and outpatient correction model in linear space for acute encephalitis

Figure S14: Processing steps for outpatient encounter data

Figure S15: System of differential equations that captures relationships between rates and compartments

Figure S16: Dismod-MR 2.1 sequence of estimation cascade

Figure S17: SF-12 composite scores and disability weights for 60 health states with fitted loess regression

Figure S18: DALY burden estimation for GBD 2023

Section 1: GBD Overview

Section 1.1: Global Burden of Diseases, Injuries, and Risk Factors Study 2023

The Global Burden of Diseases, Injuries, and Risk Factors Study (GBD) is a collaborative research effort aimed at estimating morbidity and mortality from a comprehensive set of diseases, injuries, and risk factors. The GBD Collaborator Network draws on the expertise of over 14,000 contributors from around the world. For this paper, we estimated numbers and rates of incidence, prevalence, years lived with disability (YLDs), and disability-adjusted life-years (DALYs) for the years 1990–2023; we estimated deaths and years of life lost (YLLs) for 1980–2023 by age, sex, and location.

To reduce computing power and time, we reduced the number of draws (or computations) per process to 250, from 500 in GBD 2021. Based on simulation testing, we determined that a change in the number of draws did not impact final mean estimates, nor lead to inappropriately narrow uncertainty estimates.

A simple R simulation illustrates how the estimated UI is only weakly dependent on the number of draws (500 or 250). Correctly using draws gives intervals of about the same width, and which interval is wider depends on random variation (as it should). Some code to illustrate this is given below. Running the example below (which simulates 250 draws and 500 draws around a simple normal distribution) provides an uncertainty width estimated from 250 draws and 500 draws. Because of the random nature of this code (and because we are not setting a seed), the values will change and are random each time. Running this repeatedly, the reader will see the widths remain similar depending on whether there are 250 draws or 500 draws, and the option that leads to the narrower UI width changes each time it is run.

```
# Sample 250 draws from N(0,1)

sample_250 <- rnorm(250, mean = 0, sd = 1)

ci_250 <- quantile(sample_250, probs = c(0.025, 0.975))

width_250 <- diff(ci_250)


# Sample 500 draws from N(0,1)

sample_500 <- rnorm(500, mean = 0, sd = 1)

ci_500 <- quantile(sample_500, probs = c(0.025, 0.975))

width_500 <- diff(ci_500)


# Create a simple table with the widths

ci_widths <- data.frame(Sample_Size = c(250, 500), CI_Width = c(width_250, width_500))

# Print the table
```

print(ci_widths)

Other notes

Section 1.2: Geographical locations of the analysis

We produced estimates for 204 countries and territories that were grouped into 21 regions and seven super regions (table 1). The seven super-regions are central Europe, eastern Europe, and central Asia; high income; Latin America and the Caribbean; north Africa and the Middle East; south Asia; southeast Asia, east Asia, and Oceania; and sub-Saharan Africa. In GBD 2023, we continue to analyse at subnational levels countries that were added in previous cycles, including Brazil, China, Ethiopia, India, Indonesia, Iran, Italy, Japan, Kenya, Mexico, New Zealand, Nigeria, Norway, Pakistan, Russia, the Philippines, Poland, South Africa, and the USA. All analyses are at the first level of administrative organisation within each country except for New Zealand (by Māori ethnicity), and the Philippines (by provinces). To meet data use requirements, in this publication we present subnational estimates for Brazil, India, Indonesia, Japan, Kenya, Mexico, and the USA; given space constraints, these results are presented in Appendix 3 instead of the main text. Subnational estimates for China are included in maps but are not reported in appendix tables. Subnational estimates for other countries will be released in separate publications, although please note that we only release estimates for a subset of these countries, per agreements with country partners.

At the most detailed spatial resolution, we generated estimates for 843 unique locations. As was done in GBD 2021, in GBD 2023, we continue to use the set of locations defined as standard locations and non-standard locations. Standard GBD locations are defined as the set of all subnationals belonging to countries where data quality is high and with populations over 200 million, in addition to all other countries. Standard locations include the subnationals for China, India, the USA, and Brazil, but not Indonesia; data for China, India, the USA, and Brazil are also included at the country level. All other countries with subnational estimates are defined as non-standard locations.

Section 1.3: Statement of GATHER compliance

This study complies with the Guidelines for Accurate and Transparent Health Estimates Reporting (GATHER) recommendations.¹ We have documented the steps involved in our analytical procedures and detailed the data sources used. See table 2 for the GATHER checklist. The GATHER recommendations may be found here: <http://gather-statement.org/>.

Section 1.4: GBD cause hierarchy

The GBD cause and sequelae list is organized hierarchically (see table 3) to accommodate different purposes and needs of various users.

The first two levels aggregate causes into general groupings. At Level 1 there are three cause groups: communicable, maternal, neonatal, and nutritional diseases (Group 1 diseases); non-communicable diseases (Group 2); and injuries (Group 3). These Level 1 aggregates are subdivided at Level 2 of the hierarchy into 22 cause groupings (e.g., neonatal disorders, neurological disorders, and transport injuries). The disaggregation into Levels 3 and 4 contains the finest level of detail for causes captured in GBD 2023. The greatest detail available for some causes, such as anxiety disorders or rheumatoid

arthritis, is at Level 3 of the hierarchy, while other specific causes are at Level 4 of the hierarchy with an aggregate category at Level 3 (for example, depressive disorders at Level 3, which encompasses major depressive disorders and dysthymia at Level 4). Sequelae of diseases and injuries are organised at Levels 5 and 6 of the hierarchy. In GBD, sequelae are defined as distinct, mutually exclusive categories of health consequences that can be directly attributed to a cause. For example, both neuropathy and blindness due to diabetic retinopathy are sequelae of diabetes; stroke and ischaemic heart disease are not, as these consequences cannot be categorically ascribed to diabetes in an individual despite good evidence for increased risk of these outcomes. The finest detail for all sequelae estimated in GBD is at Level 6 and is aggregated into summary sequelae categories (Level 5) for causes with large numbers of sequelae. Examples include the grouping of the infectious disease episodes and long-term sequelae of meningitis. For GBD 2023 there are 3704 mutually exclusive and collectively exhaustive sequela, 2106 cause sequelae and 1598 injuries sequelae, and thus our YLD estimates at each level of the hierarchy sum to the total of the level above. Prevalence and incidence aggregation is estimated at the level of individuals who may have more than one sequela or disease and therefore are not additive.

The GBD cause list continues to evolve to reflect the policy relevance, and public health and medical care importance of the causes of major losses of health. The cause and sequelae list expanded based on input from the Scientific Council and GBD collaborator network. For GBD 2023, the causes of death cause list has increased to 292 causes, from the 288 causes in GBD 2021. The non-fatal cause list has expanded from 365 causes in GBD 2021 to 370 causes in GBD 2023. The total number of fatal and non-fatal causes combined for GBD 2023 is 375. As in GBD 2021, we made no estimates for YLDs for just five causes, either because no disability is possible (as is the case with sudden infant death syndrome); because disability may occur rarely but at levels too low for accurate estimation given the data (as for aortic aneurysm); or because the disability is captured by the complicating causes that led to that cause of death (as for indirect maternal deaths, late maternal deaths, and maternal deaths aggravated by HIV/AIDS).

Section 1.5: Data input sources overview

GBD 2023 synthesises a large and growing number of data input sources including surveys, censuses, vital statistics, and other health-related data sources. The data from these sources are used to estimate morbidity; illness, and injury; and attributable risk for 204 countries and territories from 1990 to 2023; mortality deaths are estimated from 1980 to 2023. The input sources are accessible through an interactive citation tool available in the GHDx.

Citations for specific GBD components, causes and risks, and locations can be found through the Data Input Sources Tool in GHDx: <http://ghdx.healthdata.org/gbd/2020/data-input-sources>. This tool allows users to view and access GHDx records for input sources and export a comma-separated value (CSV) file that includes metadata, citations, and information about where the data were used in GBD. As required by GATHER, additional metadata for input sources are available through the citation tool as well.

Section 1.6: Funding sources

This publication and the research it presents was funded by the Gates Foundation (award OPP1152504); Queensland Department of Health, Australia; the New Zealand Ministry of Health; the Norwegian Institute of Public Health; St. Jude Children's Research Hospital; UK Department of Health and Social

Care; and Bloomberg Philanthropies. The funders of the study had no role in study design, data collection, data analysis, data interpretation, or writing of the report. All authors had full access to all data in the study and had final responsibility for the decision to submit for publication.

Section 1.7: Abbreviations

ARC	annualized rate of change
ASFR	age-specific fertility rate
ACMR	all-cause mortality rate
BMI	Body Mass Index
CMNN	Communicable, maternal, neonatal, and nutritional diseases
CoD	causes of death
CODEm	Cause of Death Ensemble modelling
COMO	comorbidity correction
COPD	Chronic obstructive pulmonary disease
CSMR	cause-specific mortality rates
CV	coefficient of variation
DALYs	disability-adjusted life-years
DisMod-AT	disease model-Bayesian age-time
DisMod-MR	disease model-Bayesian meta-regression
DW	disability weight
EDU15+	education for those 15 years old and older
EMR	excess mortality rate
GATHER	Guidelines for Accurate and Transparent Health Estimates Reporting
GBD	Global Burden of Diseases, Injuries, and Risk Factors Study
GHDx	Global Health Data Exchange
GPR	Gaussian process regression
HALE	healthy life expectancy
HAT	human African trypanosomiasis
ICD	International Classification of Diseases
ICG	ICD groups
IFD	in-facility delivery proportion
IHME	Institute for Health Metrics and Evaluation
LASSO	least absolute shrinkage and selection operator
LDI	lag-distributed income
LOESS	locally estimated scatterplot smoothing
MAD	median absolute deviation
MCCD	Medical Certification of Causes of Death
MEPS	Medical Expenditure Panel Survey
MICS	Multiple Indicators Survey
MR-BRT	Meta-regression—Bayesian, regularised, trimmed
NESARC	National Epidemiologic Survey on Alcohol and Related Conditions
NSMHWB	Australian National Survey of Mental Health and Wellbeing
NTDs	neglected tropical diseases
RSME	root mean square error
SARS-CoV 2	Severe acute respiratory syndrome coronavirus 2
SD	Standard deviation
SID HCUP	State Inpatient Database

SDI	Social Demographic Index
ST-GPR	spatiotemporal Gaussian process regression
TFR	total fertility rate
TFU25	younger than 25 years old (fertility rate)
UI	uncertainty interval
UK	United Kingdom
UI	uncertainty interval
USA	United States of America
WHO	World Health Organization
YLDs	years lived with disability
YLLs	years of life lost

Section 2: Non-fatal outcome estimation

The GBD 2023 non-fatal estimation process describes the steps necessary to estimate incidence, prevalence, and YLDs for disease and injury sequelae in GBD 2023. Conceptually, the estimation effort is divided into eight major components: (1) compiling data sources through data identification and extraction; (2) data adjustments; (3) estimation of prevalence and incidence by cause and sequelae by using DisMod-MR 2.1, or alternative modelling strategies for select cause groups; (4) estimation by impairment; (5) severity distributions; (6) incorporation of disability weights (DWs); (7) comorbidity adjustment; and (8) the estimation of YLDs by sequelae and causes. Section 6 contains additional detail specific to each non-fatal disease, impairment, and injury, and their sequelae. Non-fatal modelling strategies vary significantly between causes.

Section 2.1: Data sources, identification, and extraction

Figure S1 is a map showing the number of countries and territories with non-fatal data for a given year, aggregated by GBD Region. We aggregate all “Communicable, maternal, neonatal, and nutritional diseases” in red, “Non-communicable diseases” in blue, and “Injuries” in green.

Figure S1. Number of countries and territories with non-fatal data by disease group for a given year by GBD region

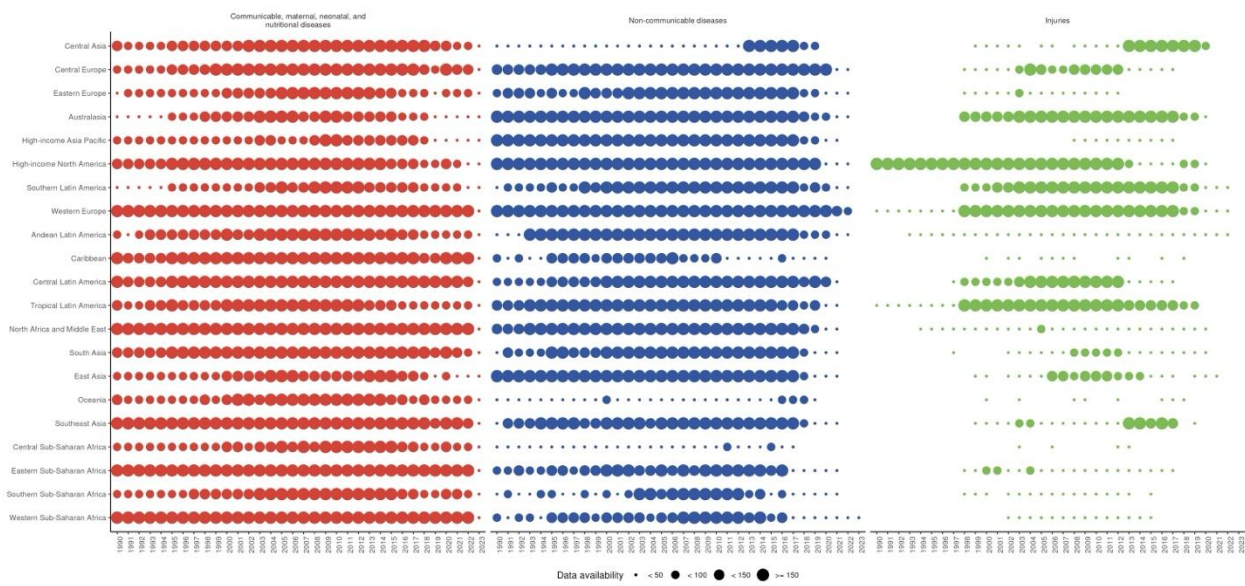
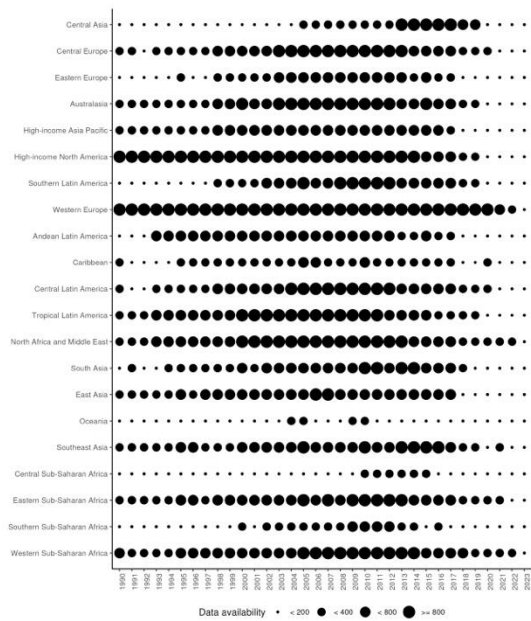


Figure S2 shows a map showing the number of countries and territories with non-fatal data for a given year, aggregated by GBD Region, across all the estimated causes in the GBD.

Figure S2. Number of countries and territories with non-fatal data for a given year by GBD region



Section 2.1.1: Systematic reviews

For GBD 2023, updated systematic reviews were conducted for 122 causes and risk factors. For other disease sequelae, only a small fraction of the existing data appears in the published literature, and other sources predominate, such as survey data, disease registers, notification data, or hospital inpatient data. As was done in past rounds of GBD, data were systematically screened from household surveys archived in the GHDx (<http://ghdx.healthdata.org/>), including Demographic and Health Surveys, Multiple

Indicator Cluster Surveys (MICS), Living Standards Measurement Surveys, and Reproductive Health Surveys. Other national health surveys were identified on the basis of survey series that had yielded usable data for past rounds of GBD, sources suggested to us by in-country GBD collaborators, and surveys identified in major multinational survey data catalogues such as the International Household Survey Network and the WHO Central Data Catalog, as well as through country Ministry of Health and Central Statistical Office websites. Citations for all data sources used for non-fatal estimation in GBD 2023 are provided in searchable form through a web tool (<http://ghdx.healthdata.org/>). A description of the search terms used for cause-specific systematic reviews are detailed by cause in Section 6.

Section 2.1.2: Survey data preparation

For GBD 2023, survey data for which we have access to the unit record data constitutes a substantial part of the underlying data used in the estimation process. During extraction, we concentrated on demographic variables (eg, location, sex, age), survey design variables (eg, sampling strategy and sampling weights), and the variables used to define the population estimate (eg, prevalence or a proportion) and a measure of uncertainty (standard error, confidence interval or sample size, and number of cases).

Section 2.1.3: Disease registries

For GBD 2023 non-fatal estimation, disease registries were an important source for a select number of conditions such as cancers, end-stage renal disease, and congenital disorders.

Registry data is particularly key in the estimation of neoplasms when we consider the increasing attention to non-communicable diseases, particularly cancers, in low and middle-income areas of the world. The GHDx source tool (<http://ghdx.healthdata.org/data-type/disease-registry>) provides a comprehensive list of registry data used in GBD estimation processes.

Section 2.1.4: Case notifications

Case notifications, active screening, intervention coverage studies, and surveillance contributed to estimates of infectious diseases. If data were available, we extracted it from survey and administrative microdata; otherwise, data were extracted from published literature and reports. For many infectious diseases and neglected tropical diseases (NTDs), we used cases for which notification was made by countries to the WHO and other global monitoring entities. The causes for which we used WHO case notification data included tuberculosis, measles, yellow fever, rabies, dengue, cholera, whooping cough, human African trypanosomiasis (HAT), meningitis, and other infectious diseases and NTDs, such as Ebola.

Section 2.2: Clinical data sources and methods summary

Clinical data derived from electronic health records (EHRs) played a key role in the estimation of many non-fatal diseases, impairments, and injuries in GBD 2023. Data sources were heterogeneous in granularity, comprehensiveness, and level of detail, and the methods described below were used to transform data to be comparable and complete across locations, ages, sexes, years, and causes modelled in the GBD.

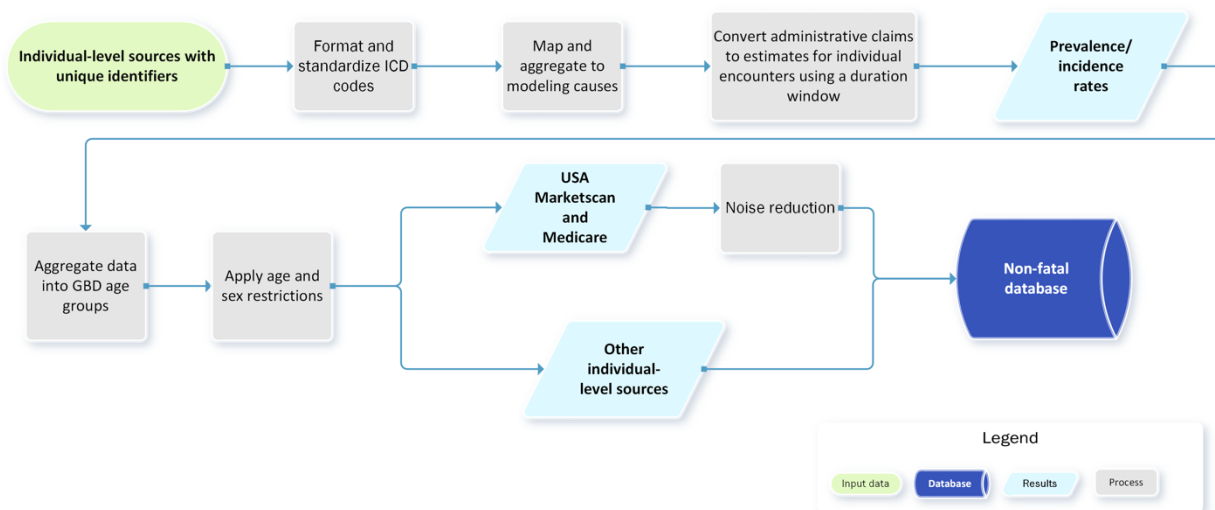
Section 2.2.1: Mapping diagnoses to GBD diseases and injuries

Most clinical sources are coded using the International Classification of Diseases (ICD) system that we map to GBD-defined diagnosis groups. ICD-9 and ICD-10 codes are mapped to what are termed “ICD code groups” (ICGs) with a many-to-one relationship, which simplifies the disease categorization and reduces complexity. ICGs are then mapped to a disease or injury modelling entity used by GBD modelers.

Mapping of ICD codes is not exhaustive as some causes in the GBD cause hierarchy do not use EHR data. We designate whether each modelling entity is processed in terms of incidence or prevalence, depending on the nature of the disease and the expected pattern of treatment. Table 13 shows the ICD codes used for non-fatal modelling by GBD cause, impairment, and injury.

Section 2.2.2 Individual-level clinical data sources with unique identifiers in GBD 2023

Figure S3. Processing steps for individual-level clinical data sources with unique identifiers



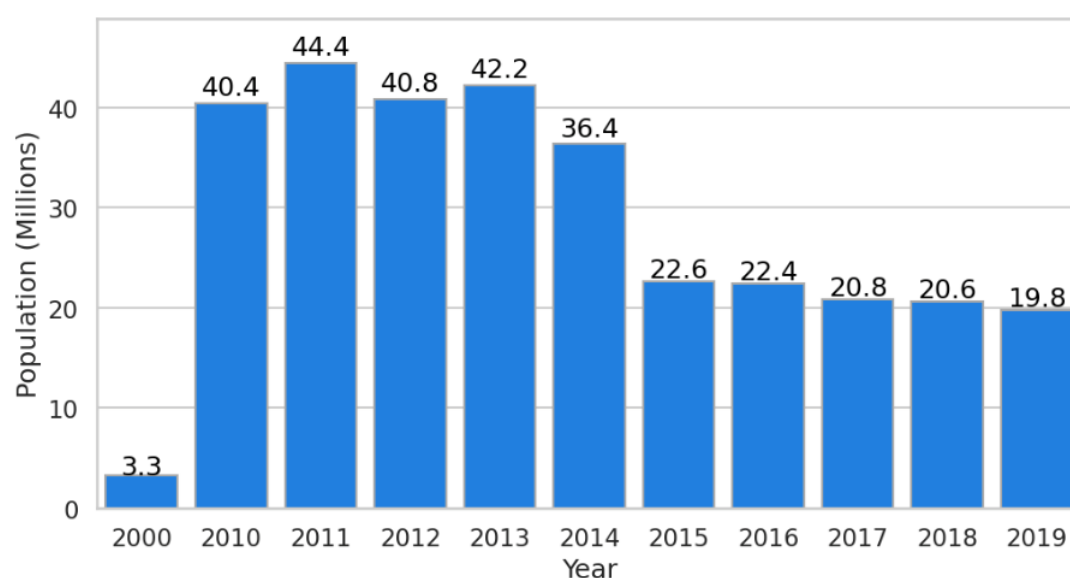
Poland National Health Fund Patient claims, United States Merative MarketScan Research Database claims, and United States Medicare claims were processed. These sources contain every health service encounter for the populations they covered, including hospital inpatient records, with same-day inpatient setting visits and overnight admissions, and outpatient, with general practitioner, emergency department (or urgent care), and specialty visits. We designate which health services each modelling entity requires data for, including, for example, inpatient only or inpatient *and* outpatient data. We mapped encounter-level data to ICD diagnoses to the diseases, impairments, and injuries modelled in the GBD. GBD causes are processed as “prevalence” or “incidence” based on the specification of the research team responsible for modelling the cause. Prevalent conditions are identified as any primary or non-primary diagnosis on any inpatient or outpatient claim within the year of interest. To reduce noise from spurious coding practices, a minimum of two outpatient claims for the same individual are required in a calendar year to count as a prevalent case. Incidence of a disease, impairment, or injury was calculated based on a duration window which varied by cause. Any individual who had multiple diagnoses for the same condition within the duration window is counted as a single incident case, and additional diagnoses outside of the duration window are treated as new incident cases.

After mapping to cause and identifying prevalent and incident cases by cause, for USA MarketScan and Medicare sources, we applied a noise reduction Poisson model to smooth trends over age and time and ensure compliance with data usage agreements by sharing modelled rates. More detail on the populations covered by each source is outlined below.

United States Merative MarketScan Research Database claims

For GBD 2023, we accessed data derived from the Truven database of USA private health insurance and Medicare private supplemental insurance for the years 2000 and 2010-2019.

Figure S4. Merative MarketScan Research Databases Enrollee Population by Year



For each of these individuals, claims representing every health service encounter were used and all episodes of care were linked to individuals by unique identifiers. For the GBD, we subset the population in the MarketScan database to individuals with a full year of insurance coverage or those who were born or died in the year of interest, to ensure the sample includes all healthcare utilization for a given individual in that year.

United States Medicare claims

Medicare data from the Centers for Medicare and Medicaid Services (CMS), provided by the Research Data Assistance Center (ResDAC), was added for GBD 2023. Medicare is a federal health insurance program for people aged 65 and older, but also younger individuals with specific eligible conditions. Only data for enrollees 65 and older were processed for years 2000, 2010, and 2014-2016. Specific

information on insurance eligibility by month allowed us to produce denominators of enrollees that can be a fraction of the year and calculated as enrollee-months. For GBD conditions that use inpatient-only data in their models, claims for Part A enrollees were used (figure S5) and notably does not include enrollees and their claims from Part C (also known as Medicare Advantage). For GBD causes that use data for both inpatient stays and outpatient encounters, a representative 5% sample of enrollees (figure S5) and their Part B claims from Medicare Carrier files were used, which include physician services and outpatient care. Claims for Part C enrollees were also not available for the 5% sample.

Figure S5. Medicare Part A Enrollee Population by Year

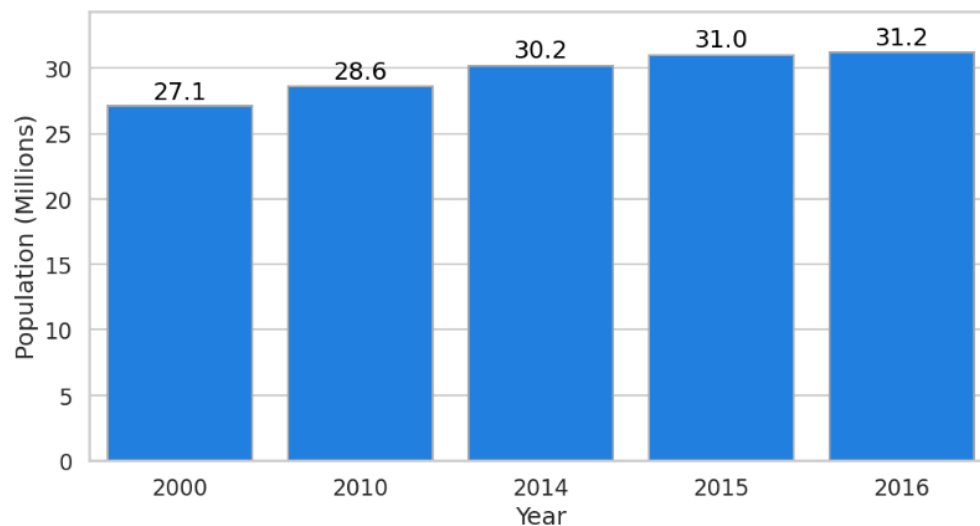
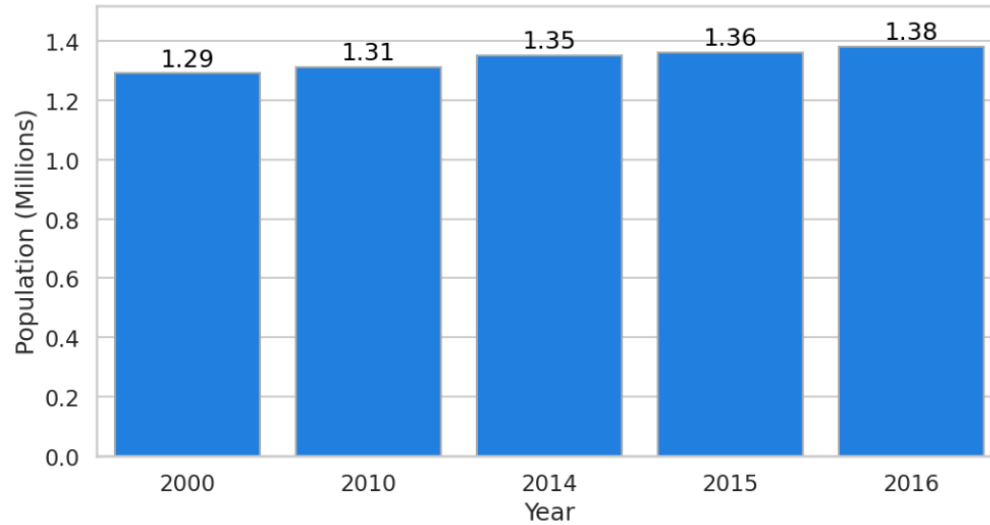


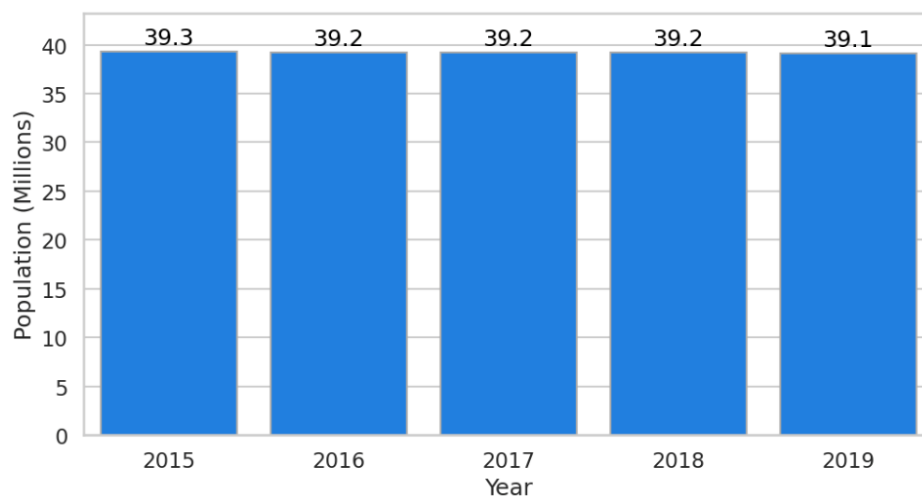
Figure S6. Medicare Part A+B Enrollee Population by Year from 5% Enhanced Sample



Poland National Health Fund Patient claims

Anonymized, individual-level claims data from Poland were accessed through an existing collaboration and institutional partnership with the Agency for Health Technology Assessment and Tariff System (AOTMiT). The data is derived from the National Health Fund (Narodowy Fundusz Zdrowia) database in Poland and is representative of every service encounter and episode of care in the public health care system (close to 92% population coverage) from 2015 to 2019, and for each region (voivodeship) in the country.

Figure S7. Poland Enrollee Population by Year



Other individual-level data sources, but without unique identifiers

Fee-for-service claims data from Singapore and Taiwan were also processed for GBD 2023. Tabulated inpatient-only claims data from Singapore for the years 1991-2019 were derived from the MediClaims Database and provided by the Ministry of Health of Singapore. The MediClaims data processed for the GBD is inclusive of all inpatient admissions in the country's public and private hospital facilities, and for all patients covered under MediShield Life, MediSafe, and MediFund, with admissions aggregated at the national level. Similarly, Taiwan National Health Insurance claims for the year 2016, derived from the National Health Insurance Research Database (NHIRD) and covering all residents in Taiwan under a universal single-payer health care system, was used. The NHIRD is representative of the whole population for Taiwan and covers both inpatient admissions and outpatient encounters.

Calculating uncertainty

Uncertainty in individual-level data sources was calculated using Wilson's approximation, utilizing sample size derived from enrollment data (i.e. MarketScan, Medicare) or GBD population estimates (i.e. Poland) depending on the source. Uncertainty in outpatient encounter data (Section 2.2.6) was also calculated using Wilson's approximation and GBD population. Uncertainty for aggregate inpatient data (Section 2.2.3) that are not complete for the population and use the inpatient utilization envelope came from multiplying 1000 sampled draws from the envelope and correction factor models and deriving uncertainty intervals from the upper and lower bounds. Inpatient sources that are complete for the population rely on creating 1000 draws from a Poisson distribution that are then multiplied by the correction factor draws to derive uncertainty.

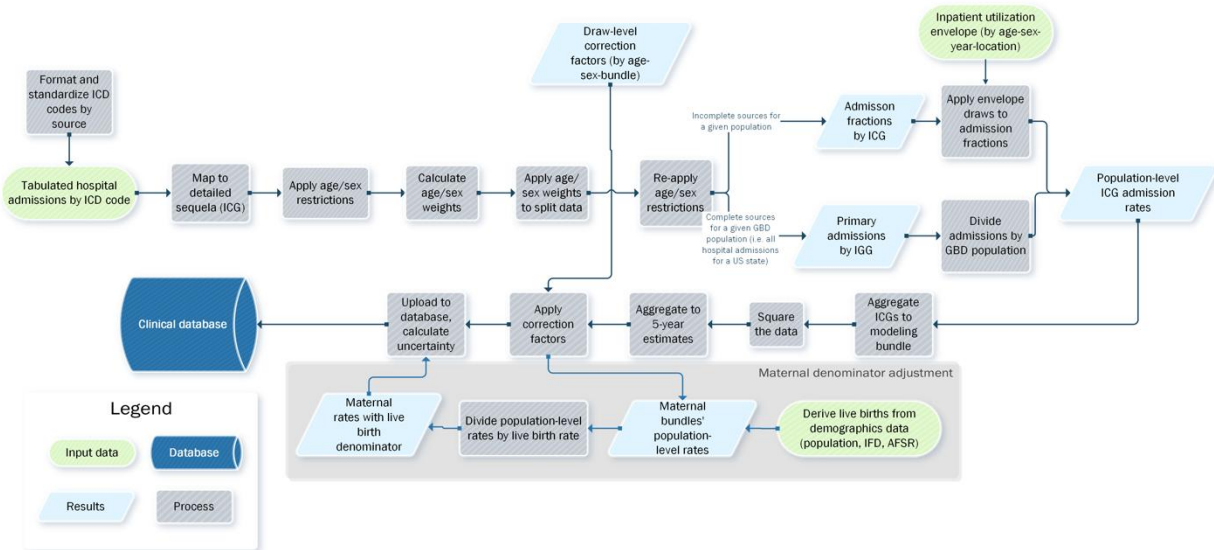
Wilson's approximation:

$$\sigma^2 = \frac{\frac{p(1-p)}{n} + \frac{1.96^2}{4n^2}}{\left(1 + \frac{1.96^2}{n}\right)^2}$$

where p is prevalence and n is the sample size

Section 2.2.3: Aggregate inpatient hospital admission sources used in GBD 2023

Figure S8. Processing steps for aggregate inpatient hospital admission sources



Hospital inpatient records data were extracted from 2817 location-years in 44 countries. For inpatient sources, a case of disease was defined as an overnight admission in a hospital or a stay of >24 hours, depending on how length of stay is represented in each source. ICD coding was standardized across sources and versions of ICD. Counts of admissions with a primary diagnosis of each cause were extracted from all sources and propagated through the aggregate hospital admissions data processing steps. For inpatient sources, a case of disease was defined as an overnight admission in a hospital or a stay of >24 hours, depending on how length of stay is represented in each source. We mapped each primary diagnosis code for a given admission to its modelling entity using the clinical ICD map. Non-primary diagnostic detail, where available, was included through the correction factor modelling process (Section 2.2.5). Admissions were then aggregated to create cause fractions, defined as the number of admissions for a given disease, impairment, or injury divided by total admissions for that age, sex, location, and year. In cases where a source is not complete for the country or subnational population, the inpatient utilization envelope was used and applied to the cause fractions (outlined in Section 2.2.4). For sources that are complete for the population, rates are created with numerators being primary admissions for a given condition and denominators being GBD population estimates.

Section 2.2.4: Estimation of the inpatient utilization envelope

The inpatient utilization envelope provided all-cause inpatient admission rate estimates from 1990 to 2019 for all age groups including <1 year detail, males and females, and 934 locations within the GBD 2023 framework. To inform nonfatal burden estimation, we applied those rates to upscale the number of cause-specific admission counts from several aggregate inpatient sources (table 4) that were not complete for the GBD location they represent and therefore did not capture all admissions in a given population. This approach enabled inclusion of these nonrepresentative sources covering 12 more countries.

Table 4. Clinical inpatient sources requiring the envelope adjustment

Location name	Subnational detail	Years of data
Argentina Public Hospital Discharge Statistics	No	2010, 2011
Armenia Hospital Data	No	2016
Botswana Health Management Data System (HMDS)	No	2007, 2008, 2009
Indonesia Integrated Hospital Data – Sistem Informasi Rumah Sakit (SIRS)	Yes	2013
Iran Hospital Data	No	2001-2010
Japan Diagnosis Procedure Combination Database	Yes	2010, 2011, 2012, 2013, 2014, 2015
Kenya National Inpatient Morbidity and Mortality Statistics	No	1999
Mexico Institutions of Health Sector Hospital Discharges	Yes	2004, 2018, 2019
Nepal Hospital Inpatient Discharges	No	2010, 2011, 2013, 2015
Portugal – European Hospital Morbidity Database	No	2015
Türkiye Diagnosis-Related Group Hospital Inpatient Database	No	2011, 2012
Viet Nam Hospital Data	No	2013

Case definition

We defined a hospital admission as hospitalization in a formal health care facility for, at least, an overnight stay. The inpatient utilization rate was the average number of admissions per person per year. We excluded inpatient admissions of healthy newborns and, wherever possible, excluded same-day discharges in accordance with the case definition.

Input data

We included sources either covering the complete population in a given location or using a nationally or subnationally representative sample of the population. We extracted the total inpatient admission counts per year. The inputs comprised administrative data, national government reports, scientific literature, and surveys from 1990 to 2019.

Table 5. Inclusion and exclusion criteria for data sources used in the inpatient utilization envelope model

Inclusion criteria	Exclusion criteria
Case definitions: overnight stay, discharge	
Data types: Administrative data, Report, Scientific Literature, Survey	
Population coverage: complete in a location at the national or subnational level	Population coverage: incomplete for a given location at the national or subnational level
Time period covered: 1990 – 2019	

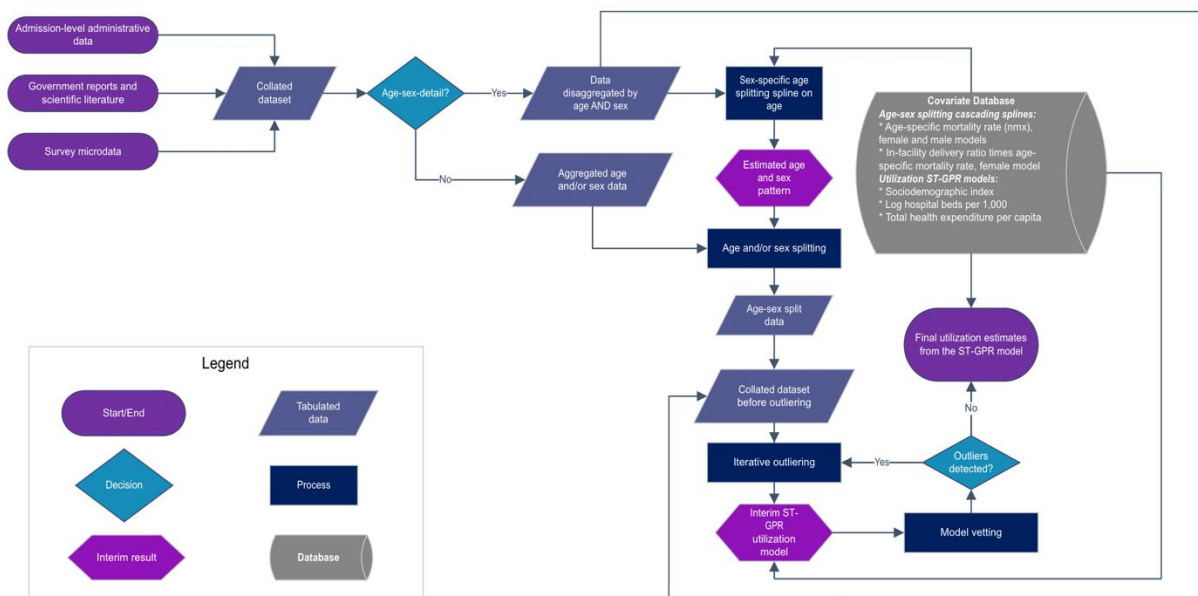
Recall: Annual admission counts / 12-month recall period	Recall: Less or over 12-month recall
Unit of measurement: total number of admissions for any cause excluding healthy newborns and day cases	Unit of measurement: total number of admitted persons.

The data search and extraction were completed in several steps. Firstly, we processed and included the available individual admission-level administrative inpatient data covering the entire populations in a country or a subnational location aggregating it by age group and sex. Secondly, we searched the GHDx for government reports concluding the search on December 12th, 2023. We restricted the search to the following GBD regions: Andean Latin America, the Caribbean, Central Asia, Central Latin America, Central Sub-Saharan Africa, Eastern Sub-Saharan Africa, Southern Sub-Saharan Africa, Western Sub-Saharan Africa, East Asia, Southeast Asia, Southern Latin America, Eastern Europe, High-income Asia Pacific, North Africa and the Middle East, Oceania, and South Asia. We applied no language restrictions and required all returned records to have tabulated reports. We screened the titles translated to English using a set of keywords, including "Health", "Statistical", "Yearbook", "Annual", "Report", "Statistics", "Bulletin", "Hospital", "Discharges", "Inpatient". Thirdly, we complemented the retrieved reports with external web searches on public government websites addressing missingness after the GHDx review, which resulted in the addition of government reports for 33 countries previously not catalogued in the GHDx. We also included three known multinational administrative data sources, including Eurostat, European Health for All database (HFA-DB), and the Organization for Economic Co-operation and Development (OECD) Health Statistics further expanding coverage. Finally, we used a subset of surveys used to produce utilization estimates in GBD 2021 reporting the number of admissions per person in the past 12 months. We applied no language restrictions to our search and required all returned records to contain either microdata or tabulated reports.

Data analysis

First, we split data points covering wider age bins and/or both sexes into GBD standard age/sex categories using inpatient utilization age patterns produced with a Bayesian spline cascade. We visually inspected the resulting age/sex specific data set to identify outliers. Finally, we modeled average annual per capita inpatient utilization rates for 934 locations, 25 age groups, and for males and females, from 1990 to 2019, where the 25 age categories were early neonatal (0-6 days), late neonatal (7-27 days), 1 to 5 months, 6-11 months, 12 to 23 months, 2 to 4 years, 5 to 9 years, followed by 5-year increments up to 90 to 94 years, and 95 years and older. We modelled per capita inpatient utilization rate with Spatiotemporal Gaussian Process Regression (ST-GPR) adjusting for total health expenditure per capita, log hospital beds per 1,000 population, and Sociodemographic Index (SDI).

Figure S9. Overview process of estimation of hospital utilization envelope



Changes from GBD 2021 to GBD 2023

This iteration of inpatient utilization models has seen several significant updates increasing the robustness and performance of the model. First, we newly incorporated over 2,000 location-years of government reports, adding on to and replacing lower-quality surveys. Second, we added over 300 location-years of aggregated individual-admission data with extensive age and sex detail where we could ensure adherence to the case definition of inpatient utilization, which could not always be readily confirmed in other data sources. Surveys that previously required crosswalking due to deviating case definition were excluded from the analysis in favor of widely available government reports. Third, we developed country-year-specific age patterns for each sex that we used to split aggregate data points, whereas the previous model relied on a global age pattern that could not capture variations across locations and years due to differences in population structure and health care systems. Fourth, we systematically excluded healthy newborns from the input data with <1 year age detail, which resulted in more reliable utilization estimates for infants. Fifth, we replaced the Healthcare Access and Quality Index (HAQi) indicator with Sociodemographic Index (SDI) to adjust the utilization model to remove potential co-linearity between HAQi and healthcare utilization and control for differences in SDI between countries. Finally, we eliminated spatial smoothing in the ST-GPR model due to the lack of evidence of any correlation in healthcare system specifications within the GBD location hierarchy.

Applying the inpatient utilization envelope to incomplete inpatient sources

We applied the inpatient utilization envelope to inpatient sources incomplete for the population they represent (table 4). To do so, we calculated age- and sex-specific cause fractions by dividing the number of cause-specific admissions by total inpatient admissions in the source. Then, we calculated the all-cause number of inpatient admissions by multiplying the total per capita inpatient utilization rate from the envelope by population of the respective demographic group. Finally, we multiplied the cause

fractions by the total admission count, which amounted to the total number of cause-specific admissions in a given population.

We calculated total cause-specific admissions per age, sex, location, and year from every clinical source as follows:

$$CFr = \frac{A_{cause}}{n}$$

$$A_{tot} = AR * N$$

$$TA_{cause} = CFr * A_{tot}$$

where

CFr – cause fraction

A – admissions

n – sample size in a source

AR – estimated admission rate from the envelope

N – GBD population

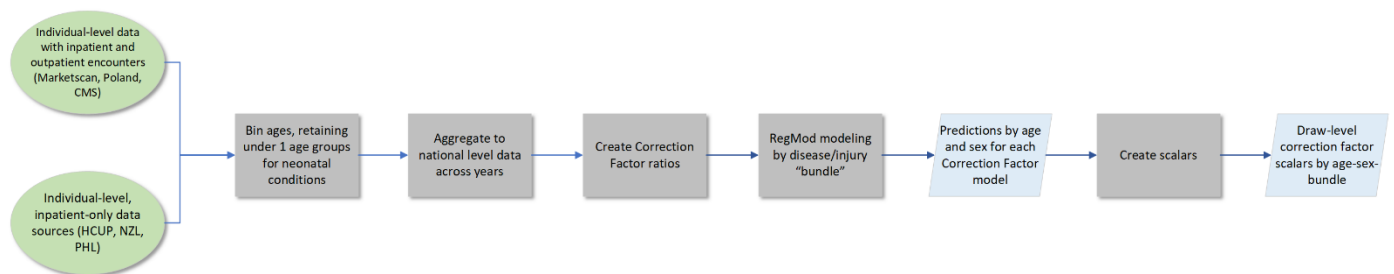
TA – total admissions

per age, sex, location, and year.

Section 2.2.5: Correction Factors

We performed three corrections on the aggregate-level inpatient hospital admissions data sources to standardize all inpatient hospital admissions data to the same definition of care and to account for cases of disease that were not captured in these sources. These data adjustments include: deduplication of multiple admissions per individual for the same cause or condition, within one year (for prevalent conditions) and within a cause-specific duration (for incident causes), accounting for cases of disease that were non-primary diagnoses and not captured in the aggregate-level data, and accounting for outpatient admissions also not present in the aggregate-level inpatient admissions data. These corrections were applied to the aggregate-level inpatient admissions data as scalars, resulting in four incidence and prevalence estimates: (1) un-corrected inpatient admissions, primary diagnosis only; (2) inpatient admissions by individual encounter, deduplicated for readmissions, primary diagnosis only; (3) inpatient admissions by individual encounter, deduplicated for readmissions, accounting for all diagnoses; and (4) inpatient admissions and outpatient encounters by individual, deduplicated for readmissions, accounting for all diagnoses. Estimate 4, where all three corrections are applied, was used for most causes, with exceptions based on the nature of the disease; in instances where outpatient care or non-primary diagnosis were not expected, estimates 3 or 2 were used instead.

Figure S10. Overview of Correction Factor process



Correction Factor ratios for the estimates 2 and 3 above were calculated using all clinical sources that had individual-level data with unique patient identifiers and both primary and non-primary diagnosis positions. These sources include those described in “Section 2.2.2. Individual-level data sources with unique identifiers in GBD 2023”: United States Merative MarketScan claims (2010-2019, in plots referenced as ‘MarketScan’), Taiwan National Health Insurance claims (2016, ‘TWN’), Poland National Health Fund Patient claims (2015-2019, ‘POL’), United States Medicare claims (2000, 2010, 2014-2016, ‘Medicare’), as well as inpatient admissions data from the Philippines Health Insurance PhilHealth (2013-2016, 2019, ‘PHL’), New Zealand National Minimum Dataset (2000-2019, ‘NZL’), and the United States HCUP State Inpatient Databases (2003-2009, 2016, 2018, 2019, ‘HCUP’). Of these sources, only MarketScan, Medicare, Poland, and the Taiwan claims data included a link between inpatient and outpatient care and could be used in estimate 4 described above.

To create a correction for readmissions, to go from total admissions to single cases of disease, we calculate the ratio of the number of unique primary diagnoses among inpatient admissions for an individual to the total number of primary diagnoses among inpatient admissions for each cause by age, sex, and source. For a correction for missing non-primary diagnoses, we calculate the ratio of the number of deduplicated primary diagnoses among inpatient admissions to the number of deduplicated primary and non-primary diagnoses among inpatient admissions by age, sex, and source. To model a correction for both missing outpatient admissions and non-primary diagnoses, we calculate the ratio of the number of deduplicated primary diagnoses among inpatient admissions to the number of deduplicated primary and non-primary diagnoses among inpatient admissions and outpatient encounters by age, sex, and source. The data aggregates that make up these ratios can be found in Table 6a below and the ratios can be found in Table 6b. These three sets of ratios are always proportions between 0 and 1. The data that goes into each of the ratios is aggregated over year and subnational location and is specific by age group, sex, and source.

Table 6a. Data aggregates that go into correction factor ratios with various data processing steps applied

Name	Description	D	N	O
dno1	Deduplicated primary diagnoses, among inpatient admissions	X		

dno12	Deduplicated primary and non-primary diagnoses, among inpatient admissions	X	X	
dno123	Deduplicated primary and non-primary diagnoses, among inpatient admissions and outpatient encounters	X	X	X

Here, *D* means deduplication has been applied, *N* means non-primary diagnoses are included, and *O* means outpatient encounters are included.

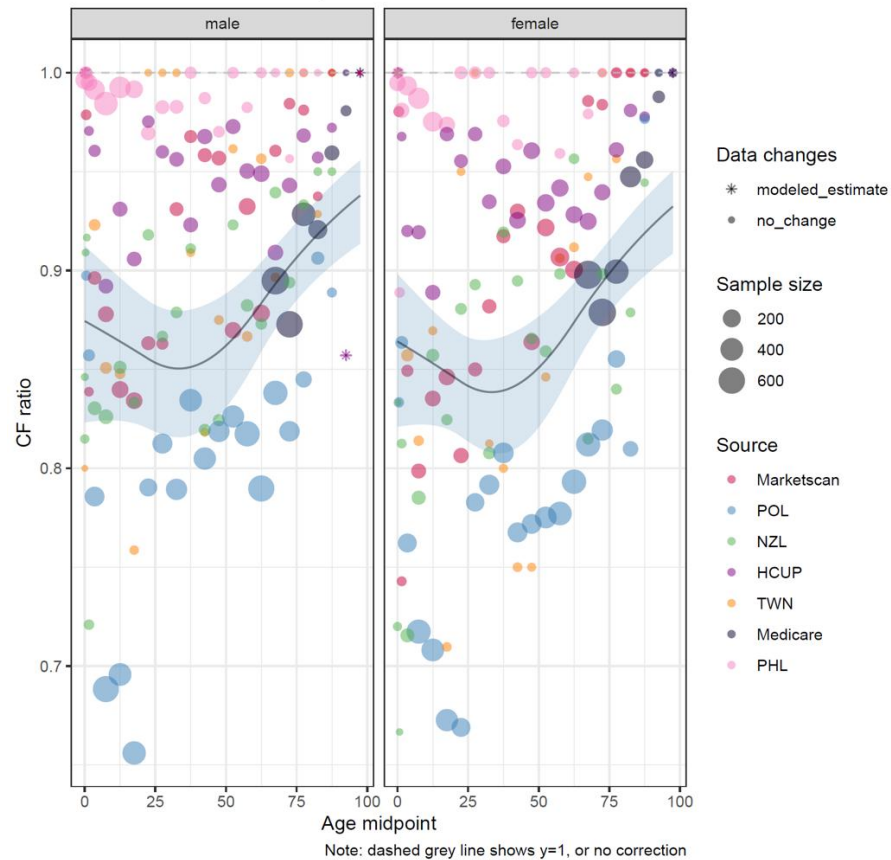
Table 6b. Ratios for correction factor models

Correction Factor	Corrects for	Ratio for modeling
CF1	Readmissions	dno1/dno0
CF2	Non-primary Dx	dno1/dno12
CF3	Non-primary Dx and outpatient	dno1/dno123

For each GBD cause, each of the Correction Factor ratios were modeled in RegMod ([see here](#) for more information on RegMod) as a binomial model with a fixed effect on sex and a B-spline on age, as described by the equation below. For the spline on age, there were three quadratic knots spaced according to data density with linear tails. All models were conducted in logit-space in order to bound the results by 0 and 1. To get predictions from the models, 1000 draws were taken from the posterior distribution, with an uncertainty interval calculated from 0.025 and 0.975 quantiles in the draw-level results and point estimates calculated directly from model coefficients. Figures S11-13 show examples of the model results for acute encephalitis.

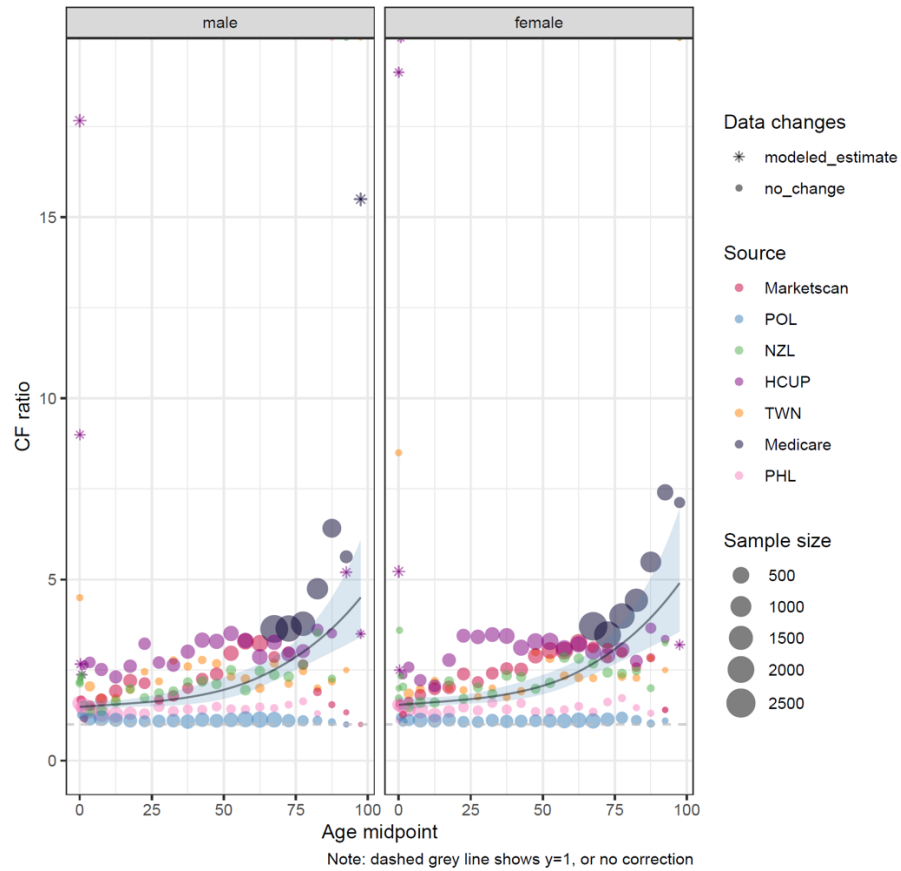
$$E[y|sex_i, age_i] = \text{logit}^{-1}(\beta_0 + \beta_1 sex_i + \text{spline}(age_i))$$

Figure S11. Readmissions correction model in linear space for acute encephalitis



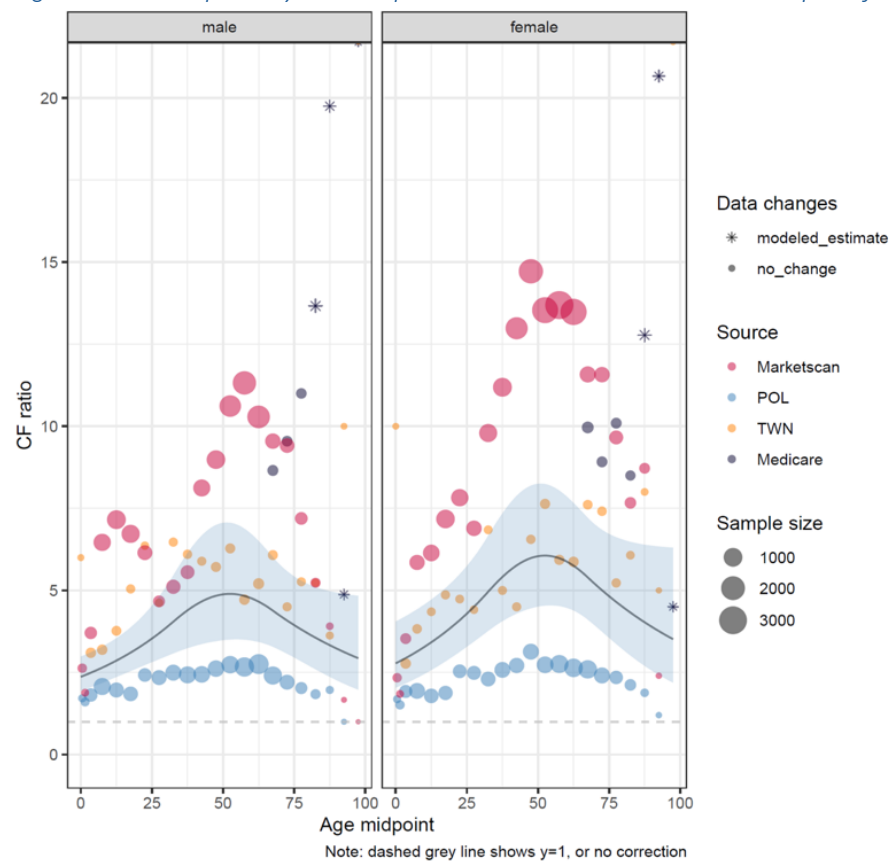
“Data changes” shows where the data has been modelled to be visualized according to data privacy restrictions. Note that these changes are not present when the model is fit, but only for visualization purposes.

Figure S12. Non-primary correction model in linear space for acute encephalitis



“Data changes” shows where the data has been modelled to be visualized according to data privacy restrictions. Note that these changes are not present when the model is fit, but only for visualization purposes. The data shown represents the inverse of the non-primary correction model to more clearly visualize the scalar applied to correct the inpatient data.

Figure S13. Non-primary and outpatient correction model in linear space for acute encephalitis



“Data changes” shows where the data has been modelled to be visualized according to data privacy restrictions. Note that these changes are not present when the model is fit, but only for visualization purposes. The data shown represents the inverse of the non-primary, outpatient correction model to more clearly visualize the scalar applied to correct the inpatient data.

For the readmissions correction, the model predictions were applied to inpatient data as a scalar, by age and sex. The readmissions correction is always a value between 0 and 1, so the effect was a decrease in inpatient admissions from total inpatient admissions to just incident encounters. For the non-primary diagnosis correction, the readmissions correction is first applied, followed by the inverse of the predictions from the non-primary correction model, increasing the estimate to include non-primary inpatient cases. For the outpatient and non-primary correction, the readmissions correction is also first applied and then followed by the inverse of the predictions from the outpatient and non-primary model, increasing the estimate to include both non-primary inpatient cases and outpatient cases. Table 7 outlines how the corrections are applied as the scalars described here to create different estimates. The corrections are applied at the draw-level over 1000 draws, the estimates are produced by taking the median of the draws, and the uncertainty intervals are produced by taking the 97.5 and 2.5 percentiles of the draws.

Table 7. How correction factors are applied to create various incidence and prevalence estimates

Estimate	Correction(s) applied	Scalar to apply	Scalar equation
Inpatient admissions by individual encounter, deduplicated for readmissions, primary diagnosis only	Readmissions	CF1	dno1/dno0
Inpatient admissions by individual encounter, deduplicated for readmissions, accounting for all diagnoses	Readmissions and non-primary Dx	CF1 * 1/CF2	dno1/dno0 * dno12/dno1
Inpatient admissions and outpatient encounters by individual, deduplicated for readmissions, accounting for all diagnoses	Readmissions, non-primary Dx, and outpatient	CF1 * 1/CF3	dno1/dno0 * dno123/dno1

Determination of maternal causes used separate cause fractions and a different scalar calculated from a maternal hospital admissions rate instead of the inpatient hospital envelope using the equation

$$\left(\frac{\text{events}}{\# \text{ of total hospital visits}} \right) * \left(\frac{\text{hospital visits}}{\text{live births}} \right) * \left(\frac{\text{births}}{\text{population}} \right)$$

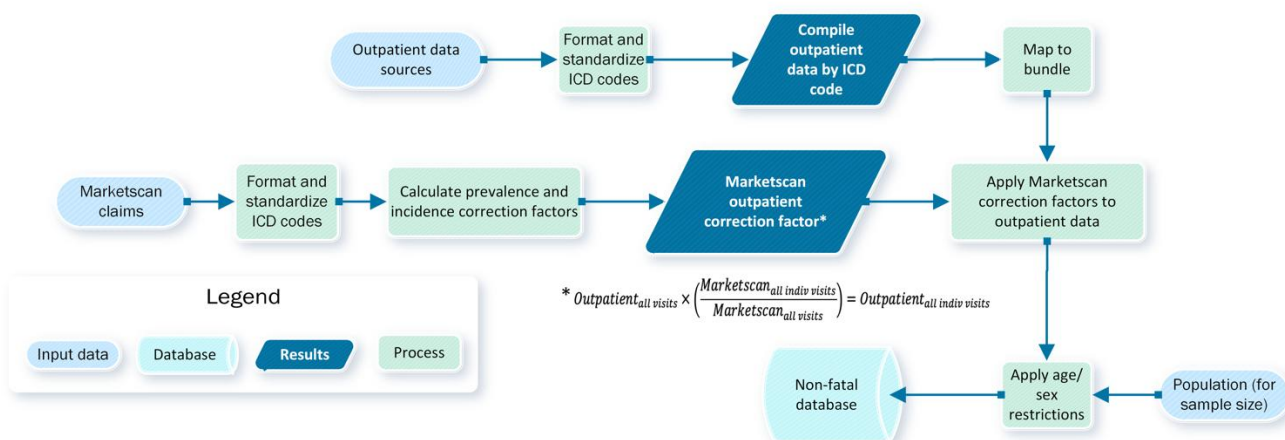
Determination of injuries used a separate correction factor from those described above which adjusted data that was only E-coded by data that contained E-codes and N-codes (nature of injury codes) with the following equation

$$\frac{1}{\frac{\text{E-code primary dx}}{\text{E-code any dx+N-code any dx}}}$$

A final adjustment was applied to each of the above estimates. The HAQ Index was used to account for differences in access and quality of health care across time and space. The HAQ Index adjustment was applied by dividing the above estimates by a scalar ranging from 0 to 100, where 0 represents the first percentile of observed access and quality and 100 the 99th percentile.

Section 2.2.6: Outpatient encounter data

Figure S14. Processing steps for outpatient encounter data



Outpatient encounter data that could not be linked to comprehensive inpatient records were available from the USA and Sweden for 109 location-years. No changes were made in the processing of outpatient data from GBD 2019, except for updates to the ICD mapping to GBD diseases and injuries. As with the aggregate inpatient sources, a scalar was calculated by using MarketScan claims data to adjust for multiple visits per individual within one year (for prevalent conditions) and within a cause-specific duration (for incident causes).

We have also processed and included electronic health records from general practices (GP data), provided by the Netherlands Institute for Health Services Research (Nivel), covering about 10% of the Dutch population, with approx. 1.75 million individuals registered with a GP each year. The data set includes incidence and prevalence numbers per 1,000 persons in the Netherlands, based on ICPC-1 coding system. From the ICPC-1 coding system, we were able to directly map over 100 codes to GBD disease and injury categories. The Dutch GP data were processed into ‘illness episodes’ by Nivel, using a three-year window: reporting calendar year + two previous calendar years. What this means is that the data received contains any outstanding episodes of care (for chronic conditions) *and* any new conditions recorded over a three-year period. For chronic conditions lasting one year or more, we have made a correction to the prevalence estimates for the GBD, subtracting out half of the incident cases in a year.

Section 2.2.7: New data added for GBD 2023

In addition to Medicare claims, several new sources were provided for non-fatal modelling as well as additional years of existing GBD clinical sources. Table 8 below summarizes the data added in this GBD round with relevant characteristics of the data source. Sources that include unique identifiers are either processed (A) at IHME in a secure environment, or (B) by collaborators before being shared for use in the GBD. Where available, 2020 and 2021 years of clinical data were processed and reviewed but not used directly in non-fatal models. Dramatic decreases in utilization during the COVID-19 pandemic were observed and further analyses are needed to disentangle the effects of care seeking behaviour on incidence and prevalence rates as well as true changes in rates for these conditions across the time series.

Table 8. Source-years of clinical data added for GBD 2023

Source	Years Added	Includes unique identifiers	Includes outpatient encounters	Complete for population
United States Merative MarketScan Research Database	2018, 2019	Yes	Yes	No
Poland National Health Fund Patient Claims	2019	Yes	Yes	Yes
Singapore MediClaims Database	2018, 2019	Yes	No	Yes
Mongolia H-Info Health Facility Data*	2019	No	Yes	Yes
Russia Statistical Yearbook	2018, 2019	No	Yes	Yes
South Korea Medical Services Claims Data*	2018, 2019	Yes	Yes	Yes
Austria Hospital Inpatient Discharges	2015-2018	No	No	Yes
Brazil Hospital Information System (SIH)	2017-2019	No	No	Yes
Chile Hospital Discharge Information System	2019	No	No	Yes
Georgia Hospital Data	2016, 2018, 2019	No	No	Yes
Germany Inpatient Admissions	2000-2019	No	No	Yes
Italy Hospital Inpatient Discharges	2019	No	No	Yes
Mexico Institutions of Health Sector Hospital Discharges*	2004-2019	No	No	No
Netherlands Nivel Primary Care Database	2011-2019	No	Yes – GP visits only	No
New Zealand National Minimum Dataset	2018, 2019	No	No	Yes
Philippines Health Insurance (PhilHealth) Data	2019	Yes	No	Yes
United States State Inpatient Databases (HCUP – US SID)	2016-2019^	Yes^	No	Yes

*New source for GBD 2023

^Select states only

Section 2.3 Data Adjustments

Section 2.3.1: Crosswalking

Crosswalking refers to the process of adjusting data for known biases. An observation is considered biased if it differs in a systematic way from the standard GBD definition of the modeled parameter. Examples include self-reported rather than doctor-diagnosed measures of disease incidence, case ascertainment using disparate diagnostic tests, sampling from non-representative populations (e.g. blood donors), and observations derived from different types of surveillance systems, among others. If

the difference between an alternative measurement method and the GBD definition is systematic, we can model it as a function of covariates and use this model to predict the degree of adjustment needed for a given alternative or non-standard observation. The result of crosswalking is that GBD models can incorporate data from a wider range of sources.

Specifically, crosswalking involves:

1. Finding pairs of alternative and reference (e.g. self-reported and measured) observations that match on relevant criteria (e.g. age, sex, location and year);
2. Taking the difference between these observations in log or logit space, to ensure that the crosswalk adjustment remains bounded correctly;
3. Running a meta-regression model that estimates this difference potentially as a function of covariates;
4. Predicting how much each alternative data point in the original dataset should be adjusted; and
5. Applying the adjustment.

Section 2.3.2: Bias adjustment for alternative case definitions and study methods

In GBD 2023 we continued the practice started in GBD 2019 of crosswalking non-fatal and risk exposure data to account for alternative case definitions or study methods. The adjustments were applied prior to entering data into our main analytical tools of DisMod-MR 2.1 and ST-GPR, ensuring that all data inputs were expressed on a consistent scale. We also used this approach to convert data presented for both sexes to a male and female equivalent. The starting point was to explicitly state the reference case definition and study method and identify alternative definitions and study characteristics that fall within our inclusion criteria.

We compiled data from both within-study comparisons (ie, data that used alternative and reference definitions in the same population) and between-study comparisons (ie, data that used an alternative definition in one population and a reference definition in another population that overlap in location, time, age, and sex) of different case definitions. For between-study comparisons, we allowed a maximum calendar year difference between studies of five years. Where validation studies (ie, those carried out at the introduction of a new set of diagnostic criteria comparing to previous criteria) were available, we extracted data on the comparison of alternative to reference. For quantities of interest with multiple alternative definitions/methods we also looked for pairs comparing two alternatives.

If both between and within study pairs were available, we examined whether there was a systematic difference between these. If there was a significant difference, we made judgement call as to whether within-study or between study data comparisons were most appropriate. In general, this was the within-study data. However, there were important measurement or conceptual reasons for choosing between-study data. For example, for crosswalks between self-reported height and weight compared to measured height and weight, between-study comparisons may be preferable if respondents knew they would be measured and, therefore, were less likely to misreport their height and weight.

To quantify the degree of bias for an alternative data source, we calculated the difference between matched pairs of alternative and reference observations and used this quantity as the dependent variable in a mixed effect meta-regression model. The model could include any number of covariates to capture how bias might vary as a function of other variables, like age or sex. Predictions from the model were then used to convert alternative observations to their equivalent reference values. For GBD 2021 and later rounds, we developed an open-source Python package to facilitate the process of modeling and applying bias adjustments (ihmeuw-msca, 2023).

To choose covariates for the model, we examined whether there were systematic differences in the adjustments by key demographics (age, sex, geographic location, year) and other potential factors that may lead to variation in the degree of bias adjustment. We did this when there was a strong rationale, eg, biological plausibility, for variation by such characteristics. After fitting the model, for predicted adjustment factors that were not statistically significant, we still applied the adjustments if there was a conceptual reason to believe that the alternative definition is biased. This expands the variance of data points using a non-standard case definition or study method, effectively reducing their influence in subsequent modeling steps.

Section 2.3.3: Example bias adjustment calculation

As an example, we provide mathematical notation for a bias adjustment to a data source that measures prevalence using a non-standard case definition. We have pairs of alternative and reference observations (denoted i) that match on age, sex, location, and time period combination (denoted j). The degree of bias varies as a function of age and sex. Because the parameter of interest is prevalence, which is bounded by 0 and 1, we calculate the logit-scale difference between alternative and reference observations in a given matched pair:

$$y_{i,j} = \text{logit}(p_{i,j}^{alt}) - \text{logit}(p_{i,j}^{ref})$$

In preparing the data for this calculation, if the values of either the reference or alternative were zero, we aggregated values across age groups until both values had non-zero observations. We used the delta method to compute the standard error of the reference and alternative measures in logit space. The standard error of the logit-scale difference was computed as the square root of the sum of the variances of each data point in a pair.

If the parameter had instead been bounded by only 0, like incidence, we would have calculated the log-scale difference. From simulations we found that the two methods provide almost identical results for quantities that after adjustment do not exceed a value of 0.5 (eg, prevalence or proportion). The logit-scale difference method much better dealt with higher values and avoided prevalence or proportions to exceed one.

As a next step in this hypothetical example, we modeled the differences as the dependent variable in a mixed-effects meta-regression model with age and sex as covariates:

$$y_{i,j} = \beta_0 + \beta_1 \text{age}_i + \beta_2 \text{sex}_i + u_j + \epsilon_i$$

$$u_j \sim N(0, \gamma)$$

$$\epsilon_i \sim N(0, \sigma_i^2)$$

We then used the linear predictor of this model to predict the degree of bias adjustment needed for the various age and sex combinations among the alternative observations:

$$\hat{\delta}_{a,s} = f(\text{age}, \text{sex}) = \hat{\beta}_0 + \hat{\beta}_1 \text{age} + \hat{\beta}_2 \text{sex}$$

To adjust a particular alternative observation $p_{a,s}^{alt}$ we subtracted the adjustment factor in logit space, and the inverse logit transformation was applied to the result to convert back to natural units:

$$p_{a,s}^{adjusted} = \text{logit}^{-1}(\text{logit}(p_{a,s}^{alt}) - \hat{\delta}_{a,s})$$

The uncertainty for the adjusted logit-scale prevalence includes:

- uncertainty of the original observation in logit space,
- uncertainty from the posterior distribution of the predicted adjustment, and
- random intercepts in the meta-regression model (denoted γ above).

The variances from the three components were summed and then transformed into natural unit space using the delta method.

Section 2.3.4 Network Analysis

When there were multiple alternative case definitions or study methods, we used network analysis to leverage the additional information provided by indirect comparisons. For example, if A is the reference and B and C are two alternatives, the comparison of C versus A would be considered a direct comparison to the reference. This case was the subject of the previous section. In contrast, the combination of A versus B and B versus C provides an indirect comparison of the alternative C against the reference A. Or in other words, the inclusion of B-versus-C comparisons in the dataset provides additional information with which to estimate the difference between C and A.

Implementing a network analysis requires careful construction of the design matrix, or the dataset we pass to the mixed effects meta-regression model. Continuing the example with reference A and alternatives B and C, the design matrix for a network analysis with no covariates is created as follows:

- Create k dummy variables where k are all definitions/methods other than A (eg, $k = B, C$)
- Code dummy k as
 - o 1 if the first term of the logit-scale difference is k ;
 - o -1 if k is second term of the logit-scale difference;
 - o 0 otherwise

For example:

Study	Comparison	DummyB	DummyC
1	logit(B)-logit(A)	1	0
2	logit(B)-logit(A)	1	0
3	logit(C)-logit(A)	0	1

4	logit(C)-logit(A)	0	1
5	logit(C)-logit(B)	-1	1
6	logit(C)-logit(B)	-1	1

The coding structure outlined above assumes that all case definitions are mutually exclusive. In some cases, however, individual case definitions are composed of different sub-components or dimensions. For example, case definitions may vary by the type of symptoms that a respondent experiences as well as the recall period over which those symptoms are experienced. In the presence of sparse data, it may be difficult to find both direct and indirect comparisons of all individual case definitions. In these cases, an alternative approach is to assume different dimensions of case definitions have a multiplicative effect. In other words, the effect of recall period has the same relative effect across different categories of symptoms reported by respondents. To implement this coding scheme:

- Create k dummy variable columns for each case definition dimension.
- For each dummy variable k :
 - Add 1 if k is a component of the first term in the logit-scale difference.
 - Subtract 1 if k is a component of the second term in the logit-scale difference.

Network analysis is a feature of the open source Python package for conducting bias adjustments (ihmeuw-msca , 2023) mentioned earlier. The package abstracts away the need to create the design matrix manually as in this example and can incorporate an arbitrary number of alternative definitions and covariates.

Section 2.3.5 Age sex splitting

Before modelling, we ran a DisMod-MR 2.1 model with data disaggregated by age to estimate countries' age-pattern and then applied the estimated age-pattern to split aggregated all-age data into the 5-year age groups preferred for ST-GPR modelling. This procedure was done by calculating a constant, k , which was the ratio of the aggregated all-age data point, $\mu_{all\ age}$, to the all-age estimated utilisation rate from the DisMod-MR 2.1 model, $\widehat{\mu_d}$

$$k = \frac{\mu_{all\ age}}{\widehat{\mu_d}}$$

The constant, k , was then multiplied by age-specific utilisation rates from the DisMod-MR 2.1 model. Observation-specific uncertainty and uncertainty from the estimated age-pattern were both propagated into the uncertainty for a given post-splitting data point. The split data were then incorporated into the final DisMod-MR 2.1 model.

Section 2.4: Spatiotemporal Gaussian process regression (ST-GPR) modelling

The input data were modelled by using ST-GPR to allow for smoothing over age, time, and location in locations that were missing complete datasets. The flowchart showing the analytic steps can be found elsewhere.⁶ The approach is a stochastic modelling technique that is designed to detect signals amidst noisy data. It also serves as a powerful tool for interpolating non-linear trends.^{8,9} Unlike classical linear models that assume that the trend underlying data follows a definitive functional form, GPR assumes that the specific trend of interest follows a Gaussian process, which is defined by a mean function $m(\cdot)$

and a covariance function $Cov(\cdot)$. For example, let $p_{c,a,s,t}$ be the prevalence, in normal, log, or logit space, observed in country c , for age group a , and sex s at time t :

$$(p_{c,a,s,t}) = g_{c,a,s}(t) + \epsilon_{c,a,s,t}$$

where

$$\begin{aligned} \epsilon_{c,a,s,t} &\sim Normal(0, \sigma_p^2) \\ g_{c,a,s}(t) &\sim GP\left(m_{c,a,s}(t), Cov\left(g_{c,a,s}(t)\right)\right). \end{aligned}$$

The derivation of the mean and covariance functions, $m_{c,a,s}(t)$ and $Cov\left(g_{c,a,s}(t)\right)$, along with a more detailed description of the error variance (σ_p^2), is described below.

Section 2.4.1 Estimating mean functions

We estimated mean functions by using a two-step approach. To be more specific, $m_{c,a,s}(t)$ can be expressed, depending on the prevalence transformation, as:

$$\log(p_{c,a,s}(t)) = X_{c,a,s}\beta + h(r_{c,a,s,t})$$

$$\text{logit}(p_{c,a,s}(t)) = X_{c,a,s}\beta + h(r_{c,a,s,t})$$

$$p_{c,a,s}(t) = X_{c,a,s}\beta + h(r_{c,a,s,t})$$

where $X\beta$ is the summation of the components of a hierarchical mixed-effects linear regression, including the intercept and the product of covariates with their corresponding fixed-effect coefficients. Some models were run as hierarchical mixed-effects linear regressions with random effects on the levels of the location hierarchy. For most mixed-effects models, random effects were only used in the fit, not in the prediction. The second part of the equation, $h(r_{c,a,s,t})$, is a smoothing function for the residuals, $r_{c,a,s,t}$, derived from the linear model. Cause-specific methods details can be found in appendix section 6.

Although the linear component captures general trends over time, much of the data variability may still not be adequately accounted for. To address this, we fit a locally weighted polynomial regression (locally estimated scatterplot smoothing, or LOESS) function $h(r_{c,a,s,t})$ to systematically estimate this residual variability by borrowing strength across time, age, and space patterns (the spatiotemporal component of ST-GPR).^{10,11} The time adjustment parameter, defined by λ , aims to borrow strength from neighboring time points (ie, the prevalence in this year is highly correlated with prevalence in the previous year but less so further back in time). The age-adjustment parameter, defined by ω , borrows strength from data in neighboring age groups. The space-adjustment parameter, defined by ξ , aims to borrow strength across the hierarchy of geographical locations. The spatial and temporal weights are combined into a single space-time weight to allow the amount of spatial weight given to a particular

point $r_{c,a,s,t}$ to fluctuate given the data availability at each time t and location-level l in the location hierarchy.

Let $w_{c,a,s,t}$ be the final weight assigned to observation $r_{c,a,s,t}$ with reference to a focal observation r_{c_0,a_0,s_0,t_0} . We first generated a temporal weight $t.w_{c,a,s,t}$ for smoothing over time, which was based on the scaled distance along the time dimension of the two observations (Ng M) :

$$t.w_{c,a,s,t} = \frac{1}{e^{\lambda|t-t_0|}}$$

Next, we generated a spatial weight to smooth over geography. Specifically, we defined a geospatial relationship by categorizing data based on the GBD location hierarchy (table 1). ζ acts as a scalar on a given datapoint given its proximity to the target location:

$$t.w_{c,a,s,t} = \zeta^{|c-c_0|}$$

For example, estimating a country, would use the following weighting scheme:

- Country data: $\zeta^0 = 1$
- Regional data not from the country being estimated: ζ^1
- Data from other regions in the same super region: ζ^2
- Global data from other super regions: ζ^3

Under the spatial weighting specification, typical values of ζ range from [0.001, 0.2], where ζ can be interpreted as the amount to down weight regional datapoints compared to country datapoints for a given estimating country. For example, for a given datapoint $r_{c,a,s,t}$ and $\zeta = 0.01$, a datapoint not within country c but within the same region r as $r_{c,a,s,t}$ would be assigned $\frac{1}{100}$ the weight of a datapoint within the country.

The spatial and temporal weights were then multiplied and summed across each level of the location hierarchy and normalised for each time period t . This procedure allowed the space-time weight to implicitly take into account the amount of data available at the country vs. region vs. super-region level and attribute spatial weight accordingly.

Given a normalisation constant,

$$K_i = \sum_{c \in C} s.w_{c,t} * t.w_{c,t} + \sum_{c \in R} s.w_{c,t} * t.w_{c,t} + \sum_{c \in SR} s.w_{c,t} * t.w_{c,t}$$

the final space-time weight would then equal

$$w'_{c,a,s,t} = \frac{s.w_{c,t} * t.w_{c,t}}{K_i}$$

Finally, we calculated the weight $w''_{c,a,s,t}$ to smooth over age, which is based on a distance along the age dimension of two observations. For a point between the age a of the observation $r_{c,a,s,t}$ and a focal observation r_{c_0,a_0,s_0,t_0} , the weight is defined as follows:

$$w''_{c,a,s,t} = \frac{1}{e^{\omega|a-a_0|}}$$

The final weights were then computed by simply multiplying the space-time weights and age weights and normalising so all weights for a given time period t sum to 1. A full derivation of weights for each category, assuming the location being estimated was a country, follows:

- 1) If the observation $r_{c,t}$ belongs to the same country c_0 of the focal observation r_{c_0,t_0} :

$$w_{c,a,s,t} = \frac{(w'_{c,a,s,t} w''_{c,a,s,t})}{\sum_{c=c_0} (w'_{c,a,s,t} w''_{c,a,s,t})} \quad \forall c = c_0$$

- 1) If the observation $r_{c,t}$ belongs to a different country than the focal observation r_{c_0,t_0} , but both belong to the same region R :

$$w_{c,a,s,t} = \frac{(w'_{c,a,s,t} w''_{c,a,s,t})}{\sum_{c \neq c_0} (w'_{c,a,s,t} w''_{c,a,s,t})} \quad \forall c \neq c_0 \cap R[c] = R[c_0]$$

- 2) If the observation $r_{c,t}$ belongs to the same super region SR but to both a different country c_0 and a different region $R[c_0]$ than the focal observation r_{c_0,t_0} :

$$w_{c,a,s,t} = \frac{(w'_{c,a,s,t} w''_{c,a,s,t})}{\sum_{c \neq c_0} (w'_{c,a,s,t} w''_{c,a,s,t})} \quad \forall c \neq c_0 \cap R[c] \neq R[c_0] \cap SR[c] = SR[c_0]$$

- 3) If the observation $r_{c,t}$ is from a different super region than the focal observation r_{c_0,t_0} (ie, all other data currently not receiving a weight):

$$w_{c,a,s,t} = \frac{(w'_{c,a,s,t} w''_{c,a,s,t})}{\sum_{c \neq c_0} (w'_{c,a,s,t} w''_{c,a,s,t})} \quad \forall c \neq c_0 \cap R[c] \neq R[c_0] \cap SR[c] \neq SR[c_0]$$

Observations could be down weighted by a factor of 0.1, usually because they were not geographically representative at the unit of estimation. Details of reasons for down weighting can be found in cause-specific modeling summaries. The final weights were then normalised such that the sum of weights across age, time, and geographic hierarchy for a reference group was 1.

Section 2.4.1: Estimating error variance

σ_p^2 represents the error variance in normal or transformed space including the sampling variance of the estimates and prediction error from any crosswalks performed. First, variance was systematically imputed if the data extraction did not include any measure of uncertainty. When some sample sizes for data were available, missing sample sizes were imputed as the 5th percentile of available sample sizes.

Missing variances were then calculated as $\sigma_p^2 = \frac{p*(1-p)}{n}$ for proportions or were predicted from the mean by using a regression for continuous values. When sample sizes were entirely missing and could not be imputed, the 95th percentile of available variances at the most granular geographic level (ie, first country, then region, etc.) were used to impute missing variances. For proportions where $p*n$ or $(1-p)*n$ is <20, variance was replaced by using the Wilson Interval Score method.

Next, if prevalence was modelled as a log transformation, the error variance was transformed into log-space by using the delta method approximation as follows:

$$\sigma_p^2 \cong \frac{\sigma_{p'}^2}{p_{c,a,s,t}^2}$$

where $\sigma_{p'}$ represents the error variance in normal space. If prevalence was modelled as a logit transformation, the error variance was transformed into logit-space by using the delta method approximation as follows:

$$\sigma_p^2 \cong \frac{\sigma_{p'}^2}{(p_{c,a,s,t} * (1 - p_{c,a,s,t}))^2}$$

Finally, prior to GPR, an approximation of non-sampling variance was added to the error variance. Calculations of non-sampling variance were done on normal-space variances. Non-sampling variance was calculated as the variance of inverse-variance weighted residuals from the space-time estimate at a given location-level hierarchy. If there were <10 data points at a given level of the location hierarchy, the non-sampling variance was replaced with that of the next highest geography level with >10 data points.

Section 2.4.2: Estimating the covariance function

The final input into GPR is the covariance function, which defines the shape and distribution of the trends. Here, we have chosen the Matern-Euclidian covariance function, which offers the flexibility to model a wide spectrum of trends with varying degrees of smoothness. The function is defined as follows:

$$M(t, t') = \sigma^2 \frac{2^{1-\nu}}{\Gamma(\nu)} \left(\frac{d(t, t')\sqrt{2\nu}}{l} \right)^\nu K_\nu \left(\frac{d(t, t')\sqrt{2\nu}}{l} \right)$$

where $d(\cdot)$ is a distance function; σ^2 , ν , l , and K_ν are hyperparameters of the covariance function specifically σ^2 is the marginal variance, ν is the smoothness parameter that defines the differentiability of the function, l is the length scale, which roughly defines the distance between which two points become uncorrelated, and K_ν is the Bessel function. We approximated σ^2 by taking the normalised median absolute deviation $MADN(r'_c)$ of the difference, which is the normalised absolute deviation of the difference of the first-stage linear regression estimate from the second-stage spatiotemporal smoothing step for each country. We then took the mean of these country-level MADN estimates for all countries with 10+ country-years of data to ensure that differences between first- and second-stage estimates had sufficient data to truly convey meaningful information on model uncertainty. We used the

parameter specification $\nu=2$ for all models. The scale parameter l used for each cause is reported in appendix sections 3.4 and 4.12.

Section 2.4.3: Prediction using GPR

We integrated over $g_{c,t}(t_*)$ to predict a full time series for country c , age a , sex s , and prediction time t_* as follows:

$$p_{c,a,s}(t_*) \sim N\left(m_{c,a,s,t}(t_*), \sigma_p^2 I + \text{Cov}\left(g_{c,a,s,t}(t_*)\right)\right)$$

Random draws of 1000 samples were obtained from the distributions above for every country for a given indicator. The final estimated mean for each country was the mean of the draws. In addition, 95% UIs were calculated by taking the 2.5 and 97.5 percentile of the sample distribution. The linear modelling process was implemented by using the lmer4 package in R, and the ST-GPR analysis was implemented through the PyMC2 package in Python.

Section 2.4.4: Subnational scaling and aggregation

To ensure internal consistency of the estimates between countries and their respective subnational locations, national estimates were either created by population-weighted aggregation or subnational estimates were adjusted by population-weighted scaling to the national estimates, depending on the data coverage of a given country compared to that of its subnational locations. For example, if data coverage was better at the national level than at its corresponding subnational locations for a given country and cause across age, sex, and time, estimates were rescaled to be consistent with the national level. Conversely, if data coverage was better at the subnational level, estimates for its parent country were generated through population-weighted aggregation of subnational estimates. Estimates can also be scaled within logit space. Scaling in logit space ensures that subnational estimates of proportion models do not exceed one after being rescaled to the national estimate.

Section 2.5: MR-BRT meta-regression modelling

Section 2.5.1 MR-BRT Overview

MR-BRT is a meta-regression modeling tool developed at IHME. In contrast to other types of regression, meta-regression incorporates uncertainty in the dependent variable; each observation comes with its own standard error. This characteristic is important when the input data are results of scientific studies that are reported with uncertainty. Observations with greater uncertainty are given less weight in the model. To describe variation in the parameter of interest, MR-BRT can incorporate both fixed and random effects. Fixed effects include binary and continuous covariates as in a traditional regression model. Random effects describe group-level variation and are often used to characterize differences between studies beyond what is captured by measured covariates.

Section 2.5.2 MR-BRT Formula

Formally, a linear mixed effects meta-regression as implemented in MR-BRT can be described as:

$$y_{ij} = \beta_0 + \beta_1 x_1 + \dots + \beta_n x_n + u_j + \epsilon_{ij}.$$

The variable y_{ij} refers to the value of observation i in study j ; it is typically expressed in log or logit space to ensure that model predictions remain within logical constraints, for example that relative risks cannot be negative. The terms $\beta_0 + \beta_1 x_1 + \dots + \beta_n x_n$ comprise the linear predictor, including both the intercept and the effects of any number of covariates. The term u_j is a random intercept corresponding to study j . The full set of random intercepts is assumed to follow a normal distribution where γ is the variance of between-study heterogeneity. Random effects can be estimated for continuous covariates as well, in which case they are called random slopes. The term ϵ_{ij} refers to the stochastic error corresponding to observation i in study j , and the set of values are assumed to follow a normal distribution in which observation-specific standard errors are known prior to modeling. This linear mixed effects formulation of the model covers most features MR-BRT. Features that involve nonlinear optimization techniques like the ratio model (described below) extend this framework and are described formally elsewhere.¹²

Section 2.5.3 MR-BRT Features

MR-BRT – as suggested by its full name “Meta-Regression with Bayesian priors, Regularization and Trimming” – comes equipped with several capabilities that expand upon the classical mixed effects meta-regression model:

Section 1: Bayesian priors can be applied to any estimated coefficient, enabling information from outside the dataset to be considered in the process of fitting the model. A Uniform prior sets hard bounds on the allowed values of an estimated coefficient. A Gaussian prior acts as a suggestion for the estimated value of a coefficient, with the standard deviation of the specified Gaussian distribution determining the strength of the prior.

Section 2: LASSO variable selection, also known as L1 regularization, can be implemented by specifying Laplace priors with mean 0 on the β coefficients. Similarly, ridge regression, also known as L2 regularization, can be implemented by specifying Gaussian priors with mean 0 on the β coefficients.

Section 3: Trimming is a method for identifying and removing the effects of outliers. Users define the proportion of points to be excluded and the algorithm determines which ones to exclude. Because the trimming algorithm is an integrated part of the model’s likelihood function, MR-BRT identifies outliers and estimates the β coefficients simultaneously during the fitting process.

Section 4: A spline term may be used to describe the nonlinear effect of a covariate. MR-BRT implements a B-spline, or basis spline. Users have control over the flexibility of the estimated curve by specifying the number of knots, location of knots, spline degree (i.e. cubic or quadratic), linearity in the tail segments, convexity, concavity, or a monotonicity constraint requiring the spline to be non-decreasing or non-increasing.

Section 5: Pairs of exposure intervals may be used as an independent variable using a method known as the “ratio model”. This feature is most often used when the epidemiological literature reports relative risks corresponding to a reference exposure range (e.g. BMI = [18,22]) and an alternative exposure range (e.g. BMI = [30,35]). It is usually used in conjunction with a spline to capture the nonlinear effect of the exposure. The ratio model works by integrating over the span of each interval and taking the ratio as part of the likelihood function.¹²

The source code for MR-BRT is publicly available on GitHub as the Python package `mrtool` (ihmeuw-msca, 2023). The `mrtool` package builds upon the open source mixed effects package `LimeTr` (<https://github.com/zhengp0/limetr>). For a full technical description of MR-BRT and the underlying mathematics.¹²

Section 2.6: DisMod-MR 2.1 estimation

Section 2.6.1: Estimation of sequelae and causes

The most extensively used estimation method is the Bayesian meta-regression method DisMod-MR 2.1. For some causes, such as HIV/AIDS or measles, disease-specific natural history models have been used for which the underlying three-state model in DisMod-MR 2.1 (susceptible, cases, dead) is insufficient to capture the complexity of a disease process. For some diseases with a range of sequelae differentiated by severity, such as COPD or diabetes mellitus, DisMod-MR 2.1 was used to meta-analyse the data on overall prevalence with separate DisMod-MR 2.1 models of the proportions of cases with different severity levels or sequelae. Likewise, DisMod-MR 2.1 was used to meta-analyse data on the proportions of liver cancer and cirrhosis due to underlying etiologies such as hepatitis B, hepatitis C, and alcohol use disorders.

Section 2.6.2: DisMod-MR 2.1 description

Until GBD 2010, non-fatal estimates in burden of disease assessments were based on a single data source on prevalence, incidence, remission, or a mortality risk selected by the researcher as most relevant to a particular location and time. For GBD 2010, we set a more ambitious goal: to evaluate all available information on a disease that passes a minimum quality standard. That required a different analytical tool that would be able to pool disparate information presented for varying age groupings and from data sources by using different case definitions. The DisMod-MR 1.0 tool used in GBD 2010 evaluated and pooled all available data, adjusted data for systematic bias associated with case ascertainment methods that varied from the reference and produced estimates by world regions with UIs by using Bayesian statistical methods. For GBD 2013, the improved DisMod-MR 2.0 increased computational speed, which allowed computations to be consistent between all disease parameters at the country rather than the region level. The hundred-fold increase in speed of DisMod-MR 2.0 was partly due to a more efficient rewrite of the code in C++, but also due to switching to a model specification of log rates rather than a negative binomial model used in DisMod-MR 1.0. In cross-validation tests, the log rates specification worked as well as or better than the negative binomial specification. In GBD 2015, we also improved how country covariates differentiate non-fatal estimates for diseases with sparse data. The coefficients for country covariates are re-estimated at each level of the cascade but with the default set to constraining estimates of covariates to the upper and lower bound of the global fit. For a given location, country coefficients are calculated by using both data and prior information available for that location. In the absence of data, the coefficient of its parent location is used to utilise the predictive power of our covariates in data-sparse situations.

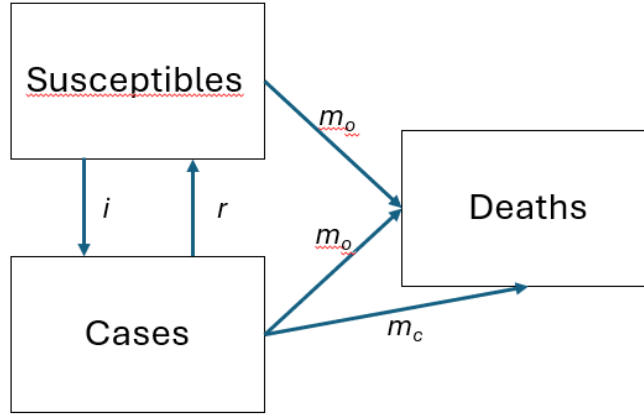
For GBD 2016, the computational engine (DisMod-MR 2.1) remained substantively unchanged from GBD 2015. We updated the age prediction sets to include age groups 80–84 years, 85–89 years, 90–94 years, and 95 years and older to comply with changes across all functional areas of the GBD.

In GBD 2017, we continued to use DisMod-MR 2.1 because no substantial changes were made. Updates to computation include extending the terminal prediction year to 2017 and additional subnational units in Ethiopia, Iran, New Zealand, Norway, and the Russian Federation.

In GBD 2019, 2021 and 2023, no substantial changes were made to DisMod-MR 2.1, but we made more substantial changes to how we use the tool. First, we added the years 2019, 2020, and 2021, 2022 and 2023 as additional years of estimation. Second, we also included the option again to have random effects on cause-specific mortality rates (CSMR) and EMR. This functionality had been dropped a couple of GBD rounds earlier. Third, as we did all our adjustments for alternative case definitions and study methods as well as adjustments to combined-sex data points prior to entering data into DisMod-MR 2.1, we no longer used the functionality in DisMod-MR 2.1 to estimate coefficients for study and sex covariates and instead, processed data prior to entry in DisMod-MR 2.1 as described in the sections above on crosswalking and sex splitting. Fourth, based on simulation testing conducted in GBD 2019 we found that coverage improved, and errors reduced when passing down priors with a wider setting of minimum coefficient of variation (which determines the uncertainty around priors and hence how ‘informative’ the priors are) than had generally been used in past GBD iterations. We settled on a default value of 0.8 where in the past values of 0.4 or less had been more commonly used. We made some exceptions for highly prevalent conditions where a lower minimum coefficient of variation (CV) setting achieved the task of making priors less informative, but not completely uninformative.

DisMod-MR 2.1 simultaneously fits several epidemiologic measures using a Bayesian, nonlinear, mixed-effects regression model, which produces estimates of a consistent set of epidemiologic measures in a compartmental disease model, meaning that they are the solution to the set of differential equations that specify the model. The compartmental model assumes a steady state, and is often illustrated with the following figure S15, which we recreate here from Flaxman et al. (2015).¹³ The system of differential equations that captures the relationships between rates and compartments follows the figure S15.

Figure S15. System of differential equations that captures relationships between rates and compartments



A three-compartment model of a disease in a population consisting of susceptibles, i.e. those in the population without disease, cases of the disease and deaths. Individuals move from susceptible to case by the incidence rate i , and can return to susceptible by remission rate r . Cases are susceptible to a mortality rate, m_i , that reflects the excess mortality in cases of the disease. Cases and susceptibles are subject to a mortality rate from other causes, m_o . Each quantity in figure S15 is a function of age and time, governed by the following differential equations:

$$\frac{d}{d\tau} S(a + \tau, t + \tau) = -(i + m_o)S + rC$$

$$\frac{d}{d\tau} C(a + \tau, t + \tau) = iS - (r + m_o + m_c)C$$

where:

$S = S(a, t)$ = stock of susceptibles

$C = C(a, t)$ = stock of cases, i.e. those in population with disease

$i = i(a, t)$ = incidence hazard for susceptibles S

$r = r(a, t)$ = remission ('cure') hazard for individuals with the disease C

$m_o = m_o(a, t)$ = without condition mortality hazard for susceptibles S

$m_c = m_c(a, t)$ = excess mortality hazard for individuals with the disease C

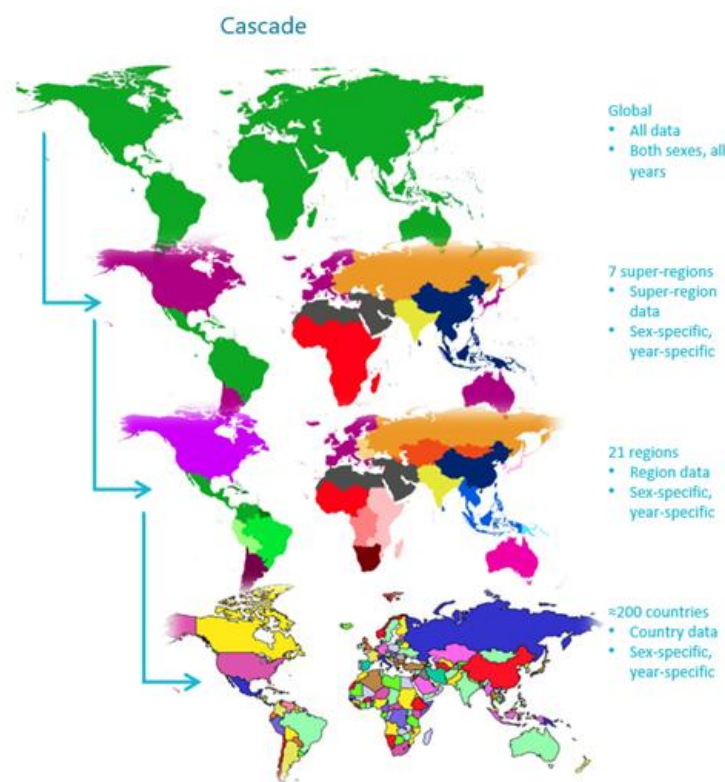
a and t denote age and time.

A Monte Carlo Markov Chain simulation starts with the prevalence at birth, and samples the posterior distribution integrating the observed values for all observations of all epidemiologic quantities, ensuring that susceptibles and cases sum to population size, and that the partial derivatives of each differential equation in the system with respect to time zero. For most diseases the birth prevalence is set to zero.

For disease that can be present at birth, a non-zero birth prevalence is specified. Population sizes for all combinations of year, age, sex, and location are input to DisMod-MR 2.1 from the latest GBD estimates, which are published iteratively and available by download from the GBD results tool.

The sequence of estimation occurs at five levels: global, super-region, region, country and, where applicable, subnational location. The super-region priors are generated at the global level with mixed-effects, non-linear regression by using all available data; the super-region fit, in turn, informs the region fit, and so on down the cascade. For some countries, additional models are fit for subnational locations. At each step in this cascade, only the data appropriate to that year, sex and geography are used to update the prior and yield the posterior. This is illustrated in figure S16 below.

Figure S16. DisMod-MR 2.1 sequence of estimation cascade



Once fitting is complete for the finest granularity, results are summed, at the draw level, weighting by population size, to produce final results for larger geographic units from the finest results.

Analysts can choose to branch the cascade in terms of time and sex at different levels depending on data density. The default used in most models is to branch by sex after the global fit but to retain all years of data until the lowest level in the cascade is reached.

The computational engine is limited to three levels of random effects; we differentiate estimates at the super-region, region, and country level. In GBD 2013, the subnational units of China, the United Kingdom and Mexico were treated as “countries” to enable a random effect to be estimated for every location with contributing data. However, the lack of a hierarchy between country and subnational units meant that the fit to country data contributed as much to the estimation of a subnational unit as the fits for all other countries in the region. We found inconsistency between the country fit and the aggregation of subnational estimates when the country’s epidemiology varied from the average of the region. Adding an additional level of random effects required a prohibitively comprehensive rewrite of the underlying DisMod-MR engine. Instead, we added a fifth layer to the cascade, with subnational estimation informed by the country fit and country covariates, plus an adjustment based on the average of the residuals between the subnational location’s available data and its prior. This technique mimicked the impact of a random effect on estimates among subnationals.

In GBD 2017 and 2019 GBD rounds we calculated priors on excess mortality and entered these as data points by matching sex-specific prevalence data with an age width of 20 or less with the corresponding CSMR for the same location and year. For stability, we excluded calculation of EMR for prevalence data points of less than 1 in a million. EMR is simply calculated as CSMR divided by prevalence. As with previous GBD years, for diseases with an average duration of less than a year (as indicated by a setting of remission greater than one), we ran an initial global model to get an equivalent prevalence and used the following formula to calculate EMR:

$$EMR = \frac{CSMR * (remission + (ACMR - CSMR) + EMR_{pred})}{incidence}$$

where,

ACMR is the all-cause mortality rate

EMR_{pred} is the EMR fit from an initial global DisMod model

Despite using the log of LDI or the HAQ Index as a covariate with a prior that the coefficient had to be negative, we found many disease models with an implausible distribution of mortality to prevalence (or incidence) ratios implying lower case fatality in locations with lower HAQ Index than in countries with higher HAQ Index. This likely signals an inconsistency between fatal and non-fatal data inputs. For GBD 2019, we decided to run regressions on EMR data (calculated as described above) first using MR-BRT with HAQ Index as a predictor. In general, we tend to think that CSMR estimates are more robust than non-fatal data because of much greater data availability and a lesser task in adjusting cause death data for garbage coding than the complex task of adjusting non-fatal data sources for alternative case definitions and study methods. To indicate that we would reduce the random effects on EMR and the minimum coefficient of variation for priors on EMR being created at each next level down the cascade.

Section 2.6.3: DisMod-MR 2.1 likelihood estimation

Analysts have the choice of using a Gaussian, log-Gaussian, Laplace, or Log-Laplace likelihood function in DisMod-MR 2.1. The default offset-log-Gaussian equation for the data likelihood is

$$-\log[p(y_j|\psi)] = \log(\sqrt{2\pi}) + \log(\delta_j + s_j) + \frac{1}{2} \left(\frac{\log(a_j + \eta_j) - \log(m_j + \eta_j)}{\delta_j + s_j} \right)^2$$

where,

y_j is the value of the point estimate of the epidemiologic measure (prevalence, incidence, remission, excess mortality rate, with-condition mortality rate, cause-specific mortality rate, relative risk or standardized mortality ratio) for observation j

Φ denotes all model random variables

η_j is the offset value, *eta*, for a particular “integrand” (prevalence, incidence, remission, excess mortality rate, with-condition mortality rate, cause-specific mortality rate, relative risk, or standardised mortality ratio)

a_j is the adjusted measurement for data point j , defined by

$$a_j = e^{(-u_j - c_j)} y_j$$

where:

u_j is the total “area effect” (i.e., the sum of the random effects at three levels of the cascade: super-region, region and country); random effects at the same level of geography are normally distributed with a group mean of zero; and

c_j is the total covariate effect (i.e., the mean combined fixed effects for sex, study level, and country level covariates), defined by

$$c_j = \sum_{k=0}^{K[I(j)]-1} \beta_{I(j),k} \hat{X}_{k,j}$$

with SD

$$s_j = \sum_{l=0}^{L[I(j)]-1} \zeta_{I(j),l} \hat{Z}_{l,j}$$

where:

k denotes the mean value of each data point in relation to a covariate (also called x-covariate)

$I(j)$ denotes a data point for a particular integrand j

$\beta_{(ij),k}$ is the multiplier of k^{th} covariate for the i^{th} integrand

$\hat{X}_{k,j}$ is the value of covariate k corresponding to observation j

I denotes the SD of each data point in relation to a covariate (also called z-covariate)

$\zeta_{(ij),k}$ is the multiplier of the I^{th} z-covariate for the j^{th} integrand

δ_j is the standard deviation of the adjusted measurement for observation j , after adjustment of y_j for location random effects and fixed effect covariates relevant to that integrand

$$\delta_j = \log \left[y_j + e^{(-u_j - c_j)} \eta_j + c_j \right] - \log \left[y_j + e^{(-u_j - c_j)} \eta_j \right]$$

where:

m_j denotes the predicted value of the integrand corresponding to observation j based on underlying age-specific rates, without accounting for location random effects (u_i) or fixed effects for predictive covariates (c_j). That is to say, it is an integration of the fitted spline values for the age-range referred to by observation j , and defined by:

$$m_j = \frac{1}{B(j) - A(j)} \int_{A(j)}^{B(j)} I_j(a) da$$

where:

$A(j)$ is the lower bound of the age range for a data point

$B(j)$ is the upper bound of the age range for a data point

I_j denotes the function of age corresponding to the integrand for data point j

Section 2.7: DisMod-AT estimation

DisMod-AT was developed to meet the need for a modeling tool which can estimate disease progression over age and time simultaneously. This capability is crucial because many diseases exhibit age, time, and cohort dynamics, which cannot be modeled independently of each other.^{14–16} By including both age and time in its model simultaneously, DisMod-AT is able to estimate dynamics for all three, producing much improved estimates of diseases for which they are strongly linked and advancing the state of the art for disease modeling.

The compartmental model of disease progression used in previous versions of DisMod (Section 2.6: DisMod-MR 2.1 estimation) is also used by DisMod-AT, but with the addition of time t :

$$S'(a, t) = -(\iota(a, t) + \omega(a, t))S(a, t) + \rho(a, t)C(a, t)$$

$$C'(a, t) = \iota(a, t)S(a, t) - (\rho(a, t) + \chi(a, t) + \omega(a, t))C(a, t)$$

with the initial conditions $C(0, t) = P(0, t)$ and $S(0, t) = 1 - P(0, t)$, where:

$S(a, t)$ is the fraction of the population which is susceptible to the condition,

$C(a, t)$ is the fraction of the population with the condition,

$P(0, t)$ is birth prevalence rate; $\iota(a, t)$ is incidence rate,

$\rho(a, t)$ is remission rate; $\chi(a, t)$ is excess mortality rate, and

$\omega(a, t)$ is mortality rate from other causes.

DisMod-AT's key advance from DisMod-MR 2.1 is to explicitly consider the equations' age- and time-dependence simultaneously, whereas DisMod-MR 2.1 considered them time-independent. By taking time to be the time of birth for a population cohort, c , such that $f(a, t) = f(a, c + a) = f_c(a)$, these equations can be written:

$$S'_c(a) = -(\iota_c(a) + \omega_c(a))S_c(a) + \rho_c(a)C_c(a)$$

$$C'_c(a) = \iota_c(a)S_c(a) - (\rho_c(a) + \chi_c(a) + \omega_c(a))C_c(a)$$

DisMod-AT solves this system in age and time simultaneously by utilizing interior-point optimization methods rather than the Markov chain Monte Carlo method in DisMod-MR.^{17–19} This advance is what allows DisMod-AT to estimate age and time dynamics in disease, which is impossible if the equations are considered independently of either age or time.

DisMod-AT uses the same geographical cascade structure as DisMod-MR, carrying estimation strength across geography using Bayesian priors (Section 1.2: Geographical locations of the analysis). First, DisMod-AT estimates disease at the global level drawing from all the world's data and passes that estimate to the seven super-regions as their prior after adjustment by the random effect for each super-region and the beta values of predictive covariates that were chosen of any of the primary rates (remission, incidence and excess mortality) in the model. Then, regions are fit using super-region estimates to inform their priors, followed by fits for country-level locations using region estimates to inform their priors, and concluding with fits for sub-national locations, where present, using country estimates to inform their priors.

The model for each estimated quantity, integrands in GBD publication nomenclature, is a likelihood function of the form:²⁰

$$\mathcal{P}(y_i|u, \theta) = e^{D(y_i, A_i(u, \theta), \delta_i(\theta))}$$

where y_i is the i^{th} measurement (data) value, u are the random effects, θ are the fixed effects (i.e. of predictive covariates), D is the probability density function corresponding to y_i , specified as a configuration option by the analyst, $A_i(u, \theta)$ is the i^{th} average integrand.²¹

$$A_i(u, \theta) = \frac{1}{b_i - a_i} \frac{1}{d_i - c_i} \int_{a(i)}^{b(i)} \int_{c(i)}^{d(i)} \frac{w_i(a, t)}{\bar{w}_i} I_i(a, t) e^{E_i(a, t)} dt da$$

$I_i(a, t)$ is the i^{th} integrand function, in terms of $\iota(a, t)$, $\rho(a, t)$, $\chi(a, t)$, $\omega(a, t)$, $S(a, t)$, and $C(a, t)$, \bar{w}_i is the age-time weighting:

$$\bar{w}_i = \frac{1}{b_i - a_i} \frac{1}{d_i - c_i} \int_{a(i)}^{b(i)} \int_{c(i)}^{d(i)} w_i(a, t) dt da$$

$E_i(a, t)$ is the effect for the i^{th} measurement value, in terms of the fixed effects θ :

$$E_i(a, t) = \sum_{j \in K_i} x_{i,j} (\beta_j(a, t) + \Delta \beta_j(a, t))$$

$\beta_j(a, t)$ is a piecewise linear function giving the effect for covariate $j \in K_i$, $x_{i,j}$ is the value of covariate $j \in K_i$, K_i are the set of covariates on the i^{th} integrand; $\delta_i(\theta)$ is the transformed standard deviation corresponding to y_i :

$$\delta_i(\theta) = \begin{cases} \log(y_i + \eta_i + \sigma_i(\theta)) - \log(y_i + \eta_i) & \text{if log density} \\ \sigma_i(\theta) & \text{otherwise} \end{cases}$$

η_i is a small number which offsets log densities to prevent infinities, $\sigma_i(\theta)$ is the standard deviation corresponding to y_i :

$$\sigma_i(\theta) = \begin{cases} \Delta_i + \varepsilon_i(\theta) \\ \Delta_i(1 + \varepsilon_i(\theta)) \\ \sqrt{\Delta_i^2 + \varepsilon_i(\theta)} \\ \Delta_i \sqrt{1 + \varepsilon_i(\theta)} \end{cases}$$

Δ_i is a configurable option specifying a minimum coefficient of variation on the measurement, $\varepsilon_i(\theta)$ is the estimated average noise effect:

$$\varepsilon_i(\theta) = \frac{1}{\bar{w}_i} \int_{a(i)}^{b(i)} \int_{c(i)}^{d(i)} w_i(a, t) \sum_{j \in K_i} x_{i,j} \gamma_j(a, t) dt da$$

and $\gamma_i(a, t)$ is a piecewise linear function for the covariate multiplier values.

Analysts specify the set of age points and of time points for which DisMod-AT estimates disease rates as configuration options. DisMod-AT interpolates between these points bilinearly when required. In tandem, analysts also specify Bayesian priors on the change in estimated rates between age or time points, configuring how much the model expects estimated rates to vary over each age or time increment. These priors can be specified for each rate in the compartmental model and may be specified over subsets of age or time points. Additionally, analysts have the option to specify Bayesian priors on covariate fixed effects, random effects, measurement value effects, and measurement noise effects.

We selected four causes—type 1 diabetes, major depressive disorder (MDD), anxiety disorders, and autism spectrum disorders—for our initial roll-out of DisMod-AT in GBD 2023. These causes were chosen over concerns that DisMod-MR 2.1 may not have been adequately capturing trends in the underlying data. For GBD 2021, the impact of the COVID-19 pandemic on MDD and anxiety disorders had to be quantified via a custom analysis after analysis in DisMod-MR 2.1. While we have not conducted a formal

comparison of results generated by DisMod-AT versus DisMod-MR for these conditions, DisMod-AT has improved our ability to model temporal trends and generate more robust estimates, especially where non-linear age or cohort effects are present. DisMod-AT has been more sensitive to the increase in prevalence of depressive and anxiety disorders and autism spectrum disorder which were evident in epidemiological data for some regions, and has allowed us to more accurately model known adult epidemiology for type 1 diabetes, despite a paucity of data for adult age groups. DisMod-AT allowed the impact of the COVID-19 pandemic to directly influence the incidence of both MDD and anxiety disorders via a location-level covariate, avoiding the need for post-analysis adjustments required of DisMod-MR 2.1. The transition to DisMod-AT is ongoing, and its broader implementation across causes will be accompanied by systematic validation exercises.

Section 2.8: Impairment and underlying cause estimation

For GBD 2023, as in GBD 2021, GBD 2019, GBD 2017 and GBD 2016, we estimated the country-age-sex-year prevalence of nine impairments. Impairments in GBD are conditions or specific domains of functional health loss that are spread across many GBD causes as sequelae and for which there are better data to estimate the occurrence of the overall impairment than for each sequela based on the underlying cause. These impairments included anaemia, epilepsy, hearing loss, heart failure, intellectual disability, infertility, vision loss, Guillain-Barré syndrome, and pelvic inflammatory disease. Overall impairment prevalence was estimated by using DisMod-MR 2.1. We constrained cause-specific estimates of impairments, as in the 19 causes of blindness, to sum to the total prevalence estimated for that impairment. Anaemia, epilepsy, hearing loss, vision loss, heart failure, and intellectual disability were estimated at different levels of severity. Estimates were made separately for primary infertility (those unable to conceive), secondary infertility (those having trouble conceiving again), and whether the impairment affected men and/or women. In the case of epilepsy, we determined the proportions with idiopathic and secondary epilepsy as well as the proportions with severe and less severe epilepsy by using mixed effects regressions. The sparse data for the proportion of seizure-free, treated epilepsy were pooled in a random effects meta-analysis. DisMod-MR 2.1 models produced country-, age-, sex-, and year-specific severity levels of hearing loss and vision loss. Because of limited information on the severity levels of intellectual disability, we assumed a similar distribution of severity globally based on random effects meta-analysis of IQ-specific data for the overall impairment. This assumption was supplemented by cause-specific severity distributions for chromosomal causes and iodine deficiency; the severity of intellectual disability included in the long-term sequelae of causes including neonatal disorders, meningitis, encephalitis, neonatal tetanus, and malaria was estimated in combined health states of multiple impairments such as motor impairment, blindness, and/or seizures (R, 2015). We changed the name of the intellectual disability impairment to specify that estimates reflect cases arising during the developmental period, which we have defined as ages under 20 years. The severity of heart failure was derived from our Medical Expenditure Panel Surveys (MEPS) analysis and therefore was not specific for country, year, age, or sex.

A detailed description of the methods of each impairment can be found in Section 6 of this appendix.

Section 2.8.1: Impairment squeeze

For the impairments epilepsy, intellectual disability, and blindness, mentioned above in Step 4, we often have better information regarding the total prevalence of the impairment rather than the prevalence of said impairment due to its various causes. For example, we have more data and a better idea of the total number of blind individuals (which we refer to herein as the blindness “envelope”) in the world than we do the number of individuals who are blind due to a specific cause like retinopathy of prematurity or cataract. We achieve this consistency by either squeezing or inflating the individual sequela prevalence values so that their sums fit into each appropriate envelope. Blindness, epilepsy, and/or intellectual disability appear in various combinations with motor impairment levels as sequelae for a number of neonatal disorders and infectious diseases like malaria and neonatal tetanus (“Moderate motor impairment with blindness and epilepsy due to neonatal tetanus”, for example). This presents an extra challenge because any squeeze or inflation of one of the impairments making up a sequela affects the others.

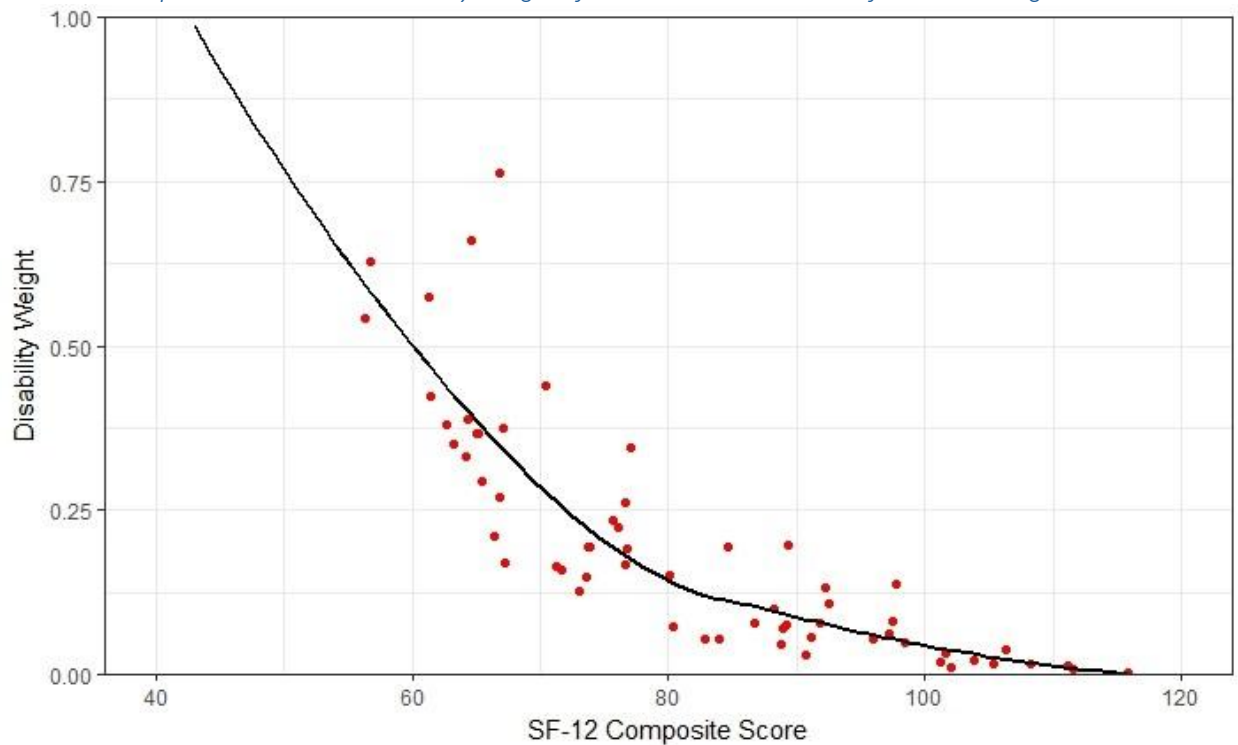
We set rules on how to do these adjustments sequentially. First, when the envelope of an impairment is smaller than the sum of all contributing causes, we redistribute the excess prevalent cases of combined impairment sequelae onto the sequelae that only have motor impairment (at a mild, moderate, or severe level) within the same cause grouping. Second, we apply the adjustments in a particular order such that we always fit at least one of the envelopes exactly where the other one or two envelopes may be exceeded by some amount. We first enforce a fit to the epilepsy impairment envelope, then intellectual disability, and last, blindness. Thus, the epilepsy envelope always matches exactly, whereas the intellectual disability and blindness envelopes may occasionally be exceeded on a draw-by-draw basis.

Sequelae were defined in terms of severity for 236 causes. We generally followed the same approach for estimating the distribution of severity we used in GBD 2021. In cases in which severity was related to a particular impairment, such as mild, moderate, and severe heart failure due to ischaemic heart disease or pulmonary arterial hypertension, the analysis was driven by impairment estimation methods. Severity levels for causes such as chronic kidney disease, epilepsy and COPD were modelled using DisMod-MR 2.1 or ST-GPR, whereas we performed meta-analyses to estimate the allocation of severity for causes such as rheumatoid arthritis, and multiple sclerosis. For dementia, we changed from using meta-analysis of three age categories to a more flexible model in MR-BRT using a spline on age. That allowed us to increase the number of studies informing severity from 7 to 67. For gallbladder and biliary diseases, we performed a meta-analysis of six community-based studies of the proportion of cases of gallbladder disease identified by ultrasonography who are symptomatic. In previous rounds, inpatient admission for gall bladder and biliary disease as a primary diagnosis were taken to represent symptomatic cases. For many causes, we continue to have inadequate data on severity from surveys or the epidemiological literature. For those diseases, we made use of three population surveys: the MEPS (Medical Expenditure Panel Survey) 2000–2014, the [US] National Epidemiological Survey on Alcohol and Related Conditions (NESARC)²² 2000–2001 and 2004–2005, and the Australian National Survey of Mental Health and Wellbeing of Adults (NSMHWB) 1997 (Mental Health and Wellbeing: Profile of Adults, Australia, 1998)²³. Each dataset contained individual-level measurements of functional health status made by using the 12-Item Short Form Health Survey (SF-12) as well as diagnostic information on the causes affecting each individual.

To use the data collected by measuring the distribution of severity with the SF-12, the individual SF-12 summary scores were mapped to an equivalent DW. A convenience sample of respondents was asked to complete SF-12 for the hypothetical individual living in a health state described by using a selection of 60 of the 235 health states with their lay descriptions from the GBD DW surveys reflecting the full range of severity. Each of these health states has a measured DW associated with it on a zero to one scale. We collected 2783 usable responses in total.

The final relationship between SF-12 score and DW is depicted in Figure S17:

Figure S17. SF-12 composite scores and disability weights for 60 health states with fitted loess regression



1.

To generate a smooth mapping from SF-12 combined scores to the GBD DW space, we used locally estimated scatterplot smoothing regression on the random effects for each health state. DWs were capped to remain between 0 and 1. All SF-12 survey data were thus transformed into DW space. The second stage of the analysis was to build models predicting the transformed SF-12 scores as a function of the number of causes suffered by each individual. First, variable selection was performed by using least absolute shrinkage and selection operator (LASSO) regression to penalize the regression coefficients of highly correlated causes. The tuning parameter, λ , controls the strength of the least-squares penalty. When $\lambda=0$, LASSO regression returns the same results as ordinary least-squares regression. Higher values of λ impose a stronger penalty and constrain a greater number of model parameters to 0. A ten-fold cross-validation was used to find the value of the λ that minimized the mean cross-validated error. This process resulted in a λ value of 0.0013 and eliminated 10 causes from the

analysis. Transformed SF-12 scores into the DW scale for the remaining 190 causes were then modelled for each measure m of each individual i over n total causes in the survey as follows:

$$\text{logit}(DW)_{im} = \beta_0 + \beta_1 \text{Condition}1_{im} + \dots + \beta_n \text{Condition} n_{im}$$

This equation effectively assumes that comorbid causes act to change SF-12 scores in a multiplicative fashion rather than an additive fashion.

To estimate the comorbidity-corrected effect of each cause (ie, in isolation) on total disability, we compared the predicted DW without the cause of interest (counterfactual DW) with the predicted DW including the cause of interest. Following the multiplicative comorbidity equation, the joint effect can be written

$$\text{Condition specific DW} = 1 - \frac{1 - \text{predicted DW}_m}{1 - \text{counterfactual DW}_m}$$

The mean of this cause-specific effect over all observations is the population marginal effect of a cause.

Using the model above, we estimate a counterfactual DW – the total individual DW excluding the effect of the cause of interest. We compared the observed distribution of functional health status with this counterfactual distribution to determine the marginal effect of the cause of interest. In other words, we estimated the health state for each individual and for each cause as the cumulative individual weight minus the effects of all comorbid causes.

$$\text{Health state DW} = 1 - \frac{1 - \text{individual cumulative DW}_m}{1 - \text{counterfactual DW}_m}$$

The estimation strategy for health state-specific severity distributions for which there are multiple severity categories involved binning individuals' weights into severity cut-offs (eg, mild, moderate, and severe) for which DWs were derived. These bins were defined by using results from the GBD Disability Weights Studies²⁴ for causes that had multiple health states defined. Cut-offs for the severity group were the midpoints between DWs of the health state and cases distributed into severity bins accordingly. For example, individuals with a health state DW above the mid-point between the mild DW and the moderate DW for a particular condition would be assigned the moderate sequela. Cases were considered asymptomatic if the counterfactual weight was equal to or greater than the individual cumulative weight. The proportion of cases of a condition assigned to each level of severity for that condition was then used as the severity distribution of the condition for prevalence estimates to be apportioned accordingly into severity-specific prevalence estimates.

Section 2.9: Disability weights

To compute YLDs for a particular health outcome in a given population, the number of people living with that outcome is multiplied by a disability weight (DW) that represents the magnitude of health loss associated with the outcome. DWs are measured on a scale from 0 to 1; 0 implies a state equivalent to full health, and 1, a state equivalent to death.

Section 2.9.1: GBD 2010 Disability Weights Measurement Study

For GBD 2010, a primary data collection effort focused on measuring health loss rather than welfare loss by using a standardised approach of simple comparison questions directed to the general public across diverse communities.

Multi-country household surveys were conducted between Oct 28, 2009 and June 23, 2010 in five countries (Bangladesh, Indonesia, Peru, Tanzania, and the USA) selected to provide diversity across culture, language, and socioeconomic status.

Personal face-to-face computer-assisted interviews were conducted for all household surveys except for the survey in the US, which was conducted by computer-assisted telephone interview. Households were randomly selected by using a multistage stratified sampling design for which the probability of selection was proportional to the population size. In all cases, samples were designed to be representative of a given geographical area and, in the USA, to provide national representation.

For every contacted household, an adult respondent age 18 years or older was randomly selected by the survey program by means of the Kish approach. For face-to-face interviews, as many as three visits were made to selected households to establish contact. When a respondent was identified, as many as three return visits were made to do the survey at a time when the respondent was available. For the US telephone surveys, repeated calls were made up to seven times.

A web-based survey was posted at a dedicated URL between July 26, 2010 and May 16, 2011. The survey was initially available in English and subsequently available in Spanish and Mandarin. Recruitment of respondents occurred through several channels, such as news items and editorials in scientific journals, announcements at scientific meetings, postings on websites of institutions participating in the GBD, and social networking and communication mobilisation channels as well as direct contact with individuals and groups with known global health interests by tapping into the professional networks of the study investigators and their colleagues. Participants in the web-based survey were required to be ages 18 or older. Household surveys obtained oral informed consent from all participants; written informed consent was obtained from participants in the web survey. Ethical review board approval was obtained from each household survey site and the University of Washington, Seattle, WA.

Standardised survey instruments were developed to obtain comparative assessments of the full array of disease and injury sequelae, parsimoniously captured in 220 unique health states. Lay descriptions of health states formed the basis for all comparisons. These descriptions used simple, non-clinical vocabulary that emphasised the major functional consequences and symptoms associated with each health state. Development of these descriptions involved an iterative process of detailed consultation with experts participating in the GBD 2010 study; the goal was to capture the most relevant details of each health state while avoiding ambiguity and ensuring consistency. When possible, health states were grounded in standard clinical classifications systems. For example, the Canadian Cardiovascular Society grading scale was referenced for descriptions of stages of angina,²⁵ and the New York Heart Association

functional classification was referenced for severity of heart failure.²⁶ Pilot testing indicated that the lay descriptions in face-to-face interviews should not exceed 30 words.

A paired comparison question formed the basis of all surveys. The questions in the survey were framed with the following statement, “A person’s health may limit how well parts of his body or mind work. As a result, some people are not able to do all of the things in life that others may do, and some people are more severely limited than others. I am going to ask you a series of questions about different health problems. In each question, I will describe two different people...” Descriptions of two hypothetical people, each with a particular health state, were presented to respondents who were then asked which person they regarded as healthier. Health pairs in all surveys were selected by a randomizing computer algorithm. In the five household surveys, paired comparisons were presented for a subset of 108 health states pertaining to chronic conditions. The framing of chronic and acute conditions is different as they were presented as causing life-long or temporary health loss. We chose to only field health states that could be framed as lasting a lifetime in the household surveys as we hypothesized that presenting differently framed comparisons would be difficult to convey in face-to-face interviews. In the web survey, we considered this more feasible because respondents could read and refer to the framing of the question for each pair-wise comparison. All 220 health states were thus evaluated in the web survey.

In addition, the web survey included questions relating to population health and health programs specifically—such as “Imagine two different health programs. The first program prevented 1000 people from getting an illness that causes rapid death. The second program prevented 2000 people from getting an illness that is not fatal but causes lifelong health problems resulting in moderate to severe disability. Which program would you say produced the greater overall health benefits?” This information was used to anchor the results from the pair-wise comparisons on the 0–1 DW scale.

[Section 2.9.2: GBD 2013 European disability weights measurement study](#)

The GBD 2010 DWs were critically dependent on the ways that outcomes were described to survey respondents. Descriptions for health states were designed to balance validity and parsimony, and this approach necessarily meant that some details of different health states had to be omitted. Because lay descriptions were developed collaboratively through individual expert groups organised around a particular set of health issues, some amount of variability in language and detail inevitably occurred. Criticisms and suggestions for improvement came from a number of commentators on the GBD 2010 DWs measurement study.^{27–29}

GBD 2013 expanded the list of disease and injury causes and sequelae mapped to 235 unique health states. Additional data for the European Disability Weights Measurement Study were collected between September 23, 2013, and November 11, 2013, in Hungary, Italy, the Netherlands, and Sweden. The initiation of these surveys was connected to a project sponsored by the European Centre for Disease Prevention and Control.³⁰ The four selected countries were chosen to be representative of the four regions of Europe (east, south, middle, and north) in terms of age, sex, and education of the respondents. Respondents were recruited from standing internet panels in each country on the basis of quota sampling with reference to age, sex, and education in such a way as to maintain the population representativeness of these characteristics. Eligible participants were 18–65 years old and were

preselected in the Netherlands, where the age, sex, and education of respondents were already known, or in the other three countries, invited to participate via a web-link and then selected on the basis of their individual characteristics.

The protocol for the European DWs measurement study followed the protocol that was developed and implemented in the GBD 2010 DWs measurement study. Lay descriptions for some health states that lacked mention of an important symptom or for which consistency of wording across different levels of severity had been noted were reworded. The European DWs measurement study included 255 health states, of which 183 were used in the analyses of GBD 2013. Those 183 consisted of 135 of the 220 health states that were included in the European DWs measurement study with unmodified lay descriptions and 30 from GBD 2010 for which alternative lay descriptions were included. DWs were estimated for additional sequelae that were incorporated into GBD 2013 but had not been included in GBD 2010.

Finding high correlation in resulting DW values between the country surveys and the web survey, we analysed the results of all surveys together. We ran probit regression analyses on the answers to the pair-wise comparison questions by using dummies for each health state with a value of 1 for the first state in a pair, -1 for the second state in a pair, and 0 for all states other than the pair. This method formalizes the intuition that if two health states in a pair produce similar health loss, the answers are likely to be evenly split; a pair of health states with very different health loss get many more responses favouring one over the other. The statistical methods infer the distances between values attached to different health states based on the frequencies of responses to the paired comparisons.

A second analytic step is needed to anchor the resulting estimates onto the 0–1 DWs scale, where 0 equals no loss of health, with 1 meant to represent loss equivalent to death. We anchored results from the probit regression analysis onto the 0–1 scale by using population health equivalence data from the GBD 2010 web survey by using a linear regression of the probit coefficients from the analysis of paired comparisons on the logit-transformed DW estimates derived from interval regression of the population health equivalence responses. Using numerical integration, we then estimated mean values for DWs on the natural 0–1 scale. Uncertainty was estimated by bootstrapping with 1,000 samples. For a complete overview of disability weights applied to the Global Burden of Disease Study, see the GBD 2013 article by Salomon and colleagues.²⁴

Section 2.10: Comorbidity correction (COMO)

The final stage in the estimation of YLDs is a micro-simulation, which adjusts for comorbidity. We refer to this micro-simulation process as “COMO” (for comorbidity correction). We estimated the co-occurrence of different diseases by simulating 20,000 individuals in each location-age-sex-year combination as exposed to the independent probability of having any of the sequelae included in GBD based on prevalence. We tested the contribution of dependent and independent comorbidity in the US MEPS data and found that independent comorbidity was the dominant factor even though well-known examples of dependent comorbidity exist, such as clustering of conditions like diabetes and stroke or anxiety and alcohol use disorders. Age was the main predictor of comorbidity such that age-specific micro-simulations accommodated most of the required comorbidity correction.³¹

The two components necessary for the computation of YLDs and are the two inputs into COMO: 1) prevalence of each disease sequela and 2) DWs. The prevalence values of causes are primarily produced by using DisMod-MR 2.1 and, for causes with multiple sequelae, subsequently apportioned into sequela-specific prevalence based on available estimates of the severity distribution. The estimation of DWs and severity distributions have been described earlier in this appendix.

The micro-simulation, as performed for each age-sex-location-year, can best be represented as a four-step process. First, simulated individuals (simulants) are exposed to independent probabilities of having each sequela, where the probability is equal to the prevalence estimate. For each simulant, the probability of having a disease sequela is equal to the estimated prevalence. Each simulant is determined to have or not have the disease sequelae based on a draw from a binomial distribution. From this simulation, simulants end up with any number of sequelae, from 0 up to the theoretical maximum given their demographics. Second, the DW for each simulant is estimated on the basis of the disease sequelae that they have acquired. The formula for the cumulative DW for a simulant is one minus the multiplicative sum of one minus each DW present

$$Simulant\ DW_l = 1 - \prod_{k=i}^j (1 - DW_k)$$

Where:

DW_k is the DW for the k^{th} disease sequela that the simulant l has acquired.

Once the simulant DW is computed, the DW attributable to each sequela for the simulant is calculated by using the following formula:

$$ADW_{lk} = \frac{DW_k}{\sum_{k=i}^j DW_k} * Simulant\ DW_l$$

Where:

ADW_{lk} is the attributable DW for disease sequela k in simulant l ,

DW_k is the DW for disease sequela k .

Simulant DW_l is the DW for simulant l from the combination of all sequelae that they have acquired. This formula apportions the overall simulant DW to each condition in proportion to the DW of each condition in isolation.

Finally, YLDs per capita in an age-sex-country-year are computed by taking the sum of the attributable DWs for a disease sequela across simulants.

$$YLD\ Rate_k = \frac{\sum_{l=1}^n ADW_{lk}}{n}$$

The actual number of YLDs from disease sequela k in an age-sex-location-year is then computed as the YLD rate k times the appropriate age-sex-location-year population.

By repeating the simulation process for each age-sex-country-year 250 times, the uncertainty in the prevalence of each disease sequela and the DW is propagated into the final comorbidity corrected YLD

results. We selected 20,000 simulants for each age-sex-location-year group on the basis of simulation testing, which has shown that results are stable for YLDs at this number of simulants even in the younger age groups when prevalence is relatively low. Mean results for YLDs that reflect 10 million simulants (20,000 simulants multiplied by 250 iterations to capture uncertainty) are very stable in each age-sex-location-year. For any given location-year-age-sex group, a cause aggregate prevalence values were calculated as $1 - \prod(1 - \text{prevalence})$.

Section 2.11: YLD computation, uncertainty, and residual YLDs

We computed YLDs by sequela as prevalence multiplied by the DW for the health state associated with that sequela. The uncertainty ranges reported around YLDs incorporate uncertainty in prevalence and uncertainty in the DW. To do this, we take the 250 samples of comorbidity-corrected YLDs and 250 samples of the DW to generate 250 samples of the YLD distribution. We assume no correlation in the uncertainty in prevalence and DWs. The 95% uncertainty interval is reported as the 0.025 and 0.975 quantiles of the distribution. DW draws are not year specific; UIs for YLDs at different points in time for a given disease or sequela are correlated because of the shared uncertainty in the DWs. For this reason, changes in YLDs over time can be significant even if the UIs of the two estimates of YLDs largely overlap. Prevalence UIs are used to determine significance of change in YLDs over time since DW draws are year agnostic.

Section 2.11.1: Residual YLDs

Despite expanding our list of causes and sequelae in successive GBD iterations, many diseases remain for which we do not explicitly estimate disease prevalence and YLDs. Less common diseases and their sequelae were included in 34 residual categories (table 10). For 22 of these residual categories, epidemiological data on incidence or prevalence were available, so these were modelled accordingly. For 13 residual categories, epidemiological data on incidence and prevalence were not available, but sufficient CoD data allowed for CoD estimates. For these residual categories, we estimated YLDs by multiplying their YLL estimates by the ratio of YLDs to YLLs from the Level 3 causes in the same disease category that were explicitly modelled. This scaling was done for each country-sex-year. This approach made the simplifying assumption that the residual diseases caused disability proportionate to the ratio of disability to mortality in explicitly modelled diseases. We did not include causes with large disability but no or little mortality in estimating these ratios. For example, we estimated the YLDs from other neurological disorders from the YLD to YLL ratios for dementia, multiple sclerosis, and Parkinson's disease but did not include the YLDs from headaches and epilepsy in the ratio. Detailed information on how YLDs for residual causes were estimated are available in their respective cause writeups in section 6.

Section 2.12: Birth prevalence

A number of conditions are present at birth and quantifying them is important in fully describing the epidemiology of diseases within populations. These include many conditions included in the GBD cause group of neonatal disorders, infections that are transmitted from mother to child either transplacentally or during birth, and congenital birth defects arising either *de novo* or from maternal exposures. Although these conditions were included in the underlying models informing previous GBD iterations,

we developed a system for reporting them for the first time in GBD 2017; a list of these causes is reported in table 11.

Mathematically (ie, in the models), conditions present at birth are equivalent to “birth prevalence.” However, we report these as “incidence” in recognition of the way that GBD defines incidence as a new case of a disease or injury entering the population. To process these results for publication in GBD, we used a three-step process. First, the number of cases at birth was calculated as birth prevalence rate multiplied by number of live births for each location, sex, and year. Second, the number of cases present at birth were summed with incident cases during the early neonatal period (calculated as the 0-to-6-days incidence rate times the 0-to-6-days population), and the early neonatal incidence rate was recalculated by re-dividing by the 0-to-6-days population. Third, incidence rates for aggregate age groups were re-calculated by using the revised incidence figures for the early neonatal period. Causes included in reporting are all of those for which birth prevalence has been estimated as part of existing modelling processes. Although extensive, this list should not be considered exhaustive of all of the conditions that can be present at birth.

Section 3: SDI

Section 3.1: SDI definition

The Socio-demographic Index (SDI) is a composite indicator of background social and economic conditions that influence health outcomes in each location. In short, it is the geometric mean of 0 to 1 indices of total fertility rate (TFR) for those younger than 25 years old (TFU25), mean education for those 15 years old and older (EDU15+), and lag-distributed income (LDI) per capita. After calculating SDI, values were multiplied by 100 for a scale of 0 to 100.

Section 3.2: Development of revised SDI indicator

SDI was originally constructed for GBD 2015 by using the Human Development Index (HDI) methodology, wherein a 0 to 1 index value was determined for each of the original three covariate inputs (TFR in ages 15 to 49 years, EDU15+, and LDI per capita) by using the observed minima and maxima over the estimation period to set the scales.³²

In response to feedback from collaborators and the evolution of the GBD, we have refined the indicator with each GBD cycle. Beginning in GBD 2017, along with our expanded estimation of age-specific fertility, we replaced TFR with TFU25 as one of the three component indices. The TFU25 provides a better measure of women’s status in society because it focuses on ages at which childbearing disrupts the pursuit of education and entrance into the workforce. In addition, we observed that in highly developed countries, the TFU25 has tended to decline consistently over time despite rebounds in TFR driven by increasing fertility at older ages. The concordance correlation coefficient between SDI based on the GBD 2016 method and the updated method for GBD 2017 was 0.981.

During GBD 2016, we moved from using relative index scales to using absolute scales to enhance the stability of SDI interpretation over time because we noticed that the measure was highly sensitive to the addition of subnational units that tended to stretch the empirical minima and maxima.²¹ We selected the minima and maxima of the scales by examining the relationships each of the inputs had with life

expectancy at birth and under-5 mortality and by identifying points of limiting returns at both high and low values if they occurred before theoretical limits (eg, a TFU25 of 0) were reached.

Thus, for each covariate input, an index score of 0 represents the minimum level of each covariate input past which selected health outcomes can get no worse, and an index score of 1 represents the maximum level of each covariate input past which selected health outcomes cease to improve. As a composite, a location with an SDI of 0 would have a theoretical minimum level of sociodemographic development relevant to these health outcomes, and a location with an SDI of 1 (before multiplying by 100 for reporting) would have a theoretical maximum level of sociodemographic development relevant to these health outcomes.

We computed the index scores underlying SDI as follows:

$$I_{cly} = \max \left(\frac{C_{ly} - C_{low}}{C_{high} - C_{low}}, 0.005 \right)$$

Where:

I_{cly} is the index for covariate C , location l , and year y and is equal to the difference between the value of that covariate in that location-year and the lower bound of the covariate divided by the difference between the upper and lower bounds for that covariate

If the values of input covariates fell outside the upper or lower bounds, they were mapped to the respective upper or lower bounds. We also note that the index value for TFU25 was computed as $1 - I_{TFU25ly}$ because lower TFU25s correspond to higher levels of development and thus higher index scores. For GBD 2023, we expanded the computation of SDI to 925 locations spanning the time period 1950–2023.

The composite SDI is the geometric mean of these three indices for a given location-year. The cut-off values used to determine quintiles for analysis were then computed by using country-level estimates of SDI for the year 2023, excluding countries with populations less than 1 million.

For GBD 2023, final SDI values were multiplied by 100 for reporting, in order to improve understanding of and broader engagement with the values. Final reporting values are on a 0 to 100 scale.

Example calculation

We present the equation used to calculate SDI for a hypothetical country in the year 2023:

$$TFU25 = 1.09; \text{ Mean educ yrs pc} = 8.23; \ln LDI = 9.60$$

$$I_{TFU25} = 1 - \frac{1.09 - 0}{3 - 0} = 0.637$$

$$I_{Educ} = \frac{8.23 - 0}{17 - 0} = 0.484$$

$$I_{lnLDI} = \frac{9.60 - 5.52}{11.00 - 5.52} = 0.744$$

$$SDI = \sqrt[3]{I_{TFU25} \cdot I_{Educ} \cdot I_{lnLDI}} = \sqrt[3]{.637 \cdot .484 \cdot .744} = 0.611$$

$$I_{lnLDI} = \frac{9.58 - 5.52}{11.00 - 5.52} = 0.741$$

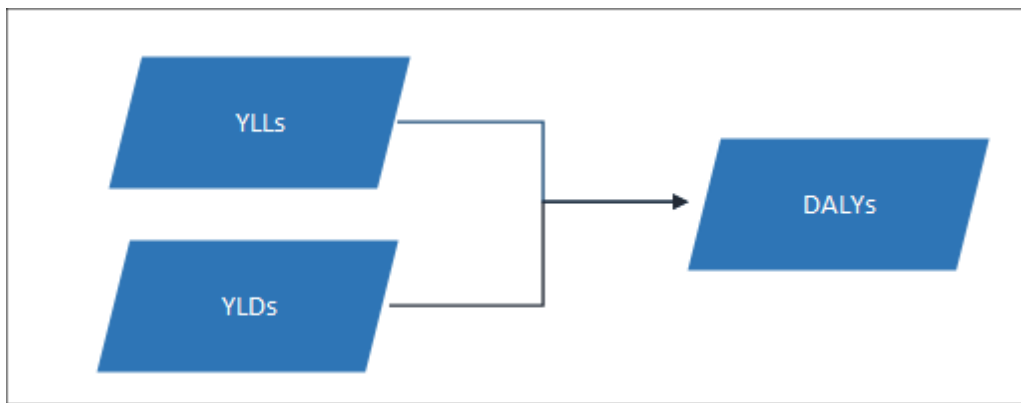
$$SDI = \sqrt[3]{I_{TFU25} \cdot I_{Educ} \cdot I_{lnLDI}} = \sqrt[3]{.855 \cdot .543 \cdot .741} = 0.701$$

GBD 2023 reporting $SDI = 0.701 * 100 = 70.1$

Section 4: Estimation process for DALYs

To estimate DALYs for GBD 2023, we started by estimating cause-specific mortality and non-fatal health loss. For each year for which YLDs have been estimated, we computed DALYs by adding YLLs and YLDs for each age-sex-location (Figure S18). Uncertainty in YLLs was assumed to be independent of uncertainty in YLDs. We calculated 250 draws for DALYs by summing the first draw of the 250 draws for YLLs and YLDs and then repeating for each subsequent draw. 95% UIs were computed by using the 25th and 975th ordered draw of the DALY uncertainty distribution. We calculated DALYs as the sum of YLLs and YLDs for each cause, location, age group, sex, and year.

Figure S18. DALY burden estimation for GBD 2023



Section 5: HALE

The first step to calculating healthy life expectancy for a population (defined by sex, country, and year) was to compute average health of individuals for every age group in that population. We combined information about prevalences for all sequelae and their associated disability weights and accounted for comorbidity with a Monte Carlo simulation approach. We made the assumption that comorbidities were independent within each age group. We created simulations where individuals were exposed to each

sequela with a probability equal to the estimated prevalence of that sequela in each age group. This created a simulated population where the frequencies of many possible multi-morbidities were consistent with the underlying estimates of prevalence. We define 1 minus the disability weight as the positive health associated with each sequela. The combined health for a simulated individual was the product of these positive health values for all relevant sequelae in the presence of multiple sequelae.

Average health values are computed as 1 minus the YLD per person in a population, which are then used to compute health adjusted person years. We incorporated average health values into the life table using Sullivan's method. First, we multiplied values in the nL_x (average person-years lived within an age interval starting at age x) column of the life table by the corresponding average health value in that interval. We recalculated the rest of the life table using the adjusted nL_x values. Sullivan's method began with an adjusted estimate of health adjusted life years within the terminal age interval (equal to nL_x multiplied by the average health value for the terminal age group) and subsequent calculations we produced estimates by iterating through younger age intervals, summing the health-adjusted person-years with all age intervals above the current age interval to generate health adjusted person years lived above a certain age (adjusted T_x) for each age group. After calculating adjusted T_x for all age groups, HALE was calculated by dividing the adjusted T_x for each age group by the proportion of hypothetical birth cohort still alive at age x .³³

References

- 1 Stevens GA, Alkema L, Black RE, *et al.* Guidelines for Accurate and Transparent Health Estimates Reporting: the GATHER statement. *The Lancet* 2016; **388**: e19–23.
- 2 Lim SS, Vos T, Flaxman AD, *et al.* A comparative risk assessment of burden of disease and injury attributable to 67 risk factors and risk factor clusters in 21 regions, 1990-2010: a systematic analysis for the Global Burden of Disease Study 2010. *Lancet* 2012; **380**: 2224–60.
- 3 GBD 2015 Disease and Injury Incidence and Prevalence Collaborators. Global, regional, and national incidence, prevalence, and years lived with disability for 310 diseases and injuries, 1990-2015: a systematic analysis for the Global Burden of Disease Study 2015. *Lancet* 2016; **388**: 1545–602.
- 4 GBD 2016 Disease and Injury Incidence and Prevalence Collaborators. Global, regional, and national incidence, prevalence, and years lived with disability for 328 diseases and injuries for 195 countries, 1990-2016: a systematic analysis for the Global Burden of Disease Study 2016. *Lancet* 2017; **390**: 1211–59.
- 5 GBD 2017 Disease and Injury Incidence and Prevalence Collaborators. Global, regional, and national incidence, prevalence, and years lived with disability for 354 diseases and injuries for 195 countries and territories, 1990-2017: a systematic analysis for the Global Burden of Disease Study 2017. *Lancet* 2018; **392**: 1789–858.
- 6 GBD 2019 Diseases and Injuries Collaborators. Global burden of 369 diseases and injuries in 204 countries and territories, 1990-2019: a systematic analysis for the Global Burden of Disease Study 2019. *Lancet* 2020; **396**: 1204–22.
- 7 GBD 2021 Diseases and Injuries Collaborators. Global incidence, prevalence, years lived with disability (YLDs), disability-adjusted life-years (DALYs), and healthy life expectancy (HALE) for 371 diseases and

- injuries in 204 countries and territories and 811 subnational locations, 1990-2021: a systematic analysis for the Global Burden of Disease Study 2021. *Lancet* 2024; **403**: 2133–61.
- 8 Vasudevan S, Ramos F, Nettleton E, Durrant-Whyte H. Gaussian process modeling of large-scale terrain. *Journal of Field Robotics* 2009; **26**: 812–40.
 - 9 Rasmussen CE. Gaussian Processes in Machine Learning. In: Bousquet O, von Luxburg U, Rätsch G, eds. *Advanced Lectures on Machine Learning: ML Summer Schools 2003, Canberra, Australia, February 2 - 14, 2003, Tübingen, Germany, August 4 - 16, 2003, Revised Lectures*. Berlin, Heidelberg: Springer, 2004: 63–71.
 - 10 Ng M, Fleming T, Robinson M, *et al*. Global, regional, and national prevalence of overweight and obesity in children and adults during 1980-2013: a systematic analysis for the Global Burden of Disease Study 2013. *Lancet* 2014; **384**: 766–81.
 - 11 Ng M, Freeman MK, Fleming TD, *et al*. Smoking prevalence and cigarette consumption in 187 countries, 1980-2012. *JAMA* 2014; **311**: 183–92.
 - 12 Zheng P, Barber R, Sorensen RJD, Murray CJL, Aravkin AY. Trimmed constrained mixed effects models: formulations and algorithms. *J Comput Graph Stat* 2021; **30**: 544–56.
 - 13 Flaxman AD, Vos T, Murray CJL. *An integrative metaregression framework for descriptive epidemiology*. Seattle: University of Washington Press, 2015.
 - 14 Bell A, Jones K. The impossibility of separating age, period and cohort effects. *Soc Sci Med* 2013; **93**: 163–5.
 - 15 Glenn ND. Cohort analysts’ futile quest: statistical attempts to separate age, period and cohort effects. *American Sociological Review* 1976; **41**: 900–4.
 - 16 Goldstein H. Age, period and cohort effects — a confounded confusion. *Journal of Applied Statistics* 1979; published online Jan 1. DOI:10.1080/02664767900000003.
 - 17 Borwein J, Lewis A. *Convex Analysis and Nonlinear Optimization*. New York, NY: Springer, 2006 DOI:10.1007/978-0-387-31256-9.
 - 18 More JJ, Sorensen DC. Newton’s method. Argonne National Lab. (ANL), Argonne, IL (United States), 1982 DOI:10.2172/5326201.
 - 19 Wächter A, Biegler LT. On the implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming. *Math Program* 2006; **106**: 25–57.
 - 20 data_like - dismod_at documentation. https://dismod-at.readthedocs.io/latest/data_like.html (accessed March 11, 2025).
 - 21 avg_integrand - dismod_at documentation. https://dismod-at.readthedocs.io/latest/avg_integrand.html (accessed March 11, 2025).

- 22 Hasin DS, Grant BF. The National Epidemiologic Survey on Alcohol and Related Conditions (NESARC) Waves 1 and 2: review and summary of findings. *Soc Psychiatry Psychiatr Epidemiol* 2015; **50**: 1609–40.
- 23 Henderson S, Andrews G, Hall W. Australia's Mental Health: An Overview of the General Population Survey*. *Aust N Z J Psychiatry* 2000; **34**: 197–205.
- 24 Salomon JA, Haagsma JA, Davis A, *et al.* Disability weights for the Global Burden of Disease 2013 study. *Lancet Glob Health* 2015; **3**: e712–723.
- 25 Campeau L. The Canadian Cardiovascular Society grading of angina pectoris revisited 30 years later. *Can J Cardiol* 2002; **18**: 371–9.
- 26 Dolgin M. Nomenclature and criteria for diagnosis of diseases of the heart and great vessels, 9th ed. Boston: Little, Brown, 1994.
- 27 Nord E. Disability weights in the Global Burden of Disease 2010: unclear meaning and overstatement of international agreement. *Health Policy* 2013; **111**: 99–104.
- 28 Taylor HR, Jonas JB, Keeffe J, *et al.* Disability weights for vision disorders in Global Burden of Disease study. *Lancet* 2013; **381**: 23.
- 29 Voigt K, King NB. Disability weights in the global burden of disease 2010 study: two steps forward, one step back? *Bull World Health Organ* 2014; **92**: 226–8.
- 30 Kretzschmar M, Mangen M-JJ, Pinheiro P, *et al.* New methodology for estimating the burden of infectious diseases in Europe. *PLOS Medicine* 2012; **9**: e1001205.
- 31 Vos T, Flaxman AD, Naghavi M, *et al.* Years lived with disability (YLDs) for 1160 sequelae of 289 diseases and injuries 1990–2010: a systematic analysis for the Global Burden of Disease Study 2010. *Lancet* 2012; **380**: 2163–96.
- 32 GBD 2015 Mortality and Causes of Death Collaborators. Global, regional, and national life expectancy, all-cause mortality, and cause-specific mortality for 249 causes of death, 1980–2015: a systematic analysis for the Global Burden of Disease Study 2015. *Lancet* 2016; **388**: 1459–544.
- 33 GBD 2017 DALYs and HALE Collaborators. Global, regional, and national disability-adjusted life-years (DALYs) for 359 diseases and injuries and healthy life expectancy (HALE) for 195 countries and territories, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017. *Lancet* 2018; **392**: 1859–922.

Section 6: Cardiovascular non-fatal methods

PLEASE NOTE:

For the manuscript, “Global, Regional and National Burden of Cardiovascular Diseases and Risk Factors in 204 countries and territories, 1990-2023,” we have only included modelling methods descriptions for the cardiovascular diseases included in our analysis.

Contents

1. Acute endocarditis	67
2. Acute myocarditis	71
3. Atrial fibrillation and flutter	74
4. Heart failure impairment	84
5. Ischaemic heart disease	97
6. Ischaemic stroke, intracerebral haemorrhage, and subarachnoid haemorrhage	120
7. Lower extremity peripheral arterial disease	161
8. Non-rheumatic valvular heart diseases	170
9. Other cardiovascular and circulatory diseases	179
10. Pulmonary arterial hypertension	181
11. Rheumatic heart disease	186