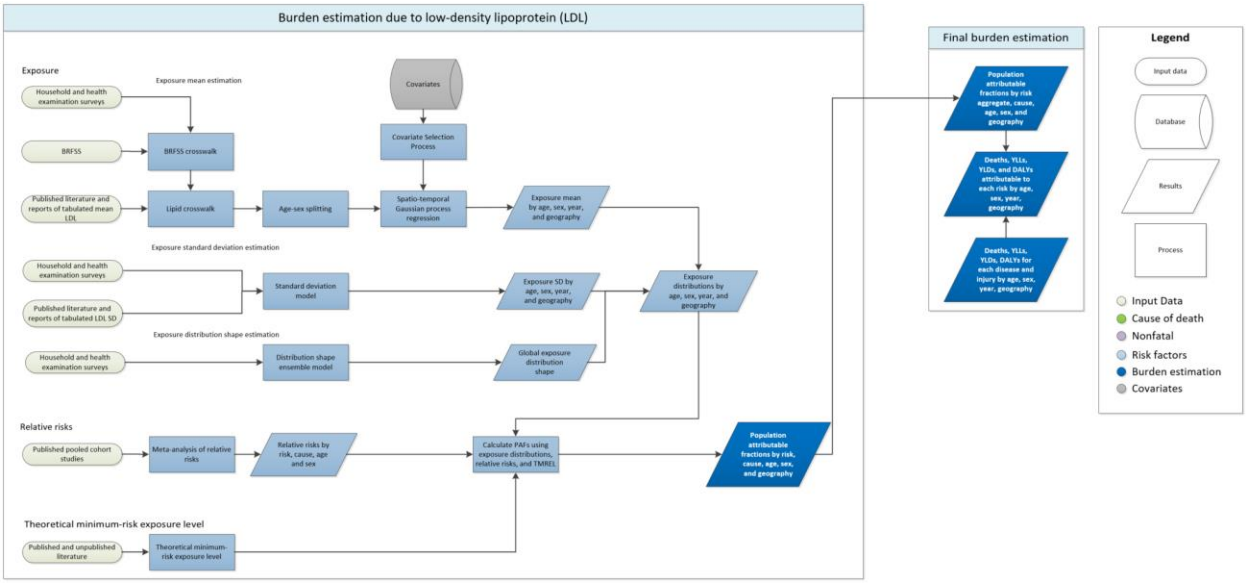
High LDL cholesterol

Flowchart



Input data and methodological summary

Definition

Exposure

In earlier iterations of the GBD study, including GBD 2010, 2015, and 2016, we estimated the burden attributable to total cholesterol (TC).[1] Beginning in GBD 2017, we modelled blood concentration of low-density lipoprotein cholesterol (LDL) in units of mmol/L.[2] We used data on blood levels for LDL, TC, triglyceride (TGL), and high-density lipoprotein cholesterol (HDL) from literature and from household survey microdata and reports. We adjusted data for TC, TGL, and HDL using the correction approach described in the "Lipid Crosswalk" section below.

Input data

Exposure

For GBD 2021, a systematic review of data sources included in the LDL exposure model was not performed. However, we updated our original list of sources using the Global Health Data Exchange (GHDx) repository[3] and added six new data sources including STEPS surveys from Vietnam and Zambia, and national surveys from the UK, South Africa, Singapore, Ethiopia, and Colombia. The total counts of the data inputs used for GBD 2021 are shown in Table 1. Details of inclusion and exclusion criteria and data processing steps follow.

Inclusion criteria

Studies were included if they were population-based and reported measurements of total LDL, TC, HDL, and/or TG from blood tests or if LDL was calculated using the Friedewald equation.[4] We assumed the

data were representative of the location if the geography or population chosen was not related to the diseases.

Outliers

All data were used in the modelling process unless an assessment of data strongly suggested that the data were biased. A candidate source was excluded if the quality of study did not warrant a valid estimate because of selection (non-representative populations) or if the study did not provide methodological details for evaluation. In a small number of cases, a datapoint was an outlier candidate if it deviated significantly from other datapoints within the respective country or region or the level was implausibly low or high based on expert judgement and other country data.

Data extraction

Where possible, individual-level data on LDL estimates were extracted from survey microdata, and these were collapsed across demographic groupings to produce mean estimates in the standard GBD five-year age-sex groups. If microdata were unavailable, information from survey reports or from literature were extracted along with any available measure of uncertainty, including standard error, uncertainty intervals, and sample size. Standard deviations were also extracted. Where LDL was reported split by groups other than age, sex, location, and year (eg, by diabetes status), a weighted mean was calculated.

Table 1: Data inputs for exposure for LDL cholesterol

	Countries with data	New sources	Total sources
Exposure	145	6	714

Relative risk

In all versions of GBD up to and including 2019[2,5] we estimated relative risks for TC (total cholesterol) and cardiovascular disease. These risks were derived from a meta-regression analysis of two combined epidemiological studies: the Asia Pacific Cohort Studies Collaboration (APCSC) and the Prospective Studies Collaboration (PSC).[6] For GBD 2017 and 2019, we use the relative risks for TC and cardiovascular disease to approximate the age-specific relative risks for LDL based on the knowledge that the relative risks for LDL and TC exhibit substantial similarity. Additionally, this approach relied on the strong linear correlation observed between TC and LDL at the individual level.

For GBD 2021, we revisited the underlying 52 randomised trials used in a previously published systematic review and meta-analysis of intensive LDL-lowering treatment for the prevention of major vascular events conducted by Wang and colleagues[7] and re-analysed the evidence on LDL as a risk factor of ischaemic heart disease and ischaemic stroke.

The original systematic review was done in accordance with Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines for meta-analyses of interventional studies.[8] Details of the methods, including the PRISMA diagram are reported elsewhere.[7] In brief, using the search terms *"low-density lipoprotein cholesterol" or "cholesterol" or "lowering" or "statins" or "ezetimibe" or "proprotein convertase subtilisin kexin type 9 antibody"*, the following databases were searched: MEDLINE, Embase, and the Cochrane Central Register of Controlled Trials from inception to August 1, 2018. A second search was repeated on June 15, 2019. For the purpose of the GBD final analysis, the search was updated on July 7, 2020, and the search results were updated accordingly. Table 2 shows the

search strings and search strategy used for each outcome included in this analysis. The PRISMA flow diagram is shown below.

Table 2: Search strings used in the literature review for LDL cholesterol and cardiovascular outcomes

Outcome	Search strategy	Search String	Number of studies selected
Cardiovascular outcomes	Past systematic review plus literature review of trials	Search string adapted from Wang et al. Updated search string: (hydroxymethylglutaryl coenzyme a reductase inhibitor[tw] OR statin[tw] OR statins[tw] OR ezetimibe[tw] OR PCSK9 inhibitor[tw] OR PCSK9 antibody[tw] OR Proprotein convertase subtilisin kexin type 9 antibody[tw]) AND (LDL-C[tw] OR Low-density lipoprotein cholesterol[tw] OR cholesterol[tw] OR lipid lowering[tw]) AND randomized controlled trial[Publication Type] AND (2019/06/15[PDAT] : 2020/07/07[PDAT])	38

Inclusion criteria were published randomised controlled trials of treatment to reduce cholesterol using statins, ezetimibe, or PCSK9 inhibitors, with at least 1000 patient-years of follow-up, and that reported cardiovascular outcomes of interest. All trials comparing therapy versus no therapy, more-intensive versus less-intensive intervention, or higher versus lower doses of a medication were eligible for inclusion. Primary and secondary prevention trials were included. Trials were excluded if patients were followed up for less than six months or treatment was compared with medications other than cholesterol-lowering drugs or placebo. Trials including enrolling patients on hemodialysis were excluded. The total counts of the relative risks data sources used for GBD 2021 are shown in Table 3.

Table 3: Data inputs for relative risks for LDL cholesterol

	Countries with data	New sources	Total sources
Relative risks for ischaemic heart disease	-	38	38
Relative risks for ischaemic stroke	-	30	30
Total relative risks (unique sources)	-	38	38

No countries reported given the assumption of global relative risks.

Data processing

For GBD 2021, there were no changes to data processing or methods used to estimate the LDL exposure model.

Lipid crosswalk

Total cholesterol consists of three major components: LDL, HDL, and TGL. LDL is often calculated for an individual using the Friedewald equation,[4] shown below:

$$LDL = TC - \left(HDL + \frac{TGL}{2.2}\right)$$

We used this relationship at the individual level to impute the mean LDL for a study population when only data on TC, HDL, and TGL were available. Because studies report different combinations of TC, HDL, and TGL, we constructed a single regression to utilise all available data to evaluate the relationship between each lipid and LDL at the population level. We used the following regression:

$$LDL = ind_{tc}\beta_1 TC - \left(ind_{hdl}\beta_2 HDL + ind_{tgl}\beta_3 TGL\right) + \sum \alpha_l I_l$$

Where ind_{tc}, ind_{hdl}, and ind_{tgl} are indicator variables for whether data are available for a given lipid, I_l is an indicator variable for a given set of available lipids l, and α_l is a unique intercept for each set of available lipid combinations. For example, for sources that only reported TC and HDL, $\alpha_{l=TC,HDL}$ should account for the missing lipid data, ie, TGL. The form of this regression allows us to estimate the betas for each lipid using all available data. As a sensitivity analysis, we also ran separate regressions for each set of available lipids and found that the single regression method had much lower root-mean-squared error. We found almost no relationship between LDL and HDL or TGL when TC was not available, so only studies that reported TC were adjusted to LDL.

Incorporating USA prevalence data

Survey reports and literature often report information only about the prevalence, but not the level, of hypercholesterolemia in the population studied. These sources were not used to model LDL, except for data from the Behavioral Risk Factors Surveillance System (BRFSS) because of the availability of a similarly structured exam survey covering the identical population, the National Health and Nutrition Examination Survey (NHANES). BRFSS is a telephone survey conducted in the USA for all counties. It collects self-reported diagnosis of hypercholesterolemia. These self-reported values of prevalence of raised TC in each age group, sex, USA state, and year were used to predict a mean TC for the same strata with a regression using data from the NHANES, a nationally representative health examination survey of the USA adult population. The regression was:

$$\mathrm{TC}_{\mathrm{l,a,t,s}} = \beta_0 + \beta_1 \mathrm{prev}_{\mathrm{l,a,t,s}}$$

where $\mathrm{TC}_{\mathrm{l,a,t,s}}$ *is the location, age, time, and sex specific mean total cholesterol and* $\mathrm{prev}_{\mathrm{l,a,t,s}}$ *is the location, age, time, and sex specific prevalence of raised total cholesterol. The coefficients for both models are reported in Table 4.*

Table 4: Coefficients in the sex-specific USA states total cholesterol prediction models

Term	Male model	Female model
Intercept	4.23	4.36

Prevalence	6.25	5.22

Out-of-sample root-mean-square error (RMSE) was used to quantify the predictive validity of the model. The regression was repeated ten times for each sex, each time randomly holding out 20% of the data. The RMSEs from each holdout analysis were averaged to get the average out-of-sample RMSE. The results of this holdout analysis are reported in Table 5. Total cholesterol estimates were crosswalked to LDL using the lipid crosswalk reported above.

Table 5: Out of sample RMSEs of the sex-specific USA states TC prediction models

	Male model	**Female model**
Out-of-sample RMSE	0.21 mmol/L	0.20 mmol/L

Age and sex splitting

Prior to modelling, data provided in age groups wider than the GBD five-year age groups were processed using the approach outlined in Ng and colleagues.[9] Briefly, age-sex patterns were identified using person-level microdata (60 sources) and estimate age-sex-specific levels of TC from aggregated results reported in published literature or survey reports. To incorporate uncertainty into this process and borrow strength across age groups when constructing the age-sex pattern, we used a model with auto-regression on the change in mean LDL over age groups:

$$\mu_a = \mu_{a-1} + \omega_a$$
$$\omega_a \sim N(\omega_{a-1}, \tau)$$

Where μ_a is the mean predicted value for age group *a*, μ_{a-1} is the mean predicted value for the age group previous to age group *a*, ω_a is the difference in mean between age group *a* and age group *a-1*, ω_{a-1} is the difference between age group *a-1* and age group *a-2*, and τ is a user-input prior on how quickly the mean LDL changes for each unit increase in age. We used a τ of 0.05 mmol/L for this model. Draws of the age-sex pattern were combined with draws of the input data needing to be split to calculate the new variance of age-sex-split datapoints.

Modelling strategy

Exposure

For GBD 2021, there were no changes in the modelling strategy used to estimate LDL exposure. Exposure estimates were produced from 1980 to 2020 for each national and subnational location, sex, and each five-year age group starting from 25. As in GBD 2019, we used a spatiotemporal Gaussian process regression (ST-GPR) framework to model the mean LDL at the location, year, age, and sex level. Details of the ST-GPR method used in GBD 2021 can be found elsewhere in the appendix.

Covariate selection

The first step of the ST-GPR framework requires the creation of a linear model for predicting LDL at the location, year, age, sex level. Covariates for this model were selected in two stages. First, a list of variables with an expected causal relationship with LDL was created based on significant association found within high-quality prospective cohort studies reported in the published scientific literature. Covariates included in the first step were mean body-mass index (BMI), prevalence of obesity, age- and sex-specific SEV for low fruit, age- and sex-specific SEV for low nuts and seeds, age- and sex-specific SEV

for low vegetables, Healthcare Access and Quality (HAQ) Index, and Socio-demographic Index (SDI). The second stage in covariate selection was to test the predictive validity of every possible combination of covariates in the linear model, given the covariates selected above. This was done separately for each sex. Predictive validity was measured without sample root-mean-squared error.

In GBD 2016, the linear model with the lowest root-mean-squared error for each sex was then used in the ST-GPR model. Beginning in GBD 2017, we used an ensemble model of the 50 models with the lowest root-mean-squared error for each sex. This allows us to utilise covariate information from many plausible linear mixed-effects models. The 50 models were each used to predict the mean LDL for every age, sex, location, and year, and the inverse-RMSE-weighted average of this set of 50 predictions was used as the linear prior. The relative weight contributed by each covariate is plotted by sex in Figure 2. The results of the ensemble linear model were used for the first stage in an ST-GPR model. The result of the ST-GPR model are estimates of the mean LDL for each age, sex, location, and year.

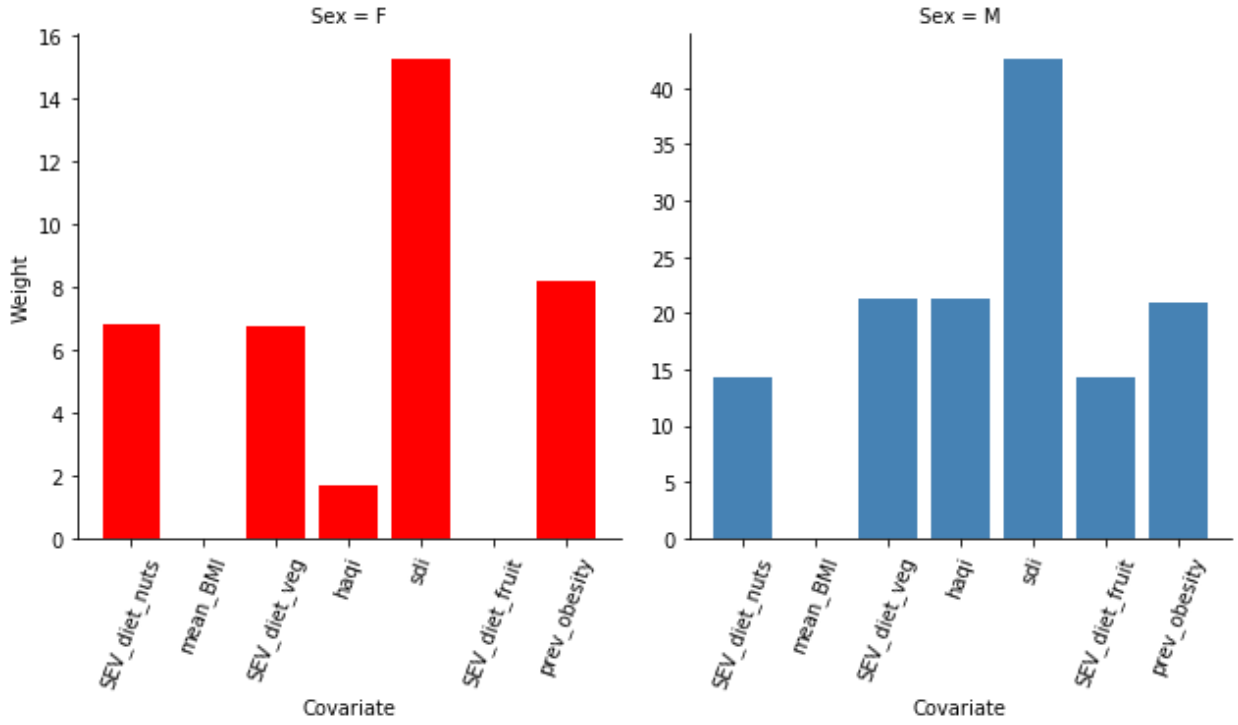
Estimate of standard deviation

The standard deviation of LDL within a population was estimated for each national and subnational location, sex, and five-year age group starting from age 25 using the standard deviation from person-level and some tabulated data sources. Person-level microdata accounted for 3009 of the total 4001 rows of data on standard deviation. The remaining 992 rows came from tabulated data. Tabulated data were only used to model standard deviation if they were sex-specific and five-year-age-group-specific and reported a population standard deviation LDL. The LDL standard deviation function was estimated using a linear regression:

$$\log\left(\mathrm{SD}_{c,a,t,s}\right) = \beta_0 + \beta_1 \log\left(\mathrm{mean\_LDL}_{c,a,t,s}\right) + \beta_2 \mathrm{sex} + \sum_{k=3}^{17} \beta_k \mathrm{I}_{A[a]}$$

where $\mathrm{mean\_LDL}_{c,a,t,s}$ is the country-, age-, time-, and sex-specific mean LDL estimate from ST-GPR and $\mathrm{I}_{A[a]}$ is a dummy variable for a fixed effect on a given five-year age group.

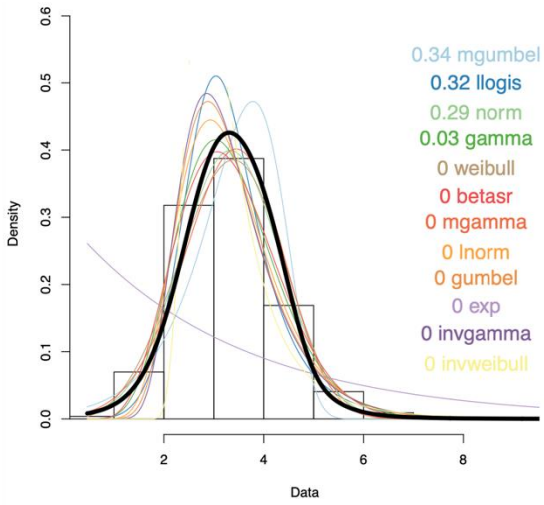
Figure 1: Covariates relative weights



Distribution shape modelling

The shape of the distribution of LDL was estimated using all available person-level microdata sources, which was a subset of the input data into the modelling process. The distribution shape modelling framework for GBD 2021 is detailed elsewhere in the appendix. Briefly, an ensemble distribution created from a weighted average of distribution families was fit for each individual microdata source, separately by age and sex. The weights for the distribution families for each individual source were then averaged and weighted to create a global ensemble distribution for each sex. Figure 2 shows the final ensemble distribution for both sexes combined.

Figure 2. Global ensemble distribution fit and distribution-specific weights for LDL



Theoretical minimum risk exposure level

For GBD 2017, based on a meta-analysis of randomised trials that showed that outcomes can be improved even at low levels of LDL cholesterol, below 1.3 mmol/L,[10] the TMREL for LDL was define as a uniform distribution between 0.7 and 1.3 mmol/L. This value remained unchanged for GBD 2019. For GBD 2021, we used the LDL values reported in the randomised trials used to estimate the LDL relative risks to update the TMREL as well. We first identify the LDL levels from the reference exposure (or control) and alternate (or comparison) groups of the studies for all outcomes associated to LDL. Then, we compute the midpoint of the lower and upper bounds of the reference exposure groups and define the TMREL as a uniform distribution with lower/upper bounds of 0.9 and 1.4. The lower bound was given by the percentile 15 of the midpoints calculated above, and the upper bound was given by the percentile 15 of only the upper bounds of the reference exposure groups. We calculated these ranges for each RO pair, and then took the weighted average of the ranges using global cause-specific deaths as the weights.

The TMREL was defined as a uniform distribution rather than a fixed value to represent the uncertainty regarding the level at which the scientific evidence was consistent with adverse effects of exposure.

Relative risk

For GBD 2021, we switched from using DisMod-MR 2.1[11] to meta-regression—Bayesian, regularised, Trimmed (MR-BRT)[12] to estimate the effect sizes and generate a dose–response curve for each of the outcomes associated with LDL (ischaemic heart disease and ischaemic stroke). This new approach enabled us to incorporate random effects across studies accounting for between-study heterogeneity, data sparsity, and publication bias. Relative risks (RR) for ischaemic heart disease and ischaemic stroke were modelled with log (RR) as the dependent variable and LDL exposure values as the independent variable. Due to data sparsity, and given that most of the studies included in the meta-regression do not report information disaggregated by stroke subtypes (ie, ischaemic vs. haemorrhagic stroke), we decided to combine data sources that reported “stroke” and “ischaemic stroke” as outcomes in a single model, assuming that physiologically LDL can be associated with ischaemic stroke only. Further technical details supporting estimation of non-log-linear risk curves using relative risks, trimming, and general meta-analysis models are detailed elsewhere[13] and can be found in a different section of this appendix.

We implemented the Fisher scoring correction to the heterogeneity parameter, which corrects for data-sparse situations. In such cases, the between-study heterogeneity parameter estimate may be 0, simply from lack of data. The Fisher scoring correction uses a quantile of gamma, which is sensitive to the number of studies, study design, and reported uncertainty. In addition, we have added methodology that can detect and flag publication bias. The approach is based on the classic Egger's regression strategy,[14] which is applied to the residuals of the model. In the current implementation, we do not correct for publication bias, but flag the risk–outcome pairs where the risk for publication bias is significant. For this analysis, no risk of publication bias was detected for LDL and related outcomes, as shown in the funnel plots in figures 3 and 4. Given the data limitations. we assumed that the estimated RRs were universal for all countries and sex categories and were the same for incidence and mortality.

Figure 3. Low-density lipoprotein cholesterol and ischaemic heart disease log relative risk (a) and residuals by estimated standard deviation (b)

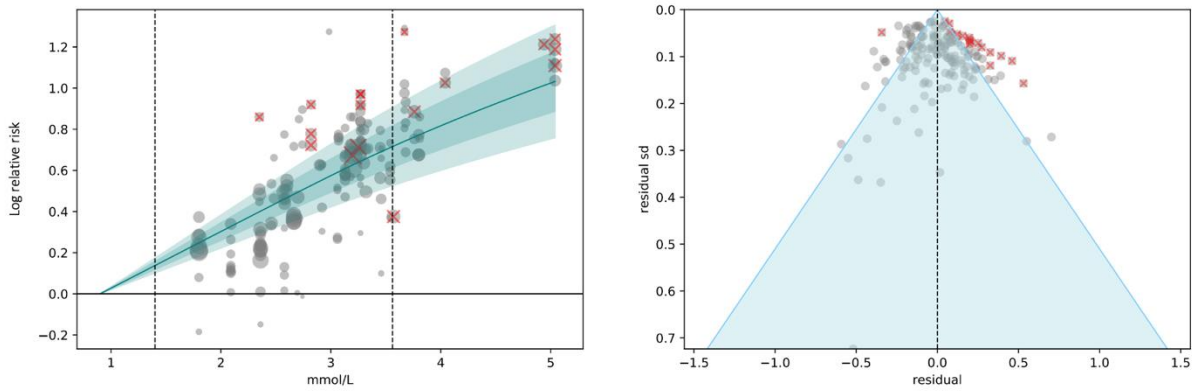
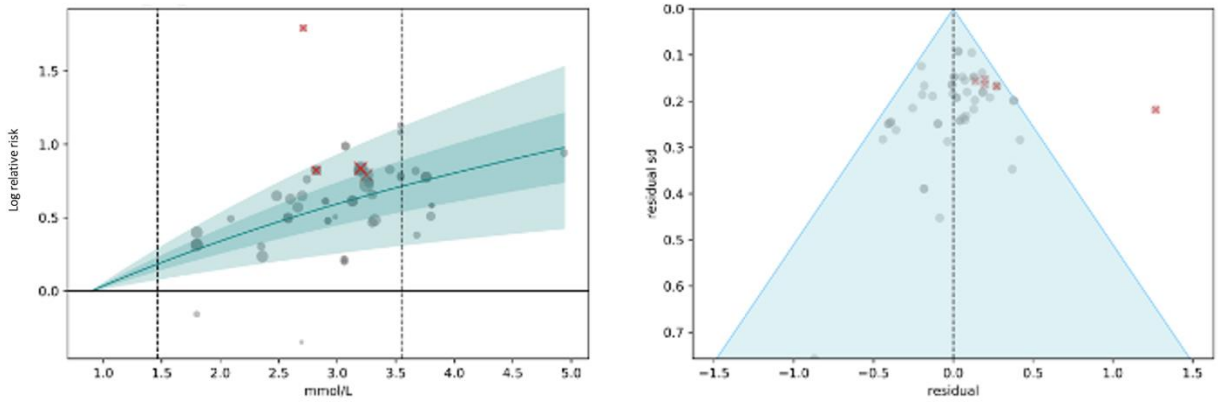


Figure 4: Low-density lipoprotein cholesterol and ischaemic stroke log relative risk (a) and residuals by estimated standard deviation (b)



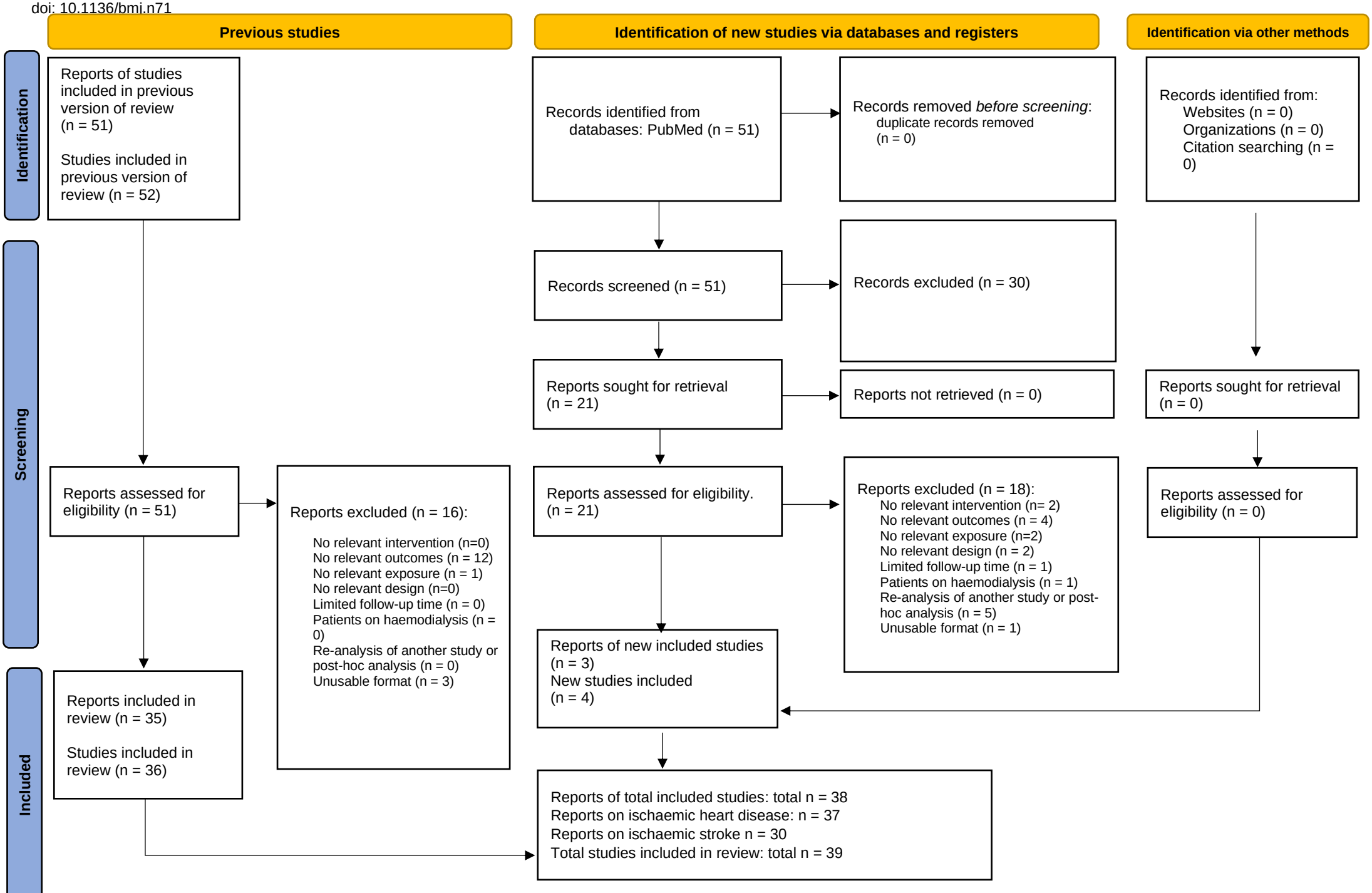
Figures 3 and 4. The risk curves are computed relative to an LDL cholesterol value of 0.9 mmol/L. In panel (a) the dark line indicates mean relative risk across LDL cholesterol exposure levels; the light and dark shading show 95% uncertainty intervals with and without between-study heterogeneity, respectively; the size of the datapoints corresponds to the inverse of the standard error, with those trimmed during the model fitting process marked by a red x; and the dashed lines represent the 15th percentile of the reference exposure and the 85th percentile of the alternative exposure. To visualise log-relative-risk points in panel (a), we plotted each datapoint with the x-value at the midpoint of the alternative group and the y-value offset by the difference between the reported and predicted log risk. Panel (b) depicts a customised funnel plot, with the x-axis representing residuals between predicted and observed relative risks, and the y-axis representing uncertainty from both measurement error and between-study heterogeneity.

To account for the heterogeneity of the effect size by age and given the limitations of both the available data and MR-BRT in terms of lack of age-specific data and estimates, we estimated cause-specific age attenuation factors using a second MR-BRT model with log (RR) as the dependent variable and age as an independent variable, including data for TC from the APCSC and the PSC cohorts only reported by Singh and colleagues. We then applied these cause-specific attenuation factors to the corresponding RR curve using the mid age at event observed in these two cohort studies (60–64 years) as the reference group to finally generate RR for standard five-year GBD age categories starting at age 25. With this new methodology, we removed the previous assumption that there is not a protective effect of LDL and stroke after age 70.

In future iterations of GBD, we plan to update the MR-BRT tool to be able to incorporate a second spline on age and generate more accurate age-specific RR curves.

PRISMA 2020 flow diagram. Study selection flow diagram for meta-analysis of LDLc and ischaemic heart disease and ischaemic stroke combined

From: Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ* 2021;372:n71. doi: 10.1136/bmj.n71



Citations

1 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. *The Lancet* 2016; **388**: 1659–724.

2 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.

3 Global Health Data Exchange | GHDx. http://ghdx.healthdata.org/ (accessed Oct 8, 2021).

4 Friedewald WT, Levy RI, Fredrickson DS. Estimation of the concentration of low-density lipoprotein cholesterol in plasma, without use of the preparative ultracentrifuge. *Clin Chem* 1972; **18**: 499–502.

5 Murray CJL, Aravkin AY, Zheng P, *et al.* 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.

6 Singh GM, Danaei G, Farzadfar F, *et al.* The Age-Specific Quantitative Effects of Metabolic Risk Factors on Cardiovascular Diseases and Diabetes: A Pooled Analysis. *PLOS ONE* 2013; **8**: e65174.

7 Wang N, Fulcher J, Abeysuriya N, *et al.* Intensive LDL cholesterol-lowering treatment beyond current recommendations for the prevention of major vascular events: a systematic review and meta-analysis of randomised trials including 327 037 participants. *Lancet Diabetes Endocrinol* 2020; **8**: 36–49.

8 Moher D, Liberati A, Tetzlaff J, Altman DG, Group TP. Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. *PLOS Med* 2009; **6**: e1000097.

9 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. *The Lancet* 2014; **384**: 766–81.

10 Boekholdt SM, Hovingh GK, Mora S, *et al.* Very Low Levels of Atherogenic Lipoproteins and the Risk for Cardiovascular Events: A Meta-Analysis of Statin Trials. *J Am Coll Cardiol* 2014; **64**: 485–94.

11 Vos T, Lim SS, Abbafati C, *et al.* Global burden of 369 diseases and injuries in 204 countries and territories, 1990–2019: a systematic analysis for the Global Burden of Disease Study 2019. *The Lancet* 2020; **396**: 1204–22.

12 Zheng P, Barber R, Sorensen RJD, Murray CJL, Aravkin AY. Trimmed Constrained Mixed Effects Models: Formulations and Algorithms. *J Comput Graph Stat* 2021; **30**: 544–56.

13 Sorensen (rsoren@uw.edu) R. Health Metrics Toolbox. https://ihmeuw-msca.github.io/index.html (accessed Oct 8, 2021).

14 Bias in meta-analysis detected by a simple, graphical test | The BMJ. https://www.bmj.com/content/315/7109/629.full (accessed Oct 8, 2021).