Institute for Health Metrics and Evaluation

Data Release Information Sheet

Data Summary

Dataset name: Global Burden of Disease Study 2019 (GBD 2019) Particulate Matter Risk Curves

Date of release: January 8, 2021

Date of update: March 18, 2021 ["Note on Scaling These Curves to the TMREL" was added to the "Additional Data Information" section]

Summary:

The Global Burden of Disease Study 2019 (GBD 2019), coordinated by the Institute for Health Metrics and Evaluation (IHME), estimated the burden of diseases, injuries, and risk factors for 204 countries and territories and selected subnational locations.

This dataset provides estimates for relative risks due to exposure to particulate matter for ischemic heart disease, stroke, chronic obstructive pulmonary disease, lung cancer, acute lower respiratory infection, type 2 diabetes mellitus, as well as birthweight and gestational age shifts, low birthweight (<2500g), and pre-term births (<37 weeks). The input data used to create the estimates are also provided. These splines are generated using the MR-BRT meta-regression tool and input data from epidemiologic studies of exposure to ambient air pollution, household air pollution from the use of solid fuels, and secondhand tobacco smoke.

Acknowledgements

Contributing organizations:

- Global Burden of Disease Collaborative Network

Funders:

- Bill and Melinda Gates Foundation (BMGF)

Suggested Citation:

Global Burden of Disease Collaborative Network. Global Burden of Disease Study 2019 (GBD 2019) Particulate Matter Risk Curves. Seattle, United States of America: Institute for Health Metrics and Evaluation (IHME), 2021.

Additional Data Information

This dataset provides the input data and risk curves for all of the particulate matter (PM2.5) outcomes in GBD 2019.

Statistical Draws Files: Each file contains exposures from 0 to 2500, rounded to 2 significant digits (i.e. 0, 0.01, 0.02, … , 1.1, 1.2 ,… , 11, 12,… , 110, 120, … , 1100, 1200, … 2500). For each exposure value, there are 1000 estimates of the effect.

Summary Files: Each file contains the mean (summary) estimate of the 1000 draws, and confidence intervals (CI).

Brief Methodological Notes

One curve is estimated for each cause for both incidence and mortality with the exception of CVD causes (ischemic heart disease and stroke). For these, separate curves are also estimated for each 5-year age group. All files represent relative risks except birthweight (continuous shift in grams, bw) and gestational age (continuous shift in weeks, ga).

In GBD 2019, the curve was generated without incorporating the the oretical minimum risk exposure level (TMREL) estimate. The TMREL was then used when generating RR estimates with the following equations:

For exposure, *X,*

$$RR_X = \begin{cases} 1 \ , & X \leq tmrel \\ \dfrac{MRBRT(X)}{MRBRT(tmrel)} \ , & X > tmrel \end{cases}$$

and

$$shift_X = \begin{cases} 0 \ , & X \leq tmrel \\ MRBRT(X) - MRBRT(tmrel) \ , & X > tmrel \end{cases}$$

With this methodological update, users can choose whatever value of the counterfactual is suitable to their own analysis. In GBD 2019, a uniform distribution from 2.4 to 5.9 micrograms/meter cubed was used. Also included in this dataset: the 1000 samples of this distribution used in the GBD 2019 analysis.

Note about the input data files: The fields used for fitting the curves include conc (95th percentile), conc_den (5th percentile), and log_rr & log_se (mean and standard error of the log relative risk scaled to the relevant exposure range) or shift & shift_se (mean and standard error of the shift scaled to the relevant exposure range for birthweight and gestational age). These datasets include active smoking (ier_source="AS"), but active smoking was NOT included in the final MR-BRT models.

Much more detailed methodological information is included in the file
IHME_GBD_2019_PM_RISK_MODELING_DESCRIPTION_Y2021M01D06.PDF

Note on Scaling These Curves to the TMREL

The files contain RR values >1 for exposure levels below the TMREL upper boundary because this curve has not yet been scaled to the TMREL. To scale the RR predictions to the TMREL, there are three strategies you could use.

1. The first is to follow GBD 2019 methods directly, using a draw-based approach:
   - 1000 draws were predicted of the PM2.5 exposure level for a given pixel
   - Each PM2.5 draw was assigned to a TMREL draw (modeled as a uniform distribution between 2.4 and 5.9 ug/m3, resulting in 1000 TMREL draws total)
   - Each PM2.5 draw and each TMREL draw was assigned to its own MRBRT RR draw, resulting in 1000 MRBRT (pm2.5) draws and 1000 MRBRT (TMREL) draws
     - MRBRT predictions were assigned draw-wise: for example, the 1st PM2.5 draw was assigned to the 1st draw of the MRBRT prediction at that exposure; the 2nd PM2.5 draw was assigned the 2nd draw of the MRBRT prediction at that exposure…
   - To calculate the RR for each exposure draw
     - For pairs where MRBRT (pm2.5) > MRBRT (TMREL), the resulting RR (exposure) was calculated as MRBRT (pm2.5) / MRBRT (TMREL)
     - For pairs where MRBRT (pm2.5) <= MRBRT (TMREL), the RR (exposure) was assigned a value of 1

Note that because PAFs were calculated using a draw-based approach, some exposure draws within the TMREL range actually do end up having an RR value >1. This occurs when a PM2.5 exposure value < the TMREL upper boundary is paired with a smaller TMREL draw.

2. The second approach is to scale the MRBRT curve in its entirety before assigning RR predictions to PM2.5 exposure values. You could do this by dividing each of the 1000 RR draws for each exposure level by RR predictions for 1000 draws of the TMREL. You could also divide each of the 1000 RR draws for each exposure level by 1000 draws of the RR at the TMREL midpoint (4.15 ug/m3) only, though this is a simplification.

3. The third (further simplified) approach is to assign an RR of 1 to any exposure less than the upper bound of the TMREL distribution, and divide the rest of the RR predictions by 1000 draws of the RR at the TMREL midpoint. This would likely produce slightly smaller PAFs than those obtained using a draw-based approach because some of the exposure draws that fall within the TMREL bounds end up having a RR > 1 (as described above).

The R code that we used to calculate PAFs for GBD 2019 can be found here:
https://github.com/ihmeuw/ihme-modeling/blob/master/gbd_2019/risk_factors_code/air/paf/03_calc_rr_paf.R

http://ghdx.healthdata.org/gbd-2019/code/risk-1

Data Files Information

File Inventory

File Name	Description	Version Date
IHME_GBD_2019_PM_RISK_DRAWS.zip	Statistical draws data. ZIP archive contains 38 CSV files.	January 6, 2021
IHME_GBD_2019_PM_RISK_SUMM.zip	Summary/mean estimates data. ZIP archive contains 38 CSV files.	January 6, 2021
IHME_GBD_2019_PM_RISK_INPUTS.zip	Inputs data. ZIP archive contains 40 CSV files.	January 6, 2021
IHME_GBD_2019_PM_RISK_TMREL_DRAWS_Y2021M01D06.CSV	Theoretical minimum risk exposure level (TMREL) draws (1000)	January 6, 2021
IHME_GBD_2019_PM_RISK_SOURCE_LIST_Y2021M01D06.CSV	All ambient and household sources used in generating risk curves	January 6, 2021
IHME_GBD_2019_PM_RISK_EXTRACTION.zip	Extraction sheets for studies measuring exposure to: • Household air pollution (HAP) • Outdoor or ambient air pollution (OAP) • OAP with adverse reproductive outcomes (ARO) • Secondhand smoke (SHS) ZIP archive contains 4 XLSM files	January 6, 2021
IHME_GBD_2019_PM_RISK_MODELING_DESCRIPTION_Y2021M01D06.PDF	Ambient particulate matter pollution modeling description. (Drawn from Methods Appendix for "Global burden of 87 risk factors in 204 countries and territories, 1990–2019: a systematic analysis for the Global Burden of Disease Study 2019"	January 6, 2021
IHME_GBD_2019_PM_RISK_INFO_SHEET_Y2021M01D06.PDF	Data Release Information Sheet	March 18, 2021

Variable Information

Cause/Condition List

The following cause or condition variables appear in the file names for the input, statistical draws, and summary estimate files, and in the file names themselves for the input data and summary data.

Cause/Condition Variable	Label
bw	Birth weight
cvd_ihd	Ischemic heart disease
cvd_ihd_25	Ischemic heart disease, ages 25 to 29
cvd_ihd_30	Ischemic heart disease, ages 30 to 34
cvd_ihd_35	Ischemic heart disease, ages 35 to 39
cvd_ihd_40	Ischemic heart disease, ages 40 to 44
cvd_ihd_45	Ischemic heart disease, ages 45 to 49
cvd_ihd_50	Ischemic heart disease, ages 50 to 54
cvd_ihd_55	Ischemic heart disease, ages 55 to 59
cvd_ihd_60	Ischemic heart disease, ages 60 to 64
cvd_ihd_65	Ischemic heart disease, ages 65 to 69
cvd_ihd_70	Ischemic heart disease, ages 70 to 74
cvd_ihd_75	Ischemic heart disease, ages 75 to 79
cvd_ihd_80	Ischemic heart disease, ages 80 to 84
cvd_ihd_85	Ischemic heart disease, ages 85 to 89
cvd_ihd_90	Ischemic heart disease, ages 90 to 94
cvd_ihd_95	Ischemic heart disease, ages 95 to 99
cvd_stroke	Stroke
cvd_stroke_25	Stroke, ages 25 to 29
cvd_stroke_30	Stroke, ages 30 to 34
cvd_stroke_35	Stroke, ages 35 to 39
cvd_stroke_40	Stroke, ages 40 to 44
cvd_stroke_45	Stroke, ages 45 to 49
cvd_stroke_50	Stroke, ages 50 to 54
cvd_stroke_55	Stroke, ages 55 to 59
cvd_stroke_60	Stroke, ages 60 to 64
cvd_stroke_65	Stroke, ages 65 to 69
cvd_stroke_70	Stroke, ages 70 to 74
cvd_stroke_75	Stroke, ages 75 to 79
cvd_stroke_80	Stroke, ages 80 to 84
cvd_stroke_85	Stroke, ages 85 to 89
cvd_stroke_90	Stroke, ages 90 to 94
cvd_stroke_95	Stroke, ages 95 to 99
ga	Gestational age
lbw	Low birthweight
lri	Lower respiratory infections
neo_lung	Tracheal, bronchus, and lung cancer
ptb	Preterm birth
resp_copd	Chronic obstructive pulmonary disease

Cause/Condition Variable	Label
t2_dm	Diabetes mellitus type 2

Input Data

Variable	Variable Label	Variable Definition
underlying_nid	Underlying Node ID	GHDx identifier for some studies--used for meta-analyses, certain database sources, and in some other specific cases this is required.
nid	Node ID	GHDx identifier for each study.
source_type	Study Type	Type of epidemiological study (e.g., prospective cohort, case-control, etc.).
location_name	Location Name	Concatenation of location name, ISO3 code, and, for subnational locations, IHME-specific location identification number.
location_id	Location ID	IHME-specific location identification number.
ihme_loc_id	IHME Location ID	ISO3 code, and, for subnational locations, ISO3 code plus IHME-specific location identification number.
study	Name of Study	If cohort does not have a unique name, study is identified by author, publication year, and location.
ier_source	Type of Pollution	Type of pollution: outdoor (ambient) air pollution (OAP), household air pollution (HAP).
ier_cause	Cause Name	IHME-specific code for cause.
case_cutpoint	Case Cutpoint	For categorical gestational age and birthweight, what is the cutoff point?
bw_min	Birthweight (Minimum)	Restrictions on the sample population in grams (g).
bw_max	Birthweight (Maximum)	Restrictions on the sample population in grams (g).
ga_min	Gestational Age (Minimum)	Restrictions on the sample population in weeks.
ga_max	Gestational Age (Maximum)	Restrictions on the sample population in weeks.
year_start	Start Year	First year of data collection (year of baseline); if not given, subtract 3 years from publication year.
year_end	End Year	Last year of data collection (follow-up); if not given, subtract 1 year from publication year.

Variable	Variable Label	Variable Definition
year_id	Year ID	When only one year was provided for the study period this was used for year of data collection.
measure	Measure	Epidemiological measure used to denote effect size (odds ratio, relative risk, hazard ratio, etc.).
conc_increment	Concentration (Increment)	Unit increase of PM2.5 pollution corresponding to reported effect size (measure).
conc_mean	Concentration (Mean)	Mean of study population PM2.5 exposure distribution.
conc_den	Concentration (Density)	Density of study population PM2.5 exposure distribution.
conc	Concentration	Measures of study population PM2.5 exposure distribution.
shift	Shift	Mean reported shift in gestational age or birthweight.
shift_lower	Shift (Lower Bound)	2.5% reported shift in gestational age or birthweight.
shift_upper	Shift (Upper Bound)	97.5% reported shift in gestational age or birthweight.
shift_se	Shift (Standard Error)	Standard error of reported shift in gestational age or birthweight.
weight	Weight	Weight given to extracted value, calculated as the inverse variance of the estimate.
shift_unit	Shift (Unit)	Mean reported shift in gestational age or birthweight converted to 1 unit (µg/m3) change
shift_unit_lower	Shift (Unit, Lower Bound)	Mean reported shift in gestational age or birthweight converted to 1 unit (µg/m3) change.
shift_unit_upper	Shift (Unit, Upper Bound)	Mean reported shift in gestational age or birthweight converted to 1 unit (µg/m3) change.
shift_unit_se	Shift (Unit, Standard Error)	Mean reported shift in gestational age or birthweight converted to 1 unit (µg/m3) change.
exp_rr	Relative risk	Mean of reported relative risk.
exp_rr_lower	Relative Risk (Lower Bound)	2.5% percentile of reported relative risk.
exp_rr_upper	Relative Risk (Upper Bound)	97.5% percentile of reported relative risk.
log_rr	Log Relative Risk	Mean of (reported) log relative risk.
log_rr_lower	Log Relative Risk (Lower Bound)	2.5% percentile of (reported) log relative risk.

Variable	Variable Label	Variable Definition
log_rr_upper	Log Relative Risk (Upper Bound)	97.5% percentile of (reported) log relative risk.
log_se	Log Relative Risk (Standard Error)	Standard error of (reported) log relative risk.
exp_rr_unit	Relative Risk (Unit)	Mean reported relative risk converted to 1 unit (µg/m3) change.
exp_rr_unit_lower	Relative Risk (Unit, Lower Bound)	2.5% percentile of reported relative risk converted to 1 unit (µg/m3) change.
exp_rr_unit_upper	Relative Risk (Unit, Upper Bound)	97.5% percentile of reported relative risk converted to 1 unit (µg/m3) change.
log_rr_unit	Log Relative Risk (Unit)	Mean (reported) log relative risk converted to 1 unit (µg/m3) change.
log_rr_unit_lower	Log Relative Risk (Unit, Lower Bound)	2.5% percentile of (reported) log relative risk converted to 1 unit (µg/m3) change.
log_rr_unit_upper	Log Relative Risk (Unit, Upper Bound)	2.5% percentile of (reported) log relative risk converted to 1 unit (µg/m3) change.
log_unit_se	Log Relative Risk (Unit, Standard Error)	Standard error of (reported) log relative risk converted to 1 unit (µg/m3) change.
cv_subpopulation	Covariate: Sub-population	0 if generalizable--general population with reasonable exclusions, 1 analysis of subgroup such as high-risk group.
cv_exposure_population	Covariate: Population Exposure	0 for exposure measured/modeled at the individual (postal code or </= 1 km radius) and 1 for population level exposure.
cv_exposure_selfreport	Covariate: Exposure Self-report	0 for measured exposure and 1 for self-report.
cv_exposure_study	Covariate: Study Design	0 for exposure measured multiple times (time component in model) and 1 for only at baseline.
cv_outcome_selfreport	Covariate: Outcome Self-reported	0 outcome based on death certificate or medical record and 1 if self-report.
cv_outcome_unblinded	Covariate: Outcome Unblinded	0 if assessment of outcome is blind to exposure or vice versa and 1 if unblinded.
cv_reverse_causation	Covariate: Reverse Causation	0 no risk of reverse causation and 1 if there is a risk.
cv_confounding_nonrandom	Covariate: Non-random Confounding	0 for randomized study and 1 for non-randomized.
cv_counfounding.uncontroled	Covariate: Uncontrolled Confounding	0 for randomization or outcome controlled for all major known

Variable	Variable Label	Variable Definition
		confounders and age, sex, education, income, and smoking. 1 for age, sex, and other critical determinants. 2 for only age and sex.
cv_selection_bias	Covariate: Selection Bias	0 for greater than 95% follow-up, 1 for follow up of 85-95%, and 2 for less than 85% follow up. Case-control studies should be scored based on the percentages of cases and controls for which exposure data could be ascertained.
education	Education	Was education controlled for at individual level, ecological level, both, or neither?
income	Income	Was income controlled for at individual level, ecological level, both, or neither?
median_age_fup	Median age of cohort	Median age of the cohort during the follow-up period.
incidence	Disease measure	Disease measure of
child	Study population HAP--lri studies that measured exposure for children (edited)	For lower respiratory infection studies, 1 if study population was children 0 if adults.
new_gbd2019	New data extraction for GBD 2019	New data extraction for GBD 2019 cycle (1 if yes, 0.5 if data were re-extracted, 0 if included in prior GBD releases)

Statistical Draws Data

Variable	Variable Label	Variable Definition
exposure	Exposure	Exposures from 0 to 2500, rounded to 2 significant digits.
draw_#	Draw Number	Number of each statistical draw in model for a particular PM2.5 value.

Summary Data

Variable	Variable Label	Variable Definition
label	Cause	Cause for which splines have been generated.

Variable	Variable Label	Variable Definition
exposure_spline	Exposure Value	Given value of PM2.5 pollution (µg/m3) for which draws have been generated.
mean	Mean	Mean effect size estimate.
lower	95% Uncertainty Interval (Lower Bound)	2.5% percentile effect size estimate.
upper	95% Uncertainty Interval (Upper Bound)	97.5% percentile size estimate.

Additional Information

Terms and Conditions

http://www.healthdata.org/about/terms-and-conditions

Contact information

To request further information about this dataset, please contact IHME:

Address:
Institute for Health Metrics and Evaluation
Population Health Building/Hans Rosling Center
3980 15th Ave. NE, Seattle, WA 98195 USA
UW Campus Box #351615

Telephone: +1-206-897-2800

Fax: +1-206-897-2899

Email: data@healthdata.org

Website: www.healthdata.org

These files may be updated periodically, so we appreciate hearing feedback or additional information about how these data are being used.