

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

## Appendix 2: Cause of death estimation methods

### Preamble

This appendix contains methodological details about the Global Burden of Disease cause of death estimation process. Parts of this section have been reproduced from the supplementary appendix of “Global burden of 292 causes of death in 204 countries and territories and 660 subnational locations, 1990–2023: a systematic analysis for the Global Burden of Disease Study 2023,” as cited below:

Naghavi et al. Global burden of 294 causes of death in 204 countries and territories and 660 subnational locations, 1990–2023: a systematic analysis for the Global Burden of Disease Study 2023. *Lancet* (*In press*).

## Preamble

This appendix provides further methodological detail for “Global, Regional and National Burden of Cardiovascular Diseases and Risk Factors in 204 countries and territories, 1990-2023.” This study complies with the Guidelines for Accurate and Transparent Health Estimates Reporting (GATHER) recommendations.<sup>1</sup> 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.

Portions of this appendix have been reproduced or adapted from appendices for GBD 2021 Causes of Death Collaborators,<sup>2</sup> and GBD 2021 Demographics Collaborators.<sup>3</sup> References are provided for reproduced or adapted sections.

## Table of Contents

<b>Preamble</b> .....	2
Section 1: List of appendix tables .....	5
Section 2: GBD overview .....	6
Section 2.1: Geographical locations of the analysis.....	6
Section 2.2: Time period of the analysis.....	6
Section 2.3: GBD cause list .....	6
Section 2.4: Statement of GATHER compliance .....	7
Section 2.5: Data input sources overview .....	7
Section 2.6: Funding sources.....	7
Section 2.7: Abbreviations.....	7
Section 3: GBD 2023 causes of death database .....	8
Section 3.1: CoD data identification <sup>2</sup> .....	8
Section 3.2: Verbal autopsy <sup>2</sup> .....	10
Section 3.3: Standardise input data (step 1) <sup>2</sup> .....	13
Section 3.4: Map to GBD cause list (step 2) <sup>2</sup> .....	15
Section 3.5: Age-sex splitting (step 3) <sup>2</sup> .....	16
Section 3.6: Correction for miscoding of Alzheimer’s and other dementias, Parkinson’s disease, and atrial fibrillation and flutter (step 4) <sup>3</sup> .....	17
Section 3.7: Redistribute (Step 5) <sup>2</sup> .....	19
Section 3.8: Correction for Misclassification of COVID-19 deaths.....	24
Section 3.9: HIV/AIDS misclassification correction (step 6) <sup>2</sup> .....	26
Section 3.9: Scale strata to province (step 7) <sup>2</sup> .....	26
Section 3.10: Restrictions post-redistribution (step 8) <sup>2</sup> .....	27
Section 3.11: Drop VR country-years or mark as non-representative (step 9) <sup>2</sup> .....	27
Section 3.12: Cause aggregation (step 10) <sup>2</sup> .....	27
Section 3.13: Remove shocks and HIV/AIDS maternal adjustments (step 11) <sup>2</sup> .....	28
Section 3.14: Noise reduction (step 12) <sup>2</sup> .....	32
Section 3.15: Cause of death database and outlier identification (step 13) <sup>2</sup> .....	35
Section 3.16: Causes of death data star-rating calculation <sup>2</sup> .....	35
Section 4: Causes of death modelling methods.....	37
Section 4.1: CODEm <sup>2</sup> .....	37
Section 4.2: Causes modelled outside of CODEm <sup>2</sup> .....	40
Section 5: COVID-19 estimation .....	43

Section 6: Central computation <sup>2</sup> .....	45
Section 6.1: Imported cases .....	45
Section 6.2: CoDCorrect .....	45
Section 6.3: Years of life lost calculation .....	46
Section 6.4: GBD world population age standard.....	46
Section 7: References .....	47
Section 8: Tables and figures .....	51
Section 9: CoD cause-specific modelling descriptions .....	57

## Section 1: List of appendix tables

See section 8 for all the appendix tables listed here.

### Appendix tables:

Appendix Table S1: GBD cause hierarchy with levels.

Appendix Table S2: Total number of site years by cause and source type for GBD 2023.

## Section 2: GBD overview

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.

### Section 2.1: Geographical locations of the analysis

We produced estimates for 204 countries and territories that were grouped into 21 regions and seven super-regions (table S1). 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, Sweden, the UK, and the USA. All analyses are at the first level of administrative organisation within each country except for New Zealand (by Māori ethnicity), Sweden (by Stockholm and non-Stockholm), the UK (by local government authorities), and the Philippines (by provinces). To meet data use requirements, in this publication we present subnational estimates for Brazil, Ethiopia, Indonesia, Iran, Japan, Kenya, Mexico, Norway, Pakistan, South Africa, Sweden, the UK, and the USA; given space constraints, these results are presented in Appendix 2 instead of the main text. Additionally, subnational estimates for China, India, Nigeria, and Russia 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 925 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 2.2: Time period of the analysis

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.

### Section 2.3: GBD cause list

The GBD cause and sequelae list for causes of death is organised hierarchically 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 (eg, 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 3499 mutually exclusive and

collectively exhaustive sequelae, 2089 cause sequelae, and 1410 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 286 causes in GBD 2021. The non-fatal cause list has expanded from 364 causes in GBD 2019 to 365 causes in GBD 2023. The total number of fatal and non-fatal causes combined for GBD 2023 is 371. 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 2.4: Statement of GATHER compliance

This study complies with the Guidelines for Accurate and Transparent Health Estimates Reporting (GATHER) recommendations.<sup>1</sup> We have documented the steps in our analytical procedures and detailed the data sources used. See appendix 1 for the GATHER checklist relevant to the Global Burden of Cardiovascular Disease paper. The GATHER recommendations can be found at the GATHER website under [GATHER Statement](#).

## Section 2.5: 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 2021; cause-specific mortality is 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 GBD Sources Tool in the GHDx: <http://ghdx.healthdata.org/gbd-2021/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 data use agreements allow. As required by GATHER, additional metadata for input sources are available through the citation tool as well.

## Section 2.6: Funding sources

This publication and the research it presents was funded by the Gates Foundation (OPP1152504); Queensland Department of Health, Australia; the National Health and Medical Research Council; the Norwegian Institute of Public Health; the New Zealand Ministry of Health, UK Department of Health and Social Care, St. Jude Children's Research Hospital. 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 2.7: Abbreviations

ART	antiretroviral therapy
BTL	basic tabulation list
CDC	United States Centers for Disease Control & Prevention
CHAMPS	Child Health and Mortality Prevention Surveillance
CoD	causes of death
CODem	Cause of Death Ensemble modelling
CRVS	civil registration and vital statistics system
DALYs	disability-adjusted life-years

DHS	Demographic and Health Survey
DisMod-MR	disease model-Bayesian meta-regression
DSP	disease surveillance points
EMR	excess mortality rate
EUREG	European Registry
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
HDSS	Hierarchical Data Storage System
ICD	International Classification of Diseases
IHME	Institute for Health Metrics and Evaluation
INDEPTH	International Network for the Demographic Evaluation of Populations and their Health
InterVA	Interpreting Verbal Autopsy
LMER	linear mixed effects regression
MCCD	Medical Certification of Causes of Death
MITS	minimally invasive tissue sample
MMR	maternal mortality ratio
NORDCAN	database of cancer statistics for the Nordic countries
OPRM	other pandemic-related mortality
PAF	population attributable fraction
PCVA	Physician-Certified Verbal Autopsy
PHMRC	Population Health Metrics Research Consortium
PWC	piecewise continuous
RSME	root mean square error
SD	standard deviation
SRS	Sample Registration System
ST-GPR	spatiotemporal Gaussian process regression
UI	uncertainty interval
UNAIDS	Joint United Nations Programme on HIV and AIDS
UN	United Nations
USA	United States of America
USSR	Union of Soviet Socialist Republics
VA	verbal autopsy
VR	vital registration
WHO	World Health Organization
YLDs	years lived with disability
YLLs	years of life lost

## Section 3: GBD 2023 causes of death database

All available data on causes of death (CoD) are standardised and pooled into a single database used to generate cause-specific mortality estimates by age, sex, year, and geography. Appendix figures 1 and 2 show the high-level view of data inputs, analytical steps, and outputs of the CoD analysis framework. Section 3 of this appendix provides details on each step in the development of the CoD database as illustrated in appendix figure 1.

### Section 3.1: CoD data identification<sup>2</sup>

#### Section 3.1.1: Overview of data types

The CoD database contains eight types of data sources: vital registration (VR), verbal autopsy (VA), cancer registry, police records, sibling history, surveillance, survey/census, and minimally invasive tissue sample (MITS) diagnoses. In countries with VR systems with high completeness and low garbage coding, vital registration is often the primary



source of data for causes of death, but police data, surveillance data, and open-source databases are also used for select causes, such as injuries, violence, and maternal causes of death. Less than half of the world's population has deaths captured in a VR system; therefore, in countries with incomplete VR systems, vital statistics for causes of death may be supplemented with other data types (appendix figure 3).

### **Section 3.1.2: ICD detail**

A majority of the CoD data are VR data obtained from the World Health Organization (WHO) Mortality Database; a compilation of data submitted to WHO by individual countries. VR is also obtained from country-specific mortality databases operated by official offices. Each cause is coded directly to the three- or four-digit ICD-coded cause of death, when possible, whereas cause codes in data tabulated by International Classification of Disease (ICD) are coded to aggregated cause groups. The CoD database contains 2608 country-years of detailed data from 1980 to 2018, which includes underlying CoD coded with three- to four-digit codes by country, year, sex, and age groups. Detailed causes are coded to one of the following ICD-detail coding systems: ICD-8, ICD-9, or ICD-10. Each coding system has a similar cause hierarchy and cause list that has continually developed over time. ICD-10 is the current standard and the most exhaustive cause list. Within the cause lists, five-digit codes, where available, are truncated to four-digit codes to condense the lists. Updates to ICD-detail occur biannually as WHO releases new versions or as country collaborators provide additional data. Updates to data from WHO increasingly include ICD-10 CoD data as it is the most current classification of CoD, while updates to ICD-8 and ICD-9 detailed lists are less common. In the case of overlapping data, preference is given to data from pre-determined country collaborations, which are updated annually.

### **Section 3.1.3: ICD tabulations list**

The ICD tabulation lists include the ICD-8 List A (ICD-8A and ICD-8B), ICD-9 Basic Tabulation List (BTL), ICD-10 Mortality Tabulation, Russia Tabulation; the former Union of Soviet Socialist Republics (USSR) Tabulation, Russia Tabulation from 1989-2001, Russia Tabulation after 2001, and India Medical Certification of Cause of Death (MCCD) and China Disease Surveillance Points (DSP) ICD-9 and ICD-10. These data sources make up 1096 country-years from 1980 to 2017 in the CoD database. All are condensed versions of the ICD-8, ICD-9, and ICD-10 detail lists with some differences in the format of cause lists depending on the data source. ICD-8A, ICD-9 BTL, and ICD-10 Mortality Tabulation CoD are assigned to subtotal groups (referred to as chapters) and cause groups respective to ICD-detail groups. Additionally, ICD-9 BTL includes ICD-9 detail codes for some cancers and a custom tabulation scheme for the former USSR countries. The Russia Tabulation lists, and India MCCD cause lists, each have custom nomenclatures based on ICD-detail cause codes.

Two of the drawbacks in using tabulation lists are discrepancies in the accuracy of death counts and lack of detail due to aggregated cause groups. There are instances where the sum of deaths in chapter subtotals are not equal to the sum of cause groups within the chapter. To account for any missing or duplicate deaths reported within the cause groupings, death counts are systematically adjusted by calculating the differences between subtotals and sub-causes within the cause groups. Any differences are assigned to a remainder cause group. To account for the lack of cause code detail, select cause groups are disaggregated (Step 1.1) to create a complete cause list. Updates to ICD tabulation lists obtained from WHO occur less frequently compared to ICD-detailed lists as more countries are reporting deaths in ICD-detail. In instances of overlapping data, preference is given first to detailed collaborator data, followed by detailed WHO data, then tabulated collaborator data, and finally tabulated WHO data.

### **Section 3.1.4: China Disease Surveillance Points /China Center for Disease Control and Prevention**

The two primary sources of data for China are surveillance data from the China Disease Surveillance Points (DSP) system and VR data collected by the Chinese Center for Disease Control and Prevention (CDC). In the China DSP data, deaths were reported across 145 disease surveillance points used from 1991 to 2003, 161 disease surveillance points from 2004 to 2012, and 605 disease surveillance points from 2013 to 2017. While China DSP with ICD-10 coding is considered sample VR data, it provides national coverage and cause detail. Thus, it receives similar processing and treatment to the China CDC VR from 2008 to 2016. From 2008 to 2017, all of the deaths and CoD information from the DSP system and other system points throughout China were collected and reported via the Mortality Registration and Reporting System, an online reporting system of the Chinese CDC. For these years of DSP, we process using a country-specific cause map. For DSP ICD-9, which contains aggregated causes, we use a tabulated cause map. For DSP ICD-10, we disaggregate causes using the ICD code detail in the CDC data. From

2018 to 2021, the format of the DSP data changed to aggregated causes and eastern, central, and western regions. These four years, we used previous years of formatted DSP data from 2010 to 2017 to split the aggregated causes and regions into ICD10 codes and provinces. The deaths in these data are reported at the strata level, a metric that is specific to China. Counties are stratified by urban and rural classification, but definitions of urbanity vary across counties. In Step 7, we use a method developed to scale up deaths from strata level to the province level.

### **Section 3.1.5: Sample registration system**

Sample registration systems are expanding in several countries and are key sources of data in Indonesia and India. The Sample Registration System (SRS) is a dual-record system wherein a resident part-time enumerator continuously records births and deaths in each household within the sample unit every month. A full-time SRS supervisor thereafter independently collects the vital events along with other related details for each of the preceding six-month periods during the calendar year. Cause of death in SRS systems is recorded using either the detailed ICD-10 coding list, or the INDEPTH ICD-10 short list.<sup>2</sup>

### **Section 3.1.6: India Medical Certification of Cause of Death**

The India MCCD has data for the urban parts of the majority of the states and union territories beginning in 1980. Deaths reported in this data source have been medically certified and are considered VR data. The CoD are reported in a tabulation list with a unique numbering scheme that conforms to ICD-9 and ICD-10 detail codes, which must be disaggregated. MCCD is state-split to fill in data gaps (Step 1.2 state splitting) prior to age-sex splitting. Because SRS is widely considered a more credible assessment of CoD in India, we chose to use MCCD data only in certain cases for modelling with Cause of Death Ensemble modelling (CODEm). We preserved MCCD data in the database for two primary reasons. First, where the three midpoint years of SRS data resulted in the loss of a clear time trend, as was the case for maternal mortality, we chose to preserve MCCD in addition to SRS. Second, MCCD has an advantage over SRS in cases where VA is not a valid instrument for ascertaining CoD, like encephalitis and dengue fever. In these cases, we kept MCCD over SRS.

## **Section 3.2: Verbal autopsy<sup>2</sup>**

### **Section 3.2.1: Verbal autopsy coded to ICD-10 and other lists**

In countries without VR systems, VA studies are a viable data source to inform CoD. Data are obtained by trained interviewers who use a standardised questionnaire to ask relatives about the signs, symptoms, and demographic characteristics of family members deceased within a year. CoD is assigned based on the answers to the questionnaires using a variety of methods, explained in more detailed in section 3.2.2.

VA data are highly heterogeneous: studies use different instruments, different cause lists (from single causes to full ICD cause lists), different methods for assigning CoD, different recall periods, and different age groups. Cultural differences may also affect the interpretation of specific questions. VAs are likely accurate in assigning CoD to road injury or homicide but less accurate for causes requiring medical certification, such as cardiovascular causes. Studies may also occur as standalone assessments or as part of an extended network, such as the International Network for the Demographic Evaluation of Populations and their Health (INDEPTH) Network<sup>2</sup>— a continuous surveillance source with several Demographic Surveillance Systems sites that collect data coded to ICD-detail causes.

### **Section 3.2.2: Methods of ascertaining cause of death from VA questionnaires**

Verbal autopsy (VA) is one of the most common tools used to obtain data on causes of death in countries that lack a functional civil registration and vital statistics (CRVS) system. As such, a wide variety of methods have been developed for ascertaining cause of death based on a VA interview. The historical standard and still one of the most popular methods is Physician-Certified Verbal Autopsy (PCVA), in which two physicians review the interview responses and formulate the most likely cause of death independently. A third physician may arbitrate the result if there is disagreement. This method has some important advantages, including incorporating the knowledge and expertise of local physicians, but is often expensive and slow, includes the variability inherent to human judgement, and has a high opportunity cost in terms of physicians' time spent delivering health care.

In response to these challenges, a variety of automated CoD ascertainment methods have been developed based on algorithmic or probabilistic approaches.<sup>4,5</sup> Two popular automated methods are Tariff 2.0/SmartVA and Interpreting Verbal Autopsy (InterVA). The Tariff method is an algorithm designed and validated by the Population Health Metrics Research Consortium (PHMRC) and IHME. It is based on the association between individual signs and symptoms and causes of death and is trained on the PHMRC gold standard validation study database, a database of VAs collected for deaths that are linked to hospital records in India, the Philippines, Mexico, and Tanzania, where the hospital records provide the true cause of death.<sup>6</sup> The InterVA (Interpreting Verbal Autopsy) method is a Bayesian method, based on the probability of answering yes to a given line item conditional on the true cause of death. These conditional probabilities are determined using a mix of expert opinion and reference data.<sup>7</sup>

A 2014 study by the PHMRC and IHME found that InterVA showed markedly lower performance in predicting causes of death in the PHMRC gold standard validation database when compared to three other automated CoD ascertainment methods, including Tariff.<sup>8</sup> As a result, InterVA modelled VA data are vetted with special care and frequently excluded for causes other than injuries and maternal causes, which are easier for most VA CoD ascertainment methods to diagnose.

Each GBD round, a systematic review of verbal autopsy literature data is conducted to ascertain and include new VA studies using PubMed and Google Scholar. The PubMed search typically returns fewer articles, while Google Scholar returns more articles but tends to be less precise for capturing raw VA data. For GBD 2023, African Journals Online was also included in the literature search with the goal of addressing the dearth of VA sources from African regions, with a review of sources in French.

The search strings for the literature search are below.

Google Scholar search strings:

"verbal autopsy" AND ("cause of death" OR "causes of death" OR "reason for death" OR "reasons for death" OR "cause-specific mortality"), 2021-2021,2022-2022, 2023-2023

PubMed search strings:

("verbal autopsy") AND ("2019/04/24 2021/05/15"[Date - Publication] : "3000"[Date - Publication])

AJOL search strings:

("verbal autopsy")

Inclusion criteria:

- Mentions verbal autopsy
- Mentions causes of death or cause fractions
- References data collection or raw/crude data
- Data must provide a viable, non-biased CoD distribution at an acceptable GBD location level: for example, we do not use global or regional data.

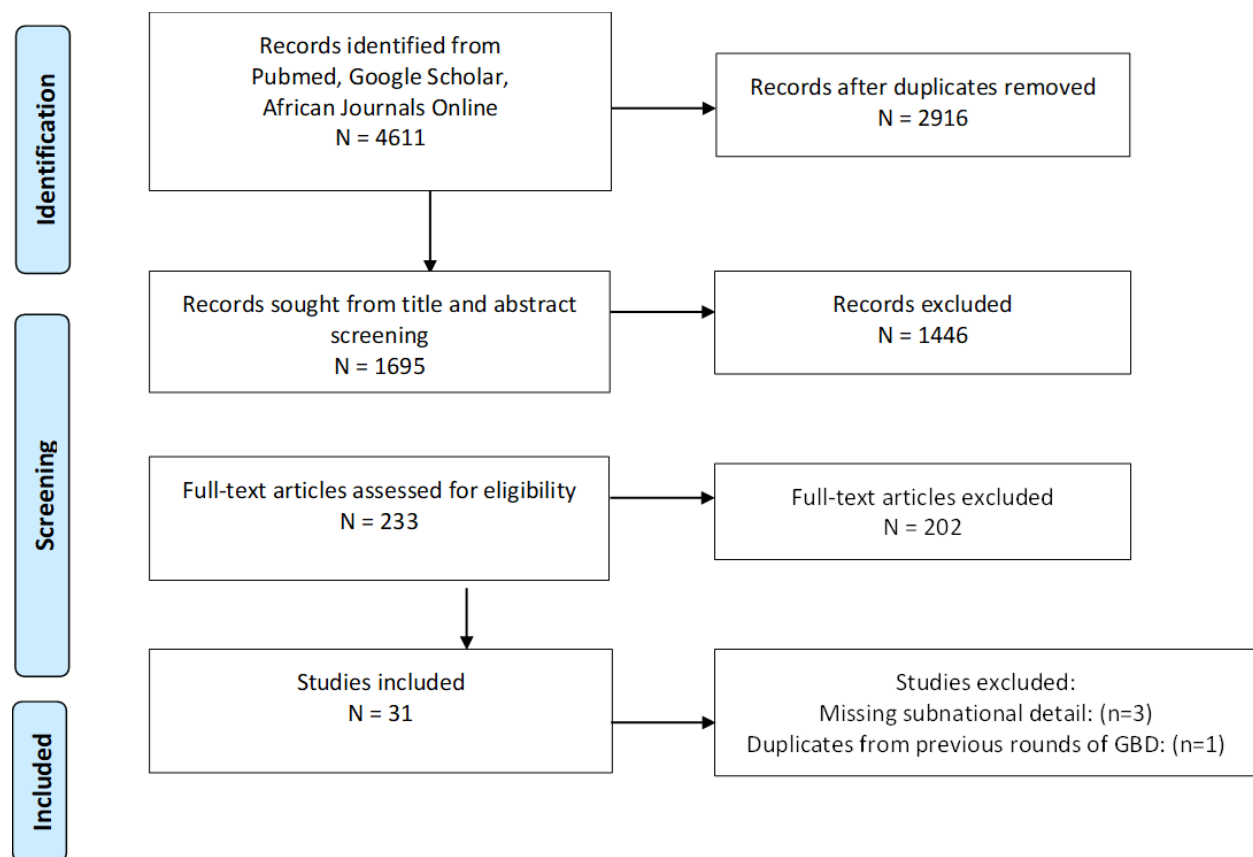
Exclusion criteria:

- Studies already included in previous rounds of GBD
- Non-representative demographics, aka “sub-populations” (eg, HIV-positive people, ICU/NICU patients)
  - This does not include non-representative locations, eg, certain villages, cities, provinces, states, etc. within a GBD location
- Hospital-based sampling
- Methods/theory papers
- Review papers
  - We never extract from review papers, but instead trace the underlying studies and extract the dataset from those
- Papers published by the GBD

- Protocols
- Papers that only report on stillbirths
- Papers that cover exclusively pre-1980
- VA that uses the InterVA assignment method (except for injury and maternal-related deaths)
- Small sample sizes (<50 deaths)
  - Some exceptions have been made for studies from data-sparse countries
- Deaths for one cause without a denominator, for example, only cancer deaths without total deaths
- National level where we model subnationally (eg, South Africa, India, Indonesia)
- Locations reported that are higher than country level (eg, results for several countries together cannot be used)
- Adjusted/modelled estimates (ie, not raw counts) – if the numbers have uncertainty intervals, this is a good sign that they are actually modelled estimates
- Studies where the denominator is not “all deaths due to all causes,” eg, datasets where the denominator is all tuberculosis deaths

Figure A. PRISMA diagram for systematic review of verbal autopsy data included in GBD 2023

PRISMA Flow Diagram



### Section 3.2.3: Other data types

#### Section: 3.2.3.1: Maternal mortality data

In locations with low-quality or no VR, maternal mortality metrics can be found in surveillance, surveys, census, and sibling history data sources. The best data have death counts due to maternal causes and the total number of deaths for

women within the reproductive ages of 10–54 by year. If a data source is missing these components, creating a complete cause list is necessary by using livebirths and all-cause mortality deaths.<sup>3</sup> The livebirths are used to scale down the envelope to match the coverage of the raw data. Though death counts are the preferred metric, maternal mortality is often measured by using the maternal mortality ratio (MMR), which is easily converted to deaths by using livebirths. The China Maternal and Child Surveillance data are adjusted by scaling data from the strata to the province level (Step 7).

#### Section: 3.2.3.2: Surveys and censuses reporting fraction of deaths due to selected injuries

Surveys and censuses are often used in countries with less-developed VR systems; in countries with adequate VR, surveys and censuses are supplementary. Much like VAs, the CoD validity is a concern because of lack of medical certification at the time of death. For these data sources, we keep only causes related to maternal mortality and injuries. The remaining causes are accounted for as a remainder of total deaths in the sample size. One example of these surveys is DHS sibling history. In these surveys, females are interviewed regarding their sisters and survival status. Further questions are asked regarding the deceased siblings such as whether the sibling was pregnant or was within six weeks post-partum at the time of death. From these results, we can discern maternal mortality after applying correction factors including the Gakidou-King weights<sup>9</sup> and removing incidental HIV deaths.

#### Section 3.2.4: Police records

In most countries, police and crime reports are an important source of information for some types of injury deaths, notably road injuries and interpersonal violence. Our police data come from reports on road traffic and crime trends. The police reports used in this analysis were obtained from published studies, national agencies, and institutional surveys such as the United Nations (UN) Crime Trends survey and the UN Office on Drugs and Crime Global Study on Homicides. We assessed whether police reports were likely to be complete and to cover the entire country by comparing police trends with those seen in VR. Data are excluded in instances where police data for road traffic injuries are significantly lower than the VR. Inclusion methods for police data are based on data quality and completeness. For countries with high-quality data, with a star rating of 4 or 5, we utilise police records if the death counts provided are higher than those in the VR. For countries with lower-quality data, with a star rating of 0–3 stars, we always include police records. Police data that meet our inclusion criteria and provide complete coverage are uploaded to the database for use in road injuries and interpersonal violence deaths estimation.

#### Section 3.2.5: Population-based cancer registries

##### Section 3.2.5.1: Cancer registries with incidence

Data on cancer incidence were sought from population-based cancer registries as well as from databases that include multiple registries, including Cancer Incidence in Five Continents, NORDCAN, and EUREG. Cancer registries were identified through the membership list of the International Association of Cancer Registries, through the GBD Collaborator Network, or through publications. Registries were excluded if they were not representative of the coverage population, if the data were limited to years prior to 1980, if the source did not provide details on the population covered, or if the list of cancer types included was not comprehensive for the age group covered. Beginning in GBD 2019, childhood cancer-specific population-based cancer registry data were sought and included.

##### Section 3.2.5.2: Cancer registries with incidence and high-quality mortality data

In addition to incidence, some high-quality cancer registries also report cancer mortality data. These data were also extracted and used as inputs to the mortality-to-incidence model.

### Section 3.3: Standardise input data (step 1)<sup>2</sup>

The input data to the CoD database are received in various formats and must be standardised to run through central CoD machinery to then upload to the database. Raw data inputs come from data sources such as mortality databases, literature reviews, or reports. Usable data sources must have a clear sample size of the number of deaths in the population and ideally, exhaustive cause lists. The complexity of the data cleaning process varies drastically across data sources. For VR microdata with the location, age, sex, year, and ICD-coded cause of every death, very little effort is necessary to standardise it into a consistent structure. Other sources such as VA, sibling history, and surveillance data may require careful review to accurately extract scans of hardcover CoD reports into spreadsheets that can be transformed and standardised based on the standard process described below.

At this point, data are assigned source identifiers so that they can be linked to the GHDx and cited appropriately. Any aggregate age and sex categories are flagged for age-sex splitting. The methods of cause-of-death assignment and data collection are reviewed to determine which source type to assign; for example, we distinguish sibling history data from surveys with a VA module. Only data at the most detailed level of the GBD location hierarchy are used.

Documentation from the source is reviewed to determine if the population is representative of the location or only a subset of the population in that location. Data sources representing a subset of the population are flagged as non-representative; this flag is used by CODEm to increase the variance associated with such datapoints.

Finally, diagnostics are reviewed at this stage to avoid sending cleaning errors downstream. We review cause-specific deaths for each demographic group to ensure the data are reasonable. For example, it is unlikely that male breast cancer deaths are higher than female breast cancer deaths or deaths from neonatal causes occur in age groups over 1 year. All death totals are compared with the sum of cause-specific deaths to ensure the observed deaths are accounted for and sample size is complete.

### **Section 3.3.1: Disaggregation (step 1.1)**

Causes of death in tabulated code systems are specified using tabulated codes, which consist of ranges of detailed codes from the corresponding detailed code systems. These tabulated codes represent groups of related causes, some of which are too general to be mapped directly to estimated GBD causes.

To correct for this, we use a set of age-, sex-, and super-region-specific proportions during data preparation to split tabulated codes into GBD causes and garbage codes that can then be mapped to a GBD cause. These proportions were calculated in GBD 2016 based on detailed VR from code systems ICD-10, ICD-9 detail, and ICD-8 detail. The target causes and garbage codes for a given tabulated code were determined based on the detailed codes that were missing from the tabulated cause list. For any cause and demographic group where we lacked ICD detail, global proportions were used.

As an example, the ICD-10 tabulated list includes a generic tabulated code for “Other intestinal infectious diseases.” In order to make this cause coding compatible with standard GBD cause mapping, it requires splitting onto the detailed causes it represents. To do this, we first define the list of target causes for which the tabulated code will be split. This includes diarrhoea, typhoid fever, and paratyphoid fever in this example. Next, detailed ICD-10 data are used to generate the proportions by which the tabulated code will be split across the target causes. Proportions are generated using all years of available detailed ICD-10 data by super-region, age, and sex. For example, in ICD-10 detail data, in Latin America and the Caribbean, for all-age males, 94.5% of deaths within “Other intestinal infectious diseases” were coded to diarrhoea. Thus, in ICD-10 tabulation data, 94.5% of the “Other intestinal infectious diseases” code were assigned to diarrhoea for countries within this super-region and demographic group.

### **Section 3.3.2: State splitting (step 1.2)**

Two sources for CoD estimation in India are the MCCD report, which reports medically certified deaths from health facilities in mostly urban areas,<sup>10</sup> and the SRS, which collects information via VA about one-half of 1% of the total population in India, including both urban and rural areas, from 8853 sampling units as of 2014.<sup>11</sup> For MCCD, missing data impedes estimation of trends at the state level. We used a first-order, log-linear model of the four-way contingency table of deaths by sex, age, state, and year to estimate the missing state-years. We fit the model to all available data for MCCD separately for each cause, including state-specific all-age measurements and age-specific national measurements. From this, we produced estimates for each combination of sex, age, state, and year. We then used these estimates wherever the raw data did not include sex-specific, age-specific, and state-specific death counts.

For MCCD, the model was fit separately for ICD-10-based and ICD-9-based reports by using the tabulated cause list present in the data.

### **Section 3.3.3: Calculate non-maternal deaths (step 1.3)**

In cases when maternal mortality metrics do not include both deaths due to maternal causes and deaths due to non-maternal causes for women of reproductive age, livebirths and all-cause mortality estimates can be used to calculate deaths, which is the same method used in sibling history data. Many studies report maternal deaths as the MMR.

MMR is the number of maternal deaths per 100 000 livebirths and can be used to calculate deaths when it has been derived from primary data and not estimated. Maternal deaths were calculated by using MMR and livebirths; if livebirths were missing, we substituted livebirth estimates and used the following equation:

$$\text{Maternal deaths} = \frac{\text{MMR}}{100,000} \times \text{Live births}$$

If a study was non-representative, we extracted sample size and livebirths from that study. After maternal deaths were calculated, we used the difference from all-cause mortality estimates to determine non-maternal deaths.

A more accurate and data-inclusive method of calculating maternal and non-maternal deaths incorporates coverage and splits deaths for a range of years into individual years. If there were livebirths in the study, we adjusted the coverage.

$$\text{Coverage} = \frac{\text{Live births}}{\text{GBD estimated live births}}$$

After coverage was calculated, totals deaths were scaled to be more representative. This gives a more accurate death count since the envelope assumes representative coverage. We then calculated non-maternal deaths by using all-cause mortality as an all-cause total.

$$\text{Maternal envelope with coverage} = \text{Maternal envelope} \times \text{Coverage}$$

An additional adjustment can be applied to maternal data spanning over a range of consecutive years, which allows for more data inclusion. The years within specified year ranges are separated into individual years, and total deaths within the year range were split between each individual year by using the fixed proportions of maternal deaths from VR in that particular country. We used only VR data to inform the proportions because it was both high-quality and representative.

### Section 3.4: Map to GBD cause list (step 2)<sup>2</sup>

In GBD 2023, we used 439 maps to translate causes found in the input data to the GBD 2023 cause list. This included 31 maps for VR data, 314 for VA data sources, and 98 for other data types. The largest and most universal maps used were those for ICD-9 and ICD-10 VR data. The input data causes varied from three- to four-digit ICD codes to custom cause lists with cause names such as “cholera” or “hepatitis”. Our mapping process enabled us to compare these various data sources across demographic groups, by mapping all ICD codes to the mutually exclusive and collectively exhaustive list of GBD causes.

A crucial aspect of enhancing the comparability of data for cause of death is to deal with uninformative, so-called garbage codes. Garbage codes are codes to which deaths were assigned that cannot or should not be considered as the underlying cause of death, for example: heart failure, ill-defined cancer site, senility, ill-defined external causes of injuries, and septicaemia. Additionally, any codes not specific enough to assign to a detailed cause of death is considered a garbage code, such as unspecified infectious disease or unspecified injury. In GBD 2019, we developed additional maps to translate ICD codes found in the input data that are non-underlying causes to appropriate target codes based on the levels of the GBD cause list. These garbage codes were mapped to Class 1–4 of the GBD cause list according to the following criteria:

**Class 1** includes all garbage codes for which a Level 1 GBD causes (communicable, maternal, neonatal, and nutritional diseases; injuries; and non-communicable diseases) cannot be directly assigned. For example, the underlying causes of “sepsis” or “peritonitis”, if not specified in the data, could be an injury, a non-communicable disease, or a type of communicable disease. In these cases, deaths will be redistributed across all three of these Level



1 causes. In addition, deaths coded to impossible or ill-defined causes of death (including “senility” and “unspecified causes”) fall into this category, as they will be redistributed onto all causes.

**Class 2** includes all garbage codes that can be assigned to one Level 1 cause in the GBD cause list. This would include deaths coded to “unspecified injuries” (X59), which are redistributed onto all injuries.

**Class 3** includes all garbage codes for which we know the Level 2 CoD and can redistribute onto one Level 3 cause. This includes deaths coded to causes such as “unspecified cardiovascular disease”, which falls within the Level 2 cause “cardiovascular diseases”, as well as those coded to “unspecified cancer site”, which falls within the Level 2 cause “neoplasms”.

**Class 4** includes all garbage codes for underlying causes of death that can be redistributed within a Level 4 cause. This includes garbage codes such as “unspecified stroke” or “unspecified road injuries.”

### Section 3.5: Age-sex splitting (step 3)<sup>2</sup>

Different sources, particularly VA studies, report deaths for a wide range of age groups with varying intervals. For the analysis of CoD, we mapped these different age intervals to the GBD standard set of age groups. While the proportions informing this mapping were updated in GBD 2023, the approach to undertake this mapping was the same as in the prior GBD studies (GBD 2021, GBD 2019, GBD 2017, GBD 2016, GBD 2015, GBD 2013, and GBD 2010).

In the process of assembling a consolidated demographic database, we found that the aggregation of age groups is perhaps the strongest source of inconsistency. By convention, such data are reported in broad age groupings such as 0–4, 5–14, and 15–49, or with both sexes together. The issue of comparability between age-sex groups arose when assembling the GBD CoD database. We developed a tool called age-sex splitting that takes aggregated age groupings and the “both sexes combined” grouping and divides them into what their constituent age groups would likely have been if respective cause-specific and country-specific age distributions had been used. The analytical framework for GBD includes six age categories for infants and children under 5: early neonatal (0–6 days), late neonatal (7–27 days), 1–5 months, 6–11 months, 12–23 months, and 2–4 years, and 19 age categories for those over 5: 5–9 years, 10–14 years, and so forth, proceeding in five-year age groups until the terminal age group of 95 years and older. We treat unknown ages and sexes in the same manner we treated the “all ages combined” age category and “both sexes combined” sex group. Through this process, we were able to directly compare all data sources on even terms.

The approach to age splitting is based on the following formula. The key assumption underlying this formula is that the relative risk of death by age group compared to a reference age group is invariant across populations. Although this assumption is likely violated in specific cases, a strong biologically based pattern of the relative risk of death for a cause by age is observed for most causes. The basic formula is as follows:

$$D_a = R_a N_a \left( \frac{D_a^{a+x}}{\sum_a^{a+x} (R_a N_a)} \right)$$

Where:

$D_a$  = the number of deaths from a cause in age group  $a$

$R_a$  = global cause-specific mortality rate of age group  $a$

$N_a$  = the country-year-sex-specific population in age group  $a$

$D_a^{a+x}$  = the number of deaths in the age group  $a$  to  $a+x$

With the assumption of invariant relative risks of death by age with respect to a reference age group, this equation can be used, along with population distribution by age, to split an aggregate number of deaths for the age groups  $a$  to  $a+x$



into specific deaths for each age group within the aggregate interval.

$$D_{as} = R_{as} N_{as} \left( \frac{D_{as}^{a+x,s}}{\sum_{a+x} (R_{as} N_{as})} \right)$$

Where:

$D_{as}$  = the number of deaths from a cause in age group  $a$ , sex  $s$

$R_{as}$  = global cause-specific mortality rate of age group  $a$ , sex  $s$

$N_{as}$  = the country-year-sex-specific population in age group  $a$  for sex  $s$

$D_{a,s}^{a+x,s}$  = the number of deaths in the age group  $a$  to  $a+x$  for sex  $s$

In some cases, deaths are reported for an aggregate age group for both sexes combined. The task in this case is more complicated, but the same principle can be applied. In this case, we assumed that the relative risks of death by age and sex are constant.

This equation can be used to split data aggregated by age and sex. The assumption, however, of invariant relative risks across age and sex is a stronger assumption. Fortunately, data pooled across sexes are less common in the published or unpublished CoD data.

The relative risk of death in a particular age group for a given sex is derived from the global distribution of cause-specific mortality rates found in available VR data. Location-years from the following code systems are used, provided they report the requisite age detail and sex detail: ICD-7, ICD-8, ICD-9 BTL, ICD-10 tabulated, ICD-9, and ICD-10. Upon compiling these data, we mapped them to GBD causes and retained any observation coded to either cause Level 3 or a most detailed cause. Cause aggregation was then performed to ensure accurate death totals at all levels of the cause hierarchy. Thus, age distributions were generated for most detailed causes, as well as all Level 1, 2, and 3 causes. In the event that a cause does not have a distribution, it will be split using the distribution of its nearest parent cause.

We next adjusted separately for estimated adult and child VR completeness. Location-year-age-sex-cause-specific deaths and population were then aggregated across all location-years to produce cause-specific mortality rates by age and sex. These were used to determine the risk of death at any age relative to any reference age group, as shown in the above equations.

### Section 3.5.1: Remap unmodelled age-sex combinations

Occasionally, data sources include deaths by a cause for which consensus exists that death is rare for the sex and/or age. For example, some number of deaths may be attributed to cervical cancer in males, or to maternal causes in children younger than 10 years. We have constructed a list of age-sex-cause combinations we do not model. Some sex/cause combinations cannot be modelled because the input data for causes of death is limited to binary male and female sex groups and thus constrains the combinations of cause and gender that can be modelled for the GBD. Because of these limitations to our input data, when we encounter a death in an unmodelled age-sex-cause combination, cause/sex combinations occur, we reassign them to a related cause or garbage package. For example, IBD may be remapped to other digestive disorders.

## Section 3.6: Correction for miscoding of Alzheimer's and other dementias, Parkinson's disease, and atrial fibrillation and flutter (step 4)<sup>3</sup>

### Section 3.6.1: Objective

For certain causes of death, mortality rates reported in VR systems are impossible to reconcile with observed trends in disease prevalence and excess mortality. For dementia,<sup>12–20</sup> Parkinson’s disease,<sup>21,22</sup> and atrial fibrillation and flutter, these disparities can largely be attributed to death certification practices which lead to under-coding or over-coding in many country-years. We sought to address the known bias in CoD data by first identifying the proportion of all deaths that should be assigned to these causes and next determining the GBD causes and garbage groups to which these deaths are being incorrectly assigned.

In past GBD iterations, we estimated Alzheimer’s disease and other dementias, Parkinson’s disease, and atrial fibrillation and flutter on the basis of longitudinal prevalence and excess mortality data to help account for changing patterns in death certification and corresponding implausible time trends in many VR sources. This method was first implemented for Alzheimer’s disease and other dementias in GBD 2013. We added atrial fibrillation and flutter to the causes modelled in GBD 2015 and Parkinson’s disease to the causes modelled in GBD 2016 by using this strategy. All of these causes were processed in CoDCorrect in a manner that was agnostic to the likely targets of misclassification, which inappropriately led to changes in mortality estimates for causes unrelated to these three in GBD 2015. For GBD 2016, we improved this process by completing a literature review to identify the causes of death most closely associated with Parkinson’s and Alzheimer’s diseases<sup>3,10,11,17</sup> and limiting the CoDCorrect adjustments to include only those causes. For GBD 2017, we refined this approach further by using multiple CoD data to determine the GBD causes and garbage codes from which we move deaths as well as the pattern of misclassification.

### Section 3.6.2: Correction process

Changes in coding practices for Alzheimer’s diseases and other dementias, Parkinson’s disease, and atrial fibrillation and flutter result in spatiotemporal mortality trends that are incompatible with prevalence and case-fatality trends.<sup>23</sup> These changes in coding practices are believed to be the result of shifting consensus in cause of death certification, meaning there is a bias in vital registration (VR) data that needs correction. For Parkinson’s disease and atrial fibrillation and flutter, we first estimated excess mortality from prevalence and CoD data in countries with the highest ratio of cause-specific mortality to prevalence, which represents the greatest willingness to code to an under-coded cause. Then, using DisMod-MR 2.1 (see Section 4.2.3), we derived estimates of cause-specific mortality rates from available prevalence surveys as well as the estimates of excess mortality rate, applied across all countries and over time. We divide this value by the all-cause mortality rate to determine the fraction of overall mortality to attribute to each under-coded cause. For dementia, the modelling process was redesigned in 2019 to no longer depend on vital registration data from the highest dementia mortality locations. Instead, we used relative risk data from cohort studies to calculate total number of excess deaths due to dementia, and end-stage disease proportions from linked hospital to death records to subset these deaths to the proportion of excess deaths with end-stage conditions, which we attributed to dementia. Finally, we used log-linear interpolation to interpolate final estimates of death due to dementia for the entire time series and saved as a custom CoD model.

To ascertain the causes from which we would move deaths to under-coded causes, we leveraged multiple CoD data from the USA—by looking to the combinations of intermediate and immediate causes (ie, chain causes) present on death certificates with an under-coded cause listed as underlying and identifying other death certificates with similar or identical chain causes, we can determine the expected pattern of miscoded deaths.

The first stage in this process is to parse out country-years of data where we believe coding practices to be relatively stable. For dementia, this “gold standard” dataset features USA 2010–2015; for Parkinson’s, USA 2005–2015; and for atrial fibrillation and flutter, USA 2014–2015. We then collect all deaths in those years with the under-coded cause listed as underlying and remove any mention of the under-coded cause from the death certificate. Next, for each unique chain identified in this manner, we search the entire time series of USA data (1980–2015) to identify the distribution of underlying causes that share that chain. The premise here is that if the diagnosis of dementia, Parkinson’s, or atrial fibrillation and flutter were missed, the other causes listed on the death certificate would have been the basis for certification. We then reallocate the under-coded cause deaths in the gold standard years by chain based on that alternative underlying cause distributions from the full time series.

Upon iterating through all unique chains, we are left with a counterfactual dataset excluding under-coded causes of death. Each remaining cause can be subdivided into correctly coded deaths and deaths that have been recoded from

an under-coded cause by the process described (although not all causes are necessarily targeted by the recoding algorithm). We then calculate the ratio of recoded deaths to total deaths by cause, age, and sex in our counterfactual dataset. This ratio represents the proportion of each cause that we believe to be miscoded Parkinson’s disease, Alzheimer’s disease and other dementia, or atrial fibrillation and flutter in a counterfactual scenario of 100% under-reporting for these causes.

We multiply the ratios derived from the multiple cause data by the cause-specific deaths in each VR dataset to determine the local pattern of miscoding. In this way, the method is sensitive to the observed epidemiology of a given country and year. Then, we calculate the deficit in under-coded cause mortality for each location, year, age, and sex by taking the difference in the expected cause fraction for a given under-coded cause based on prevalence and excess mortality compared to the proportion of deaths actually certified by the VR system. Finally, we scale the cause-specific miscoded deaths to match the deficit and then move them accordingly. We assumed that misclassification of actual dementia and Parkinson’s deaths in past years occurred only for reported causes of deaths that were plausibly the direct result of dementia or resulted from misdiagnosis of other organic brain diseases based on clinical expert judgement. A similar assumption is used for atrial fibrillation and flutter, for which only cardiovascular causes and ill-defined garbage codes are considered. Finally, while we assumed that Parkinson’s disease and atrial fibrillation and flutter are strictly under-coded in the VR, we assumed that Alzheimer’s disease and other dementias can be under- or over-coded. Anywhere where our expected cause fraction for Alzheimer’s disease and other dementia was lower than the observed, we removed deaths from this cause and redistributed them as non-specific garbage.

Because the deaths being reallocated vary by location-year, we need a mechanism to ensure plausible limits to how many deaths are extracted from each GBD cause and garbage code. To achieve this, we first run the above-mentioned algorithm on all 5-star VR data (see Section 3.16 for an explanation of the star data quality rating system). Then, we determine the 95th percentile of the proportion of deaths moved to an under-coded cause for each GBD cause and garbage code group by age and sex across location-years among these data. Those values are subsequently stored and applied as the limits for deaths moved by this process for all other VR data.

## Section 3.7: Redistribute (Step 5)<sup>2</sup>

A crucial aspect of enhancing the comparability of data for CoD is to deal with uninformative, so-called garbage codes. Garbage codes, ie, codes that are not specific enough, are an immediate or intermediate CoD, or impossible CoD, should not be considered as the underlying cause of death—for example: heart failure, ill-defined cancer site, senility, ill-defined external causes of injuries, and septicaemia. The methods for redistributing these garbage-coded deaths are outlined in detail in Johnson et al,<sup>24</sup> and the primary algorithm for redistributing deaths assigned to these codes has not changed since GBD 2013.

### Section 3.7.1: Redistribute HIV-related garbage codes (step 5.1)

Because of the disparate nature of HIV/AIDS mortality across space and time, dynamic redistribution of HIV/AIDS-related garbage codes was needed. To inform this redistribution, we generated target proportions for each garbage group by age band (under 1 month, 1–59 months, 5–19 years, 20–49 years, 50–59 years, 60–69 years, 70–79 years, and 80 years and older), five-year time interval, and sex. The garbage groups either target HIV or a remainder target. The allotment of deaths to either of these is based on the regional increase in the mortality rate of all codes in the group relative to the rates seen from 1980 to 1984—an increase greater than 5% is assumed to be HIV/AIDS-related, and the proportion of those deaths exceeding 5% are redistributed to HIV/AIDS. Any increase less than or equal to 5% is then assigned to the remainder target.

### Section 3.7.2: Regress garbage codes versus non-garbage codes (step 5.2)

For each redistribution package, we defined the “universe” of data as all deaths coded to either the package’s garbage codes or the package’s redistribution targets for each country, year, age, and sex. We then ran a regression based on the following equation separately for each target group and sex:

$$TG_{crt} = \alpha + \beta_1 Gar_{crt} + \beta_2 Age_{crt} Gar_{crt} + \theta_r Gar_{crt} + \gamma_r + \varepsilon_{ct}$$

Where:

$TG_{crt}$  = percentage of deaths within the given garbage code's universe that were coded to a given target group, by country

$Gar_{crt}$  = percentage of deaths within the given garbage code's universe that were coded to a given set of garbage codes

$Age_{crt}$  = age interaction term for the fixed effect on the interaction of garbage and age

$\alpha$  = constant

$\beta_1$  = slope coefficient describing the association between  $Gar_{crt}$  and  $TG_{crt}$

$\beta_2$  = slope coefficient describing the association between the interaction  $Age_{crt}Gar_{crt}$  and  $G_{crt}$

$\gamma_r$  = region-specific random intercept (or super-region if the random effect on region is not significant)

$\theta_r$  = region-specific random slope (or super-region if the random effect on region is not significant)

$\varepsilon_{ct}$  = standard error, normally distributed and calculated by bootstrapping

This regression was adjusted from GBD 2013 to include fixed effects on the interaction of garbage and age to ensure smooth age patterns. We made this decision after investigating diagnostic visualisations that showed unlikely gaps between proportions assigned to different age groups.

Once proportions were produced for each country, sex, age, and target group, certain adjustments were made to conform our packages to the best medical evidence available. In some cases, we implemented restrictions on the proportions that the regressions could yield. For example, we did not allow any redistribution onto “Chagas disease” outside of Latin America and the Caribbean or “self-harm” under the age of 15 years. In other cases, we capped the proportion for some targets to the level that would be produced from proportional redistribution; for example, “haemoglobinopathy” and “haemolytic anaemia” were restricted to the level of proportional redistribution in the redistribution of “left heart failure”. Occasionally, further adjustments were made on a case-by-case basis per country, age, sex, and target group to suppress the impact of outliers based on existing epidemiological evidence and expert judgment.

In GBD 2019, we updated the regressions for stroke and diabetes. We dropped the proportion of garbage from the regression formula and ran regression on high-quality, low proportion garbage data (4/5 stars, <50% GC). We also included all covariates included in the CODEm models for both stroke and diabetes.

### **Section 3.7.3: Development of an algorithm for redistribution of garbage codes based on multiple CoD data**

Multiple CoD data are a form of individual record causes of death data that include an underlying CoD along with other causes in the death chain, including intermediate and immediate causes. By analysing this type of data, we can sometimes find the true underlying CoD in other CoD data where the underlying cause is a garbage code or a mis-assigned CoD.

As of GBD 2019, this method has been expanded and used in redistribution of the following intermediate causes: sepsis, embolism (pulmonary and arterial), heart failure (left, right, and unspecified), acute kidney injury, hepatic failure, acute respiratory failure, pneumonitis, and unspecified central nervous system disorders. Using multiple CoD records for the USA, Mexico, Brazil, Taiwan (province of China), Italy, Canada, New Zealand, Austria, South

Africa, and Colombia, we identified the fraction of deaths where the underlying cause of death and the intermediate cause was in the causal chain. Using a mixed effect linear regression, we estimated the fraction of intermediate-cause-related deaths by underlying GBD cause. These fractions were multiplied by the GBD 2023 CoDCorrect result to calculate the number of deaths intermediate cause-related deaths for each GBD cause. Lastly, we calculated the “intermediate cause fraction”, with total intermediate-cause-related deaths as the denominator, by age, sex, location, year, and GBD cause. These fractions were used to redistribute the intermediate-cause-related deaths to a GBD cause. An example is given below for sepsis, where  $a, s, l, y, c$  denotes a given age group, sex, location, year, and underlying cause of death:

1.  $sepsis\ fraction = \beta_{HAQ\ Index} + \beta_{age\ group} + \beta_{sex} + Y_{cause} + \varepsilon$
2.  $sepsis\ deaths_{a,s,l,y,c} = sepsis\ fraction_{a,s,l,y,c} * GBD\ deaths_{a,s,l,y,c}$
3.  $total\ sepsis\ deaths_{a,s,l,y} = \sum_c sepsis\ deaths_{a,s,l,y,c}$
4.  $fraction\ of\ sepsis\ to\ redistribute_{a,s,l,y} = \frac{sepsis\ deaths_{a,s,l,y,c}}{total\ sepsis\ deaths_{a,s,l,y}}$

To redistribute X59 and Y34 (unspecified injuries) deaths, we used a multi-step approach that utilised the pattern of nature of injury codes in the causal chain in the multiple CoD data. First, we looked at deaths where X59, Y34, and GBD injuries causes were the underlying cause of death and got the pattern of nature of injury codes in the chain. We then derived a cause-specific redistribution proportion based on the probability of a given pattern being coded to X59/Y34 or a GBD injuries cause and summing up these proportions for all patterns. An example below is given for X59:

5.  $P_{(pattern_j|UCoD\ X59)} = \frac{\#\ of\ pattern_j\ deaths\ |UCoD\ X59}{\sum_{j=0}^m (\#\ of\ pattern_j\ deaths\ |UCoD\ X59)}$
6.  $P_{(GBD\ injuries\ cause_i|pattern_j)} = \frac{\#\ of\ UCoD\ GBD\ injuries\ cause_i\ deaths\ |pattern_j}{\sum_{i=0}^n (\#\ of\ UCoD\ GBD\ injuries\ cause_i\ deaths\ |pattern_j)}$
7.  $redistribution\ proportion_{GBD\ injuries\ cause_i} = \sum_{j=0}^m (P(pattern_j|UCoDX59) * P(GBD\ injuries\ cause_i|pattern_j))$

Where:

$pattern_j$  = a given nature of injury code pattern in the chain of the multiple CoD data

UCoD X59 = a death with X59 coded as the underlying cause of death (UCoD)

UCoD GBD injuries cause<sub>i</sub> = a death with a GBD injuries causes coded as the UCoD

We applied these cause-specific redistribution proportions on the data where X59/Y34 were the underlying cause of death to get the number of X59/Y34 deaths “attributable” to each GBD injuries cause. Then, for each GBD injuries cause in the multiple CoD data, we calculated the fraction of redistributed X59/Y34 deaths over the fraction of total injuries death for that cause and modelled this intermediate cause fraction using a mixed effects linear regression similar to the one mentioned above. Like mentioned above, these fractions were then multiplied by GBD 2023 CoDCorrect results, and the cause fractions for X59 and Y34 were calculated by age, sex, location, year, and GBD injuries cause, and then used to redistribute X59 and Y34 deaths to GBD injuries causes.

Additionally, multiple CoD data were used in the correction of the mis-assignment of deaths due to drug overdoses

to unintentional other poisoning. More than 90% of these types of poisonings are due to exposure to narcotics, psychodysleptics, and other drugs, specified or unspecified. More than 97% of these poisonings by substance or drug occurred in ages 15–65 years. These are clearly not cases of accidental ingestion of substances but rather deliberate ingestion and unintentional poisoning. Using multiple CoD records from the USA, Mexico, Brazil, Taiwan (province of China), Italy, Colombia, Australia, and various European countries from 1980 to 2017, we selected all deaths with underlying causes coded to X40–X44 (Table A below). Table B shows the combination of other potential causes that can be found in the multiple CoD data for these underlying causes, and table A shows the ICD-10 codes corresponding to these causes. On the basis of Table B, we proportionally redistributed mis-assigned unintentional poisoning deaths to one of these causes. The main assumption behind this algorithm is the predominance of the fatality of some substances when a combination of drugs is considered. Given the combination of different drugs and substances in these codes, opium is the main cause of fatality.<sup>25,26</sup> Other substances, like cocaine, methamphetamine, and alcohol in combination with cannabis are less likely to be dominant in fatality.<sup>27</sup>

For example, if the multiple CoD data show that 40% of deaths include opioid use disorders as an intermediate cause where the underlying cause is X40–X44, the redistribution proportion for opioid use disorders will be exactly 40% due to the dominance of the fatality of opioid use disorders compared to other drugs in the above table. Additionally, in our final results, cannabis and psychoactive and psychedelic drug use disorder deaths were mapped to other drug use disorders.

**Table A. ICD-10 codes for substances or drugs used to assign deaths coded to an underlying cause of unintentional poisoning by using multiple CoD data**

Accidental poisoning codes	All X40, X41, X42, X43, X44 codes
Opioid codes	T40.0, T40.1, T40.2, T40.3, T40.4, T40.6, F11.0, F11.1, F11.2, F11.3, F11.4, F11.5, F11.6, F11.7, F11.8, F11.9
Amphetamine codes	T43.6, F15.0, F15.1, F15.2, F15.3, F15.4, F15.5, F15.6, F15.7, F15.8, F15.9
Cocaine codes	T40.5, F14.0, F14.1, F14.2, F14.3, F14.4, F14.5, F14.6, F14.7, F14.8, F14.9
Psychoactive and psychedelic drug codes	T40.8, T40.9, F16.0, F16.1, F16.2, F16.3, F16.4, F16.5, F16.6, F16.7, F16.8, F16.9
Alcohol codes	T51.0, T51.1, T51.2, T51.3, T51.8, T51.9, F10.0, F10.1, F10.2, F10.3, F10.4, F10.5, F10.6, F10.7, F10.8, F10.9
Cannabis codes	T40.7, F12.0, F12.1, F12.2, F12.3, F12.4, F12.5, F12.6, F12.7, F12.8, F12.9

**Table B. Multiple cause of death selection algorithm used for redistributing unintentional poisoning causes of death to substance or drug use cause of death**

Selection algorithm						
	Opioids	Cannabis	Cocaine	Amphetamines	Alcohol	Psychoactive and psychedelic drugs
Opioids	Opioids	Opioids	Opioids	Opioids	Opioids	Opioids
Cannabis	Opioids	Cannabis	Cocaine	Amphetamines	Alcohol	Psychoactive and psychedelic drugs
Cocaine	Opioids	Cocaine	Cocaine	Amphetamines + cocaine	Cocaine + alcohol	Cocaine

Amphetamines	Opioids	Amphetamines	Amphetamines + cocaine	Amphetamines	Amphetamines + alcohol	Amphetamines
Alcohol	Opioids	Alcohol	Cocaine + alcohol	Amphetamines + alcohol	Alcohol	Psychoactive and psychedelic drugs
Psychoactive and psychedelic drugs	Opioids	Psychoactive and psychedelic drugs	Cocaine	Amphetamines	Psychoactive and psychedelic drugs	Psychoactive and psychedelic drugs

Multiple CoD data were only available to us for the USA, Mexico, Brazil, Taiwan (province of China), Italy, Colombia, Australia, and various European countries. Because of this limited sample, we applied the result from the multiple CoD analysis from each country to its respective super-region and used global proportions for sub-Saharan Africa. We hope for increased availability of multiple CoD data in future analyses to achieve a more precise distribution for more locations.

#### Section 3.7.4: Verbal autopsy anaemia adjustment (step 5.3)

To compensate for the over-representative cause fractions from anaemia found in VA studies, we redistributed these deaths based on the causal attribution of severe anaemia from GBD 2023. The proportions were country-year-age-sex-specific.

#### Section 3.7.5: Calculate redistribution uncertainty (step 5.4)

We categorised garbage codes into four levels in order of increasing specificity (see Section 3.4). Some garbage codes are redistributed on all causes (eg, unspecified causes of death) and others are only redistributed onto specific causes (eg, unspecified cancer). Major garbage refers to garbage codes in Levels 1 or 2. Because of the variation in redistribution, estimating uncertainty from garbage redistribution for CODEm modelling was an important goal for GBD 2019.

We assigned redistribution variance to each datapoint in the CoD database by calculating residual variance from a regression predicting the percentage of garbage-coded deaths redistributed to a cause, given the proportion of garbage codes we observed for that location, year, age, sex, cause, and the age-standardised relative rate of major garbage codes across all causes. If there is a cause that has greater residual variance, we assume greater redistribution uncertainty.

The two model inputs are the observed percentage of Levels 1, 2, and 3 garbage codes (by cause, age, sex, location, and year) in redistributed CoD data and the percentage of garbage codes in the raw data (calculated as the age-standardised mortality rate ratio of major garbage-coded deaths to all deaths in the raw data by location, year, and sex). Level 4 garbage codes were excluded from the model to avoid overestimating uncertainty in countries with high percentages of major garbage codes. Additionally, the classification of Level 4 garbage codes is not stable between successive GBD rounds—for example, “unspecified diabetes” was not a garbage code in GBD 2016, and in GBD 2017 was reclassified as a Level 4 garbage code to permit estimation of diabetes by type. These deaths are still taken into account later in the uncertainty estimation process. The model predicts the percentage of garbage-coded deaths redistributed to a cause, given the proportion of garbage codes we observed for that location, year, age, sex, cause, and the age-standardised relative rate of major garbage codes across all causes. From this model, we calculate residual variance. It is important to note that the variance here is a measurement of uncertainty of redistribution, not of the level of miscoding in the raw CoD data for a given demographic.

To calculate variance, a dataset was generated that contained percentage garbage by location, year, age, sex, and cause, where percentage garbage is determined by the equation:

$$pct_{garbage} = \frac{deaths_{redistributed} - deaths_{raw}}{deaths_{redistributed}}$$



A mixed-effect linear regression model was then fit to predict the logit percentage of deaths from redistribution by age-standardised relative rate of major garbage codes.

$$\begin{aligned} \text{logit}(pct_{garbage_{ij}}) \\ = \beta_0 + \beta_1 * \log(ASR_{majorgarbage_{ij}}) + \beta_2 * 15yearage_{ij} + \gamma_{1j} * \log(ASR_{majorgarbage_{ij}}) + u_j \\ + e_{ij}, \theta_{\{i\}} \sim N(0, \sigma^2) \end{aligned}$$

Where:

$i$  indexes dataset-location-year-age-sex-cause datapoints nested within  $j$  groups by GBD region

$ASR_{majorgarbage_{ij}}$  is age-standardised relative rate of major garbage

Residual variance, as estimated by the mean absolute deviation, was calculated for each cause, sex, and age.

The next step was to use the residual variance to calculate uncertainty around each datapoint in the CoD database. First, we calculated the percentage garbage of each datapoint by treating all deaths that could not be directly mapped to a GBD cause as garbage, including Level 4 garbage codes. Percentage garbage was calculated as

$$pct_{garbage} = \frac{deaths_{redistributed} - deaths_{corrected}}{deaths_{corrected}}$$

Where:

$deaths_{corrected}$ : deaths post misdiagnosis correction (Section 3.6)

$deaths_{redistributed}$ : deaths post redistribution (Section 3.7)

Residual variance was matched to each datapoint, and 100 draws were sampled from a normal distribution by using the cause-age-sex-specific residual variance and mean of 0. The logit transformed percentage garbage was added to each value in the distribution. Each draw was then transformed out of logit space, and the post-redistribution deaths were calculated as:

$$deaths = \frac{deaths_{corrected}}{1 - pct\_garbage}$$

Draws of deaths were processed through noise reduction before calculating the final redistribution variance passed to CODEm, which was added to the total data variance. The mean of the draws was not used as the final estimate because it was found that the logit transformation biased the distribution of cause fractions higher. Instead, only point estimates were used.

### Section 3.8: Correction for Misclassification of COVID-19 deaths

In GBD 2023, we received 83 Country years of data in 2020, 67 in 2021, and 38 in 2022. During these years, there is evidence that COVID-19 deaths were misclassified as other causes of death.<sup>28,29</sup> These misclassified COVID-19 deaths then manifest as spikes in the other causes of death that they are miscoded to. For identifying miscoded deaths we looked at 3 age bands, under 15, 15 to 54, and 55+ as well as limiting our analysis to the country level and level 3 causes. To systematically identify these spikes caused by misscoded COVID-19 in other causes, we developed a Support Vector Machine (SVM) that identified deviations from established time trends in the years 2020-2022. This SVM model looked at the data available from 2015 to 2019 for a given time series and provided a classification for the target year. Normalized features were created to inform the SVM about the nature of the data.



These features included the target year compared to every other year, the standard deviation of the data from 2015-2019, and whether the target year was the highest observed year.

After identifying spikes in mortality during the pandemic years through the use of the SVM, we had to determine if a given spike is miscoded COVID-19 or if it is a true increase in the cause. To do this, we first used a similar strategy that excess mortality estimates used where we generated a counterfactual estimate to represent a COVID-19 free world. In this analysis, the counterfactual consists of a linear regression based on the data available from 2015 to 2019. Using the counterfactual estimates, we ran a regression for each cause-location-age band where we compared the excess numbers of deaths against the observed known COVID-19 deaths to return a Pearson correlation coefficient as well as a p-value. This regression shows the relationship of how much of the spike occurred along with the presence of COVID-19. If a location-cause-age band has a positive Pearson correlation as well as a P value of under .05 then we believe that there is miscoded COVID-19, and the observation is eligible for correction.

Once a country, cause, year, sex, age band combination with excess mortality has been identified through the SVM and regressions, we then calculate an estimate for the portion of excess attributable to the misassignment of COVID-19. It is important to note that while the identification step is done using national aggregates and the age bands mentioned above, the calculation of excess deaths attributable to COVID-19 is done at the most granular level – detailed GBD age groups and subnational locations where possible. To calculate the amount of excess mortality attributable to COVID-19, we first create an estimate of expected deaths absent of any pandemic effects using a combination of two counterfactuals. Total excess is then scaled to determine the amount of total excess that is attributable to COVID-19. The details of these steps are below.

The first of the two counterfactual estimates is determined by a linear regression of pre-pandemic data. For each location, age, sex, cause combination identified, a linear regression between year and observed deaths is ran for the five years prior to the pandemic (2015 to 2019). The resulting linear equation is then used to predict forward an estimated death total independent of COVID-19 or any excess mortality for the years 2020, 2021, and 2022.

The second of the two counterfactuals is determined using a global relative rates approach, adapted from a similar method used in the correction of misclassified HIV.<sup>30</sup> This approach involves generating relative rates, which are simply ratios between observed mortality rates in a given age group and the mortality rate in age groups believed to be relatively free of COVID-19. For the years 2020, 2021, and 2022, we can generate expected death totals by comparing these ratios to those observed in pre-pandemic data. The details of this process are as followed:

1. From the five years prior to the pandemic (2015 to 2019), compute global, all year, cause specific death rates by age and sex.
2. Compute death rates for each location, year, age, sex, cause observation identified to contain excess mortality in 2020, 2021, and 2022.
3. Compute two sets of “reference death rates”, for both the global data from step 1, and the granular data from step 2. Reference death rates are calculated as the average death rate from age groups relatively absent of COVID-19. The reference age groups are 5-9, 10-14, and 15-19.

$$GlobalReferenceDR_{s,c} = Average(DR_{5-9,s,c} ; DR_{10-14,s,c} ; DR_{15-19,s,c})$$

$$ObservationReferenceDR_{l,y,s,c} = Average(DR_{5-9,l,y,s,c} ; DR_{10-14,l,y,s,c} ; DR_{15-19,l,y,s,c})$$

4. For the global rates calculated in step 1, use their respective reference death rates to calculate a “relative death rate” for each age, sex, cause combination. This is simply the ratio between the age specific rate and the reference death rate calculated in step 3.

$$RelativeDR_{a,s,c} = \frac{DR_{a,s,c}}{ReferenceDR_{s,c}}$$

5. Finally, generate a counterfactual estimate such that the relative death rate of the observation would be equivalent to the global, COVID-19 free relative death rate. This can be done by multiplying the reference death rate of the observation by population, and then scaling to the global relative death rate.

$$Counterfactual_{l,y,a,s,c} = (ReferenceDR_{l,y,s,c} * Population_{l,y,a,s}) * GlobalRR_{a,s,c}$$

The final estimate for expected deaths is then calculated as the mean of the individual counterfactuals. Total excess mortality can then be estimated by subtracting the expected death total from the observed death total. The final step of the correction is then to determine what portion of the excess mortality should be attributed to COVID-19 misclassification. To do this, we use the regression coefficient from the linear regression ran during the identification step. This coefficient describes for a one unit increase in COVID-19, the associated increase in excess mortality. Thus, we can then multiply total excess by this coefficient to estimate the amount of total excess attributable to misclassified COVID-19. This total is then subtracted from the cause of interest and re-assigned to COVID-19.

### Section 3.9: HIV/AIDS misclassification correction (step 6)<sup>2</sup>

In many location-years, certain causes of death known to be comorbid with HIV/AIDS (eg, tuberculosis, other infectious diseases) are seen to have age patterns that diverge from those observed in location-years without widespread HIV epidemics and are in fact more reflective of HIV mortality trends. To identify these instances, a global relative age pattern is generated by using all VR deaths in countries with observed HIV prevalence less than 1% in 2010 by using the following equation:

$$RR_{asc} = \frac{R_{asc}}{\bar{x}(R_{65sc}, R_{70sc}, R_{75sc})}$$

Where:

$RR_{asc}$  is the relative death rate for age group  $a$ , sex  $s$ , and cause  $c$ ;

$R_{asc}$  is the rate for that age group

$\bar{x}(R_{65sc}, R_{70sc}, R_{75sc})$  is the mean of the rates in ages 65–69, 60–74, and 75–79 for that sex and cause.

This is preferable to comparing mortality rates because we are able to isolate divergence in age pattern while accounting for varying levels of overall mortality by fixing death rates to age groups that are unlikely to be confounded by the presence of HIV. Expected deaths for an identified cause were then determined by the equation:

$$ED_{lyasc} = \bar{x}(R_{ly65sc}, R_{ly70sc}, R_{ly75sc}) \times p_{lyas} \times RR_{asc}$$

Where:

$ED_{lyasc}$  are deaths for location  $l$ , year  $y$ , age group  $a$ , sex  $s$ , and cause  $c$ ;

$\bar{x}(R_{ly65sc}, R_{ly70sc}, R_{ly75sc})$  is the mean of the rates for ages 65–69, 60–74, and 75–79 for that location-year-sex-cause;

$p_{lyas}$  is the population for that location-year-age-sex

$RR_{asc}$  is the global standard relative rate determined in the previous step for that age-sex-cause.

The expected deaths remain attributed to that particular cause, while the difference between observed and expected are reallocated to HIV/AIDS.

### Section 3.9: Scale strata to province (step 7)<sup>2</sup>

Over time, a higher proportion of deaths have been registered in China through the expansion of the DSP system and provincial/county efforts to increase CoD registration. With the expansion of coverage, it is possible that province aggregates do not accurately represent the population distribution between urban and rural areas in each year. For this reason, we stratified the data preparation by urban and rural status for each county within each province.

Stratification was based on the median level of urbanisation across counties within each province as recorded in the 2010 China census. In the provinces of Tibet and Hainan, all counties were placed into one stratum based on largely homogeneous urbanisation levels within each province. This yielded a total of 62 analytical province-strata. Macao and Hong Kong were not included in this stratification system as the VR systems there are independent from that on the mainland; no weighting scheme needs to be carried out in these complete VR systems with quality CoD data.

Within each province-strata, a larger proportion of deaths in-hospital might be reported than that of deaths outside of hospital because of the internet hospital reporting system. To avoid bias, we reweighted in-hospital and out-of-hospital deaths based on the age-sex-province-specific fraction of deaths in and out of hospital in the DSP system. DSP data have been used to establish these percentages because in these communities, there is a concerted effort to identify all out-of-hospital deaths. Province-strata death rates are combined to produce overall province death rates by weighting each stratum by population in each age-sex-year group. Province death rates are rescaled so that all-cause mortality equals the estimated death rate in each age-sex-year estimated in the life-table analysis. The Bayesian noise reduction algorithm was used to deal with zero counts and small number issues for rare causes.<sup>2</sup>

### Section 3.10: Restrictions post-redistribution (step 8)<sup>2</sup>

Some causes of death can only be reliably assigned through an autopsy by a trained physician. For example, a VA would be unlikely to reliably distinguish between ischaemic and haemorrhagic stroke.

This step ensures that the detail of the cause list at this point in the data prep process is reasonable given the detail of the original data source and the methods by which the CoD was assigned. A “bridge map” is applied over a certain set of sources to ensure that these sources do not contain causes that could not reliably be determined by the methods used. These causes, identified to be too detailed, are then aggregated to their parent cause. This correction is applied to ICD-9 detail, ICD-9 BTL, ICD-10 tabulated, ICD-8 detail, ICD-8 A, China DSP (tabulated ICD-9), India MCCD, India SRS, USSR tabulated ICD-9, the Philippine Vital Statistics Reports, Iran ICD-10 VR from the Ministry of Health and Medical Education, and all VA. An example of this would be the aggregation of all sub-types of lower respiratory infection to lower respiratory infection in ICD-9 BTL.

### Section 3.11: Drop VR country-years or mark as non-representative (step 9)<sup>2</sup>

Lozano and colleagues<sup>31</sup> describe the negative impact that low-completeness VR data could have on CoD modelling for GBD 2010. In particular, in settings where a data source does not capture all deaths in a population, the cause composition of deaths captured might be different from those that are not. However, a completeness sensitivity test found that low-completeness VR data had little impact on the cause-specific mortality trends at the global level.

For GBD 2019, we investigated the impact of these data at the country and subnational levels and determined that these data produced unlikely trends in the models affected. Despite the minimal impact on global trends, better models were produced by eliminating or marking as non-representative data with extremely low completeness. VR completeness was estimated as the number of deaths registered divided by the number of deaths estimated in the GBD mortality envelope.

For this round, VR location-years with completeness less than 50% were dropped, while location-years with completeness between 50% and 69% were marked as non-representative.

In addition, any country-year with a number of deaths registered to major garbage codes greater than 50% of the deaths registered was dropped. Major garbage coding refers to garbage codes redistributed across Levels 1 and 2 of the cause hierarchy. When we redistribute garbage codes across Levels 1 and 2 of the cause hierarchy, this is because we do not have enough information to distribute them to more detailed levels [3 and 4].

### Section 3.12: Cause aggregation (step 10)<sup>2</sup>

The cause list is organised in a top-down hierarchical format containing four levels. The first group, or Level 1, sums all causes. Following all-cause mortality are Level 2 causes, which include three broad groupings of causes of deaths: “communicable, maternal, neonatal, and nutritional diseases”; “non-communicable diseases”; and “injuries”. Within

those Level 2 groupings are finer levels used for modelling. Level 3, or parent causes, are aggregated; the mortality estimate for a parent cause in the hierarchy represents the sum of the causes under that rubric. Sub-causes within Level 3 causes—Level 4—are more detailed. For example, the parent cause “intestinal infectious diseases” contains the three sub-causes: “typhoid fever”, “paratyphoid fever”, and “other intestinal infectious diseases”. Included in the parent cause estimate are deaths mapped directly to the parent and any Level 4 sub-causes. In data where there was not enough information to assign a Level 4 cause, we aggregated to the Level 3 parent cause. Exceptions to aggregating the Level 4 sub-causes to the parent are instances when certain sub-causes are not present. The United Nations Crime Trends police data only identify homicides, and aggregating homicides to injuries would not accurately represent all injuries.

### Section 3.13: Remove shocks and HIV/AIDS maternal adjustments (step 11)<sup>2</sup>

Prior to GBD 2023, CODEm models used an HIV/AIDS- and shock-free envelope. To be comparable, cause fractions in the causes of death data were also made to be HIV/AIDS- and shock-free. New to GBD2023, CODEm models now use an envelope inclusive of HIV/AIDS, but still free of shocks. Cause fractions were uploaded to the CoD database as the number of deaths due to the cause over an adjusted sample in which the number of deaths due to “conflict and terrorism” and “exposure to forces of nature” were removed.

#### Section 3.13.1: Remove shocks from denominator where cause list includes shocks (step 11.1)

The first step to generate shock-free cause fractions was to remove any deaths from the sample that were directly coded to “conflict and terrorism”, or “exposure to forces of nature”. The cause fraction uploaded to the database can be calculated by a simple equation:

$$CF_{l,t,a,x,c} = \frac{D_{l,t,a,x,c}}{D_{l,t,a,x} - D_{l,t,a,x,war} - D_{l,t,a,x,disaster}}$$

Where:

$CF_{l,t,a,x,c}$  is the cause fraction for a location  $l$ , year  $t$ , age  $a$ , sex  $x$ , and cause  $c$

$D_{l,t,a,x,c}$  is the number of deaths observed for cause  $c$  in location  $l$ , year  $t$ , age  $a$ , and sex  $x$

$D_{l,t,a,x}$  is the total number of deaths due to all causes observed in location  $l$ , year  $t$ , age  $a$ , and sex  $x$

$D_{l,t,a,x,war}$ , and  $D_{l,t,a,x,disaster}$  are the numbers of deaths observed in location  $l$ , year  $t$ , age  $a$ , and sex  $x$  for causes “conflict and terrorism”, and “exposure to forces of nature”, respectively

Cause fractions for shock causes were also uploaded to the database for use in separate estimation processes described by Wang et al.<sup>32</sup> In this case, cause fractions followed the standard equation, with variables following the same explanation.

$$CF_{l,t,a,x,c} = \frac{D_{l,t,a,x,c}}{D_{l,t,a,x}}$$

$$CF_{l,t,a,x,mat} = \frac{D_{l,t,a,x,maternal}}{D_{l,t,a,x,maternal} + \frac{E[D_{l,t,a,x,hiv\_shock\_free}]}{E[D_{l,t,a,x}]} D_{l,t,a,x,non-maternal}}$$

$CF_{l,t,a,x,mat}$

$D_{l,t,a,x,mat}$

$D_{l,t,a,x,non-maternal}$

$E[D_{l,t,a,x}]$

$$E[D_{l,t,a,x,hiv\_shock\_free}]$$

### Section 3.13.2: HIV/AIDS correction of sibling history, census, and survey data (step 11.3)

As described in our analysis from GBD 2013, many studies have failed to find increased mortality in HIV+ pregnant mothers, but those who have advanced HIV are known to have increased baseline mortality. Prior to GBD 2013, we did not distinguish between deaths in HIV+ women that were caused by pregnancy and those for whom the pregnancy was incidental to their death. To more explicitly quantify the contribution of pregnancy to death in HIV+ women, and therefore more accurately estimate the maternal death count, we completed two additional analyses for GBD 2013 and all subsequent GBD analyses. First, we determined the population attributable fraction (PAF) of HIV/AIDS to pregnancy-related death. Second, we determined the proportion of pregnancy-related deaths in HIV+ pregnant mothers that are aggravated by pregnancy and are therefore by definition maternal deaths.

$$PAF = \frac{P(RR - 1)}{1 + P(RR - 1)}$$

Where:

*PAF* is the population attributable fraction

*P* denotes the prevalence of HIV in pregnancy

*RR* is relative risk of mortality in HIV+ vs HIV- pregnant mothers.

To recap our analysis for GBD 2013, we used the paper published by Calvert and Ronsmans<sup>33</sup> to identify sources that could inform Step 1 of our HIV-correction analysis. We independently reviewed each of the component studies in Calvert and Ronsmans' review and extracted data directly, not from the systematic review paper. We identified only one additional study that was not used in Calvert and Ronsmans' analysis. We have, however, not used all the studies included in that review. Specific details are as follows:

- 1) Figueroa-Damian et al.<sup>34</sup> was excluded for not including any postpartum deaths at all.
- 2) In the case of Ryder et al.<sup>35</sup> and Zvandasara et al.,<sup>36</sup> we excluded those deaths that occurred more than 12 months after delivery.
- 3) We excluded the results from Chilongozi et al.<sup>37</sup> from the site that did not include any HIV- patients.
- 4) Leroy et al.<sup>38</sup> was not in the bibliography. We could not locate it for review, so it was excluded.
- 5) Kourtis et al.<sup>39</sup> was extracted with adjustment of the denominator based on the average number of hospitalisations per delivery in each group.
- 6) Ticconi et al.<sup>40</sup> was excluded for being both non-representative and including subgroup data from mothers with malaria infection.

A total of 21 sources were included in our analysis of the increased mortality risk of HIV+ versus HIV- women in pregnancy.<sup>41</sup> We performed DerSimonian-Laird random effects meta-analysis to derive a pooled estimate of *RR* of death during pregnancy given HIV positivity.<sup>42</sup> The pooled effect size was 6.40 (95% uncertainty interval [UI] 3.98–10.29), which was then used to calculate an HIV *PAF* for each country, age group, and year. To determine the proportion of those HIV-related deaths that were attributable to maternal causes, we performed a second systematic literature review. This time we sought evidence for the excess mortality risk of pregnancy in those women who are

already HIV+. Most studies have failed to find such an effect, but most also did not stratify their study population by stage of HIV or ART (antiretroviral therapy) status. Only two studies did this stratification, with a pooled effect size of 1.13 (95% UI 0.73–1.77).<sup>43,44</sup>

An updated literature review to inform the relative risk of mortality in pregnancy in HIV+ versus HIV- women had 14 non-usable sources. We completed this search on May 10, 2019, using the following search strings:

(( HIV[Title/Abstract] OR "Acquired Immunodeficiency Syndrome"[Title/Abstract] OR AIDS[Title/Abstract] ) AND ( "pregnant"[Title/Abstract] OR "pregnancy"[Title/Abstract] OR "postpartum"[Title/Abstract] OR "post partum"[Title/Abstract] ) AND ( "mortality"[Title/Abstract] OR "death"[Title/Abstract] ) NOT "case report" NOT ( animals[MeSH] NOT humans[MeSH] )

AND (2016/08/15[PDat] : 3000/12/31[PDat] ) )

Prevalence of HIV in pregnant women was calculated by using the Joint United Nations Programme on HIV and AIDS (UNAIDS) Spectrum model,<sup>45</sup> a compartmental HIV progression model used to generate age-specific incidence, prevalence, and death rates from pre-calculated incidence curves and assumptions about intervention scale-up and local variation in epidemiology. For each location, we used UNAIDS' age-specific ratios of fertility in women living with HIV to fertility in women not living with HIV. In most locations, this ratio is assumed to be greater than one in women aged 15–24 years and less than one and decreasing as age increases beyond 24 years. Since Spectrum assumes fertile ages of 15–49 years, we used the ratio of HIV prevalence in pregnant women to HIV prevalence in the general population at either end of that range to extend estimates to age bands 10–14 years and 50–54 years.

Unlike GBD 2013, when we applied the PAF correction to the envelope of maternal deaths predicted by CODEm, we instead applied country-year-age-group-specific *PAF* to maternal mortality input data prior to modelling in CODEm. This ensured that both the numerator and denominator of all *CF* data were internally consistent in their exclusion of background HIV/AIDS mortality. The cause fractions for maternal deaths in sibling history, survey, and census data were therefore adjusted as follows:

$$CF_{l,t,a,x,mat_{adj}} = CF_{l,t,a,x,mat} \times (1 - Prop_{hiv_{l,t,a,x}})$$

$$Prop_{hiv_{l,t,a,x}} = PAF_{l,t,a,x,hivpos} \times (1 - rr_{mat})$$

$$CF_{l,t,a,x,mat_{hiv}} = CF_{l,t,a,x,mat} \times Prop_{maternalhiv_{l,t,a,x}}$$

$$Prop_{maternalhiv_{l,t,a,x}} = PAF_{l,t,a,x,hivpos} \times rr_{mat}$$

Where:

$CF_{l,t,a,x,mat}$  = The proportion of deaths due to all maternal causes before HIV/AIDS correction for the location, year, age, and sex.

$CF_{l,t,a,x,mat_{adj}}$  = The proportion of deaths due to maternal causes after the adjustment for the location, year, age, and sex.

$CF_{l,t,a,x,mat_{hiv}}$  = The proportion of deaths due to maternal deaths aggravated by HIV/AIDS after the adjustment for the location, year, age, and sex.

$PAF_{l,t,a,x,hivpos}$  = The PAF that describes the percentage of all maternal deaths that were HIV-related for the location, year, age, and sex.

$ProP_{hiv_{l,t,a,x}}$  = The proportion of deaths in pregnancy for the location, year, age, and sex that are estimated to be incidental deaths due to HIV/AIDS and therefore not a maternal CoD.

$ProP_{maternalhiv_{l,t,a,x}}$  = The proportion of deaths in pregnancy for the location, year, age, and sex that are estimated to be HIV+ and maternal deaths that are aggravated by HIV/AIDS.

$rr_{mat} = 0.13/1.13$  = The proportion of HIV/AIDS deaths during pregnancy that were exacerbated by the pregnancy.

### Section 3.13.3: HIV/AIDS correction of other maternal mortality data (step 11.4)

Although a specific subset of codes in ICD-10 corresponds to HIV/AIDS deaths aggravated by pregnancy, these codes are sparsely used and unreliable. We therefore adapted the method described to also correct VR and VA sources for the systematic exclusion of HIV-related maternal deaths. This correction was calculated in the same manner, by using the same input data as above, with the only difference being that HIV correction of VR and VA sources resulted in a net increase in the maternal correction factor. Maternal deaths aggravated by HIV/AIDS are calculated in the following way:

$$CF_{l,t,a,x,mat_{hiv}} = CF_{l,t,a,x,mat} \times ProP_{maternalhiv_{l,t,a,x}}$$

$$ProP_{maternalhiv_{l,t,a,x}} = \frac{PAF_{l,t,a,x,hivpos} \times rr_{mat}}{1 - PAF_{l,t,a,x,hivpos} \times rr_{mat}}$$

### Section 3.13.4: Crosswalk by code system

Over time, the International Classification of Diseases (ICD) is continually updated to support more detailed and accurate record keeping. When a new ICD version is released, it is not always directly compatible with the previous version. During this time of adoption, disjoints within the time series may develop due to the incompatibility. (source on challenges of ICD versions). For example, when ICD10 is adopted, many of the deaths that would be considered garbage codes in ICD9 are now mapped directly to an ICD code that maps to a real underlying GBD cause. This creates a disjoint at the year of adoption in both the target cause as well as related causes. In these scenarios, we trust ICD10 over ICD9 unless there is evidence supporting ICD9's performance. We also believe that ICD9 captures the trend over time of a given cause even if it has not accurately captured all the deaths. To correct for this disjoint between code-systems and utilize an older ICD's time trend we developed an adjustment factor using a method called a crosswalk,<sup>46</sup> by code-system where we compare the first 3 years of the more trusted code-system and the last 3 years of the code-system we want to adjust. This comparison produces a factor that we can apply to the entire code system we want to adjust that brings the average of the last 3 years to be equal to the average of the first 3 years of the trusted code-system.

$$\overline{CSO}_{l,a,s,c} * factor = \overline{CST}_{l,a,s,c}$$

Where:

$\overline{CSO}$  = average deaths of the last 3 years of the old code system

$\overline{CST}$  = average deaths of the first 3 years of the trusted code system

$l, a, s, c$  = Location, Age, Sex, Cause specific

This adjustment is done for specific location, age, sex, level 2 causes and produces a continuous time trend through the year of ICD adoption that conserves the pattern within the adjusted code system.

This correction to a target cause requires deaths to be added or removed from other related causes. After the initial correction has been applied to a target level 3 cause, deaths are proportionally balanced between all level 3 causes within the related level 2 parent. This ensures that no deaths are added or lost during the correction and that deaths are only being moved from causes we believe are affected by the change in code-system. We applied this correction to time series that were identified to have disjoint at the year of code-system adoption. Corrections were only kept if they did not introduce disjoint within other level 3 causes during the balancing process. In general, we did not perform crosswalk by code systems if a country-cause's level 2 cause has a disjoint present. If there is a disjoint at the level 2 cause then balancing will move the disjoint from the target cause to a child cause.

Location	Level 2 Cause	Years Adjusted
Bulgaria	Cardiovascular diseases	1990-2004
Canada	Cardiovascular diseases	1986-1999
Spain	Cardiovascular diseases	1980-1998
France	Cardiovascular diseases	1980-1999
United Kingdom	Cardiovascular diseases	1980-1999
Croatia	Cardiovascular diseases	1985-1994
Hungary	Cardiovascular diseases	1980-1995
Italy	Cardiovascular diseases	1980-2002
Mexico	Cardiovascular diseases	1980-1997
Mauritius	Cardiovascular diseases	1980-2004
United States of America	Cardiovascular diseases	1980-1998
Japan	Diabetes and kidney diseases	1980-1994
Mauritius	Diabetes and kidney diseases	1980-2004
New Zealand	Diabetes and kidney diseases	1980-1999

### Section 3.14: Noise reduction (step 12)<sup>2</sup>

To deal with problems of zero counts and stochastic variation in VR, VA, MITS diagnosed data, and maternal surveys, censuses, surveillance, and sibling histories, we use a Bayesian noise-reduction algorithm. We estimate a prior for a given series of data by running a Poisson regression to estimate the number of deaths due to each respective cause and sex with dummy variables for country, age, and year. With several notable exceptions (detailed below), these regressions are sex-, cause-, and GBD region-specific, so borrowing strength over age, year, and country is only within a given data type, GBD region, cause, and sex. The posterior estimate for each data is a weighted average between the prior and data:

$$\text{posterior cause fraction} = X \times \text{weight} + \mu \times (1 - \text{weight})$$

Where:

$X$  is the mean of the data in units of cause fractions

$\mu$  is the mean of the prior in units of cause fractions

The weight is determined by a function of the mean deaths per year in a given location-age-sex-cause for vital registration data:

$$VR\ weight_{l,a,s,c} = 1 - \frac{a}{b \times \text{mean}_y(\text{deaths}_{l,y,a,s,c}) + 1}$$



In this equation,  $a$  and  $b$  are hyper-parameters set to 0.99 and 0.1, respectively. This value of  $a$  was chosen to give the data only 1% weight when data are extremely sparse and therefore average approximately zero deaths per year for a given location, age, sex, and cause. This value of  $b$  was chosen based on a grid search to determine the optimal value of  $b$  that gives as much weight to the data as possible while still resulting in optimal smoothing of small numbers.

For non-VR, the weight is instead based on the variance of the prior,  $\tau^2$ , and the variance of the data,  $\sigma^2$ .  $\tau^2$  is estimated from the Poisson regression, taking into account the variance-covariance matrix of the regression coefficients. For the data variance,  $\sigma^2$ , we use the Wilson approximation which provides an estimate of  $\sigma^2$  even in cases with a zero count of cause-specific deaths. The weight used for non-VR is:

$$non - VR\ weight = \frac{\tau^2}{\tau^2 + \sigma^2}$$

This approach to noise reduction adjusts zero counts to non-zero values, mitigating the problem that zero counts in a log rate model or a logit cause fraction CODEm must be dropped from the regression, leading to upward bias in the estimates. This is particularly important in three settings: small high-income countries with low death counts overall; rare causes with small numbers of cause-specific deaths; and the analysis of sibling history, verbal autopsy, and survey data with small samples. For example, in sibling history data, for any given age group in any given year the number of deaths reported in the survey that are pregnancy-related or the number of deaths from all causes in that age group may be small.

There are several important exceptions to the data pooling and regression specifications outlined above. First, for country-series of VR from subnationally estimated countries, regressions are run on data from only that country with a fixed effect on subnational unit. Third, for VA data, all VA data for a given super-region are pooled together and a random effect on study-location is added to the regressions, allowing for different studies and surveillance sites to borrow strength from one another within a super-region. Unless the data are part of a time series (ie, the Matlab Health and Demographic Surveillance System and India Sample Registration System), the regression has no year component. Fourth, for VA data on malaria specifically, data are pooled into regional groups defined by the endemicity of malaria, as defined by the *Plasmodium falciparum* parasitaemia rate (PFPR). We include groups for hyper- (PFPR  $\geq 0.4$ ), meso- (PFPR 0.05 – <0.04), and hypo- (PFPR <0.05) endemic regions, separately for India; sub-Saharan Africa not including South Africa, Cabo Verde, and Mauritius, and including Yemen; and all remaining locations. Fifth, we run a separate regression for each cause in MITS diagnosed data from the Child Health and Mortality Prevention Surveillance (CHAMPS) study with fixed effects on age group and sex and a random effect on study-location. We include data from both sexes in one model and borrow strength across sex in order to combat data sparsity, and because CHAMPS includes only children under 5, an age demographic where sex differences are less extreme. Sixth, several individual VA studies are excepted from the above data pooling rules and regressed in their own separate models due to sample sizes that are too large and therefore dominate super-region-level models: Indonesia Sample Registration System (SRS), India verbal autopsy studies, and the Nepal Burden of Disease VA Study. Seventh, several VA study groups are excepted from the above regression specifications and run with individualised specifications: malaria-specific VA data from SSA are regressed with fixed effects on age group and study-location-year; VA data from the southeast Asia, east Asia, and Oceania super-region are regressed with fixed effects on age group and study-location; Matlab HDSS data are regressed with fixed effects on age group and year; and India SRS data are regressed with fixed effects on age group, year, and subnational location.

We also employ several strategies to combat data sparsity and failure to converge in noise reduction models. First, for models run on GBD regions of VR data, if a given year has only one country of data in that region, we bin this year with the preceding year in the Poisson regression. Second, for VA data, if a given cause/super-region/sex group has six or fewer observations or zero non-zero observations, we include all VA globally available for that cause/sex in a regression with fixed effects on age group and study-location and use the predictions from this global model as the prior for that cause/super-region/sex group. Third, for VA and CHAMPS data, if a model that includes random effects fails to converge, we rerun the model with the corresponding fixed effects. Fourth, after this, for all noise

reduction models, if the model fails to converge, we use a weighted average of the observed cause fractions within age groups in the model input data as the prior in noise reduction. This is essentially pooling information across study, location, and year within a grouping of data with similar data type and geography.

During noise reduction, all datapoints with cause fractions of zero are raised to a non-zero value through the Bayesian average with a non-zero prior. However, some of these values can be extremely small in magnitude (eg, on the order of  $10^{-50}$  or smaller), to the point that they would be unobservable in the raw data. Cause fractions of this magnitude would become extremely large in absolute value when log- or logit-transformed during the CODEm process, leading to poor model fit in other ranges of the input data. However, these datapoints must remain non-zero in order to include them in a log- or logit-transformed model at all and avoid upwardly biasing estimates. To address this problem, we enforce a set of non-zero cause-, age- and sex-specific minimums, known as the “non-zero floor,” on all cause fractions in the cause of death database. These floor values are chosen to be high enough to be relatively close to the bulk of the remaining data in log- or logit-space, but low enough to avoid significantly upwardly biasing the distribution of the data.

Floor values are determined based on VR data for cause/age/sex groups where we have enough data to do so accurately. We first consider VR data from all countries with greater than 50 million person-years represented in the database, after cause mapping and age-sex splitting but before any further processing. We calculate all-year cause fractions for each country, cause, age, and sex in this dataset. If the minimum cause fraction across country for a given cause/age/sex is non-zero, then we consider this cause/age/sex group to have sufficient data to determine a data-driven floor value. Otherwise, we consider this cause/age/sex to be too rare to have sufficient data to determine a minimum observable non-zero rate.

For cause/age/sex groups that have sufficient data, we set floor values based on all national-level noise-reduced VR datapoints in the CoD database that are non-zero before noise reduction. The floor value for a given cause/age/sex ( $c, a, s$ ) is the minimum cause fraction across country-years ( $l, y$ ) in this non-zero subset:

$$floor_{c,a,s} = \min_{l,y} (post\ NR\ VR\ cause\ fraction_{c,a,s,l,y}),$$

where  $cause\ fraction_{c,a,s,l,y} \neq 0$  before NR

For all other cause/age/sex groups, which have insufficient data to inform the above calculation, we set a  $floor_{c,a,s}$  value such that: (a) all floor values for a given cause add up to one death globally per year and (b) the age/sex distribution of the floor values in each cause follows the age/sex distribution of the global cause-specific mortality rates calculated in Section 2.6 for the purposes of age/sex splitting:

$$floor_{c,a,s} \times mortality\ envelope_{a,s} = \frac{R_{c,a,s} \times pop_{a,s}}{\sum_{a,s} R_{c,a,s} \times pop_{a,s}}$$

$floor_{c,a,s}$  = floor value for cause  $c$ , age group  $a$ , sex  $s$

$mortality\ envelope_{a,s}$  = total estimated global deaths in 2018 for age group  $a$ , sex  $s$

$pop_{a,s}$  = total estimated global population in 2018 for age group  $a$ , sex  $s$

$R_{c,a,s}$  = global cause-specific mortality rate of cause  $c$ , age group  $a$ , sex  $s$

This ensures that we add at most one death in the world per cause per year for causes where we have insufficient data to determine a plausible minimum observable cause fraction from the data.

### Section 3.15: Cause of death database and outlier identification (step 13)<sup>2</sup>

Death rates for different causes of death generally have a stable age pattern. In large populations, these patterns will not change very rapidly over time. We can assume a relatively stable pattern in death rates for all causes except for some epidemic diseases and specific types of injuries. Rare causes in large populations and prevalent causes in small populations usually have stochastic patterns. To correct for these stochastic patterns, we implemented a noise-reduction process, explained in Step 12.

In VR data, we infrequently find one or more datapoints for specific geography/age/sex/year combinations that lie very far from the stable pattern of death rates. In these situations, the model usually ignores the datapoint(s). If the model fails to ignore these data, dramatic jumps or drops can occur in the death rates. When no logical explanation exists for variation in the death rates to this degree, we regard the datapoint(s) as outlier(s). The selection of datapoints to regard as outliers occurs after data have been prepped for modelling, as well as during preliminary reviews of the models.

In non-VR sources, data-collection methods and data quality can vary widely from source to source. Where datapoints in each age-sex-geography-year are very sparse, extreme datapoints can have a bad effect on regional estimation. In these situations, we investigate the study's methods and consider lower-quality datapoints as outliers.

Identifying outliers in the CoD data occurs prior to finalisation of models for each cause. We do not automate the selection of outliers but investigate the source of the offending data as well as reviewing other data sources for the same cause, geography, and year. Ultimately, outliers are identified based on the judgement of the modeller and senior faculty. Outlier decisions are reversible and may be revisited.

### Section 3.16: Causes of death data star-rating calculation<sup>2</sup>

GBD estimates are most accurate when computed with a full time series of complete VR with a low percentage of garbage codes. We have developed a simple star-rating system from 0 to 5 to give a picture of the quality of data available in a given country over the full time series used in GBD estimates. Countries improve in the star rating as they increase availability, completeness, and detail of their mortality data and reduce the percentage of deaths coded to ill-defined garbage codes or highly aggregated causes (figures 5a and 5b).

We assign star ratings to rate the quality of data for any given location year. Two dimensions determine this star rating: (I) the percentage of total deaths determined to be major garbage (such as ill-defined). Causes such as “injuries” or “cancer” will also be included in major garbage percentage because this percentage includes use of highly aggregated causes; and (II) the level of completeness of death registration (percentage of total deaths captured by the death registration). These two values were used to create a “percentage well-certified” value between 0 and 1, determined as:

$$pct_{wellcertified} = Completeness \times (1 - pct_{majgarbage})$$

The mapping of percentage well certified to star rating is as followed:

$$0 \text{ star: } 0\% = pct_{wellcertified}$$

$$1 \text{ star: } 0\% < pct_{wellcertified} < 10\%$$

$$2 \text{ star: } 10\% \leq pct_{wellcertified} < 35\%$$

$$3 \text{ star: } 35\% \leq pct_{wellcertified} < 65\%$$

$$4 \text{ star: } 65\% \leq pct_{wellcertified} < 85\%$$

$$5 \text{ star: } pct_{wellcertified} \geq 85\%$$

While stars are calculated for each five-year time interval as well as the full time series from 1980 to 2020, stars in the main text are presented for the full time series only.

In the case of VA, all garbage codes are considered ill-defined because redistribution for VA is highly imprecise.

For each VA data source, percentage well-certified is:

$$pct_{wellcertified} = VerbalAutopsyAdjustment \times (1 - pct_{majgarbage})$$

Where:

$$VerbalAutopsyAdjustment = SubAdj \times RegAdj \times AgeSexCoverage$$

SubAdj is 10% for subnationally representative studies; 100% for nationally representative studies. This adjustment, while arbitrary in its specific value, reflects the bias that can be associated with studies that only cover a potentially non-representative sample of a country's population.

RegAdj is 64% for all VA data sources. This accounts for the inaccuracy of VA in assigning CoD compared to medically verified VR. The specific multiplier 0.64 is based on the chance-corrected concordance of Physician Certified Verbal Autopsy (PCVA) versus medical certification by the Population Health Metrics Research Consortium.<sup>47</sup>

Age-sex coverage is the number of deaths estimated in the GBD mortality envelope for the ages and sexes in the study for the country and year divided by the number of deaths estimated in the GBD mortality envelope for the country and year. Studies that only cover children under 5 years or maternal mortality, for example, will be highly discounted by this multiplier.

Once percentage well-certified is calculated for each location-year of VR and each VA study-year, we then combine these into one measurement for each five-year time interval and the full time series 1980–2020. For each five-year time interval, we take the maximum percentage well-certified. Then for 1980–2020, we take the average of the maximum percentages well-certified for the seven five-year time intervals. Any five-year time interval in which no data were available were given a percentage well-certified value of zero.

Prior to GBD 2019, the causes of death team used an all-ages, both-sex cause fraction to estimate the percentage of garbage-coded deaths in a given location-year. Thus, the percentage of garbage for a given location year was determined as:

$$CF_G = \frac{D_G}{D}$$

Where:

$CF_G$  represents the cause fraction of percentage garbage

$D_G$  represents total garbage-coded deaths

$D$  represents the total deaths in a given location/year.

In GBD 2019, we moved to calculating the percentage of garbage-coded deaths using an age-standardised cause fraction. The steps for creating these age-standardised cause fractions, in the case of garbage, are as follows:

1. Create both-sex, age-specific cause fractions of garbage for each age group.

2. Scale these cause fractions by a set of both-sex age weights, determined by global mortality estimates from 2010 to present. That is, weights for each GBD age group were determined as:

$$W_a = \frac{D_a}{D}$$

Where:

$W_a$  is the weight for given age group “a”

$D_a$  is the total both-sex, global deaths from 2010 to present in age group “a”

$D$  is the total both-sex, global deaths from 2010 to present across all ages.

3. Sum these weighted cause fractions across all age groups to produce the age-standardised cause fraction.

In the case of percentage garbage for a given location-year, the formula to calculate percentage garbage would be given as the sum of the weighted age-specific cause fractions across all age groups “a”:

$$CF_G = \sum_a \left( \frac{G_a}{D_a} \times W_a \right)$$

Where:

$G_a$  represents the total both-sex garbage deaths in age group “a”

$D_a$  represents the total both-sex deaths in age group “a”

$W_a$  represents the weight generated from mortality estimates for age group “a”

In GBD 2023, a buffer system was implemented to prevent frequent star rating changes due to small fluctuations in completeness and percentage of garbage round over round. If a location-year decreases in star rating, as long as its PWC value remains within 3% of the cut-off for its former star rating, it will retain the former star rating.

## Section 4: Causes of death modelling methods

### Section 4.1: CODEm<sup>2</sup>

#### Section 4.1.1: Overview of methods

Cause of death ensemble modelling (CODEm) is the framework used to model most cause-specific death rates in the GBD.<sup>48</sup> It relies on four key components:

First, all available data are identified and gathered to be used in the modelling process. Although the data may vary in quality, they all contain some signal of the true epidemiological process.

Second, a diverse set of plausible models are developed to capture well-documented associations in the estimates. Using a wide variety of individual models to create an ensemble predictive model has been shown to outperform techniques using only a single model both in CoD estimation<sup>48</sup> and in more general prediction applications.<sup>49,50</sup>

Third, the out-of-sample predictive validity is assessed for all individual models, which are then ranked for use in the ensemble modelling stage.

Finally, differently weighted combinations of individual models are evaluated to select the ensemble model with the highest out-of-sample predictive validity.

For some causes (eg, lower respiratory infections), evidence exists that the relationship between covariates and death rates might differ between children and adults. Separate models are therefore run for different age ranges, when applicable. Additionally, separate models are developed for countries with extensive, complete, and representative VR for every cause to ensure that uncertainty can better reflect the more complete data in these locations.

In order to ensure the addition of subnational locations is not driving changes in estimates, we run a global model that excludes data from non-standard locations; the resulting covariate betas are then used as priors for the true global model.

In addition to CoD modelling, we also estimate fatal discontinuities. Fatal discontinuities are events that are stochastic in nature, that cannot be modelled because they do not have a predictable time trend. The fatal discontinuities by cause are aggregated by age and sex and added to the estimated number of deaths in CoD modelling for those causes during CoDCorrect. Details on their methods can be found in Section 3.4.

#### **Section 4.1.2: Model pool development**

Because many factors may co-vary with any given CoD, a range of plausible statistical models are developed for each cause. In the CODEm framework, four families of statistical models are used: linear mixed effects regression (LMER) models of the natural log of the cause-specific death rate, LMER models of the logit of the cause fraction, spatiotemporal Gaussian process regression (ST-GPR) models of the natural logarithm of the cause-specific death rate, and ST-GPR models of the logit of the cause fraction (see the 2x2 table in Foreman et al).<sup>48</sup> For more on ST-GPR, see section 4.3.3. For each family of models, all plausible relationships between covariates and the response variable are identified. Because all possible combinations of selected covariates are considered for each family of models, multi-collinearity between covariates may produce implausible signs on coefficients or unstable coefficients. Each combination is therefore tested for statistical significance (covariate coefficients must have a coefficient with p-value <0.05) and plausibility (the coefficients must have the directions expected on the basis of the literature). Only covariate combinations meeting these criteria are retained. This selection process is run for both cause fractions and death rates, then ST-GPR and LMER-only models are created for each set of covariates. For a detailed explanation of the covariate selection algorithm, see Foreman et al.<sup>48</sup>

#### **Section 4.1.3: Data variance and uncertainty estimation**

The families of models that go through ST-GPR described in Section 4.1.2 incorporate information about data variance. The main inputs for a Gaussian process regression (GPR) are a mean function, a covariance function, and data variance for each datapoint. These inputs are described in detail in Foreman et al.<sup>48</sup> For GBD 2019, we updated this calculation to incorporate garbage code redistribution uncertainty.

Three components of data variance are now used in CODEm: sampling variance, non-sampling variance, and garbage code redistribution variance. The computation of sampling variance and non-sampling variance has not changed since previous iterations of the GBD and is also described in Foreman et al.<sup>48</sup> Garbage code redistribution variance is computed in the CoD database process described in Section 2.7 of this appendix. Since variance is additive, we calculate total data variance as the sum of sampling variance, non-sampling variance, and redistribution variance. Increased data variance in GPR results in the GPR draws not following the datapoint as closely.

#### **Section 4.1.4: Testing model pool on 15% sample**

The performance of all models (individual and ensemble) is evaluated by means of out-of-sample predictive validity tests. 30% of the data are randomly excluded from the initial model fits. These individual model fits are evaluated and ranked by using half of the excluded data (15% of the total), then used to construct the ensembles on the basis of their performance. Data are held out from the analysis on the basis of the cause-specific missingness patterns for ages and years across locations. Out-of-sample predictive validity testing is repeated 20 times for each model, which has been shown to produce stable results.<sup>48</sup> These performance tests include the root mean square error (RMSE) for the log of the cause-specific death rate, the direction of the predicted versus actual trend in the data, and the coverage of the predicted 95% UI.

#### **Section 4.1.5: Ensemble development and testing**

The component models are weighted on the basis of their predictive validity rank to determine their contribution to the ensemble estimate. The relative weights are determined both by the model ranks and by a parameter  $\psi$ , whose value determines how quickly the weights taper off as rank decreases. The distribution of  $\psi$  is described in more detail in Foreman et al.<sup>48</sup> A set of ensemble models is then created by using the weights constructed from the combinations of ranks and  $\psi$  values. These ensembles are tested by using the predictive validity metrics described in Section 4.1.4 on the remaining 15% of the data, and the ensemble with the best performance in out-of-sample trend and RMSE is chosen as the final model.

#### Section 4.1.6 Final estimation

Once a weighting scheme has been chosen, 250 draws are created for the final ensemble, and the number of draws contributed by each model is proportional to its weight. The mean of the draws is used as the final estimate for the CODEm process, and a 95% UI is created from the 0.025 and 0.975 quantiles of the draws. The validity of the UI can be checked via its coverage of the out-of-sample data; ideally, the 95% UI would capture 95% of these data. Higher coverage suggests that the UIs are too large, and lower coverage suggests overfitting.

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 4.1.7: Selection of causes for which CODEm is used**

CODEm is used to model 235 causes, described in detail in Section 3.3. However, it is unsuitable for use in modelling certain causes, including those with very low death counts, those where cause-specific death record availability is inadequate, or those for which there are marked biases or variability for CoD certification over time that cannot be fully accounted for with the current garbage code redistribution algorithms. Criteria for causes where CODEm is not used are discussed in further detail in Section 3.2.

#### **Section 4.1.8: Model-specific covariates**

Modellers select covariates to be used in CODEm, but those covariates may not be significant or in the direction specified during the covariate selection step of CODEm and will therefore not be used in the model. These covariates are listed with a ‘—’ for number of draws. Additionally, covariates may be selected by CODEm but only exist in submodels that perform poorly and may end up with zero draws included in the final ensemble. Finally, all other covariates are listed with the number of draws in the final ensemble from submodels that had the covariate.

### **Section 4.2: Causes modelled outside of CODEm<sup>2</sup>**

#### **Section 4.2.1: Overview**

A number of causes required alternative modelling strategies to those used for CODEm because they were not compatible with CODEm estimation infrastructure and processes. Such unsuitability included having very low death counts; inadequate availability of cause-specific death records; and marked biases or variability for CoD certification over time that could not be fully accounted for with current garbage code redistribution algorithms. The inclusion of these causes in CODEm often renders its out-of-sample predictive validity testing unstable, but the validity of this type of testing is a key advantage of using CODEm for CoD estimation. Alternately, CODEm simply fails to generate plausible mortality rates in the absence of enough VR or VA data when these causes are included. For GBD 2023, we used alternative modelling approaches for these causes, including negative binomial models, natural history models, sub-cause proportion models, and prevalence-based models.

#### **Section 4.2.2: Negative binomial models**

"For 78 rare causes of death, too few observed deaths were included in the CoD database to produce stable estimates. For these causes, we estimated mortality by fitting count models - either negative binomial or Poisson - to available CoD data. Descriptions of the modelling process for each of these causes follows in the next sections.

#### **Section 4.2.3: DisMod-MR 2.1**

Until GBD 2010, non-fatal estimates 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 in varying age groupings and from data sources by using different methods. The DisMod-MR 1.0 tool used in GBD 2010 evaluated and pooled all available data, adjusted data for systematic bias associated with methods that varied from the reference and produced estimates with UIs by world regions. For GBD 2013, the improved DisMod-MR 2.0 had increased computational speed, allowing computations that were 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 to changing to a model specification using 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.<sup>51</sup> For GBD 2015, the computational engine (DisMod-MR 2.1) remained substantively unchanged, but we rewrote the wrapper code that organised the flow of data and settings at each level of the analytical cascade. The sequence of estimation occurred at five levels: global, super-region, region, country, and, where applicable, subnational locations (see flow diagram of DisMod-MR 2.1 cascade that follows). The super-region priors were generated at the global level with mixed-effects, non-linear regression by using all available data; the super-region fit, in turn, informed the region fit, and so on down the cascade. The



wrapper gave analysts the choice to branch the cascade in terms of time and sex at different levels depending on data density. The default used in most models was to branch by sex after the global fit but to retain all years of data until the lowest level in the cascade. For GBD 2023, we generated fits for the years 1990, 1995, 2000, 2005, 2010, 2015, 2020, 2022, 2023, and 2024.

In updating the wrapper, we consolidated the code base into a single language, Python, to make the code more transparent and efficient and to better deal with subnational estimation. The computational engine is limited to three levels of random effects; we differentiated estimates at the super-region, region, and country levels. In GBD 2013, the subnational units of China, Mexico, and the UK were treated as countries, such that a random effect was 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 unit's available data and its prior. This procedure mimicked the impact of a random effect on estimates between subnationals.

For GBD 2015, we improved how country covariates differentiate non-fatal estimates for diseases with sparse data. The coefficients for country covariates were re-estimated at each level of the cascade. For a given location, country coefficients were calculated by using both data and prior information available for that location. In the absence of data, the coefficient of its parent location was chosen to utilise the predictive power of our covariates in data-sparse situations.

Beginning in GBD 2017, the DisMod-MR 2.1 tool was used.

#### Section 4.2.4: 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 log-Gaussian equation for the data likelihood is as follows:

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

Where:

$y_j$  is a measurement value (ie, datapoint)

$\Phi$  denotes all model random variables

$\eta_j$  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 datapoint  $j$ , defined by:

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

Where:

$u_j$  is the total area effect (ie, the sum of the random effects at three levels of the cascade: super-region, region, and country)

$c_j$  is the total covariate effect (ie, 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 standard deviation (SD)

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

Where:

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

$I(j)$  denotes a datapoint for a particular integrand,  $j$

$\beta_{I(j),k}$  is the multiplier of the  $k^{\text{th}}$  x-covariate for the  $i^{\text{th}}$  integrand

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

$l$  denotes the SD of each datapoint in relation to a covariate (also called z-covariate)

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

$\delta_j$  is the SD for adjusted measurement  $j$ , defined by

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

Where  $m_j$  denotes the model for the  $j^{\text{th}}$  measurement, not counting effects or measurement noise 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 datapoint  $j$

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

$I(j)$  denotes the function of age corresponding to the integrand for datapoint  $j$

The source code for DisMod-MR 2.1 as well as the wrapper code is available at [https://github.com/ihmeuw/ihme-modelling/tree/master/gbd\\_2017/shared\\_code/central\\_comp/nonfatal/dismod](https://github.com/ihmeuw/ihme-modelling/tree/master/gbd_2017/shared_code/central_comp/nonfatal/dismod).

#### Section 4.2.5: Natural history models

For some causes for which CoD data may be systematically biased either owing to misclassification or because the disease exists in focal communities without VR or VA studies, we have developed natural history models. In natural

history models, incidence and case-fatality ratios are modelled separately and then combined to produce estimates of cause-specific mortality.

#### Section 4.2.6: Prevalence-based models

The modelling strategies for atrial fibrillation and flutter are distinct from those used for other causes modelled as natural history models. These models use prevalence estimates and excess mortality rates (EMR) generated through DisMod-MR 2.1 rather than incidence and case-fatality rates. This approach allows us to adjust estimates to more accurately reflect the number of deaths for which atrial fibrillation was the true underlying cause of death. Further details are provided in the CoD cause-specific modelling description for atrial fibrillation and flutter.

#### Section 4.2.7: Sub-cause proportion models

For certain sub-causes for which accurate diagnoses are known to be very difficult, we first modelled the parent cause in the GBD hierarchy with CODEm and then allocated deaths to specific causes by using proportions of the parent cause for each age-sex-location-year for each sub-cause. For these causes, we identified no significant predictors in negative binomial regressions. This approach was taken because the available data on these specific causes may come from sources other than VR, such as end-stage renal disease registries, or may come from too few places to model the death rates directly. Details for each cluster of causes analysed in this way follow.

## Section 5: COVID-19 estimation

In GBD 2021, COVID-19 deaths were treated as shock deaths, and thus were not part of the HIV/AIDS- and shock-free envelope or the cause fractions used in CODEm. Like HIV/AIDS, COVID-19 was included in the shock-free envelope for GBD 2023. Modelling of COVID-19 mortality was done using a tool developed at IHME called OneMod, described in detail in Appendix 3.

Three types of data were used in the model. Corrected vital registration data, as described in Section 3.8, was included where available. For select countries (Bulgaria, Czechia, Denmark, Philippines, Türkiye, USA), we also identified recent provisional releases of vital registration data that were not complete enough for inclusion in the cause of death database but did include counts of COVID-19 deaths. Lastly, we included surveillance data that was reported during the COVID-19 pandemic (through 2022 for most countries, into 2023 for China) for location-years not covered by these prior sources. Country totals by year can be found in the table below.

	2020	2021	2022	2023
<b>Corrected VR</b>	75	64	36	1
<b>Provisional VR</b>	1	1	4	3
<b>Surveillance</b>	104	114	123	1

The provisional VR and surveillance data were not age- and sex-specific, so we first ran OneMod using only the corrected VR and used the out-of-sample predictions for each location-year to split these data by age and sex, based on the below formula, where the age- and sex- specific rate of the data,  $d_{l,y,a,s}$ , was equal to the all age rate,  $d_{l,y}$ , multiplied by the proportion of all deaths predicted by the model – the model rate,  $m_{l,y,a,s}$ , multiplied by the population,  $p_{l,y,a,s}$  – in each age and sex. All models were fit using the binomial likelihood.

$$d_{l,y,a,s} = d_{l,y} \frac{m_{l,y,a,s} p_{l,y,a,s}}{\sum_{a=0}^A \sum_{s=1}^2 m_{l,y,a,s} p_{l,y,a,s}}$$

The covariate pool provided to Rover is listed in table below. COVID-19 infection rate, variant prevalence, vaccination rate, and infection-detection rate (IDR) were all estimated by the IHME COVID-19 Forecasting Team.<sup>52</sup> The value of IDR for all corrected VR data was set to 100%. All of these except for vaccination rate were included as fixed covariates, meaning they were automatically selected. Other covariates (risk factors, comorbidities, and Healthcare Access and Quality [HAQ]) were estimated as part of the Global Burden of Disease study. Covariates

were selected by the Rover module based on an absolute t-statistic inclusion threshold of 0.5. Covariate effect sizes were estimated at the global level.

	Selected for males	Selected for females	Coefficient directional constraint
Chronic kidney disease	X	-	Positive
Cardiovascular disease	X	X	Positive
Diabetes	-	X	Positive
Healthcare Access and Quality	X	X	Negative
Obesity	X	X	Positive
Smoking	-	-	Positive
Cancer	-	-	Positive
Chronic obstructive pulmonary disease	-	-	Positive
Total COVID-19 infections	X	X	Positive
Prevalence of alpha, beta, gamma, and delta variants	X	X	-
Prevalence of omicron and BA5 variants	X	X	-
Infection-detection rate (IDR; natural log)	X	X	Positive

After estimating effect sizes in SpXMod, we then modified the data and predictions before running KReg. First, we calculated an empirical effect size for IDR in countries with corrected VR early in the pandemic and surveillance data based on the overlapping years, weighted by the total VR deaths in each year:

$$\beta_{l,a,s} = \sum_{y=2020}^Y \frac{d_{l,y}^{VR} \logit(r_{l,y,a,s}^{VR}) - \logit(r_{l,y,a,s}^{surv})}{d_l^{VR} - \ln(IDR_{l,y})}$$

Where  $r$  is the rate and  $d$  represents deaths (because the covariate was in natural log space,  $\ln(IDR_{max}) == \ln(1) == 0$ ). Using this empirical effect size for countries with corrected VR and the modelled global effect size everywhere else, we then removed the IDR effect from rates in both the predictions and the data before running KReg, making a counterfactual as if the IDR was 100% in all location-years (this would not affect the corrected VR, as the IDR in those observations was already 100%):

$$\logit(r_{l,y,a,s}^{counterfactual}) = \logit(r_{l,y,a,s}^{original}) - \beta_{l,a,s} \ln(IDR_{l,y})$$

We generated draws from KReg and, in an effort to eliminate extreme outlier draws, resampled the 2.5% lowest and highest draws by location (as defined by cumulative all-age, both-sexes deaths over all years). We then raked the draws for subnational locations for which we only had parent-level data: New Zealand, Iran, India urban/rural, Indonesia, Ethiopia, Kenya, and Nigeria.

Two of the above-described pipelines were run on the full age- and sex-split dataset – one fitting age and time effects (by sex) at each level of the hierarchy (global, super region, region, and country), and one fitting only global- and then country-level effects. All countries in regions featuring at least one location with corrected VR used the full location hierarchy model, while countries in regions without any corrected VR used the model with only global and country effects (the latter group is comprised of South Asia, East Asia, and all regions in sub-Saharan Africa). This was due to the fact that, while the super region and region effects captured important contextual variation in groups of locations with sufficient differentiability in the covariates, in the absence of any corrected data the IDR effect was absorbed by the location effects for these regions (most of which have estimated IDRs below 1%), leading to

implausibly low counterfactual estimates. Were VR to become available in all super regions, we would explore using a single model that refit the IDR effect at a level below global. Because covariate selection was done at the global level (by sex), that process was the same for both model runs. The lambdas (on both the intercept and the spline) for each location level were as follows: global 1.0, super region 2.5, region 5.0, location 10.0. The spline was linear with knots at each year.

## Section 6: Central computation<sup>2</sup>

### Section 6.1: Imported cases

Imported cases are fatalities that occur in a geographic area where a particular CoD is known to be eradicated in a specific time period or where infection cannot occur. We apply space-time restrictions to these causes in the modelling strategy for that location and time period. However, in some rare cases, deaths from these causes occur outside of restricted locations and time periods. These deaths are referred to as imported cases.

Illustrating this concept, some diseases are transmitted only in certain regions of the world. For those diseases, we restricted our models to only endemic locations, excluding non-endemic countries from the analysis. In some cases, however, deaths for these geographically restricted diseases may occur in non-endemic countries. For example, individuals may become infected with a disease while travelling or residing in an endemic location. Through travel or migration, however, they may move to a non-endemic location prior to dying from that disease. Imported cases account for these kinds of deaths.

To calculate these imported cases, we find all cases from the VRs of data-rich countries for any CoD that is otherwise geographically or temporally restricted. We then create a beta distribution from that datapoint by using the sample size of the VR for that datapoint and upload these draws as a custom CoD model. This model is then used as an input to CoDCorrect.

### Section 6.2: CoDCorrect

#### Section 6.2.1: Objective of CoDCorrect

As mentioned in the main text, the CoD models are cause-specific. As such, there is no guarantee that the sum of these models will equal the results of the all-cause mortality estimates or that model results of child causes add up to the parent model results. The CoDCorrect process is used to make the CoD and all-cause mortality estimates internally consistent by using a very simple algorithm.

#### Section 6.2.2: Algorithm and levels

The core algorithm remains the same as it did in GBD 2013. The equation can be written as follows:

$$CD_{tyasjd} = D_{tyasjd} \left( \frac{PD_{tyasjd}}{\sum_{j=1}^k D_{tyasjd}} \right)$$

Where:

$CD_{tyasjd}$  is the corrected number of deaths for a location  $l$ , year  $y$ , age  $a$ , sex  $s$ , cause  $j$ , and draw  $d$

$PD_{tyasjd}$  is the parent CoD for a location  $l$ , year  $y$ , age  $a$ , sex  $s$ , cause  $j$ , and draw  $d$

$D_{tyasjd}$  is the uncorrected number of deaths estimated from a cause-specific model for a  $l$ , year  $y$ , age  $a$ , sex  $s$ , cause  $j$ , and draw  $d$

The CoDCorrect process starts by rescaling the Level 1 causes to match the all-cause mortality estimates (used for  $PD_{lyasjd}$  in the previous equation). Level 2 causes are then rescaled to their corrected parent causes. This process continues until all levels of the hierarchy have been rescaled. Causes and their levels within the CoDCorrect hierarchy can be found in table S2.

### Section 6.2.3: Diagnostic results of CoDCorrect by cause and location

### Section 6.3: Years of life lost calculation

Years of life lost (YLLs) owing to premature mortality were computed for 925 locations and 45 years. First, we used the lowest observed age-specific mortality rates by location and sex across all estimation years from locations with total populations greater than 5 million in the most recent year of estimation to establish a theoretical minimum risk reference life table. By convention, we also exclude years before 2010 with the reasoning that the previous decade has the lowest overall mortality than any other point in history. Additionally, years 2020-2021 were excluded from consideration for the TMRLT due to the COVID-19 pandemic.

The YLL is a metric that is computed by multiplying the number of estimated deaths by the predicted life expectancy by age, sex, location, and year. The metric therefore highlights premature deaths by applying a larger weight to deaths that occur in younger age groups. We propagated uncertainty from CoDCorrected deaths for all demographics. The predicted life expectancy is calculated with two main components, a global, sex-agnostic, all-time theoretical “best” life expectancy (assuming that males and females in all countries in the world could theoretically have this life expectancy) and the average age of death from with-shock life tables for each location, sex, age group, and year.<sup>3</sup>

### Section 6.4: GBD world population age standard

Age-standardised populations in the GBD were calculated by using the GBD world population age standard. For GBD 2013, GBD 2015, and GBD 2016, the age-specific proportional distributions of all national locations from the UN Population Division World Population Prospects 2012 revision for all years from 2010 to 2035 were used to generate a standard population age structure by using the non-weighted mean across all the aforementioned country-years. For GBD 2017, we used the non-weighted mean of 2017 age-specific proportional distributions from the GBD 2017 population estimates for all national locations with a population greater than 5 million people in 2017 to generate an updated standard population age structure.<sup>55</sup> For GBD 2023, we have continued to use this method using GBD 2023 population estimates.<sup>3</sup>

## Section 7: References

- 1 Stevens GA, Alkema L, Black RE, *et al.* Guidelines for Accurate and Transparent Health Estimates Reporting: the GATHER statement. *Lancet* 2016; **388**: e19–23.
- 2 GBD 2021 Causes of Death Collaborators. Global burden of 288 causes of death and life expectancy decomposition 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**: 2100–32.
- 3 GBD 2021 Demographics Collaborators. Global age-sex-specific mortality, life expectancy, and population estimates in 204 countries and territories and 811 subnational locations, 1950-2021, and the impact of the COVID-19 pandemic: a comprehensive demographic analysis for the Global Burden of Disease Study 2021. *Lancet* 2024; **403**: 1989–2056.
- 4 Uneke CJ, Uro-Chukwu HC, Chukwu OE. Validation of verbal autopsy methods for assessment of child mortality in sub-Saharan Africa and the policy implication: a rapid review. *Pan Afr Med J* 2019; **33**: 318.
- 5 Thomas L-M, D’Ambruoso L, Balabanova D. Verbal autopsy in health policy and systems: a literature review. *BMJ Glob Health* 2018; **3**: e000639.
- 6 Serina P, Riley I, Stewart A, *et al.* Improving performance of the Tariff Method for assigning causes of death to verbal autopsies. *BMC Med* 2015; **13**: 291.
- 7 Byass P, Hussain-Alkhateeb L, D’Ambruoso L, *et al.* An integrated approach to processing WHO-2016 verbal autopsy data: the InterVA-5 model. *BMC Med* 2019; **17**: 102.
- 8 Murray CJ, Lozano R, Flaxman AD, *et al.* Using verbal autopsy to measure causes of death: the comparative performance of existing methods. *BMC Med* 2014; **12**: 5.
- 9 Gakidou E, King G. Death by survey: estimating adult mortality without selection bias from sibling survival data. *Demography* 2006; **43**: 569–85.
- 10 Government of India. Medical certification of cause of death. 2023; published online Aug 11. <https://censusindia.gov.in/census.website/node/381> (accessed Sept 29, 2023).
- 11 Government of India. ORGI SRS. 2023; published online Aug 11. <https://censusindia.gov.in/census.website/node/294> (accessed Oct 16, 2023).
- 12 Ganguli M, Rodriguez EG. Reporting of dementia on death certificates: a community study. *J Am Geriatr Soc* 1999; **47**: 842–9.
- 13 Wachterman M, Kiely DK, Mitchell SL. Reporting dementia on the death certificates of nursing home residents dying with end-stage dementia. *JAMA* 2008; **300**: 2608–10.
- 14 Macera CA, Sun RK, Yeager KK, Brandes DA. Sensitivity and specificity of death certificate diagnoses for dementing illnesses, 1988-1990. *J Am Geriatr Soc* 1992; **40**: 479–81.
- 15 Olichney JM, Hofstetter CR, Galasko D, Thal LJ, Katzman R. Death certificate reporting of dementia and mortality in an Alzheimer’s disease research center cohort. *J Am Geriatr Soc* 1995; **43**: 890–3.

- 16 GBD 2019 Collaborators. Global mortality from dementia: application of a new method and results from the Global Burden of Disease Study 2019. *Alzheimers Dement N Y N* 2021; **7**: e12200.
- 17 Todd S, Barr S, Passmore AP. Cause of death in Alzheimer's disease: a cohort study. *QJM* 2013; **106**: 747–53.
- 18 Brunnström HR, Englund EM. Cause of death in patients with dementia disorders. *Eur J Neurol* 2009; **16**: 488–92.
- 19 Gao L, Calloway R, Zhao E, Brayne C, Matthews FE, Medical Research Council Cognitive Function and Ageing Collaboration. Accuracy of death certification of dementia in population-based samples of older people: analysis over time. *Age Ageing* 2018; **47**: 589–94.
- 20 Romero JP, Benito-León J, Mitchell AJ, Trincado R, Bermejo-Pareja F. Under reporting of dementia deaths on death certificates using data from a population-based study (NEDICES). *J Alzheimers Dis JAD* 2014; **39**: 741–8.
- 21 Moscovich M, Boschetti G, Moro A, Teive HAG, Hassan A, Munhoz RP. Death certificate data and causes of death in patients with parkinsonism. *Parkinsonism Relat Disord* 2017; **41**: 99–103.
- 22 Pressley JC, Tang M-X, Marder K, Cote LJ, Mayeux R. Disparities in the recording of Parkinson's disease on death certificates. *Mov Disord* 2005; **20**: 315–21.
- 23 Duncan ME, Pitcher A, Goldacre MJ. Atrial fibrillation as a cause of death increased steeply in England between 1995 and 2010. *EP Eur* 2014; **16**: 797–802.
- 24 Johnson SC, Cunningham M, Dippenaar IN, *et al.* Public health utility of cause of death data: applying empirical algorithms to improve data quality. *BMC Med Inform Decis Mak* 2021; **21**: 175.
- 25 Barker B, Degenhardt L. Accidental drug-induced deaths in Australia 1997 - 2001. Sydney: National Drug and Alcohol Research Centre, 2003.
- 26 Roxburgh, A., Burns, L. Accidental drug-induced deaths due to opioids in Australia, 2013. Sydney: National Drug and Alcohol Research Centre, 2013.
- 27 Roxburgh A, Burns L. Cocaine and methamphetamine related drug-induced deaths in Australia, 2011. Sydney: National Drug and Alcohol Research Centre, 2015.
- 28 França EB, Ishitani LH, Abreu DMX de, *et al.* Measuring misclassification of Covid-19 as garbage codes: Results of investigating 1,365 deaths and implications for vital statistics in Brazil. *PLOS Glob Public Health* 2022; **2**: e0000199.
- 29 CDC. Excess deaths associated with COVID-19. 2023; published online Sept 28. [https://www.cdc.gov/nchs/nvss/vsrr/covid19/excess\\_deaths.htm](https://www.cdc.gov/nchs/nvss/vsrr/covid19/excess_deaths.htm) (accessed Jan 13, 2025).
- 30 Birnbaum JK, Murray CJ, Lozano R. Exposing misclassified HIV/AIDS deaths in South Africa. *Bull World Health Organ* 2011; **89**: 278–85.



- 31 Lozano R, Naghavi M, Foreman K, *et al.* Global and regional mortality from 235 causes of death for 20 age groups in 1990 and 2010: a systematic analysis for the Global Burden of Disease Study 2010. *Lancet* 2012; **380**: 2095–128.
- 32 Wang H, Abajobir AA, Abate KH, *et al.* Global, regional, and national under-5 mortality, adult mortality, age-specific mortality, and life expectancy, 1970–2016: a systematic analysis for the Global Burden of Disease Study 2016. *Lancet* 2017; **390**: 1084–150.
- 33 Calvert C, Ronsmans C. The contribution of HIV to pregnancy-related mortality: a systematic review and meta-analysis. *AIDS Lond Engl* 2013; **27**: 1631–9.
- 34 Figueroa-Damián R. Pregnancy outcome in women infected with the human immunodeficiency virus. *Salud Publica Mex* 1999; **41**: 362–7.
- 35 Ryder RW, Nsuami M, Nsa W, *et al.* Mortality in HIV-1-seropositive women, their spouses and their newly born children during 36 months of follow-up in Kinshasa, Zaïre. *AIDS* 1994; **8**: 667–72.
- 36 Zvandasara P, Saungweme G, Mlambo JT, Moyo J. Post Caesarean section infective morbidity in HIV-positive women at a tertiary training hospital in Zimbabwe. *Cent Afr J Med* 2007; **53**: 43–7.
- 37 Chilongozi D, Wang L, Brown L, *et al.* Morbidity and mortality among a cohort of human immunodeficiency virus type 1-infected and uninfected pregnant women and their infants from Malawi, Zambia, and Tanzania. *Pediatr Infect Dis J* 2008; **27**: 808–14.
- 38 Leroy V, Ladner J, Nyiraziraje M, *et al.* Effect of HIV-1 infection on pregnancy outcome in women in Kigali, Rwanda, 1992-1994. Pregnancy and HIV Study Group. *AIDS* 1998; **12**: 643–50.
- 39 Kourtis AP, Bansil P, McPheeters M, Meikle SF, Posner SF, Jamieson DJ. Hospitalizations of pregnant HIV-infected women in the USA prior to and during the era of HAART, 1994-2003. *AIDS* 2006; **20**: 1823–31.
- 40 Ticconi C, Mapfumo M, Dorrucchi M, *et al.* Effect of maternal HIV and malaria infection on pregnancy and perinatal outcome in Zimbabwe. *J Acquir Immune Defic Syndr* 1999 2003; **34**: 289–94.
- 41 Brown T, Peerapatanapokin W. The Asian Epidemic Model: a process model for exploring HIV policy and programme alternatives in Asia. *Sex Transm Infect* 2004; **80**: i19–24.
- 42 DerSimonian R, Laird N. Meta-analysis in clinical trials. *Control Clin Trials* 1986; **7**: 177–88.
- 43 Matthews LT, Kaida A, Kanters S, *et al.* HIV-infected women on antiretroviral treatment have increased mortality during pregnant and postpartum periods. *AIDS* 2013; **27 Suppl 1**: S105-112.
- 44 Westreich D, Maskew M, Evans D, Firnhaber C, Majuba P, Sanne I. Incident pregnancy and time to death or AIDS among HIV-positive women receiving antiretroviral therapy. *PLOS ONE* 2013; **8**: e58117.
- 45 Stover J. AIM: A computer program for making HIV/AIDS projections and examining the social and economic impact of AIDS. Glastonbury, CT: USAID, 2005.

- 46 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.
- 47 Lozano R, Lopez AD, Atkinson C, Naghavi M, Flaxman AD, Murray CJ. Performance of physician-certified verbal autopsies: multisite validation study using clinical diagnostic gold standards. *Popul Health Metr* 2011; **9**: 32.
- 48 Foreman KJ, Lozano R, Lopez AD, Murray CJ. Modeling causes of death: an integrated approach using CODEm. *Popul Health Metr* 2012; **10**: 1.
- 49 Bell RM, Koren Y. Lessons from the Netflix prize challenge. *ACM SIGKDD Explor Newsl* 2007; **9**: 75–9.
- 50 Bell RM, Koren Y, Volinsky C. All together now: A perspective on the NETFLIX PRIZE. *CHANCE* 2010; **23**: 24–24.
- 51 Flaxman AD, Vos T, Murray CJL. An integrative metaregression framework for descriptive epidemiology. Univ. Wash. Press. 2015; published online Oct.  
<https://uwapress.uw.edu/book/9780295991849/an-integrative-metaregression-framework-for-descriptive-epidemiology> (accessed July 7, 2023).
- 52 COVID-19 Forecasting Team. Forecasting the trajectory of the COVID-19 pandemic into 2023 under plausible variant and intervention scenarios: a global modelling study. 2023; : 2023.03.07.23286952.
- 53 Fragoso TM, Bertoli W, Louzada F. Bayesian model averaging: a systematic review and conceptual classification. *Int Stat Rev* 2018; **86**: 1–28.
- 54 COVID-19 Excess Mortality Collaborators. Estimating excess mortality due to the COVID-19 pandemic: a systematic analysis of COVID-19-related mortality, 2020-21. *Lancet* 2022; **399**: 1513–36.
- 55 GBD 2017 Population and Fertility Collaborators. Population and fertility by age and sex for 195 countries and territories, 1950-2017: a systematic analysis for the Global Burden of Disease Study 2017. *Lancet* 2018; **392**: 1995–2051.

## Section 8: Tables

Appendix 2, Table S1. Total number of site years by cardiovascular cause and source type for GBD2023										
Cause	Level	Vital Registration	Vital Registration - Sample	Verbal Autopsy	Surveillance	Survey/Census	Sibling History	MITS diagnosed	Cancer Registry	Police or Forensic Medicine Reports
Cardiovascular diseases	2	19710	953	1897	1			4		
Rheumatic heart disease	3	18904	953	187						
Ischemic heart disease	3	19203	953	1600						
Stroke	3	19705	953	1431	1					
Ischemic stroke	4	17395	953							
Intracerebral hemorrhage	4	17388	953							
Subarachnoid hemorrhage	4	17388	953							
Hypertensive heart disease	3	18264	953							
Non-rheumatic valvular heart disease	3	17558	953							
Non-rheumatic calcific aortic valve disease	4	16936	576							
Non-rheumatic degenerative mitral valve disease	4	16936	576							
Other non-rheumatic valve diseases	4	16827	576							
Cardiomyopathy and myocarditis	3	18265	953					4		
Myocarditis	4	17106	576					4		
Alcoholic cardiomyopathy	4	17106	576							
Other cardiomyopathy	4	17106	576					2		
Pulmonary Arterial Hypertension	3	16060	576							
Atrial fibrillation and flutter	3	16591	576							
Aortic aneurysm	3	17475	576							
Lower extremity peripheral arterial disease	3	16039	576							
Endocarditis	3	17468	953							
Other cardiovascular and circulatory diseases	3	18264	953							

**Appendix 2, Table S2. Top cardiovascular causes containing misclassified COVID-19 (reported) by location, 2020-2022**

Location Name	Cause Name	Amount of miscoded Covid deaths	Percent of total miscoded Covid deaths	Percent of total Covid (miscoded and reported)
Global	Ischemic heart disease	476816	34.1%	7.3%
Global	Stroke	112394	8.0%	1.7%
Global	Hypertensive heart disease	90414	6.5%	1.4%
Antigua and Barbuda	Lower extremity peripheral arterial disease	2	8.7%	1.4%
Antigua and Barbuda	Other cardiovascular and circulatory diseases	2	7.9%	1.3%
Argentina	Ischemic heart disease	3033	45.8%	2.0%
Argentina	Atrial fibrillation and flutter	631	9.5%	0.4%
Argentina	Hypertensive heart disease	387	5.8%	0.3%
Argentina	Other cardiovascular and circulatory diseases	46	0.7%	0.0%
Armenia	Ischemic heart disease	4053	61.8%	24.7%
Armenia	Cardiomyopathy and myocarditis	78	1.2%	0.5%
Armenia	Atrial fibrillation and flutter	43	0.7%	0.3%
Australia	Atrial fibrillation and flutter	328	10.5%	2.2%
Australia	Hypertensive heart disease	211	6.8%	1.4%
Australia	Stroke	160	5.1%	1.0%
Austria	Atrial fibrillation and flutter	84	5.9%	0.5%
Austria	Non-rheumatic valvular heart disease	34	2.4%	0.2%
Austria	Rheumatic heart disease	22	1.5%	0.1%
Belgium	Hypertensive heart disease	23	20.2%	0.1%
Bosnia and Herzegovina	Ischemic heart disease	729	35.2%	3.3%
Bosnia and Herzegovina	Cardiomyopathy and myocarditis	173	8.4%	0.8%
Bosnia and Herzegovina	Other cardiovascular and circulatory diseases	155	7.5%	0.7%
Bosnia and Herzegovina	Atrial fibrillation and flutter	64	3.1%	0.3%
Brazil	Stroke	18578	37.6%	2.4%
Brazil	Atrial fibrillation and flutter	711	1.4%	0.1%
Bulgaria	Hypertensive heart disease	3183	34.8%	6.9%
Bulgaria	Ischemic heart disease	2863	31.3%	6.2%
Bulgaria	Stroke	921	10.1%	2.0%
Bulgaria	Other cardiovascular and circulatory diseases	19	0.2%	0.0%
Bulgaria	Rheumatic heart disease	12	0.1%	0.0%
Canada	Ischemic heart disease	4020	54.5%	6.8%
Canada	Hypertensive heart disease	618	8.4%	1.0%
Chile	Ischemic heart disease	475	25.8%	0.9%
Chile	Hypertensive heart disease	285	15.5%	0.5%
Chile	Atrial fibrillation and flutter	124	6.8%	0.2%
Chile	Cardiomyopathy and myocarditis	26	1.4%	0.0%
Colombia	Ischemic heart disease	26014	59.4%	13.0%
Colombia	Hypertensive heart disease	1053	2.4%	0.5%
Colombia	Stroke	1011	2.3%	0.5%
Costa Rica	Hypertensive heart disease	38	5.5%	0.5%
Costa Rica	Atrial fibrillation and flutter	16	2.4%	0.2%
Costa Rica	Non-rheumatic valvular heart disease	12	1.8%	0.1%
Costa Rica	Endocarditis	9	1.3%	0.1%
Croatia	Cardiomyopathy and myocarditis	432	54.5%	3.1%
Croatia	Non-rheumatic valvular heart disease	73	9.2%	0.5%
Croatia	Lower extremity peripheral arterial disease	58	7.3%	0.4%
Cuba	Ischemic heart disease	14963	37.6%	31.3%
Cuba	Stroke	2746	6.9%	5.7%
Cuba	Hypertensive heart disease	2588	6.5%	5.4%

Location Name	Cause Name	Amount of miscoded Covid deaths	Percent of total miscoded Covid deaths	Percent of total Covid (miscoded and reported)
Cuba	Atrial fibrillation and flutter	520	1.3%	1.1%
Cyprus	Atrial fibrillation and flutter	44	11.1%	3.8%
Cyprus	Ischemic heart disease	38	9.7%	3.3%
Cyprus	Endocarditis	9	2.4%	0.8%
Cyprus	Aortic aneurysm	7	1.9%	0.6%
Cyprus	Cardiomyopathy and myocarditis	5	1.2%	0.4%
Czechia	Atrial fibrillation and flutter	248	12.2%	0.6%
Czechia	Non-rheumatic valvular heart disease	104	5.1%	0.3%
Czechia	Other cardiovascular and circulatory diseases	33	1.6%	0.1%
Denmark	Lower extremity peripheral arterial disease	23	23.9%	1.7%
Denmark	Stroke	20	20.2%	1.4%
Ecuador	Ischemic heart disease	13010	49.1%	17.3%
Ecuador	Stroke	1134	4.3%	1.5%
Ecuador	Atrial fibrillation and flutter	456	1.7%	0.6%
Ecuador	Hypertensive heart disease	372	1.4%	0.5%
Estonia	Hypertensive heart disease	557	41.4%	13.1%
Estonia	Non-rheumatic valvular heart disease	30	2.2%	0.7%
Estonia	Other cardiovascular and circulatory diseases	28	2.1%	0.7%
Estonia	Atrial fibrillation and flutter	18	1.4%	0.4%
Finland	Ischemic heart disease	39	8.8%	2.0%
France	Atrial fibrillation and flutter	124	22.2%	0.2%
France	Hypertensive heart disease	10	1.9%	0.0%
Georgia	Hypertensive heart disease	675	23.2%	3.7%
Georgia	Rheumatic heart disease	113	3.9%	0.6%
Georgia	Non-rheumatic valvular heart disease	45	1.5%	0.2%
Georgia	Cardiomyopathy and myocarditis	36	1.2%	0.2%
Georgia	Other cardiovascular and circulatory diseases	22	0.7%	0.1%
Germany	Hypertensive heart disease	11172	36.4%	5.6%
Germany	Rheumatic heart disease	706	2.3%	0.4%
Germany	Atrial fibrillation and flutter	488	1.6%	0.2%
Germany	Cardiomyopathy and myocarditis	362	1.2%	0.2%
Greece	Cardiomyopathy and myocarditis	154	29.8%	2.8%
Greece	Hypertensive heart disease	153	29.6%	2.7%
Grenada	Ischemic heart disease	50	52.2%	16.1%
Grenada	Stroke	7	7.6%	2.3%
Grenada	Hypertensive heart disease	6	6.4%	2.0%
Grenada	Other cardiovascular and circulatory diseases	3	3.6%	1.1%
Grenada	Aortic aneurysm	2	2.2%	0.7%
Guam	Ischemic heart disease	4	5.2%	1.1%
Guam	Hypertensive heart disease	2	2.7%	0.6%
Guatemala	Ischemic heart disease	11032	43.6%	21.1%
Guatemala	Stroke	1742	6.9%	3.3%
Hungary	Hypertensive heart disease	1114	26.0%	2.4%
Hungary	Atrial fibrillation and flutter	230	5.4%	0.5%
Iceland	Non-rheumatic valvular heart disease	23	33.1%	7.1%
Iceland	Other cardiovascular and circulatory diseases	3	4.1%	0.9%
Ireland	Ischemic heart disease	53	13.3%	2.2%
Ireland	Endocarditis	9	2.2%	0.4%
Ireland	Hypertensive heart disease	8	2.1%	0.3%
Israel	Endocarditis	8	3.7%	0.2%

Location Name	Cause Name	Amount of miscoded Covid deaths	Percent of total miscoded Covid deaths	Percent of total Covid (miscoded and reported)
Italy	Hypertensive heart disease	3517	12.6%	2.0%
Italy	Stroke	3299	11.8%	1.9%
Italy	Atrial fibrillation and flutter	559	2.0%	0.3%
Italy	Aortic aneurysm	102	0.4%	0.1%
Japan	Ischemic heart disease	5743	7.6%	4.0%
Japan	Stroke	3231	4.3%	2.2%
Kazakhstan	Ischemic heart disease	11945	30.4%	20.1%
Kazakhstan	Stroke	6229	15.8%	10.5%
Kazakhstan	Cardiomyopathy and myocarditis	2934	7.5%	4.9%
Kazakhstan	Hypertensive heart disease	1526	3.9%	2.6%
Latvia	Ischemic heart disease	641	47.5%	10.0%
Latvia	Hypertensive heart disease	218	16.1%	3.4%
Latvia	Cardiomyopathy and myocarditis	104	7.7%	1.6%
Latvia	Atrial fibrillation and flutter	59	4.4%	0.9%
Latvia	Stroke	56	4.2%	0.9%
Latvia	Pulmonary Arterial Hypertension	11	0.8%	0.2%
Lebanon	Ischemic heart disease	889	13.1%	5.6%
Lebanon	Stroke	694	10.2%	4.3%
Lebanon	Hypertensive heart disease	312	4.6%	2.0%
Lithuania	Stroke	604	22.0%	4.1%
Lithuania	Ischemic heart disease	578	21.0%	3.9%
Lithuania	Hypertensive heart disease	555	20.2%	3.8%
Lithuania	Atrial fibrillation and flutter	71	2.6%	0.5%
Luxembourg	Stroke	5	3.6%	0.4%
Malaysia	Endocarditis	15	100.0%	3.5%
Malta	Ischemic heart disease	6	9.6%	2.3%
Malta	Other cardiovascular and circulatory diseases	5	8.2%	2.0%
Mauritius	Ischemic heart disease	413	20.7%	13.5%
Mauritius	Stroke	52	2.6%	1.7%
Mauritius	Atrial fibrillation and flutter	45	2.3%	1.5%
Mauritius	Pulmonary Arterial Hypertension	16	0.8%	0.5%
Mauritius	Non-rheumatic valvular heart disease	13	0.7%	0.4%
Mexico	Ischemic heart disease	121757	47.7%	16.6%
Mexico	Hypertensive heart disease	2479	1.0%	0.3%
Mexico	Stroke	2125	0.8%	0.3%
Micronesia (Federated States of)	Stroke	4	17.4%	8.1%
Micronesia (Federated States of)	Atrial fibrillation and flutter	2	7.3%	3.4%
Mongolia	Stroke	127	20.4%	4.6%
Mongolia	Ischemic heart disease	67	10.8%	2.4%
Mongolia	Hypertensive heart disease	55	8.9%	2.0%
Netherlands	Atrial fibrillation and flutter	444	34.4%	0.9%
Netherlands	Stroke	162	12.6%	0.3%
Netherlands	Non-rheumatic valvular heart disease	50	3.9%	0.1%
Nicaragua	Ischemic heart disease	5777	41.2%	40.6%
Nicaragua	Hypertensive heart disease	433	3.1%	3.0%
Nicaragua	Other cardiovascular and circulatory diseases	184	1.3%	1.3%
North Macedonia	Ischemic heart disease	872	34.2%	7.0%
North Macedonia	Stroke	62	2.4%	0.5%
North Macedonia	Atrial fibrillation and flutter	53	2.1%	0.4%
North Macedonia	Non-rheumatic valvular heart disease	22	0.9%	0.2%

Location Name	Cause Name	Amount of miscoded Covid deaths	Percent of total miscoded Covid deaths	Percent of total Covid (miscoded and reported)
Northern Mariana Islands	Ischemic heart disease	10	27.2%	15.2%
Northern Mariana Islands	Stroke	4	10.5%	5.9%
Northern Mariana Islands	Atrial fibrillation and flutter	2	4.7%	2.6%
Norway	Hypertensive heart disease	622	23.6%	9.0%
Norway	Ischemic heart disease	293	11.1%	4.2%
Norway	Stroke	40	1.5%	0.6%
Oman	Hypertensive heart disease	23	10.5%	0.3%
Palestine	Hypertensive heart disease	25	6.0%	0.4%
Paraguay	Ischemic heart disease	393	14.6%	1.8%
Paraguay	Atrial fibrillation and flutter	20	0.8%	0.1%
Peru	Ischemic heart disease	9380	64.4%	8.5%
Peru	Hypertensive heart disease	642	4.4%	0.6%
Peru	Atrial fibrillation and flutter	122	0.8%	0.1%
Philippines	Ischemic heart disease	32271	25.7%	11.0%
Philippines	Hypertensive heart disease	26221	20.9%	8.9%
Philippines	Stroke	20412	16.2%	7.0%
Poland	Ischemic heart disease	35400	67.3%	16.0%
Poland	Hypertensive heart disease	4081	7.8%	1.8%
Poland	Stroke	1423	2.7%	0.6%
Poland	Other cardiovascular and circulatory diseases	372	0.7%	0.2%
Puerto Rico	Stroke	645	33.4%	9.4%
Puerto Rico	Hypertensive heart disease	261	13.5%	3.8%
Puerto Rico	Ischemic heart disease	257	13.3%	3.8%
Qatar	Stroke	65	23.1%	6.7%
Qatar	Ischemic heart disease	60	21.2%	6.1%
Republic of Korea	Stroke	2291	34.4%	17.8%
Republic of Korea	Ischemic heart disease	1013	15.2%	7.9%
Russian Federation	Ischemic heart disease	114284	43.6%	11.3%
Russian Federation	Stroke	24333	9.3%	2.4%
Russian Federation	Cardiomyopathy and myocarditis	11604	4.4%	1.1%
Russian Federation	Atrial fibrillation and flutter	2059	0.8%	0.2%
Russian Federation	Hypertensive heart disease	1389	0.5%	0.1%
Saint Lucia	Hypertensive heart disease	8	51.3%	51.3%
Saint Lucia	Atrial fibrillation and flutter	2	11.7%	11.7%
Saint Vincent and the Grenadines	Other cardiovascular and circulatory diseases	4	9.6%	9.6%
Serbia	Ischemic heart disease	399	6.6%	0.7%
Serbia	Atrial fibrillation and flutter	161	2.7%	0.3%
Serbia	Stroke	47	0.8%	0.1%
Singapore	Hypertensive heart disease	635	30.6%	15.7%
Singapore	Ischemic heart disease	406	19.6%	10.0%
Singapore	Stroke	299	14.4%	7.4%
Singapore	Lower extremity peripheral arterial disease	71	3.4%	1.8%
Singapore	Aortic aneurysm	46	2.2%	1.1%
Singapore	Cardiomyopathy and myocarditis	44	2.1%	1.1%
Slovakia	Ischemic heart disease	2221	32.7%	8.9%
Slovakia	Atrial fibrillation and flutter	201	3.0%	0.8%
Slovakia	Lower extremity peripheral arterial disease	28	0.4%	0.1%
Slovenia	Aortic aneurysm	12	10.0%	0.3%
Slovenia	Atrial fibrillation and flutter	7	6.4%	0.2%
Spain	Hypertensive heart disease	459	9.1%	0.4%

Location Name	Cause Name	Amount of miscoded Covid deaths	Percent of total miscoded Covid deaths	Percent of total Covid (miscoded and reported)
Switzerland	Other cardiovascular and circulatory diseases	10	5.0%	0.1%
Tunisia	Stroke	5176	15.8%	9.1%
Tunisia	Ischemic heart disease	3825	11.7%	6.8%
Tunisia	Hypertensive heart disease	1464	4.5%	2.6%
Ukraine	Ischemic heart disease	2679	27.5%	13.2%
Ukraine	Atrial fibrillation and flutter	200	2.0%	1.0%
Ukraine	Other cardiovascular and circulatory diseases	123	1.3%	0.6%
United Arab Emirates	Cardiomyopathy and myocarditis	32	18.8%	2.6%
United Arab Emirates	Endocarditis	4	2.5%	0.3%
United Kingdom	Hypertensive heart disease	239	16.4%	0.1%
United Kingdom	Rheumatic heart disease	198	13.5%	0.1%
United Kingdom	Endocarditis	121	8.3%	0.1%
United Kingdom	Ischemic heart disease	42	2.9%	0.0%
United Kingdom	Aortic aneurysm	28	1.9%	0.0%
United States Virgin Islands	Ischemic heart disease	7	15.9%	4.5%
United States Virgin Islands	Cardiomyopathy and myocarditis	4	7.9%	2.3%
United States of America	Ischemic heart disease	44814	27.0%	4.0%
United States of America	Hypertensive heart disease	22261	13.4%	2.0%
United States of America	Stroke	14571	8.8%	1.3%
Uruguay	Aortic aneurysm	2	3.2%	0.8%



## Section 9: CoD cause-specific modelling descriptions

**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	59
2. Alcoholic cardiomyopathy	61
3. Aortic aneurysm	63
4. Atrial fibrillation and flutter	65
5. Cardiomyopathy and myocarditis	70
6. Cardiovascular diseases	72
7. Hypertensive heart disease	75
8. Intracerebral haemorrhage	77
9. Ischaemic heart disease	79
10. Ischaemic stroke	82
11. Lower extremity peripheral arterial disease	84
12. Myocarditis	86
13. Non-rheumatic valvular heart diseases	88
14. Other cardiomyopathy	91
15. Other cardiovascular and circulatory diseases	93
16. Pulmonary arterial hypertension	96
17. Rheumatic heart disease	101
18. Stroke	103
19. Subarachnoid haemorrhage	105