

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

Appendix 4: Risk factor estimation methods

Preamble

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

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

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.¹ It includes detailed tables and information on data in an effort to maximise transparency in our estimation processes and provide a comprehensive description of analytical steps. We intend this appendix to be a living document, to be updated with each iteration of the Global Burden of Disease Study.

Portions of this appendix have been reproduced or adapted from the appendices of Lim et al 2012,² GBD 2015 Risk Factors Collaborators,³ GBD 2016 Risk Factors Collaborators,⁴ GBD 2017 Risk Factor Collaborators,⁵ GBD 2019 Risk Factors Collaborators,⁶ and GBD 2021 Risk Factor Collaborators.⁷ References are provided for reproduced or adapted sections.

Contents

List of methods appendix tables.....	5
Section 1: GBD overview	6
Section 1.1: Global Burden of Diseases, Injuries, and Risk Factors Study 2023	6
Section 1.2: Geographical locations of the analysis.....	6
Section 1.3: Time period of the analysis	6
Section 1.5: GBD risk factor hierarchy	6
Section 1.6: List of abbreviations	6
Section 1.7: Data input sources overview.....	9
Section 1.8: Funding sources	9
Section 2: Risk factor estimation	9
Overview	9
Step 1. Effect size estimation.....	10
Section 2.1.1: Criteria for inclusion of risk–outcome pairs ⁷	10
Section 2.1.2: Overview of the effect size estimation pathway	11
Section 2.1.3: Collate relative risk data	11
Section 2.1.4: Estimating the shape of the risk–outcome relationship.....	16
Section 2.1.6: Estimating the burden of proof risk function.....	22
Section 2.1.7: Evaluating potential for publication or reporting bias.....	23
Section 2.1.8: MR-BRT and temperature	23
Step 2. Exposure estimation ⁵	24
Section 2.2.1: Collate exposure data.....	24
Section 2.2.2: Adjust exposure data.....	25
Section 2.2.3: Estimate exposure	27
Step 3. TMREL ⁵	38
Step 4. Estimate population attributable fractions ⁵	38
Step 5. Estimate summary exposure values ⁵	39
Step 6. Mediation ⁵	40
Section 2.6.1: Summary.....	40
Section 2.6.2: Calculating the burden of multiple risk factors.....	41
Section 2.6.3: Computing mediation factors using linear relationships	42
Section 2.6.4: Adjusting for mediation	42
Section 2.6.5: Calculating mediation factor	43

Section 2.6.6: Piecewise aggregation (Pattern 3)	44
Section 2.6.7: Uncertainty of aggregated and mediated PAFs	46
Section 2.6.8: Important assumptions in aggregating risk factors and including mediation	46
Step 7. Estimate attributable burden ⁵	46
Section 3: Cardiovascular decomposition analysis of DALYs	47
Section 4: References	48
Section 5: Risk factor appendix tables	52
Section 6: GBD 2023 risk factor-specific modelling descriptions	53

List of methods appendix tables

Table S1. GBD cardiovascular risk hierarchy with levels

Section 1: GBD overview

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

The Global Burden of Diseases, Injuries, and Risk Factors Study (GBD) is a collaborative research effort aimed at estimating morbidity and mortality from a comprehensive set of diseases, injuries, and risk factors. The GBD Collaborator Network draws on the expertise of over 14,000 contributors from around the world. For this paper, we estimated risk factor exposure levels, relative health risk by exposure, and risk-attributable burden by age, sex, and location from 1990 to 2023.

Section 1.2: Geographical locations of the analysis

We produced estimates for 204 countries and territories that were grouped into 21 regions and seven super-regions. 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, the Philippines, Poland, Russia, South Africa, the United Kingdom, and the United States of America. All analyses are at the first level of administrative organisation within each country except for New Zealand (by Māori ethnicity), the Philippines (by provinces), and the UK. To meet data use requirements, in this publication we present subnational estimates for Brazil, India, Indonesia, Japan, Kenya, Mexico, and the USA; given space constraints, these results are presented in Appendix 3 instead of the main text. Subnational estimates for China are included in maps but are not reported in appendix tables. Subnational estimates for other countries will be released in separate publications, although please note that we only release estimates for a subset of these countries, per agreements with country partners.

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

Section 1.3: Time period of the analysis

A complete set of risk-specific exposures, relative risks (RRs), theoretical minimum risk exposure levels (TMRELs), and population attributable fractions (PAFs) were computed for the years 1990 to 2023.

Section 1.5: GBD risk factor hierarchy

The GBD 2023 cardiovascular risk factors hierarchy and levels are summarised in table S1. The risk hierarchy is based on common features of individual risks; for example, risk factors that represent behavioural factors are grouped together.

The GBD risk factor list continues to evolve to reflect the policy relevance, public health, and medical care importance of major risk factors. 2023

Section 1.6: List of abbreviations

APCSC Asia-Pacific Cohort Studies Collaboration

ARC	annualised rate of change
BMI	body-mass index
BoP	Burden of Proof
BPRF	Burden of Proof risk functions
BMD	bone mineral density
CDC	Centers for Disease Control and Prevention
CF	correction factor
CKD	chronic kidney disease
COD	causes of death
CODEm	Cause of Death Ensemble modelling
COPD	chronic obstructive pulmonary disease
COVID-19	coronavirus disease 2019
CRA	comparative risk assessment
CSV	comma-separated values
CRA	comparative risk assessment
CSMR	cause-specific mortality rate
CVD	cardiovascular disease
DALY	disability-adjusted life-year
DHS	Demographic and Health Survey
DRI	data representativeness index
EDU15+	mean education for those aged 15 years or older
EMR	excess mortality rate
FAO	Food and Agriculture Organization
FPG	fasting plasma glucose
GAM	generalised additive model
GATHER	Guidelines for Accurate and Transparent Health Estimates Reporting
GBD	Global Burden of Diseases, Injuries, and Risk Factors Study
GHDx	Global Health Data Exchange
GoF	goodness of fit
HAP	household air pollution
ID	iron deficiency
IDA	Iron-deficiency anaemia
IER	integrated exposure response
IHD	ischaemic heart disease
ILO	International Labour Organization
IPV	intimate partner violence
IQ	intelligence quotient
JMP	Joint Monitoring Project
KS	Kolmogorov-Smirnov
LDI	lag-distributed income
LDL	low-density lipoprotein
LMICs	low- and middle-income countries
LOESS	locally estimated scatterplot smoothing
LRI	lower respiratory infection

MCMC	Markov Chain Monte Carlo simulations
MDG	Millennium Development Goal
MF	mediation factor
MICS	Multiple Indicator Cluster Surveys
MoM	method of moments
MR-BRT	meta-regression—Bayesian, regularised, trimmed
NCD	non-communicable disease
NCD-RisC	Non-communicable Disease Risk Factor Collaboration
OER	observed-to-expected ratio
PAF	population attributable fraction
PDF	probability distribution factor
PM _{2.5}	particulate matter <2.5 µm in aerodynamic diameter
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PCS	prospective cohort study
RCT	randomised controlled trial
PURE	Prospective Urban and Rural Epidemiological Study
REDCap	Research Electronic Data Capture
RMSE	root mean square error
ROS	risk–outcome score
RR	relative risk
SARS-CoV 2	Severe acute respiratory syndrome coronavirus 2
SBP	systolic blood pressure
SD	standard deviation
SDG	Sustainable Development Goal
SDI	Socio-demographic Index
SEER	Surveillance, Epidemiology, and End Results Program
SEV	summary exposure value
SHS	secondhand smoke
SIR	smoking impact ratio
SSB	sugar-sweetened beverages
ST-GPR	spatiotemporal Gaussian process regression
SVAC	sexual violence against children
TB	tuberculosis
TFU25	total fertility rate in those under 25 years old
TMREL	theoretical minimum risk exposure level
TSNA	tobacco-specific nitrosamines
UI	uncertainty interval
USD	United States dollars
WaSH	Water, sanitation, and handwashing
WCRF	World Cancer Research Fund
WHO	World Health Organization
YLDs	years lived with disability
YLLs	years of life lost

Section 1.7: Data input sources overview

GBD 2023 incorporated a large number and wide variety of input sources to estimate mortality, causes of death and illness, and risk factors for 204 countries and territories from 1990 to 2023. These input sources are accessible through an interactive citation tool available in the GHDx [<https://ghdx.healthdata.org/>].

Users can retrieve citations for a specific GBD component, cause or risk, and location by choosing from the available selection boxes. They can then view and access GHDx records for input sources and export a comma-separated value (CSV) file that includes the GHDx metadata, citations, and information about where the data were used in GBD. Additional metadata for each input source are available through the citation tool as required by the GATHER statement.

The citation tool is available online via the GBD 2023 Sources Tool in the GHDx [<https://ghdx.healthdata.org/gbd-2023/sources>].

Section 1.8: Funding sources

This publication and the research it presents were funded by the Gates Foundation (OPP1152504); Bloomberg Philanthropies; Queensland Department of Health, Australia; UK Department of Health and Social Care; the Norwegian Institute of Public Health; the New Zealand Ministry of Health; 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: Risk factor estimation

Overview

The comparative risk assessment (CRA) conceptual framework was developed by Murray and Lopez,⁸ who established a causal web of hierarchically organised risks or causes that contribute to health outcomes, which allows for quantification of risks or causes at any level in the framework. In GBD 2023, as in previous iterations of the GBD study, we evaluated a set of behavioural, environmental and occupational, and metabolic risks, in which risk–outcome pairs were included based on evidence rules. These risks were organised in four hierarchical levels, where Level 1 represents the overarching categories (behavioural, environmental and occupational, and metabolic) nested within Level 1 risks; Level 2 contains both single risks and risk clusters (such as child and maternal malnutrition); Level 3 contains the disaggregated single risks from within Level 2 risk clusters (such as low birthweight and short gestation); and Level 4 details risks with the most granular disaggregation, such as for specific occupational carcinogens, the subcomponents of child growth failure (stunting, wasting, underweight), and suboptimal breastfeeding (discontinued and non-exclusive breastfeeding). At each level of risk, we evaluated whether risk combinations were additive, multiplicative, or shared common pathways for intervention. This approach allows the quantification of the proportion of risk-attributable burden shared with another risk or combination of risks and the measurement of potential overlaps between behavioural, environmental and occupational, and metabolic risks. To date in GBD, we have not

quantified the contribution of other classes of risk factors. We do provide some insights into the potential magnitude of distal social, cultural, and economic factors through an analysis of the relationship between risk exposures and development measured by using the Socio-demographic Index (SDI) (more details in appendix 1, section 3).

Two types of risk assessments are possible within the CRA framework: attributable burden and avoidable burden. Attributable burden is the reduction in current disease burden that would have been possible if past population exposure had shifted to an alternative or counterfactual distribution of risk exposure. Avoidable burden is the potential reduction in future disease burden that could be achieved by changing the current distribution of exposure to a counterfactual distribution of exposure. Murray and Lopez identified four types of counterfactual exposure distributions: (1) theoretical minimum risk; (2) plausible minimum risk; (3) feasible minimum risk; and (4) cost-effective minimum risk.⁹ The theoretical minimum risk exposure level (TMREL) is the level of risk exposure that minimises risk at the population level. Other possible forms of risk quantification include plausible minimum risk and feasible minimum risk. Plausible minimum risk reflects the distribution of risk that is conceivably possible and would minimise population-level risk if achieved. Feasible minimum risk describes the lowest risk distribution that has been attained within a population, and cost-effective minimum risk is the lowest risk distribution for a population that can be attained in a cost-effective manner. Because no robust set of forecasts for all components of GBD is available, in this study we focus on quantifying attributable burden by using the theoretical minimum risk counterfactual distribution. According to the definition of avoidable burden, risk reversibility would be incorporated into this type of assessment because it would involve reducing risk to the counterfactual for the index year, given a history of past risk exposure. Given the focus in this study on attributable burden, risk reversibility is not a criterion used in estimation here.

In general, this analysis follows the CRA methods used since GBD 2015.³ The methods described here provide a high-level overview of the analytical logic and focus on areas of notable change from the methods employed in GBD 2015 and since GBD 2021. Here we aim to provide sufficient detail on the methods and overall structure of the estimation process. This study complies with the GATHER recommendations proposed by the World Health Organization (WHO) and others, which include recommendations on documentation of data sources, estimation methods, and statistical analysis.¹

Step 1. Effect size estimation

Section 2.1.1: Criteria for inclusion of risk–outcome pairs⁷

From GBD 2010 until GBD 2019, we included risk–outcome pairs that met the World Cancer Research Fund (WCRF) grades of convincing or probable evidence.¹⁰ In this framework, convincing evidence consists of biologically plausible associations between exposure and disease established from multiple epidemiological studies in different populations. Evidentiary studies must be substantial, include prospective observational studies, and, where relevant, randomised controlled trials (RCTs) of sufficient size, duration, and quality that show consistent effects. Probable evidence is similarly based on epidemiological studies with consistent associations between exposure and disease, but for which shortcomings in the evidence exist, such as insufficient available trials (or prospective observational studies). For GBD 2023, we retained risk–outcome pairs included in GBD 2021; the majority of these were evaluated using the Burden of Proof (BoP) methodology introduced in GBD 2021, described in more detail in section 2.1.6 below. Risk–outcome pairs previously included in GBD 2021 were retained in GBD 2023 unless subsequent BoP analysis indicated exclusion. Entirely new risk factors were added

based upon a minimal one-star Burden of Proof risk function (BPRF) rating (methods on BPRF detailed in section 2.1.6) and review and majority vote for inclusion by the GBD Scientific Council. More specifically, our inclusion criteria were that the RR estimate's 95% uncertainty interval (UI), conventionally calculated, without accounting for unexplained between-study heterogeneity, must not cross the null RR value of 1 (ie, the mean RR estimate must be significantly higher [for harmful risks] or lower [for protective risks] than 1) for a risk–outcome pair to be included in GBD. To maintain stability in included risk factors and risk–outcome pairs between GBD cycles, exclusion criteria for those pairs already included in GBD 2019 were less stringent; previously included pairs were excluded only if the conventionally calculated 90% UI crossed the null.

Section 2.1.2: Overview of the effect size estimation pathway

For most relative risks, our meta-analytic approach followed six main steps: 1) search and extract data from published studies using a standardised approach; 2) estimate the shape of the exposure versus relative risk relationship, integrating over exposure ranges in different comparison groups and avoiding the distorting effect of outliers; 3) test and adjust for systematic biases as a function of study attributes; 4) quantify remaining between-study heterogeneity while adjusting for within-study correlation induced by computing relative risks for several alternatives with the same reference, as well as the number of studies; 5) evaluate evidence for small-study effects to evaluate a potential risk of publication or reporting bias; and 6) estimate the mean risk function as well as the BPRF. The BPRF quantifies a conservative interpretation of the average risk increase across the range of exposure supported by the evidence. The BPRF is summarized across exposure to compute the risk–outcome scores (ROS), which are then mapped into five categories of risk as star ratings. Zheng and colleagues¹¹ published the technical developments required to implement this approach, which are also disseminated using open-source Python libraries.^{12,13} Implementation details for each step of the approach used to find the ROS are described below. Custom models were used for some risk factors such as temperature (see GBD 2023 risk factor–specific modelling descriptions below for more details).

Section 2.1.3: Collate relative risk data

The relative risk (RR) by level of exposure or by cause for mortality or morbidity can be found in published and unpublished primary studies or in secondary studies that summarise RRs. We collated information from primarily RCTs, cohort, and pooled cohort studies; and in some instances, case-control studies. We used these data to determine the RR for the risk–outcome pairs included in GBD 2023. For most risks, data from pooled cohorts or meta-analyses of cohorts were used; in the case of the risk of cataracts from household air pollution (HAP), cohort data were not available, and instead we used case-control data. We estimated RRs of mortality and morbidity for 88 risk factors for which we determined attributable burden by using RR and exposure. To the extent possible, we incorporated RRs from studies that controlled for confounding but not for factors along the causal pathway between exposure and outcome. For risk–outcome pairs with evidence available for only one element of mortality or morbidity, we generally assumed that the estimated RRs applied equally to both. Given evidence of statistically different RRs for mortality and morbidity, we incorporated different RRs for each. Details and citation information for the data sources used for RRs are provided in searchable form through a web tool (<http://ghdx.healthdata.org/>). Available data sources for determining RRs varied across risks. Details on how RRs were calculated for each risk can be found in section 4.

Systematic review protocol for relative risks

Task	Protocol
Develop inclusion criteria and search string	<ul style="list-style-type: none"> • Develop inclusion/exclusion criteria for systematic review based on GBD definition and expert knowledge. • Develop search string in collaboration with GBD risk factor and cause teams and UW librarian (or host institution librarian as needed).
Identify existing meta-analysis / systematic review	<ul style="list-style-type: none"> • Use PubMed to identify existing meta-analysis/systematic review for risk-outcome pair. Criteria to identify meta-analysis: <ol style="list-style-type: none"> 1. PRISMA compliant – meta-analysis follows PRISMA reporting guidelines 2. Published in quality journal – journal ranks in the top two quartiles based on: https://www.scimagojr.com/journalrank.php?area=270 [scimagojr.com] (using the most appropriate subjects categories for the R-O pair) 3. Incorporates inclusion criteria that are the same as or more inclusive of final inclusion criteria for risk-outcome pair (e.g. include 'diarrhea' as outcome; whereas final inclusion criteria specifies 'WHO diarrhea definition') 4. Most recent
Pre-registration	<ul style="list-style-type: none"> • Pre-register systematic review on PROSPERO; example linked here.
Screening studies in meta-analysis	<ul style="list-style-type: none"> • One reviewer will be required to include an article in the data extraction phase, and two reviewers will be required to exclude an article. Discussion and consultation with senior personnel will occur as needed. • Other considerations for data sparse topics – if meta-analysis yields 5 or fewer studies for inclusion: <ol style="list-style-type: none"> 1. Screen at least 2 additional published meta-analyses. 2. If still less than 5 included studies, conduct full literature review.
Updated search	<ul style="list-style-type: none"> • Conduct updated literature search from date published meta-analysis completed its search to present day (if applicable) or conduct literature review for studies published from at least 1985 to present if completing a full literature review. • Use at least three databases in search that include PubMed (https://www.pubmed.gov), Embase + one topic-specific database • Additional databases to consider: <ul style="list-style-type: none"> ○ Web of Science ○ Scopus ○ CINAHL: Cumulative Index to Nursing & Allied Health Literature ○ PsychINFO ○ Cochrane Library ○ LILACS: Latin American and Caribbean Health Sciences Literature ○ CNKI: Chinese National Knowledge Infrastructure ○ Global Index Medicus ○ SciELO: Scientific Electronic Library Online

Title/abstract screening	<ul style="list-style-type: none"> De-duplicate articles from multiple databases using DistillerSR (or similar). For reviews utilizing a single screener, a second reviewer will duplicate screen 100 studies or 10%, whichever is higher, of <i>excluded</i> articles as a quality check. Resolve conflicts with full-text screening and third reviewer (content expert) as needed. If the second reviewer identifies 2 or more studies that were incorrectly excluded after resolving conflicts with full-text screening, then retrain first reviewer on inclusion/exclusion criteria and redo full title/abstract screening from the beginning. If dual screening is employed, then this is considered to be above the minimum standard outlined in this document, so different criteria for % agreement and kappa can be used.
Full-text screening	<ul style="list-style-type: none"> One reviewer will be required to include an article in the data extraction phase, and two reviewers will be required to exclude an article. Discussion and consultation with senior personnel will occur as needed. Other considerations: use team-level input to resolve edge cases; include specific information on how to handle data related to bias covariates and outcomes during team-specific training process.
The following steps apply to all included studies from meta-analyses and updated literature reviews:	
Duplicate cohorts or case-control (if applicable)	<ul style="list-style-type: none"> Check for duplicate cohorts or case-control. If duplicates exist, select study to include based on exposure time, follow-up period, and covariates included in RR estimation (or other pre-determined criteria).
Data extraction	<ul style="list-style-type: none"> Use one reviewer to extract data using the relative risk extraction template and validations. Use a second reviewer (content expert) to check for correctness and completeness of extracted data from at least 10% or 25 articles, whichever is higher.
Documentation and training	<ul style="list-style-type: none"> Complete PRISMA flowchart (2020 version) Complete REDCap documentation. REDCap is a centralized web-based database used to document all GBD systematic reviews. Complete RR systematic review extended documentation. <ul style="list-style-type: none"> Team-specific and GBD-wide trainings related to risk factors training, bundles, study design and measures of association, Distiller SR, evidence score/burden of proof.

Bias covariates: categories, conventions, and cases precluding inclusion of bias covariates

In the BoP analysis, each source of bias is represented as a binary bias covariate, taking on the value '0' if the study has the gold standard in the bias covariate, and otherwise taking on the value '1.' The gold standard should be defined prior to data extraction. A value must be assigned for each bias covariate for each study, and each bias covariate must have at least two studies per group ('0' and '1').

The bias covariate selection methodology is entirely data driven. That means that any bias covariate included in the data must have sufficient studies with labels of ('0') and ('1'); in particular, a user cannot evaluate a bias covariate with no gold standard representatives, or only gold standard representatives.

Data sparsity issues that preclude bias covariate inclusion based on available data are:

- A bias covariate that has only zeros or ones, or only one study with the value '0' or '1,' cannot be included in the analysis. For inclusion, every bias covariate must have some studies that are the gold standard (have a value of '0'), and there must be at least two studies in both groups ('0' and '1'). Even with the use of priors, the model cannot include variables for which there is no variation in the covariate across observations.
- Redundant bias covariates cannot be included. If two or more bias covariates all have the same labels across studies, or nearly all the same labels across studies, all but one of the redundant bias covariates must be removed.

There are six categories of bias covariates, based on the GRADE criteria, that must be evaluated in a BPRF analysis (assuming sufficient data as discussed above). It is possible to include more than one potential bias covariate within each category, again assuming there is sufficient data to do this. These six study-level bias sources are listed below, with multiple examples of potential bias covariates:

1. Representativeness of the study population: assesses whether the study participants are representative of the target population, in terms of demographics and other relevant factors.

- Typically named cov_representativeness
- '0' for studies whose results are likely generalizable to the total population: sample was based on the general population with reasonable exclusions for pre-existing disease states.
- '1' for studies performed on non-representative subpopulations, e.g., a high-risk group.

2. Exposure: assesses whether the exposure (i.e., risk factor) of interest is well-defined and measured accurately in the study.

- Typically named cov_exposure_quality
- '0' for studies whose exposure measurement characteristics are considered the gold standard for that risk-outcome pair and '1' otherwise.
- May be broken into subgroups if sufficient information exists for each subgroup:
 - cov_exposure_population: '0' for individual-level exposure and '1' for population-level exposure.
 - cov_exposure_selfreport: '0' for measurements based on assays, tests, or physician observations and '1' for self-report.
 - cov_exposure_study: '0' if exposure was measured multiple times and '1' for only a baseline measurement. Case-control studies should be scored as '1' unless a detailed exposure history was solicited that allows for quantification of variation in exposure.

3. Outcome: assesses whether the outcome of interest is clearly defined, clinically relevant, and accurately measured in the study.

- Typically named cov_outcome_quality
- '0' for studies whose outcome measurement characteristics are considered the gold standard for that risk-outcome pair and '1' otherwise.
- May be broken into subgroups if sufficient information exists for each subgroup:

- cov_outcome_selfreport: '0' if outcome measurement was based on death certificates, physician diagnosis or medical records and '1' if based on self-report.
- cov_outcome_unblinded: '0' if outcome assessment is blind to the individual level of exposure and '1' if unblinded.

4. Reverse causation: assesses the possibility that the observed relationship between the exposure and outcome could be due to the outcome causing the exposure, rather than the other way around.

- Typically named cov_reverse_causation
- '0' for studies where reverse causation was accounted for (there is minimal or no risk of reverse causation).
- '1' for studies where reverse causation was not accounted for (there is a risk of reverse causation).

5. Control for confounding: assesses whether the study design and analysis adequately account for potential confounding factors that could affect the relationship between the exposure and outcome.

- Typically named cov_confounder_quality
- '0' for studies where confounding factors were accounted for.
- '1' for studies where confounding factors were not accounted for.
- May be broken into subgroups if sufficient information exists for each subgroup:
 - cov_confounder_nonrandom: '0' if the study was randomized and '1' if the study was nonrandomized.
 - cov_confounder_uncontrolled: '0' for randomization or for a non-randomized study where the outcome is controlled for all major known confounders including age, sex, education, income, and other critical determinants of the outcome. '1' for non-randomized studies with control for only some determinants.

6. Selection bias: assesses whether the study sample was selected in a way that could introduce bias and whether efforts were made to minimize bias in the study design and analysis.

- Typically named cov_selection_bias
- '0' for studies with greater than 95% follow-up and '1' for studies with less than 95% follow-up.
- Case-control studies should be scored based on the percentage of cases and controls for which exposure data could be ascertained.

Additional bias covariates, following the same general criteria for collecting and assigning values, may be considered for certain risk factors. The decision to include any additional bias covariates is up to the modeler and their team leads. The names of the additional bias covariates may be defined arbitrarily but should reflect the content of the bias.

Because our methods to control for bias (see **Testing for bias across different study designs and characteristics** in Section 2.1.4 for further details) are data-driven and rely on high-level information available about input studies to meta-analyses (eg, which studies were gold standard vs. not), they are

able to test and adjust for bias, but are unable to capture this bias in a more nuanced manner. Moreover, if the data available are insufficient to generate bias covariates with adequate spread, bias adjustment cannot be performed. We are presently exploring enhanced diagnostic tools to assess covariate coverage and leverage across models to better assess bias adjustment feasibility and to identify where additional data collection or alternative methods are necessary. More efforts will be made in future GBD iterations to systematize bias coding across modelers and research teams.

Section 2.1.4: Estimating the shape of the risk–outcome relationship

Most classic epidemiological analyses of exposure or dose–response risk relationships have either assumed the relationship between risk and outcome to be log-linear or have converted continuous exposure variables into dichotomous exposure categories. This assumption simplifies the analysis considerably. Unfortunately, while assuming a log-linear relationship is analytically convenient and allows for the use of simple open-source tools,¹⁴ it is not necessarily biologically or clinically plausible (see model validation section for more details). For some risks, such as smoking, log-relative risk of outcome flattens at higher exposures. For others, such as BMI, the log-relative risk curves are J-shaped. We therefore chose to estimate the shape of the relationship directly from the data using a regularised spline.

In GBD 2021, for continuous and dichotomous risk factors, we modelled RRs using meta-regression—Bayesian, regularised, trimmed (MR-BRT) (see details below), relaxing the log-linear assumption to allow for monotonically increasing or decreasing non-linear functions using splines, typically with degrees 2 or 3 (quadratic and cubic splines, respectively). Monotonicity assumptions, which are less stringent than the log-linearity assumed in classic analyses, serve primarily to stabilize model fits in areas of sparse data. Any artifacts created by deviations from monotonicity are identified by examination of the funnel plot residuals. Risk factors for which we undertook this reanalysis include all dietary risk factors, low physical activity, kidney dysfunction, unsafe water and sanitation, no access to handwashing facility, particulate matter air pollution, lead exposure, vitamin A deficiency, secondhand smoke, bullying victimisation, high body-mass index, high fasting plasma glucose, and high alcohol use. In GBD 2023, we completed the reanalysis for sexual violence against children, intimate partner violence, and chewing tobacco. We did not conduct this reanalysis for risk factors with direct PAFs or PAFs=1 as well as select risk factors such as occupational risk factors, low bone mineral density, iron deficiency, suboptimal breastfeeding, and low birthweight and short gestation.

Because knot placement can affect the shape of the risk function when modelling with a cubic spline, we generated a wide range of knot placements and created an ensemble across these different knot placements. We also included in the final estimation 10% trimming of the data to avoid the results being sensitive to outliers.

Many meta-analyses convert RRs to per unit increase for convenience, particularly when studies choose different categories that could not otherwise be compared. If samples in the primary studies at high levels of exposure were sufficient to inform the shape of the tail of the distribution, we applied a cap to the maximum RR by using the midpoint of the last category for which a RR was reported.

Basis splines, measurement mechanism, and shape constraints

First, we used a Bayesian regularised spline to obtain the general shape of the non-linear relationship. Basis splines represent nonlinear curves as linear combination of recursively generated basis elements.¹⁵ The basis elements were recursively generated using piecewise smooth polynomials and were roughly

localised to certain regions of the exposure variable in the data. In most cases, quadratic or cubic polynomials were used, often with linear tails in the presence of sparse data. This approach allowed the use of the common restricted cubic spline, as well as constraints on the shape of the relationship (including non-decreasing and non-increasing).

Given basis functions f_1, \dots, f_k , the final curve is obtained as a β -linear combination

$$signal = \beta_1 f_1 + \dots + \beta_k f_k.$$

Specifically, for any given exposure x , the prediction using the spline model is given by

(1)

$$signal(x) = \beta_1 f_1(x) + \dots + \beta_k f_k(x) = \langle X, \beta \rangle$$

where X is a vector containing $(f_1(x), \dots, f_k(x))$. Derivatives and integrals of splines can likewise be expressed as linear combinations of spline coefficient $s\beta$. For additional details about B-splines see Zheng et al.¹¹

Many studies of exposure–response relationships report relative risks between categories defined by intervals of exposure. The relative risk between two exposure groups is a ratio of integrals of the spline across two specified intervals, so we used this exact non-linear mechanism to inform the fit.¹¹ Data from studies usually compare outcome rates in one exposure alternative group to those in a separate reference group. In mathematical notation, such observations are given by

(2)

$$y_{ij} = \frac{\frac{1}{d_{ij} - c_{ij}} \int_{c_{ij}}^{d_{ij}} f(x) dx}{\frac{1}{b_{ij} - a_{ij}} \int_{a_{ij}}^{b_{ij}} f(x) dx},$$

where y_{ij} is the reported relative risk corresponding to measurement j in study i , $[a_{ij}, b_{ij}]$ delineates the reference group exposure interval, and $[c_{ij}, d_{ij}]$ delineates the alternative group exposure interval. When $f(x)$ is represented using a spline, each integral is a linear function of β similar to (1). The observation model (2) is then a non-linear function given as the ratio of linear functions,

(3)

$$y_{ij} = f_{ij}(\beta) := \frac{\langle X_{ij}^1, \beta \rangle}{\langle X_{ij}^2, \beta \rangle}.$$

with the associated log-relative risk given by

(4)

$$\ln(y_{ij}) = \ln(\langle X_{ij}^1, \beta \rangle) - \ln(\langle X_{ij}^2, \beta \rangle).$$

Equation (4) is the main model used to infer the spline, and it is a simple but non-linear function of the spline coefficient β .

When studying exposure–response relationships, we allow for shape constraints of the inferred mean response. For example, for some harmful risks, such as smoking and air pollution, we allow the relative risk to be specified as monotonically increasing with exposure. In order to introduce each of these

constraints, we used the fact that all derivatives of splines are linear functions of spline coefficients, similar to (1).

Monotonicity. Monotonicity constraints can be imposed using linear inequality constraints based on exemplar exposures. Given an exemplar exposure x_i , the requirement that the slope of the spline at exposure x_i be non-negative can be formulated as

$$\langle X_{ij}, \beta \rangle \geq 0$$

for a particular vector X_i . Linear inequality constraints are strictly enforced by the optimisation solver used to fit the model; see Zheng et al.¹¹

Robust trimming strategy

To make the estimation of the overall relationship less sensitive to potential outlying studies or observations within studies, we applied a robust, likelihood-based statistical approach—least trimmed squares (LTS)¹⁶—to our mixed effects models.¹¹ The goal of robust statistical methods is to ensure that estimates are robust to such outlying observations. Trimming approaches form a subclass of robust statistical methods, and LTS was originally developed in the context of linear regression.¹⁷ LTS works by classifying observations into a majority of inliers and minority of outliers, while simultaneously fitting the model with respect to which the inlier/outlier classification is made. Compared with other robust approaches, such as M-estimators,¹⁸ trimming methods are more effective in limiting influence than outliers, and have a high breakdown point¹⁹; ie, the proportion of the data that can be arbitrarily corrupted before the estimator becomes invalid.

Trimming estimators have been applied to a broad range of problems, from linear regression¹⁶ to high-dimensional sparse regression and general machine learning problems.²⁰ In the context of mixed effects models, trimming methods are the most effective robust tools currently available for meta-analysis.¹¹ In practice, the approach requires only a specified inlier proportion, which was set to 90% across all examples, ie, we fit the 90% most self-coherent datapoints.

Using this approach, we trimmed 10% of the observations as part of the model fitting process, simultaneously discovering and fitting the most self-coherent 90% of the observations.¹¹ Numerical studies in data-rich cases have shown that quality of estimation is unaffected by trimming, even when there are no outliers in the data.²⁰ In the meta-analytic regime, the 90% level is a heuristic that balances the sparsity of available data with the need to improve estimates in the presence of outliers. As noted below, this step also substantially decreased the number of risk–outcome pairs with evidence of residual publication or reporting bias.

Spline ensemble

Third, to make non-linear risk function estimates robust to knot placement, we created 50 models based on random knot placement samples. Spline estimates depend on the choice of spline parameters, including spline degree, number of knots, and knot placement. To mitigate the effect of spline parameter selection on results, we developed an ensemble approach over knot placement, so that the modeller only had to specify the spline degree and number of knots.

Given the degree and number of knots, we automatically sampled a set of knot placements for a feasible knot distribution (described below). For each knot placement, we fit a spline (including non-linear measurements, shape constraints, and trimming as discussed above), evaluated each resulting model by

computing its fit and curvature, and aggregated the final model as a weighted combination of the ensemble.

Sampling knots from simplex. We used a minimal set of rules that describe a feasible set from which to sample knots, and sample from this set uniformly. Given a number of knots, the rules specify feasible ranges for each knot and feasible gaps between knots. The set of knot placements that satisfy these four rules form a closed polyhedron (a volume in high-dimensional space delineated by hyperplanes). We calculated the vertices of the polyhedron using the double description method²¹ and uniformly sampled knot placements from within the polyhedron. Each knot placement yielded a model, fit using the trimmed constrained spline approach described above.

Ensemble performance evaluation. Once the ensemble was created, we scored the resulting risk curves using two criteria: model fit (measured using the log-likelihood) and total variation (measured using the highest-order derivative). These scores balanced competing objectives of fit and generalisability. Once we had these scores, we normalised them to the range $[0,1]$ and applied a logistic transformation. The transformation was used to make the scoring meaningful even in the presence of spurious curves in a large ensemble. We then multiplied the scores to down-weight models that were low under either criterion (fit or total variation). The final weights were normalised to sum to 1. Using a weighted combination of these metrics, we weighted the 50 models to create the ensemble model.

New non-linear covariates

Fourth, for risk–outcome pairs with non-linear relationships, we evaluated exposure levels since this information matters for non-log-linear pairs. To do this, we took advantage of the spline model and directly captured the typical data-generating mechanism. Specifically, we used the final model that we had estimated using the robust spline ensemble to generate a non-linear dose–response curve, which we encoded into new non-linear “signal” covariates that were later used to enable linear mixed effects analyses. Once the non-linear estimation was complete, the log-relative risk for each datapoint was a function of four parameters:

$$F(a_{ij}, b_{ij}, c_{ij}, d_{ij}) = \frac{\frac{1}{d_{ij} - c_{ij}} \int_{c_{ij}}^{d_{ij}} \hat{f}(x) dx}{\frac{1}{b_{ij} - a_{ij}} \int_{a_{ij}}^{b_{ij}} \hat{f}(x) dx}$$

where \hat{f} is the non-linear function obtained by estimating spline coefficients $\hat{\beta}$, see (4), $[a_{ij}, b_{ij}]$ delineates the reference group exposure interval, and $[c_{ij}, d_{ij}]$ delineates the alternate group exposure interval.

We produced non-linear covariates for fixed and random effects. The non-linear fixed effects covariate, denoted $signal^f$, is given by

(5)

$$signal_{ij}^f = F(a_{ij}, b_{ij}, c_{ij}, d_{ij}).$$

The new non-linear random effect covariate, denoted by $signal^r$, is given by

(6)

$$signal_{ij}^r = F(t, t, c_{ij}, d_{ij}),$$

where t denotes a fixed reference, eg, the theoretical minimum risk exposure level (TMREL). We used these new covariates in linear mixed effects models in further stages of analysis:

(7)

$$y_{ij} = \text{signal}_{ij}^f \beta_s + \text{signal}_{ij}^r u_i + \epsilon_{ij}$$

where $\epsilon_{ij} \sim N(0, \sigma_{ij}^2)$ are known by each observation, β_s is a scalar linear covariate multiplier on the signal^f covariate, and u_i is a random study-specific slope on the signal^r covariate with unknown variance γ . The posterior for β_s in (7) was used as a reference for the prior in bias covariate selection, described in step 3.

For our visualisations (figures 1A–6A), we plotted each datapoint with x-value at the midpoint exposure of the alternative group, and y-value corresponding to the sum of the log relative risk and estimated curve evaluated at the midpoint of the reference group. These visualisations allow the standard assessment of fit quality, with a perfect fit corresponding to the estimated non-linear relationship passing through the data.

Testing for bias across different study designs and characteristics

Following the approach of the GRADE criteria,²² we quantified common sources of bias across six general domains: representativeness of the study population, exposure assessment, outcome ascertainment, reverse causation, control for confounding, and selection bias. In the illustrative cases presented here, these variables were quantified for each study during the study extraction phase. For the set of studies on a risk–outcome association, we tested systematic variation as a function of these risk of bias variables through meta-regression. We converted the dose–response relationship identified in step 1 into a new “signal” covariate, effectively linearising the non-log-linear relationship. For each bias covariate x (coded as an indicator variable), we defined a corresponding interaction covariate (ie, an effect modifier):

$$y_{ij} = \text{signal}_{ij}^f \times (\beta_s + x_{ij}^1 \beta_1 + \dots + x_{ij}^k \beta_k) + \epsilon_{ij}$$

that modified the slope of the “signal” covariate. We then tested risks of bias of the effect modifiers through linear meta-regression. To be included, every bias covariate must have some studies that are the “gold standard” (ie, at the standard of the best studies that have been conducted) for that covariate; otherwise, it is not possible to incorporate it into the regression framework. Further, in considering potential covariates, we enforced that every categorical covariate had at least two studies in each category. Since bias covariates were already study-specific, we only considered the fixed-effects model in bias covariate selection.

We used a robust approach to test for bias that limited the risk of over-interpreting differences with limited numbers of studies. We used the Lasso^{23,24} approach—which augments the least squares loss typically solved in a linear regression by penalising the sum of absolute values of the bias covariate multipliers—to obtain a ranked list of bias covariates using the following equation:

(8)

$$\min_{\beta} \sum_{i,j} \frac{1}{2\sigma_{ij}^2} (y_{ij} - \text{signal}_{ij}^f \times (\beta_s + x_{ij}^1 \beta_1 + \dots + x_{ij}^k \beta_k))^2 + \frac{1}{2} \beta^T \Sigma^{\{-1\}} \beta + \lambda \|\beta\|_1$$

where β contains specifically bias covariate multipliers, Σ is a diagonal matrix linked to the posterior on β_s from the basic linear model (7), and the term $\lambda \|\beta\|_1$ penalises the sum of the absolute values, pushing the bias covariate multipliers β to 0, with a strength determined by λ .²⁴

We then selected bias covariates based on their Lasso ranking, obtained by sweeping through from high to low values of λ in (8). We then added the selected covariates to the linear meta-regression model one at a time, following this ranking. To stabilise the selection process and follow through on the “burden of proof” philosophy, we tested for significance of covariates using a Gaussian prior that biased all bias coefficients to 0 with a strength proportional to the posterior of the main dose–response relationship. If the coefficients were significant, they stayed in the model as the process continued; we terminated the process when the last added bias covariate was no longer significant after accounting for “signal” and higher-ranked covariates in the model. We predicted the risk function using the values of the included bias covariates that reflected the preferred level of the covariate, such as the highest level of control for confounding.

Section 2.1.5: Fitting the mean and quantifying between-study heterogeneity. Estimation of between-study heterogeneity is an important aspect of meta-analysis. It reflects the variation between studies and consistency across literature. In the following, we describe how we used the signal and bias covariates obtained in steps 2.4 and 3 to build a simple linear mixed effects model to capture the between-study heterogeneity.

After the selection procedure, we fit a final linear mixed-effects model that included the “signal” as well as selected bias covariates. Division by a common referent in the typical measurement mechanism induces correlation, (via an intercept shift in log-relative risk space); we therefore used a random intercept in the mixed-effects model to account for this induced within-study correlation. To capture the between-study heterogeneity, we used a study-specific random slope with respect to the “signal” model so that the random effect for each study effectively scaled the non-linear relative risk curve. Formally, we fit a linear mixed effects model of the form

$$y_{ij} = \text{signal}_{ij}^f \times (\beta_s + x_1\beta_1 + \dots + x_k\beta_k) + \text{signal}_{ij}^f u_i + \epsilon_{ij}$$

where $\epsilon_{ij} \sim N(0, \sigma_{ij}^2)$ are the reported observation standard errors, and u_i are random effects with a common unknown variance,

$$u_i \sim N(0, \gamma).$$

Parameters β and γ were estimated simultaneously using maximum likelihood; see Zheng et al¹¹ for more details. The fixed effects portion of the formula estimates the mean log-relative risk. We used the same prior on bias covariates in this analysis as we used in equation (8), ie, $\beta \sim N(0, \Sigma)$. For log-linear relative risks, this modelling choice reduced to the classic analysis, where the random slope with respect to exposure was equivalent to the random intercept for log-linear relative risk.

To account for the small studies problem—where in the setting of small numbers of studies, between-study heterogeneity (γ) can easily be under-estimated,²⁵ and in particular the estimate may be zero when too few studies are available—we quantified the uncertainty in heterogeneity estimation.²⁶ This estimate allowed a quantile of the heterogeneity parameter to be used, increasing the robustness of the

estimate against the small study problem. Among several alternatives in the literature,^{27,28} we used the Fisher information matrix (FIM)²⁷ to estimate the uncertainty of the between-study heterogeneity. The FIM is weakly dependent on observed data but is sensitive to the non-linear relationship, selected bias covariates, reported standard errors, and the number of studies. The final uncertainty intervals we report are composed of two components: (1) posterior uncertainty corresponding to fixed effect β_s , and (2) 95% quantile of γ , which depends on the estimate of γ and the estimate of the variance of γ using the inverse of Fisher information.

Section 2.1.6: Estimating the burden of proof risk function

The combined uncertainty for the mean, estimated between-study heterogeneity, and 95th quantile of the between-study heterogeneity obtained from the FIM estimate were used to generate a BPRF. The BPRF is defined as either the 5th (for harmful risks) or 95th (for protective risks) quantile curve closest to the line of relative risk equal to 1 (the null) and can be interpreted as the smallest harmful or protective effect at each level of exposure consistent with the available evidence.

In the range of exposures defined by the 15th and 85th percentiles of exposure levels observed for each risk across available studies, the ROS is defined as the signed value of the average log BPRF. For example, a log BPRF of 0.4 for a harmful risk (where null = 0) and a log BPRF of -0.4 for a protective risk would both have an ROS of 0.4 because the magnitude of the log relative risk is the same. In contrast, for risk-outcome pairs with a BPRF opposite the null from the mean risk (ie, the BPRF suggests that the relationship is opposite of the expected relationship—a BPRF below 1 for a harmful risk and a BPRF above 1 for a protective risk), ROS would be calculated as negative.

Network analysis

Some relative risks were modelled using network analysis. Network analysis is a special case of the mixed effects linear model that is used to compare multiple treatment effects, for example, in the case of drinking water. To explain the coding, we use an example with four treatments A, B, C, D .

For simplicity, assume A is this reference treatment. We then have the following coding.

$$AB \rightarrow B - A : \quad [1 \quad 0 \quad 0]$$

$$AC \rightarrow C - A : \quad [0 \quad 1 \quad 0]$$

$$AD \rightarrow D - A : \quad [0 \quad 0 \quad 1].$$

We see from this example that the design matrix under the basic network assumption is always full rank, since a subset of rows forms the identity matrix.

Comparisons that do not include the reference can be computed. For example,

$$\begin{aligned} BC \rightarrow C - B &= (C - A) - (B - A) \\ &= [0 \quad 1 \quad 0] - [1 \quad 0 \quad 0] \\ &= [-1 \quad 1 \quad 0] \end{aligned}$$

Using this simple algebra, we obtain the remaining codings.

$$BC \rightarrow C - B : \quad [-1 \quad 1 \quad 0]$$

$$BD \rightarrow D - B : \quad [-1 \quad 0 \quad 1]$$

$$CD \rightarrow D - C : \quad [0 \quad -1 \quad 1]$$

Each row of the design matrix \mathbf{X} is coded according to the comparison.

When doing network analysis, the design matrix \mathbf{X} does not include the intercept term ($\mathbf{1}$ column).

Section 2.1.7: Evaluating potential for publication or reporting bias

A significant association between mean effect and standard deviation may indicate potential for publication or reporting bias, or methodological differences between large and small studies, which likewise lead to biased results. Publication bias is an important issue in meta-analysis,²⁹ and a formal test is typically done in addition to visual inspection of the funnel plot to decrease the chances of flagging apparent bias due to chance alone. In the proposed approach, we checked whether the standard deviations were significant predictors of the observations in the presence of the “signal” and bias covariates. To detect publication bias, we used a data-driven approach known as Egger’s regression.³⁰ The approach detects if there is a significant correlation between the residuals and their standard deviations. When Egger’s regression failed to detect significant evidence of publication bias, we terminated the process. While we identified these pairs as having potential for publication or reporting bias, we followed the general literature and did not incorporate any correction to the risk function based on this finding.

Section 2.1.8: MR-BRT and temperature

While meta-regression of literature studies was applied to estimate relationships for risk–outcome pairs, for temperature, we conducted primary analysis of relationships with cause-specific mortality as described previously.³¹ The relative risk, RR, of mortality was calculated for each daily and mean annual temperature category in each administrative unit. For this purpose, we calculated the daily mean temperature and aggregated the daily cause-specific death counts for each administration. We then calculated mortality rates for each cause, c , location, l , and daily mean temperature, ie, temperature category, t :

$$MR_{clt} = \frac{deaths_{clt}}{person - days_{lt}}$$

With MR representing the mortality rate, $deaths$ being the absolute number of cause-specific deaths, and $person-days$ depicting the sum of the population in location, l , across all days with a daily temperature of t .

Following, we calculated the mean MR, \overline{MR} , for each cause, c , and location, l :

$$\overline{MR}_{cl} = \frac{deaths_{cl}}{person - days_l}$$

The daily temperature-specific mortality rate ratio, MRR , was then calculated as the ratio of the MR for each temperature category, location, and cause, and the average \overline{MR} :

$$MRR_{clt} = \frac{MR_{clt}}{MR_{cl}}$$

In order to aggregate the *MMRs* to the first-level administrative unit, we calculate the population-weighted mean temperature (PWMT) for each location and across all days and then pooled all *MRRs* for each combination of daily temperature and PWMT.

Step 2. Exposure estimation⁵

Section 2.2.1: Collate exposure data

Systematic reviews

For GBD 2023, we conducted updated systematic literature reviews of risk factor exposure for three risks (lead exposure, high fasting plasma glucose, and high systolic blood pressure). For other risk factors, only a fraction of the existing data appears in the published literature, and other sources predominate, such as survey, measurement, or satellite data. Data were systematically screened from household surveys archived in the GHDx (<http://ghdx.healthdata.org>), including Demographic and Health Surveys, Multiple Indicator Cluster Surveys, Living Standards Measurement Surveys, and Reproductive Health Surveys. Other national health surveys were identified based on survey series that had yielded usable data for past rounds of GBD, sources suggested to us by in-country collaborators, and surveys identified in major multinational survey data catalogs, such as the International Household Survey Network and the WHO Central Data Catalog, as well as through country ministry of health and central statistical office websites. Certain risks, such as poor diet and excessive alcohol consumption, also incorporated administrative record systems. Citations for all data sources used for risk factor estimation in GBD 2023 are provided in searchable form through a web tool (<http://ghdx.healthdata.org>). A description of the search terms employed for risk-specific systematic reviews are detailed by cause in appendix section 4.

Information on systematic reviews were managed by using Research Electronic Data Capture (REDCap) electronic data capture tools hosted at the University of Washington.³² REDCap is a secure, web-based application designed to support data capture for research studies that provides 1) an intuitive interface for validated data entry; 2) audit trails for tracking data manipulation and export procedures; 3) automated export procedures for seamless data downloads to common statistical packages; and 4) procedures for importing data from external sources.

Search terms

Search terms for updates of systematic reviews for GBD 2023 are shown by risk factor in section 4.

Survey data preparation

Survey data constitute a substantial part of the underlying data used in the estimation process. During extraction, we concentrated on demographic variables (such as location, gender, age), survey design variables (such as sampling strategy and sampling weights), and the variables used to define the population estimate (such as a prevalence or a proportion) and a measure of uncertainty (standard error, confidence interval, or sample size and number of cases).

Section 2.2.2: Adjust exposure data

Compiled several adjustments were applied to extracted exposure sources to make the data more consistent and suitable for modelling. In GBD 2023, we implemented adjustments of risk exposure data to deal with alternative case definitions or study methods prior to entering data into our main analytical tools of DisMod-MR 2.1, ST-GPR, and MR-BRT. This decision also included the adjustment of data presented for both sexes to a male and female equivalent. The starting point was to explicitly state the reference case definition and study method and identify alternative definitions and study characteristics that fall without our inclusion criteria.

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

We pooled either the logit difference between alternative and reference or the natural log of the ratio of alternative to reference. From simulations, we found that the two methods provide almost identical results for quantities that after adjustment do not exceed a value of 0.5 (eg, prevalence or proportion). The logit difference method much better dealt with higher values and avoided prevalence or proportions to exceed 1. If the values of either the reference or alternative were zero, we aggregated values across age groups until both values had non-zero observations. We used the delta method to compute the standard error of the reference and alternative measures in logit space. The standard error of the logit difference was computed as the square root of the sum of the variances of each datapoint in a pair.

Data analysis

We used a network random effects meta-regression in MR-BRT (see section 2.1.4). In a network analysis, if A is the reference and B and C are two alternatives, a comparison of A versus B and B versus C provides an indirect comparison of the alternative C against the reference A. To implement the network, we included dummy variables with a particular structure. This was implemented as follows, where A is the reference definition/method:

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

For example:

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

The coding structure outlined above in step 1 assumes that all case definitions are mutually exclusive. In some cases, however, individual case definitions are a function of different components or dimensions. For example, case definitions may vary by the type of symptoms that a respondent experiences as well as the recall period over which those symptoms are experienced. In the presence of sparse data, it may be difficult to find both direct and indirect comparisons of all individual case definitions. In these cases, an alternative approach is to assume different dimensions of case definitions have a multiplicative effect. In other words, the effect of recall period has the same relative effect across different categories of symptoms reported by respondents. To implement this coding scheme:

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

In MR-BRT, we ran random effects meta-regression of the logit difference (or log ratio) with all the k dummy variables as covariates, omitting the intercept in the meta-regression. We used a study_id

variable for be the unique combination of the NIDs of the reference and alternative studies (or alternative1 to alternative2). The coefficients on the k dummy variables represent the pooled logit difference of the k alternative definition to the reference taking into account evidence from both direct and indirect comparisons. In the example above, the coefficient on DummyA is the pooled logit difference of B minus A; the coefficient on DummyB is the pooled logit difference of C minus A. The standard error of the pooled logit difference incorporating the between-study variance was calculated as:

$$se(\text{logit}(\text{difference}_k)) = \sqrt{\text{var}_k + \gamma^2}$$

Where:

$se(\text{logit}(\text{difference}_k))$ = standard error of the pooled logit difference of alternative k to the reference

var_k = variance of the coefficient on dummy variable k

γ^2 = between-study variance

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

We also examined whether there were systematic differences in the adjustments by key demographics (age, sex, geographical location, year) and other potential factors that may lead to variation in crosswalks. This could only be done at present in a direct comparison model and not in a network. We did this when there was a strong rationale, eg, biological plausibility, for variation by such characteristics.

After obtaining the pooled logit difference or log ratio estimates, we predicted adjustments based on the statistical model, including uncertainty in the adjustment and sampling error of each datapoint. For non-significant logit differences or log ratios, we still applied the adjustments if there was a conceptual reason to believe that the alternative definition is biased. This expands the variance of these alternative definition datapoints.

Section 2.2.3: Estimate exposure

Mean exposure estimation

Once data were collected and compiled, the next step of the analytical flowchart was to apply adjustments, where necessary, to correct for bias. Examples of these adjustments include use of urban studies for lead; crosswalks between different measurements, methods, and definitions, such as for self-report of obesity and glycated haemoglobin (HbA_{1c}) for diabetes; and age-sex splitting of data, such as

for fasting plasma glucose (FPG) level, cholesterol level, and systolic blood pressure that may be reported from broad age groups.

For the GBD, we developed two modelling approaches, a Bayesian meta-regression model (DisMod-MR 2.1) and a spatiotemporal Gaussian process regression model (ST-GPR), to pool data from different sources, control and adjust for bias in data, and incorporate other types of information such as country-level covariates. DisMod-MR 2.1 and ST-GPR are mixed effect models that borrow information across age, time, and locations to synthesise multiple data sources into unified estimates of levels and trends. A detailed description of the likelihood used for estimation and a full description of improvements made for DisMod-MR 2.1 were detailed by Vos and colleagues,³³ who provided additional detail in the appendix to that paper.²⁰ The ST-GPR model has three main hyper-parameters that control for smoothing across time, age, and location. Values for these hyper-parameters were selected on the basis of cross-validation. Cross-validation tests were conducted for different combinations of the hyper-parameters for three types of models: one data-sparse model, one data-moderate model, and one data-dense model. In each test, 20% of the data were held out, and the performance of each combination of hyper-parameters was evaluated on the held-out data. For each hyper-parameter combination, ten cross-validation tests were conducted. The performance of each model in predicting the withheld 20% of the data was evaluated by using a combined measure based on root mean square error (RMSE) and uncertainty interval (UI) coverage. A detailed description of the ST-GPR process regression can be found below.

The main difference between these methods is their power to include unstructured types of data by sex and age group and their degree of flexibility. DisMod-MR 2.1 is the preferred tool in these cases because of its ability to integrate over age and adjust for different exposure definitions in the data; however, the use of Bayesian Markov Chain Monte Carlo (MCMC) simulations with large volumes of data renders the analysis computationally intensive and reduces the number of iterations that are possible. If standard age-group data are available – as is generally the case for metabolic risks – using ST-GPR becomes the preferred approach.

In some cases, we adapted our methods of modelling exposure to risks where necessary to account for complexities in the risk–outcome relationship or the need for particular handling of data, for example, dietary risks and ambient air pollution (see appendix section 4 for more detail). Additional details for adjustments or adaptations to particular risk models are provided in appendix section 4.

DisMod-MR 2.1 description

Until GBD 2010, non-fatal estimates in burden of disease assessments were based on a single data source on prevalence, incidence, remission, or a mortality risk selected by the researcher as most relevant to a particular location and time. For GBD 2010, we set a more ambitious goal: to evaluate all available information on a disease that passes a minimum quality standard. That required a different analytical tool that would be able to pool disparate information presented for varying age groupings and from data sources by using different 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 by world regions with UIs by using Bayesian statistical methods. For GBD 2013, the improved DisMod-MR 2.0 increased computational speed, which allowed computations to be consistent between all disease parameters at the country rather than the region level. The

hundred-fold increase in speed of DisMod-MR 2.0 was partly due to a more efficient rewrite of the code in C++ but also to changing to a model specification by 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.³⁴

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

For GBD 2016, the computational engine (DisMod-MR 2.1) remained substantively unchanged from GBD 2015. We changed the prediction year set to generate fits for the years 1990, 1995, 2000, 2005, 2010, and 2016. We updated the age prediction sets to include age groups 80–84 years, 85–89 years, 90–94 years, and 95+ years to comply with changes across all functional areas of the GBD. We also expanded the set of locations where subnational units are modelled; the set now includes Brazil, China, England, India, Indonesia, Japan, Kenya, Mexico, South Africa, Sweden, and the USA.

In GBD 2017, we continued to use DisMod-MR 2.1 because no substantial changes were made. Updates to computation include extending the terminal prediction year to 2017 and additional subnational units in Ethiopia, Iran, New Zealand, Norway, and Russia. Saudi Arabia was also modelled only at the national level in 2017.

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

DisMod-MR 2.1 simultaneously fits several epidemiologic measures using a Bayesian, nonlinear, mixed-effects regression model, which produces estimates of a consistent set of epidemiologic measures in a compartmental disease model, meaning that they are the solution to the set of differential equations that specify the model.

A three-compartment model of a disease in a population consisting of susceptibles, i.e. those in the population without disease, cases of the disease and deaths. Individuals move from susceptible to case by the incidence rate i , and can return to susceptible by remission rate r . Cases are susceptible to a mortality rate, m_i , that reflects the excess mortality in cases of the disease. Cases and susceptibles are subject to a mortality rate from other causes, m_o . Each quantity in figure 15 is a function of age and time, governed by the following differential equations:

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

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

where:

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

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

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

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

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

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

a and t denote age and time.

A Monte Carlo Markov Chain simulation starts with the prevalence at birth, and samples the posterior distribution integrating the observed values for all observations of all epidemiologic quantities, ensuring that susceptibles and cases sum to population size, and that the partial derivatives of each differential equation in the system with respect to time zero. For most diseases the birth prevalence is set to zero. For disease that can be present at birth, a non-zero birth prevalence is specified. Population sizes for all combinations of year, age, sex, and location are input to DisMod-MR 2.1 from the latest GBD estimates, which are published iteratively and available by download from the GBD results tool.

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

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

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

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

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

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

where,

ACMR is the all-cause mortality rate

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

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

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

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

where

y_j is a “measurement value” (ie, datapoint)

Φ denotes all model random variables

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

a_j is the adjusted measurement for 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) and

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

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

Where:

k denotes the mean value of each 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^{th} x-covariate for the I^{th} integrand

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

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

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

δ_j is the standard deviation 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^{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

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

I_j denotes the function of age corresponding to the integrand for datapoint j

Spatiotemporal Gaussian process regression

This type of regression has been used for risk factors for which the data density is sufficient to estimate a very flexible time trend. The approach is a stochastic modelling technique that is designed to detect signals amidst noisy data. It also serves as a powerful tool for interpolating non-linear trends.^{35,36} Unlike classical linear models that assume that the trend underlying data follows a definitive functional form, GPR assumes that the specific trend of interest follows a Gaussian process, which is defined by a mean function $m(\cdot)$ and a covariance function $Cov(\cdot)$. For example, let $p_{c,a,s,t}$ be the exposure, in normal, log, or logit space, observed in country c , for age group a , and sex s at time t :

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

where

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

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

Estimating mean functions

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

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

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

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

where $X\beta$ is the summation of the components of a hierarchical mixed-effects linear regression, including the intercept and the product of covariates with their corresponding fixed-effect coefficients. Some models were run as hierarchical mixed-effects linear regressions with random effects on the levels of the geographical hierarchy. For most mixed-effects models, random effects were only used in the fit, not in the prediction. The second part of the equation, $h(r_{c,a,s,t})$, is a smoothing function for the residuals, $r_{c,a,s,t}$, derived from the linear model.⁸ Descriptions of exposure transformations and which covariates were used in linear models can be found in appendix section 4, which described the risk-specific estimation approaches. Some models used a custom stage-1 estimate. Detailed information on the mixed-effect estimation process for these risks may be found in the risk-specific appendix section 4.

Although the linear component captures the general trend in exposures over time, much of the data variability may still not be adequately accounted for. To address this, we fit a locally weighted polynomial regression (locally estimated scatterplot smoothing, or LOESS) function $h(r_{c,a,s,t})$ to systematically estimate this residual variability by borrowing strength across time, age, and space patterns (the spatiotemporal component of ST-GPR).^{37,38} The time adjustment parameter, defined by λ , aims to borrow strength from neighbouring time points (ie, the exposure in this year is highly correlated with exposure in the previous year but less so further back in time). The age-adjustment parameter, defined by ω , borrows strength from data in neighbouring age groups. The space-adjustment parameter, defined by ξ , aims to borrow strength across the hierarchy of geographical locations. The spatial and temporal weights are combined into a single space-time weight to allow the amount of spatial weight given to a particular point $r_{c,a,s,t}$ to fluctuate given the data availability at each time t and location-level l in the location hierarchy.

Let $w_{c,a,s,t}$ be the final weight assigned to observation $r_{c,a,s,t}$ with reference to a focal observation r_{c_0,a_0,s_0,t_0} . We first generated a temporal weight $t.w_{c,a,s,t}$ for smoothing over time, which was based on the scaled distance along the time dimension of the two observations:³⁸

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

Next, we generated a spatial weight to smooth over geography. Specifically, we defined a geospatial relationship by categorising data based on the GBD location hierarchy. *Zeta* acts as a scalar on a given datapoint given its proximity to the target location:

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

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

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

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

country c but within the same region r as $r_{c,a,s,t}$ would be assigned $\frac{1}{100}$ the weight of a datapoint within the country.

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

Given a normalisation constant,

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

the final space-time weight would then equal

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

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

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

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

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

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

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

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

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

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

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

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

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

Estimating error variance

σ_p^2 represents the error variance in normal or transformed space including the sampling variance of the estimates and prediction error from any crosswalks performed. First, variance was systematically imputed if the data extraction did not include any measure of uncertainty. When some sample sizes for data were available, missing sample sizes were imputed as the fifth percentile of available sample sizes. Missing variances were then calculated as $\sigma_p^2 = \frac{p*(1-p)}{n}$ for proportions or were predicted from the mean by using a regression for continuous values. When sample sizes were entirely missing and could not be imputed, the 95th percentile of available variances at the most granular geographical level (ie, first country, then region, etc.) were used to impute missing variances. For proportions where $p*n$ or $(1-p)*n$ is <20, variance was replaced by using the Wilson Interval Score method.

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

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

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

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

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

Estimating the covariance function

The final input into GPR is the covariance function, which defines the shape and distribution of the trends. Here, we have chosen the Matern-Euclidian covariance function, which offers the flexibility to

model a wide spectrum of trends with varying degrees of smoothness. The function is defined as follows:

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

where $d(\cdot)$ is a distance function; σ^2 , ν , l , and K_ν are hyperparameters of the covariance function—specifically σ^2 is the marginal variance, ν is the smoothness parameter that defines the differentiability of the function, l is the length scale, which roughly defines the distance between which two points become uncorrelated, and K_ν is the Bessel function. We approximated σ^2 by taking the normalised median absolute deviation $MADN(r'_c)$ of the difference, which is the normalised absolute deviation of the difference of the first-stage linear regression estimate from the second-stage spatiotemporal smoothing step for each country. We then took the mean of these country-level MADN estimates for all countries with 10+ country-years of data to ensure that differences between first- and second-stage estimates had sufficient data to truly convey meaningful information on model uncertainty. We used the parameter specification $\nu = 2$ for all models. The scale parameter l used for each risk is reported in appendix section 4.

Prediction using GPR

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

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

Random draws of 500 samples were obtained from the distributions above for every country for a given indicator. The final estimated mean for each country was the mean of the draws. In addition, 95% UIs were calculated by taking the 2.5 and 97.5 percentile of the sample distribution. The linear modelling process was implemented by using the lmer4 package in R, and the ST-GPR analysis was implemented through the PyMC2 package in Python.

Subnational scaling and aggregation

To ensure internal consistency of the estimates between countries and their respective subnational locations, national estimates were either created by population-weighted aggregation or subnational estimates were adjusted by population-weighted scaling to the national estimates, depending on the data coverage of a given country compared to that of its subnational locations. For example, if data coverage was better at the national level than at its corresponding subnational locations for a given country and risk across age, sex, and time, estimates were rescaled to be consistent with the national level. Conversely, if data coverage was better at the subnational level, estimates for its parent country were generated through population-weighted aggregation of subnational estimates.

Estimates can also be scaled within logit space. Scaling in logit space ensures that subnational estimates of proportion models do not exceed 1 after being rescaled to the national estimate.

Fitting a distribution to exposure data

The most informative data describing the distribution of risk factors within a population come from individual-level data; additional sources of data include reported means and variances. In cases in which a risk factor also defines a disease or disease severity cut-off, such as haemoglobin level and mild,

moderate, or severe anaemia, or diabetes and FPG level, the prevalence of disease is also frequently reported. To model the distribution of any particular risk factor, we seek a family of probability density functions (PDF), a fitting method, and a model selection criterion. To make use of the most commonly available data describing most populations, we used the method of moments (MoM); the first two empirical moments from a population, the mean and the variance, were used to determine the parameters of two-parameter PDF describing the distribution of risk within any population. Exceptions to this rule are justified by context. We used the Kolmogorov-Smirnov³⁹ (KS) test to measure the goodness of fit (GoF) and compared the distance between the empirical and ensemble distributions, but in some cases, the GoF was based on the prediction error for the prevalence of disease.

We used an ensemble technique in which a model selection algorithm is used to choose the best model for each continuous risk factor.³⁹ We drew the initial set of candidate models from commonly used PDF families, including both right-skewed and left-skewed distributions. These included beta, exponential, gamma, gumbel, inverse gamma, inverse Weibull, log-logistic, lognormal, mirrored gamma, mirrored gumbel, normal, and Weibull. We fitted each PDF family candidate to each dataset by using the MoM and used the KS test as the measure of GoF. Preliminary analysis showed that the GoF ranking of PDF families varied across datasets for any particular risk factor and that combining the predictions of differently fitted PDF families could dramatically improve the GoF for each dataset. Therefore, we developed a new model for prediction by using the ensemble of candidate models, which is a weighted linear combination of all candidate models, $\{f\}$, where a set of weights $\{w\}$ is chosen such that $\sum_i w_i = 1$, and the values of the weights were determined by a second GoF criterion with its own validation process. For each risk, we pooled all available microdata and performed Nelder-Mead numeric optimisation across demographics subsets of data to derive a set of distribution-specific weights such that the average KS statistic across datasets would be minimised. The details can be summarised by 1) the summary statistics for each dataset; 2) a table showing the KS statistic for each candidate model; and 3) the weights defining the final ensemble model for each dataset. We then averaged across demographic subsets and datasets to determine the final weights for modelling the distribution of any particular risk factor.

Step 3. TMREL⁵

In this and all previous GBD studies, the counterfactual level of risk exposure used is the risk exposure that is both theoretically possible and minimises risk in the exposed population that consequently captures the maximum population attributable burden.⁹ For each risk evaluated in GBD 2023, Step 3 in the analytical flowchart involves taking the best available epidemiological evidence used to estimate relative risk by level of exposure and the lowest observed level of exposure from cohorts. This is used to identify a single level of risk exposure that minimises risk from all causes of death combined to the establish the TMREL. In principle, the TMREL for a given risk may vary by age, sex, and location if supported by clear evidence. Based on the available evidence, the TMREL itself can be uncertain, which is reflected in the 95% UIs.

Step 4. Estimate population attributable fractions⁵

Risks are categorised on the basis of how exposure was measured: dichotomous, polytomous, and continuous. High low-density lipoprotein (LDL) cholesterol level is an example of a risk measured on a continuous scale. The PAF, which represents the proportion of risk that would be reduced in a given year

if the exposure to a risk factor in the past were reduced to an ideal exposure scenario, is defined for a continuous risk factor as:⁴⁰

$$PAF_{joasgt} = \frac{\int_{x=l}^u RR_{joasg}(x)P_{jasgt}(x)dx - RR_{joasg}(TMREL_{jas})}{\int_{x=l}^u RR_{joasg}(x)P_{jasgt}(x)dx}$$

Where PAF_{joasgt} is the PAF for cause o due to risk factor j for age group a , sex s , location g , and year t . $RR_{joasg}(x)$ is the RR as a function of exposure level x for risk factor j for cause o , age group a , sex s , and location g with the lowest level of observed exposure as l and the highest as u ; $P_{jasgt}(x)$ is the distribution of exposure at x for age group a , sex s , location g , and year t ; and $TMREL_{jas}$ is the TMREL for risk factor j , age group a , and sex s .

The PAF_{joasgt} for dichotomous and polytomous risk factors for every country is defined as:

$$PAF_{joasgt} = \frac{\sum_{x=l}^u RR_{joasg}(x)P_{jasgt}(x) - RR_{joasg}(TMREL_{jas})}{\sum_{x=l}^u RR_{joasg}(x)P_{jasgt}(x)}$$

where PAF_{joasgt} is the PAF for cause o due to risk factor j for age group a , sex s , location g , and year t . $RR_{joasg}(x)$ is the RR as a function of exposure level x for risk factor j for cause o , age group a , sex s , and location g on a plausible range of exposure levels from l to u ; $P_{jasgt}(x)$ is the proportion of the population in risk group (prevalence) for age group a , sex s , location g , and year t ; and $TMREL_{jas}$ is the TMREL for risk factor j , age group a , and sex s .

Step 5. Estimate summary exposure values⁵

Summary exposure value (SEV) is the RR-weighted prevalence of exposure, a univariate measure of risk-weighted exposure, taking the value zero when no excess risk for a population exists and the value 1 when the population is at the highest level of risk. We report SEVs on a scale from 0% to 100% on which a decline in SEV indicates reduced exposure to a given risk factor and an increase in SEV indicates increased exposure.

In words, the formula is:

$$SEV = \frac{\text{average excess relative risk in a population}}{\text{maximum global, all-time excess relative risk}}$$

We first calculate SEVS for risk-cause pairs, rc , using the following equation,

$$SEV_{rc} = \frac{\frac{PAF_{rc}}{1 - PAF_{rc}}}{RR_{max,rc} - 1}$$

for each most-detailed age, sex, location, year, and outcome. PAF is the YLL (except for any outcomes which are YLD only and thus use the YLD) PAF. RR_{max} for categorical risks is the RR of the exposure category with the highest RR. For continuous risks, the RR_{max} is the RR at the 5th exposure percentile if the risk is protective, and at the 95th exposure percentile if the risk is harmful. For custom modelled

risks like ambient particulate matter pollution, HAP from solid fuels, ozone pollution, alcohol, smoking, and bullying, the modeller provides draws of RR_{max} .

Generally, RRs do not vary across time and space. However, exceptions exist, such as risks from secondhand smoke (SHS) for which the RR is based on the integrated exposure response (IER) curve. In these cases, the RR is averaged across location and year to ensure no time or space variation. If the PAF is negative, which signifies a protective effect for that outcome, the PAF is set to 0 and the SEV is then also 0 because the SEV is univariate and constrained to be a value between 0 and 1.

In most cases, risk – cause PAFs of 1 were not included in SEV calculations as the SEV function is undefined when the PAF value is 1. However, an alternate definition of SEV was used for a select set of risks for PAFs of 1: fasting plasma glucose, systolic blood pressure, and iron deficiency. For fasting plasma glucose, SEVs were set to risk prevalence above 7 mmol/L. For systolic blood pressure, SEVs were set to risk prevalence above 140 mmHg. Lastly, for iron deficiency, SEVs were set to the prevalence of moderate or severe anaemia.

Once we obtained a set of risk-cause-specific SEVs at the most-detailed risk, cause, age, sex, and location for all years, we averaged across causes to produce the final risk-specific SEV_r ,

$$SEV_r = \frac{1}{N(c)} \sum_c SEV_{rc}$$

where $N(c)$ is the total number of outcomes for a risk.

Step 6. Mediation⁵

Section 2.6.1: Summary

The portion of the burden of disease that is attributable to various combinations of risk factors or to all risk factors combined has been a topic of broad interest.⁴¹ In GBD 2010, we only aggregated the burden of risk factors for some clusters of risks, including access to improved water and sanitation, child and maternal malnutrition, tobacco smoking, alcohol use, dietary risk factors, occupational risk factors, and sexual abuse and violence. We did not aggregate air pollution and metabolic risk factors. For GBD 2013 onward, we aggregated all risk factors into three large categories—behavioural, environmental and occupational, and metabolic risks—and aggregated all GBD risk factors into a single attributable fraction for each disease and eventually for all causes of burden. Please note that mediation is conducted as a separate process and is not part of the BoP methodology. In our relative risk estimation, we include RRs that do not adjust for mediation as our goal is to capture the direct effect of a risk factor on an outcome.

Aggregating risk factors at different levels shares three essential challenges:

1. Risk factor coexistence or aggregation: for example, metabolic risk factors often occur together, or high-risk behaviours such as drug abuse and unsafe sex are related.
2. Mediation: a risk factor may affect another risk factor that lies in the physiological pathway to a disease outcome. It can be inside a cluster of risk factors, such as the effect of obesity through an increase in FPG level and later cardiovascular disease (CVD) outcomes, or between clusters of risk factors, such as the effect of fibre on cholesterol.

3. The formula used to calculate the aggregated PAF.

The aggregation method is conceptually applicable to other aggregations such as socioeconomic factors, education, homelessness, and refugee status that are being considered for inclusion in future GBD iterations. In the next section, we explain our approach to dealing with these challenges.

There are three patterns of associations between risk factors to consider (Figure C). The first concerns confounding; risk B affects risk A and outcome C (Pattern 1 in *Patterns of associations between risk factors*). In these cases, the RR for A should be adjusted for B; for example, the fruit RR is adjusted for smoking. If part of the effect of A is through B, a mediator, we do not adjust the effect of A for B. For example, we do not adjust the RR of body-mass index (BMI) for cholesterol because cholesterol lies in the biological pathway between BMI and cardiovascular outcomes (Pattern 2 in *Patterns of associations between risk factors*). The third pattern occurs when risks A and B are proxies of a third variable Z and aggregation aims to estimate the total effect of a latent variable Z on C. An example is child growth failure, which is measured by stunting, wasting, and underweight as proxies.

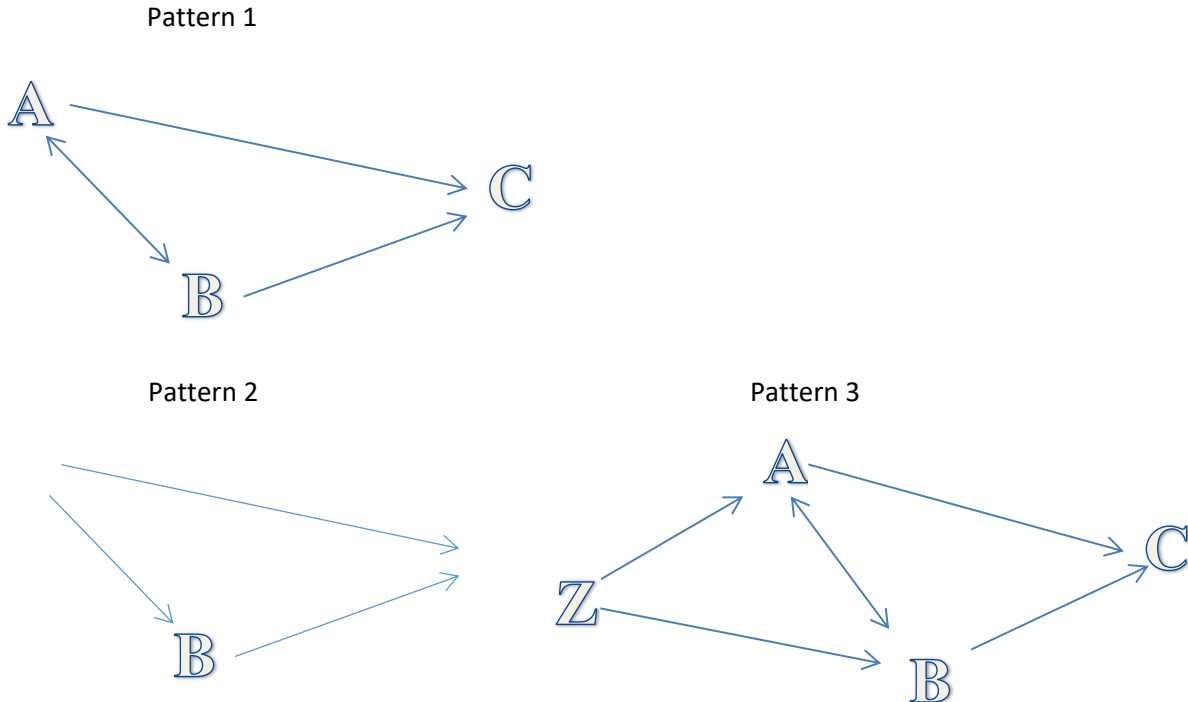


Figure C. Patterns of associations between risk factors

Section 2.6.2: Calculating the burden of multiple risk factors

Validation studies have reported congruency between the true risk associated with multiple risk factors affecting the same outcome and a multiplicative aggregation of the PAFs of the individual risk factors (formula below)⁴²

$$PAF_{1..i} = 1 - \prod_{i=1}^n (1 - PAF_i)$$

where PAF is the population attributable fraction and i is each individual risk factor.

The same validation studies also found that the overestimation from ignoring the covariance between risk factors is small. This small overestimation was important to note because few data sources exist from which we can draw information on covariance.

We endeavoured to evaluate RRs that were controlled for confounders. However, because we had to rely on the literature for many RRs, we did not always have full control over the choice of confounders controlled for in each study.

Section 2.6.3: Computing mediation factors using linear relationships

If the relationship between the distal risk and the mediating risk factor is linear, eg, an increase in BMI of 1 kg/m² leads to an increase in FPG of $\delta_{B,A}$ mmol/L, we can use the linear relationship to estimate the mediated risk and hence the mediation factor. Specifically, the relative risk of C due to A mediated by B is computed as follows:

$$RR_{C,A}^B = \int_{b_{start}}^{b_{end}} \frac{1}{RR_{C,B}(b)} R_{C,B}(\delta_{B,A}(a - a_0)_{(+)}) p(b) db$$

The linear factor $\delta_{B,A}$ may be available from the literature or may be found using root finding in certain cases, as for example when the cause is defined purely by mediator exposure, such as for diabetes.

Section 2.6.4: Adjusting for mediation

When aggregating the effects of multiple risk factors, we included a mediation factor (MF) if a part of the effect of one risk factor was included in the effect estimated for in the mediator. First, we prepared a list of possible mediations, and especially between behavioural risks and metabolic risk factors with cardiometabolic outcomes. We did not assume any mediation effect between risk factors for cancers.

Danaei and colleagues assumed that part of the effect of BMI on ischaemic heart disease (IHD) is through high systolic blood pressure (SBP), cholesterol level, and FPG.⁴³ The proportion of the BMI effect that can be explained by other metabolic risk factors is the amount of mediation. The difference between the crude RR of BMI on IHD with the RR adjusted for SBP, FPG, and cholesterol level reflects the amount of BMI effect on IHD that is mediated and already included in SBP, FPG, and cholesterol level:

$$MF = \frac{RR_{crude} - RR_{adjusted}}{RR_{crude} - 1}$$

So, to aggregate the PAF of multiple risk factors, we first calculated the part of the excess risk ($RR - 1$) of every risk factor that is not mediated, re-compute the PAF so that it only includes the non-mediated risk then aggregated PAFs by assuming they are independent.

Therefore, if MF is the mediation factor of R2 through R1, the adjusted RR for R2 including only the non-mediated component of risk is:

$$RR_{1,2} = MF_{2/1}(RR_2 - 1) + 1$$

The PAF accounting for mediation is then computed using the adjusted RR and the joint PAF computed as detailed in Section 2.7.2. For every paired risk factor and outcome, the matrix of possible mediations was calculated and used.

Section 2.6.5: Calculating mediation factor

1 – Comparing crude RR versus mediator-adjusted RR

The best example is the mediation of BMI through SBP, FPG, and cholesterol level reported by Danaei et al.⁴³ In their meta-analysis, they report the adjusted and unadjusted RR of BMI on IHD and stroke based on combined data from individual cohorts. They calculated the MF by using the following equation, and we used it directly as the MF in risk factor aggregation. Using individual-level data from cohort studies, we estimated the MF for other metabolic risk factors and some dietary risks.

$$MF = \frac{RR_{crude} - RR_{adjusted}}{RR_{crude} - 1}$$

2 – Estimating the mediation factor by pathway of the effect

For many other risk factors, no data are available to enable the use of the first method. Instead, we searched studies to estimate the effect of the risk factor on the mediator and, finally, the expected increase in IHD risk. We pooled available studies to calculate the unit increase in the mediator per unit increase in the risk factor to calculate the size of the IHD RR (Figure D).

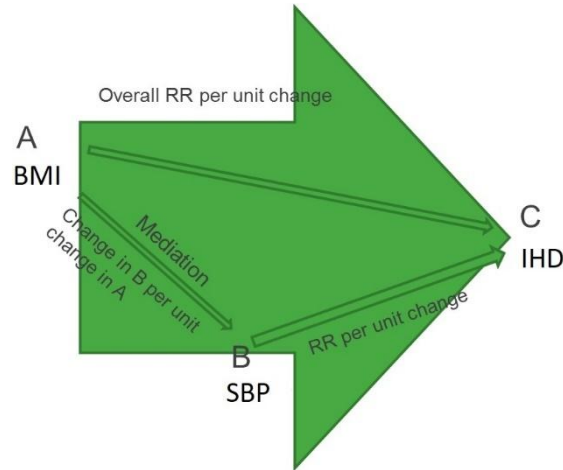


Figure D. Example of pathway between BMI, high systolic blood pressure, and ischemic heart disease

We have RRs for the effect of A on C and B on C in GBD from a meta-analysis of studies in the literature. The effect of A on B was estimated by analysis of trials.

$$RR_{ABC} = RR_{BC}^{\Delta_{AB}}$$

RR_{ABC} is the expected effect of A through B on C

RR_{BC} is the RR of each unit increase in mediator on outcome C

Δ_{AB} is the change in mediator level B per each unit change in A

If RR_{AB} is the overall effect of A on B, then:

$$MF = \frac{RR_{ABC} - 1}{RR_{AB} - 1}$$

We kept the uncertainty of each parameter to a minimum by generating and following 1000 draws of the estimates to calculate 1000 draws of the posterior distribution of the MF. We did not include risk-mediator pairs if the MF was not significant at the 5% level (more than 50 of 1000 draws were negative). We truncated the MF distribution at 1 when the whole effect of the risk factor on the outcome would be assumed to be exerted through the mediator pathway.

Some MFs equalled 1 when the whole effect was calculated through another risk factor (eg, the effect of salt through SBP) or when we assumed other risk factors were sources of the exposure (eg, fibre is provided by consuming fruit, vegetables, and whole grains, and all the beneficial effect of milk on colorectal cancer is mediated through calcium).

Air pollution

In GBD 2019, we considered mediation for particulate matter air pollution and SBP, FPG level, and cholesterol level, but in no case was the evidence strongly supportive. Review of the epidemiological evidence identified several cohort studies that reported increased prevalence and/or risk of hypertension due to long-term exposure to ambient $PM_{2.5}$, and several studies have found elevated SBP due to household solid fuel use as well. Studies of short-term exposure also reported acute elevations in blood pressure. However, there is not consensus as to whether the existing evidence with regard to the effects of long-term exposure is consistent with the current mechanistic understanding of the effect of air pollution exposure on blood pressure, and whether existing cohort studies have properly modelled that exposure.

Assumed mediations

For the risk factors with PAFs of 100%, such as FPG and diabetes, impaired kidney function and chronic kidney disease (CKD), SBP and hypertensive heart disease, alcohol and alcohol use disorders, child underweight and protein-energy malnutrition, child wasting and protein-energy malnutrition, and drug use and drug use disorders, no mediation is needed.

Section 2.6.6: Piecewise aggregation (Pattern 3)

There are three anthropometric indicators that are highly correlated: child underweight, stunting, and wasting, as shown in Figure E in this section. Available RRs for each indicator are not adjusted for the other two because these indicators are highly correlated and most of the burden occurs in an interaction. Estimating the total burden due to child growth failure, a latent variable, is difficult. The three anthropometric indicators are not independent, so the covariance between them should be considered. This consideration was the main reason that GBD 2010 only included child underweight. If covariance between these indicators is significant (as is shown in Figure E), aggregating these indicators by assuming they are independent would overestimate the total burden significantly.

To account for the high degree of correlation between CGF indicators, GBD uses a constrained optimisation method to adjust the observed univariate RRs that come out of the Burden of Proof analysis. First, we created a joint distribution of stunting, underweight, and wasting from a population of children. Second, we generated 500 RR draws for each univariate indicator and severity based on the

Burden of Proof analysis. Third, we altered these univariate RRs for the four causes (diarrhoea, LRI, malaria, and measles) and the two outcomes (mortality and morbidity) based upon interactions among the CGF indicators. An interaction occurs when the effect of one CGF indicator variable (eg, stunting) has a different effect on the outcome depending on the value of another CGF indicator variable (eg, underweight). Interaction terms alter the risk of the outcome among children with more than one indicator of CGF. These interaction terms were extracted from a pooled cohort analysis of all-cause mortality published by McDonald et al.⁴⁴ Lastly, we optimised the adjusted relative risks by minimising the error between the observed RRs (generated from the Burden of Proof analysis) and the altered RRs derived from the joint distribution and accounting for the interaction terms while ensuring that no alteration resulted in a previously identified increase in relative risk becoming protective.

For GBD 2021, we made several changes to improve the four main steps of RR adjustment. From GBD 2013 to GBD 2019, a simulated joint distribution of stunting, underweight, and wasting measures was created from the Olofin et al. meta-analysis.⁴⁵ Sources in this meta-analysis were cross-sectional Demographic and Health Surveys (DHS).⁴⁵ In GBD 2021, we created age-specific joint distributions of stunting, underweight, and wasting measures from 15 longitudinal studies (from 26 locations) in the Ki database.⁴⁶ The RR adjustment method was strengthened in GBD 2021 by constraining optimisation in two ways. Optimisation was only permitted to alter the RR for an indicator/severity in draws where the observed RR was greater than 1, and constraints were placed on the error that penalise larger alterations to the RR. These changes enabled the estimation and utilisation of age-specific adjusted RRs for GBD 2021 burden estimation. The largest change for GBD 2021 was conducting Burden of Proof analyses for each cause/outcome/risk triplet using both data from Olofin et al as well as KI data. These changes result in identifying large differences in the relationship between CGF and mortality versus morbidity as well as identifying some impact of CGF on malaria.

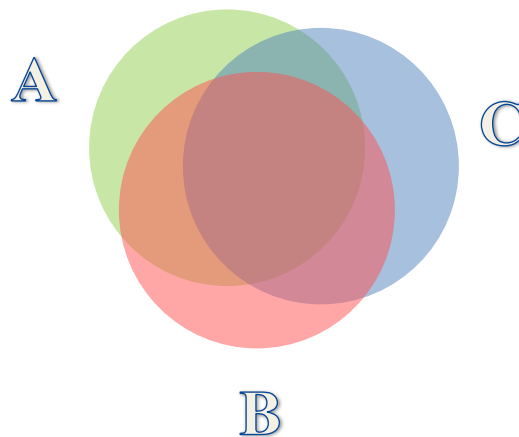


Figure E. Venn diagram demonstrating the correlation between child underweight, stunting, and wasting

After adjusting for the three risk factors, we calculated the PAFs and aggregated the underweight, stunting, and wasting burden.

Section 2.6.7: Uncertainty of aggregated and mediated PAFs

We generated 1000 draws of the posterior distribution of the MF calculated by different methods to use beside draws of other inputs to the PAF aggregation.

Section 2.6.8: Important assumptions in aggregating risk factors and including mediation

1 – The MFs or PAF adjustments are similar across countries, age, sex, and years. Although the size of mediation is probably different in different populations, few data are available to inform the covariance between different risk factors or the MF amount by age and country. For example, in some countries, the size of the mediated BMI-IHD PAF exerted through cholesterol level, as calculated by the MF, was even bigger than the total burden of cholesterol level. This finding indicated that less of the effect of BMI is mediated through cholesterol level and MFs are not similar across countries.

2 – For many risk-mediator-outcome pairs, no data are available, so we assumed the mediation is zero or one.

3 – Because the covariance between undernutrition indicators differs by location (and across time, but results were not reported), and an interaction exists between these indicators, the total burden might be underestimated.

4 – We assumed no significant covariance between PAFs, which might not be true between some risk factors, such as metabolic risk factors. Although this overestimation can be controlled by using adjusted RRs, using crude RRs for BMI and other metabolic risk factors may cause significant overestimation of the aggregated metabolic risks burden.

Step 7. Estimate attributable burden⁵

Four key components are included in the estimation of the burden attributable to a given risk factor: the metric of burden being assessed (the number of deaths, YLLs, YLDs, or DALYs [the sum of YLLs and YLDs]); the exposure levels for a risk factor; the RR of a given outcome due to exposure; and the counterfactual level of risk factor exposure. Uncertainty intervals for attributable burden were calculated using 250 draws for GBD 2023. Estimates of attributable burden as DALYs for risk–outcome pairs were generated by using the following model:

$$AB_{jasgt} = \sum_{o=1}^w DALY_{joasgt} PAF_{joasgt}$$

where AB_{jasgt} is the attributable burden for risk factor j for age group a , sex s , location g , and year t ; $DALY_{joasgt}$ is total DALYs for cause o (of w relevant outcomes for risk factor j) for age group a , sex s , location g , and year t ; and PAF_{joasgt} is the PAF for cause o due to risk factor j for age group a , sex s , location g , and year t . The proportions of deaths, YLLs, or YLDs attributable to a given risk factor or risk factor cluster were analogously computed by sequentially substituting each metric in place of DALYs in the equation provided.

Section 3: Cardiovascular decomposition analysis of DALYs

We conducted decomposition analyses of changes in deaths from 1990 to 2023. Decomposing changes in cause-specific deaths were due to changes in population growth, population age structure, exposure to all risks for a disease, and risk-deleted death rates. In this case, risk-deleted rates are the rates after removing the effect of a risk factor or combination of risk factors; in other words, observed death rates multiplied by one minus the PAF for the risk or set of risks. Our decomposition analyses draw from methods developed by Das Gupta (44) to provide a computationally tractable solution to isolating drivers of burden changes whereby all combinations of possible pathways are averaged across factors. Both the total burden and the attributable burden are determined, following the methods of Das Gupta, as a product of four factors such that:

$$T_{asgt} = (A_{sgt}B_{asgt}C_{asgt}D_{asgt})$$

where T_{asgt} represents either the total burden or the attributable burden at year t ;

A_{sgt} the all-age population size for a given sex s and geography g at year t ;

B_{asgt} is the proportion of the population in the age group for a given age group a , sex s and geography g at year t ;

C_{asgt} is the underlying rate of the outcome unrelated to the risk factor or observed rate, multiplied by $1 - PAF$ for a given age group a , sex s and geography g at year t ; and where

D_{asgt} is the ratio of total burden (or attributable burden) to the underlying rate, which reflects the risk effect for a given age group a , sex s , and geography g at year t defined as $1/(1 - PAF)$ in the case of total burden or as $PAF/(1 - PAF)$ in the case of decomposing attributable burden to a risk.

The contribution of each factor to total change in either total burden or attributable burden was determined by changing the level of one factor from time t_0 to t_1 – here 1990 to 2023 – with all other factors held constant. Thus, the effect of any of the four factors, for example A_{sgt} on the change of total burden between 1990 (A_{90}) and 2023 (A_{23}) is calculated as:

$$E_A = (A_{23} - A_{90}) \left(\frac{B_{90}C_{90}D_{90} + B_{23}C_{23}D_{23}}{4} + \frac{B_{90}C_{90}D_{23} + B_{90}C_{23}D_{90} + B_{23}C_{90}D_{90} + B_{23}C_{23}D_{90} + B_{23}C_{90}D_{23} + B_{90}C_{23}D_{23}}{12} \right)$$

Where E_A is the proportion of change due to factor A , and the subscripts for each factor in the equation denote the year for each estimate. Since the effect depends on the order of entry of the factor, we calculated the average of all combinations of the four factors. The proportion of change due to factor A_{sgt} , the age-specific population size for a given age group a , sex s , and location g at year t , is then further split, setting change in population growth equal to the percentage change in the all-age population from time t_0 to t_1 and change in population age structure to the residual, giving four factors.

The three-factor decomposition method does not work for risks for which the PAF, by definition, is 100% (such as high systolic blood pressure and hypertensive heart disease). For cardiovascular disease this only was relevant for two risk-outcome pairs, high systolic blood pressure - hypertensive disease and alcohol use and alcoholic cardiomyopathy. In these cases, we used a two-factor decomposition method that examines the contribution of population, ageing, and risk exposure. Effectively, we assumed trends in these cases are driven by exposure, not change in the risk-deleted rates.

Section 4: References

- 1 Stevens GA, Alkema L, Black RE. Guidelines for Accurate and Transparent Health Estimates Reporting: the GATHER statement. *Lancet* 2016.
- 2 Lim SS, Vos T, Flaxman AD. A comparative risk assessment of burden of disease and injury attributable to 67 risk factors and risk factor clusters in 21 regions, 1990–2010: a systematic analysis for the Global Burden of Disease Study 2010. *The Lancet* 2012; **380**: 2224–60.
- 3 Forouzanfar M, Afshin A, Alexander LT, Anderson H, Bhutta Z, Murray CJL. Global, regional, and national comparative risk assessment of 79 behavioural, environmental and occupational, and metabolic risks or clusters of risks, 1990–2015: a systematic analysis for the Global Burden of Disease Study 2015. *Lancet* 2016; **388**: 1659–724.
- 4 Gakidou E, Afshin A, Abajobir AA, *et al.* Global, regional, and national comparative risk assessment of 84 behavioural, environmental and occupational, and metabolic risks or clusters of risks, 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016. *The Lancet* 2017; **390**: 1345–422.
- 5 Stanaway JD, Afshin A, Gakidou E, *et al.* Global, regional, and national comparative risk assessment of 84 behavioural, environmental and occupational, and metabolic risks or clusters of risks for 195 countries and territories, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017. *The Lancet* 2018; **392**: 1923–94.
- 6 Murray CJL, Aravkin AY, Zheng P. Global burden of 87 risk factors in 204 countries and territories, 1990–2019: a systematic analysis for the Global Burden of Disease Study 2019. *The Lancet* 2020; **396**: 1223–49.
- 7 GBD 2021 Risk Factors Collaborators. Global burden and strength of evidence for 88 risk factors in 204 countries and 811 subnational locations, 1990–2021: a systematic analysis for the Global Burden of Disease Study 2021. *Lancet* 2024; **403**: 2162–203.
- 8 Murray CJ, Lopez AD. Global mortality, disability, and the contribution of risk factors: Global Burden of Disease Study. *Lancet* 1997; **349**: 1436–42.
- 9 Murray CJ, Lopez AD. On the comparable quantification of health risks: lessons from the Global Burden of Disease Study. *Epidemiology* 1999; **10**: 594–605.
- 10 Food, nutrition, physical activity and the prevention of cancer: a global perspective. Washington, D.C: World Cancer Research Fund & American Institute for Cancer Research, 2007.
- 11 Zheng P, Barber R, Sorensen RJ, Murray CJ, Aravkin AY. Trimmed constrained mixed effects models: formulations and algorithms. *Journal of Computational and Graphical Statistics* 2021; : 1–13.
- 12 Zheng P. limetr: limetr: linear mixed effects model with trimming. <https://github.com/zhengp0/limetr> (accessed July 28, 2021).
- 13 Zheng P. xspline: xspline: Advanced spline tools. <https://github.com/zhengp0/xspline> (accessed July 28, 2021).

- 14 Viechtbauer W. Conducting meta-analyses in R with the metafor package. *Journal of statistical software* 2010; **36**: 1–48.
- 15 de Boor C. A practical guide to splines (applied mathematical sciences, 27). New York: Springer, 2001 <https://link.springer.com/book/9780387953663> (accessed April 4, 2022).
- 16 Rousseeuw PJ, Leroy AM. Robust regression and outlier detection. John Wiley & sons, 2005.
- 17 Rousseeuw PJ. Least median of squares regression. *Journal of the American Statistical Association* 1984; **79**: 871–80.
- 18 Huber PJ. Robust Statistics. John Wiley & Sons, 2004.
- 19 Rousseeuw P. Multivariate estimation with high breakdown point. *Mathematical Statistics and Applications Vol B* 1985; : 283–97.
- 20 Aravkin A, Davis D. Trimmed statistical estimation via variance reduction. *Mathematics of Operations Research* 2020; **45**: 292–322.
- 21 Motzkin TS, Raiffa H, Thompson GL, Thrall RM. 3. The Double Description Method. In: 3. The Double Description Method. Princeton University Press, 2016: 51–74.
- 22 Guyatt GH, Oxman AD, Vist G, *et al.* GRADE guidelines: 4. Rating the quality of evidence—study limitations (risk of bias). *Journal of clinical epidemiology* 2011; **64**: 407–15.
- 23 Efron B, Hastie T, Johnstone I, Tibshirani R. Least angle regression. *The Annals of statistics* 2004; **32**: 407–99.
- 24 Tibshirani R. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)* 1996; **58**: 267–88.
- 25 Kontopantelis E, Springate DA, Reeves D. A re-analysis of the Cochrane Library data: the dangers of unobserved heterogeneity in meta-analyses. *PloS one* 2013; **8**: e69930.
- 26 Ioannidis JP, Patsopoulos NA, Evangelou E. Uncertainty in heterogeneity estimates in meta-analyses. *Bmj* 2007; **335**: 914–6.
- 27 Biggerstaff BJ, Tweedie RL. Incorporating variability in estimates of heterogeneity in the random effects model in meta-analysis. *Statistics in medicine* 1997; **16**: 753–68.
- 28 Higgins JP, Thompson SG. Quantifying heterogeneity in a meta-analysis. *Statistics in medicine* 2002; **21**: 1539–58.
- 29 Sterne JA, Egger M. Chapter 6: Regression methods to detect publication and other bias in meta-analysis. In: Rothstein H, Sutton A, Borenstein M, eds. Publication bias in meta-analysis: Prevention, assessment and adjustments. John Wiley & Sons, Ltd, 2005.
- 30 Egger M, Smith GD, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. *BMJ* 1997; **315**: 629–34.

- 31 Burkart KG, Brauer M, Aravkin AY. Estimating the cause-specific relative risks of non-optimal temperature on daily mortality: a two-part modelling approach applied to the Global Burden of Disease Study. *The Lancet* 2021; **398**: 685–97.
- 32 Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research Electronic Data Capture (REDCap) - A metadata-driven methodology and workflow process for providing translational research informatics support. *J Biomed Inform* 2009; **42**: 377–81.
- 33 GBD 2015 Diseases and Injury Incidence and prevalence Collaborators. Global, regional, and national incidence, prevalence, and years lived with disability (YLDs) for 310 acute and chronic diseases and injuries, 1990-2015: a systematic analysis for the Global Burden of Disease Study 2015. *The Lancet* Under review.
- 34 Flaxman AD, Vos T, Murray CJL, Kiyono P, editors. An integrative metaregression framework for descriptive epidemiology, 1 edition. Seattle: University of Washington Press, 2015.
- 35 Vasudevan S, Ramos F, Nettleton E, Durrant-Whyte H, Blair A. Gaussian Process modeling of large scale terrain. In: 2009 IEEE International Conference on Robotics and Automation. 2009: 1047–53.
- 36 Rasmussen CE, Williams CKI. Gaussian Processes for Machine Learning. Cambridge, Mass: The MIT Press, 2005.
- 37 Ng M, Fleming T, Robinson M, *et al*. Global, regional, and national prevalence of overweight and obesity in children and adults during 1980–2013: a systematic analysis for the Global Burden of Disease Study 2013. *Lancet* 2014; **384**: 766–81.
- 38 Ng M, Freeman MK, Fleming TD, *et al*. Smoking Prevalence and Cigarette Consumption in 187 Countries, 1980-2012. *JAMA* 2014; **311**: 183–92.
- 39 Massey Jr FJ. The Kolmogorov-Smirnov test for goodness of fit. *Journal of the American statistical Association* 1951; **46**: 68–78.
- 40 Vander Hoorn S, Ezzati M, Rodgers A, Lopez AD, Murray CJL. Estimating attributable burden of disease from exposure and hazard data. In: Comparative Quantification of Health Risks: Global and regional burden of disease attribution to selected major risk factors. World Health Organisation, 2004: 2129–40.
- 41 Preston SH. Causes and Consequences of Mortality Declines in Less Developed Countries during the Twentieth Century. In: Population and economic change in developing countries. Chicago: Univ. of Chicago Pr, 1980: 289–360.
- 42 Carnahan E, Lim SS, Nelson EC, *et al*. Validation of a new predictive risk model: measuring the impact of major modifiable risks of death for patients and populations. *The Lancet* 2013; **381**: S26.
- 43 Danaei G, Singh GM, Paciorek CJ, *et al*. The global cardiovascular risk transition: associations of four metabolic risk factors with national income, urbanization, and Western diet in 1980 and 2008. *Circulation* 2013; **127**: 1493–502, 1502e1-8.

- 44 McDonald CM, Olofin I, Flaxman S. The effect of multiple anthropometric deficits on child mortality: meta-analysis of individual data in 10 prospective studies from developing countries. *Am J Clin Nutr* 2013; **97**: 896–901.
- 45 Olofin I, McDonald CM, Ezzati M, et al. Associations of suboptimal growth with all-cause and cause specific mortality in children under five years: a pooled analysis of ten prospective studies. *PLoS ONE*; **8**: e64636.
- 46 Ki Global Health. (n.d.). <https://www.kiglobalhealth.org/>.

Section 5: Risk factor appendix tables

Table S1. GBD cardiovascular risk hierarchy with levels	
Risk	level
All risk factors	0
Environmental/occupational risks	1
Air pollution	2
Particulate matter pollution	3
Ambient particulate matter pollution	4
Household air pollution from solid fuels	4
Non-optimal temperature	2
High temperature	3
Low temperature	3
Other environmental risks	2
Lead exposure	3
Behavioural risks	1
Tobacco	2
Smoking	3
Chewing tobacco	3
Secondhand smoke	3
High alcohol use	2
Dietary risks	2
Diet low in fruits	3
Diet low in vegetables	3
Diet low in legumes	3
Diet low in whole grains	3
Diet low in nuts and seeds	3
Diet low in milk	3
Diet high in red meat	3
Diet high in processed meat	3
Diet high in sugar-sweetened beverages	3
Diet low in fibre	3
Diet low in calcium	3
Diet low in seafood omega-3 fatty acids	3
Diet low in omega-6 polyunsaturated fatty acids	3
Diet high in trans fatty acids	3
Diet high in sodium	3
Low physical activity	2
Metabolic risks	1
High fasting plasma glucose	2
High LDL cholesterol	2
High systolic blood pressure	2
High body-mass index	2
Kidney dysfunction	2

Section 6: GBD 2023 risk factor-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: a systematic analysis for the Global Burden of Disease Study 2023,” we have only included modelling descriptions for the risks relevant for our analysis.

Contents

1. Ambient particulate matter pollution	54
2. Chewing tobacco	78
3. Dietary risks	83
4. High alcohol use	122
5. High body mass index	182
6. High fasting plasma glucose	193
7. High LDL cholesterol	217
8. High systolic blood pressure	227
9. Household air pollution	244
10. Kidney dysfunction	251
11. Lead exposure	260
12. Low physical activity	273
13. Non-optimal temperature	281
14. Secondhand smoke	288
15. Smoking	302