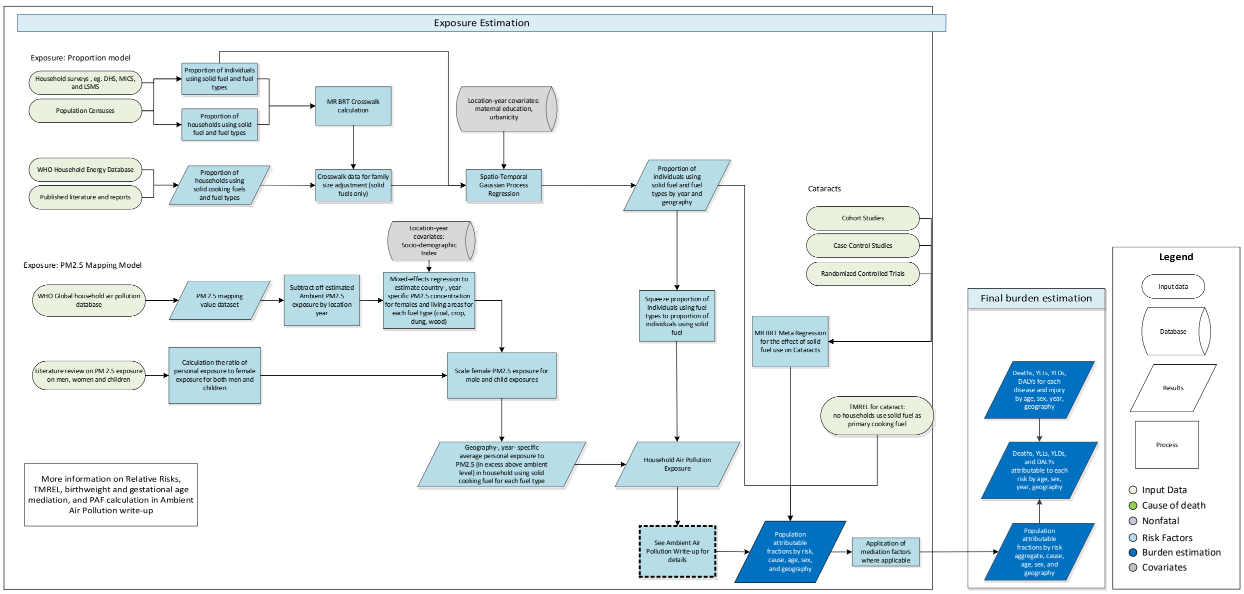
Household air pollution

Flowchart

Household Air Pollution from Solid Fuels



Input data and methodological summary

Exposure

Definition

Exposure to household air pollution from solid fuels (HAP) is estimated from both the proportion of individuals using solid cooking fuels and the level of exposure to particulate matter less than 2.5 micrometers in diameter ($PM_{2.5}$) air pollution for these individuals. Solid fuels in our analysis include wood, coal/charcoal, dung, and agricultural residues.

Input data

We extracted information on the use of solid fuels for cooking from standard multi-country survey series, including the Demographic and Health Surveys (DHS), Living Standards Measurement Surveys (LSMS), Multiple Indicator Cluster Surveys (MICS), and World Health Surveys (WHS). We also used data from censuses and country-specific survey series, such as the Kenya Welfare Monitoring Survey and South Africa General Household Survey. To fill remaining gaps in survey and census data, we downloaded the WHO Household Energy Database and updated estimates using extracted information from literature through a systematic review.[1] From this combined body of input data, each nationally or subnationally representative datapoint provided an estimate of the percentage of households or individuals using solid cooking fuels. We used studies from 1980 to 2020 to inform our time series estimates.

We excluded sources that did not distinguish specific primary fuel types, estimated fuel used for purposes other than cooking (eg, lighting or heating), failed to report standard error or sample size, reported over 15% missingness for households surveyed, reported fuel use in physical units, or were secondary sources referencing primary analyses.

Table 1: Data inputs for exposure for household air pollution.

Input data	Exposure
Site-years (total)	1117
Number of countries with data	161
Number of GBD regions with data (out of 21 regions)	20
Number of GBD super-regions with data (out of 7 super-regions)	7

Family size crosswalk

Many estimates in the WHO Energy Database and other reports quantify the proportion of households using solid fuel for cooking; however, we are interested in the proportion of individuals using solid fuel for cooking for exposure and burden assessment. To crosswalk these estimates, where available, we extracted fuel use at both the individual and household levels. We used studies that reported values for both household and individual solid fuel use and did not report a mean of 0 or 1. This resulted in 8074 source-specific pairs used as input data for the crosswalk model, which was modelled with the meta-regression—Bayesian, regularised, trimmed (MR-BRT) meta-regression tool. We applied this crosswalk only to proportion estimates for the parent solid fuel category. We did not adjust fuel-specific (coal/charcoal, crop, dung, or wood) proportion estimates due to lack of sufficient data for each individual fuel type.

Table 2: MR-BRT crosswalk adjustment factors for household air pollution exposure

Data input	Reference or alternative case definition	Gamma	Beta coefficient, logit (95% UI)*	Adjustment factor**
Proportion of individuals	Ref	0.095	---	---
Proportion of households	Alt		-0.094 (-0.097, -0.090)	1.099 (1.094–1.102)

**MR-BRT crosswalk adjustments can be interpreted as the factor the alternative case definition is adjusted by to reflect what it would have been had it been measured using the reference case definition. If the log/logit beta coefficient is negative, then the alternative is adjusted up to the reference. If the log/logit beta coefficient is positive, then the alternative is adjusted down to the reference.*

***The adjustment factor column is the exponentiated negative beta coefficient. For log beta coefficients, this is the relative rate between the two case definitions. For logit beta coefficients, this is the relative odds between the two case definitions.*

We applied this coefficient to household-only solid fuel reports with the following formula:

$prop_{individ}$ = the proportion of individuals using solid fuel for cooking, and

$prop_{hh}$ = the proportion of households using solid fuel for cooking.

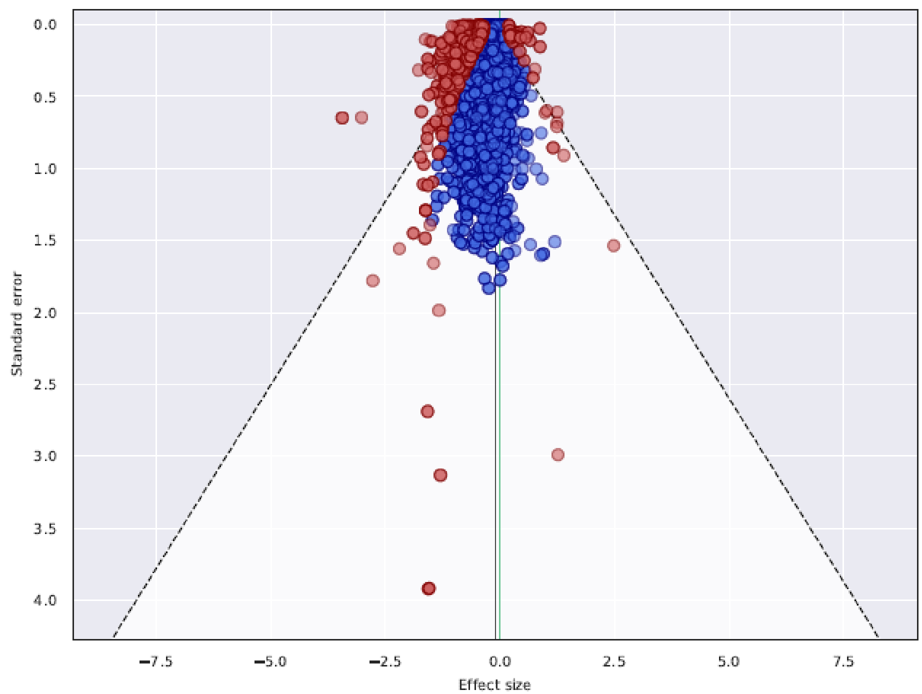
$$\log\left(\frac{prop_{individ}}{1 - prop_{individ}}\right) = \log\left(\frac{prop_{hh}}{1 - prop_{hh}}\right) - \beta$$

or

$$prop_{individ} = \frac{prop_{hh} * e^{-\beta}}{1 - prop_{hh} + prop_{hh} * e^{-\beta}}$$

As a result, household studies were inflated to account for bias in size between households that use solid cooking fuels and those that do not. Larger households are more likely to use solid fuels for cooking. The following figure depicts the 8074 datapoints that informed the crosswalk model. Red points indicate the 10% of studies trimmed as outliers during model fitting.

Figure 1: MR-BRT crosswalk for household air pollution exposure



$logit(prop_{hh}) - logit(prop_{individual})$

Modelling strategy

As in the Global Burden of Disease (GBD) Study 2019, household air pollution was modelled at the individual level using a three-step modelling strategy implementing linear regression, spatiotemporal regression, and Gaussian process regression (GPR). The full ST-GPR process is specified elsewhere in this appendix.

For GBD 2020, we updated the HAP proportion model to disaggregate estimates of solid fuel use to estimate the proportion of individuals using each of the following component fuel type categories: 1) coal or charcoal, 2) crop residue, 3) dung, and 4) wood. With this strategy, we can more finely characterise individual exposure to $PM_{2.5}$ due to solid fuel use by applying fuel-specific mapping values to fuel-specific proportion estimates. This change addresses an important limitation in our model, in that it previously assumed equal $PM_{2.5}$ exposure for all solid fuel categories.

Fuel type-specific estimates were generated by first using ST-GPR to generate location- and year-specific estimates for coal, crop, dung, and wood. ST-GPR was also used to create estimates for the parent solid fuel category, as in GBD 2019. The first step of the ST-GPR modelling process is a mixed-effect linear regression of logit-transformed proportion of individuals using solid cooking fuels. For each of the linear models, maternal education and the proportion of population living in urban areas were used as covariates. These models also included nested random effects by GBD region and GBD super-region.

Table 3: First-stage linear model and coefficients (solid model)

Variable	Beta (95% UI)
Intercept	3.36 (2.01, 4.71)
Maternal education (years per capita)	-0.55 (-0.58, -0.51)
Urbanicity (proportion of population living in urban areas)	-0.14 (-0.67, 0.39)

The four fuel-type-specific proportion estimates were then squeezed to the estimates for the overall proportion of individuals using solid fuel for cooking. For each location and year, we used the following formula, where $prop_{coal}$, $prop_{crop}$, $prop_{dung,}$ $prop_{wood}$, and $prop_{solid}$ indicate the proportion of individuals using coal, crop, dung, wood, or any type of solid fuel, respectively.

$$\text{Let } prop_{total} = prop_{coal} + prop_{wood} + prop_{crop} + prop_{dung}$$

$$S = prop_{total} \,/\, prop_{solid}$$

For each fuel category, with coal shown below as an example, the adjusted (squeezed) proportion is calculated as

$$prop_{coal}{}' = prop_{coal} \,/\, S$$

In preliminary model iterations, we mapped mixed fuel strings (eg, "wood and agricultural residues") to the category associated with highest $PM_{2.5}$ exposure to avoid underestimating HAP exposure. However, fuel-specific ST-GPR models were unstable with this approach. We therefore excluded mixed-fuel string studies from final estimates for fuel-specific proportions, though we retained these studies when modelling the proportion of overall solid fuel use.

Theoretical minimum-risk exposure level

For all HAP outcomes except cataract, burden is related to both ambient and household air pollution. These PAFs are estimated jointly and the theoretical minimum-risk exposure level (TMREL) is defined as a uniform distribution between 2.4 and 5.9 µg/m^3 $PM_{2.5}$. For cataract, the TMREL is defined as no individuals using solid cooking fuel.

Relative risks

The outcomes associated with household air pollution are lower respiratory infections (LRI), stroke, ischaemic heart disease (IHD), chronic obstructive pulmonary disease (COPD), lung cancer, type 2 diabetes, and cataract. Low birthweight and short gestation are also outcomes attributable to household air pollution through a mediation analysis. With the exception of cataract, all causes share risk curves and are calculated jointly with ambient particulate matter pollution.

Cataract relative risk meta-analysis

Prior to GBD 2019, we used the results of an external meta-analysis with a summary relative of 2.47 (95% UI 1.63–3.73) for cataract risk estimates.[2] While this effect estimate was for both sexes, in the past we estimated burden for women only because women are known to have higher HAP exposure than men. In GBD 2019, we performed our own meta-regression analysis of household air pollution and cataracts. We updated this meta-regression for GBD 2020.

We extracted all the component studies of the above meta-analysis paper but excluded one cross-sectional study. GBD risk factor analyses typically do not include cross-sectional analyses due to their weaker evidence base. In literature search conducted in GBD 2019, we found one additional paper describing different fuel types and cataracts.[3] We excluded this study because there was no comparison group without solid fuel use. We conducted an additional literature search in GBD 2020 but found no new studies to include. The following search string was used to identify studies in the PubMed database published between January 1, 2017, and July 22, 2020 (date of search).

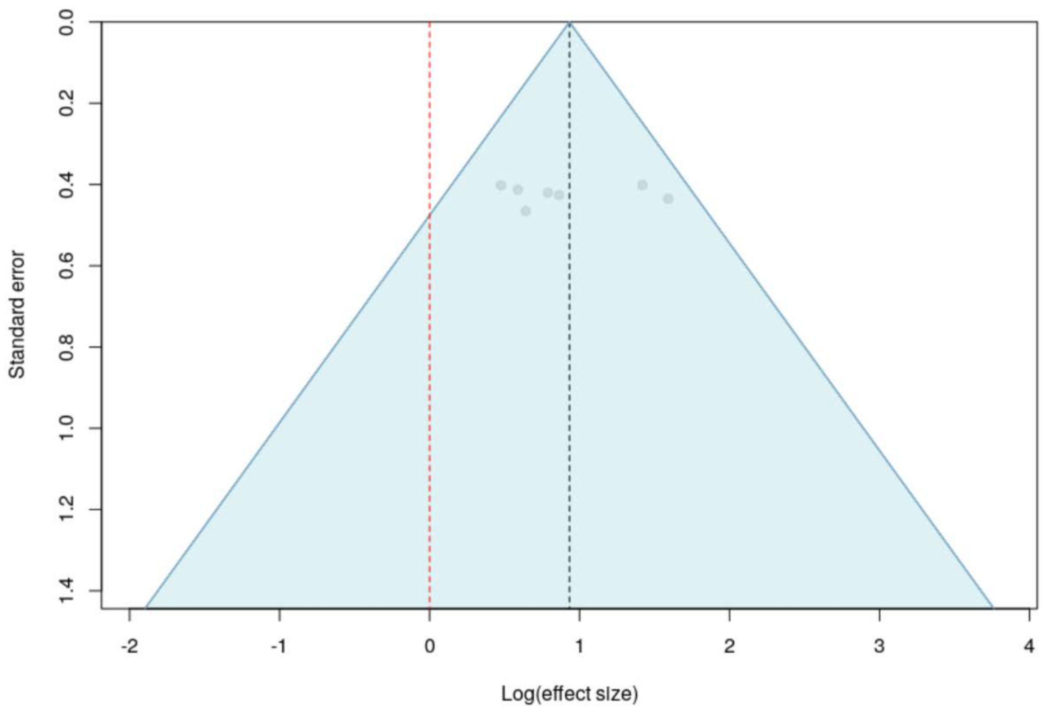
Search string: ((“Air Pollution, Indoor”[Mesh] OR “Household air”[Title/Abstract] OR “Indoor air pollution”[Title/Abstract] OR “Indoor fine particulate matter”[Title/Abstract] OR “Indoor particulate matter”[Title/Abstract] OR “Indoor air quality”[Title/Abstract] OR “Airborne particulate matter”[Title/Abstract]) AND (“Cataract”[Title/Abstract] OR “Cataracts”[Title/Abstract] OR “Cataracts”[Mesh] OR “Lens Opacities”[Mesh] OR “Lens Opacity”[Mesh] OR “Opacities, Lens”[Mesh] OR “Opacity, Lens”[Mesh] OR “Cataract, Membranous”[Mesh] OR “Cataracts, Membranous”[Mesh] OR “Membranous Cataract”[Mesh] OR “Membranous Cataracts”[Mesh] OR “Pseudoaphakia”[Mesh] OR “Pseudoaphakias”[Mesh]))

Our resulting dataset contained eight estimates from six sources in India and Nepal. We ran a MR-BRT meta-regression on these eight estimates to generate a summary effect size of 2.52 (95% UI 1.42–4.57). We did not trim any of the observations due to the relatively few input studies available compared to other GBD risk factors. We used the MR-BRT automated covariate selection process to identify significant covariates from those extracted to quantify between-study heterogeneity. Briefly, a series of loosening Lasso penalty parameters were applied to a log-linear meta-regression on all input effect size observations. Then, covariates with a non-zero coefficient were tested for significance using a Gaussian prior (significance threshold = 0.05). No significant covariates were identified. The table and figure below provide the model coefficients and a visual representation.

Table 4: MR-BRT relative risk meta-analysis for household air pollution and cataract

Covariate	**Gamma**	**Beta coefficient, logit (95% UI)**	**Beta coefficient, adjusted (95% UI)**
Intercept	0.109	0.939 (0.623–1.278)	2.56 (1.86–3.59)

Figure 2: Household air pollution and cataract risk literature funnel plot



Studies reported effect sizes for males, females, and/or both sexes. In a sensitivity analysis conducted in GBD 2019 and repeated in GBD 2020, we included a covariate for sex and found no significant difference in effect size by sex. We therefore estimate cataract as an outcome of household air pollution in both males and females.

For GBD 2020, we also implemented evidence scoring to provide an empirical measure of strength of evidence for risk-outcome pairs across risk factors in the GBD study (described in more detail elsewhere). Prior to generating an evidence score, we conducted an additional post-analysis step to detect and flag publication bias in the input data. This approach is based on the classic Egger's Regression strategy, which is applied to the residuals in our 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. Publication bias was not detected for HAP-cataract risk literature. The resulting evidence score for HAP and cataract was -0.009, which corresponds to a star rating of 1.

In GBD 2020, we also made key changes to our particulate matter risk curves. These risk curves, the mediation analysis for birthweight and gestational age, and the joint-estimation PAF approach are described in the Ambient Particulate Matter Pollution appendix.

$PM_{2.5}$ mapping value estimation

To calculate relative risks from particulate matter risk curves for individuals using solid fuels for cooking, we first estimated the $PM_{2.5}$ exposure level resulting from usage of each fuel type. Input data for the HAP mapping model included indoor and personal measurement data from the WHO Global Household Air Pollution Measurements database, which contains 196 studies with measurements from 43 countries of various pollution metrics in households using solid fuel for cooking.[4] For GBD 2020, we also added data from the PURE-AIR study published in 2020, which includes additional measurements from

120 rural locations in Bangladesh, Chile, China, Colombia, India, Pakistan, Tanzania, and Zimbabwe.[5] The final dataset included 390 estimates from 76 studies in 47 unique locations. We included 281, 81, 9, and 19 measurements for indoor exposure and personal monitors for females, children (under 5), and males, respectively. 314 estimates were in households using solid fuels, 61 in households using clean fuels (gas or electricity) only, and 15 in households using a mixture of solid and clean fuels. Of measurements from households using solid fuels, we included 40, 20, 13, 155, and 86 measurements for coal, crop, dung, wood, and mixed fuels, respectively.

The following models were used to predict log-transformed estimates of excess $PM_{2.5}$ for each individual fuel type (coal, crop, dung, wood) and for the parent solid category. Predictions for the parent solid category were used only to prepare relative risk input data for analysis, not for predicting individual exposure to $PM_{2.5}$ from solid fuel use.

Fuel types:

$$\log(excess\ PM) \sim crop + coal + dung + wood + measure\ group + 24\ hr\ measurement + LDI + (1|study)$$

Solid:

$$\log(excess\ PM) \sim solid + measure\ group + 24\ hr\ measurement + LDI + (1|study)$$

Where,

- 24-hour measurement: binary variable equal to 1 if the measurement occurred over at least a 24-hour period and not only during mealtimes
- Measure group: categorical variable indicating indoor, female, male, or children
- Solid: indicator variable equal to 1 if the measurements were among households using solid fuel only, 0.5 if the measurements represented a mix of clean and solid fuels, and 0 if the households only used clean fuels.

For previous GBD cycles, we also included the Socio-demographic Index (SDI) as a variable to predict a unique value of HAP for each location and year based on development. For GBD 2020, we updated the HAP mapping model to predict unique values from the lag-distributed income per capita (LDI). Evaluations of model fit using root mean square error (RMSE) indicated that LDI is a more suitable predictor of excess $PM_{2.5}$. We also included a random effect on study and weighted each study by the square root of its sample size.

Before modelling, we subtracted off the GBD 2019 prediction of ambient $PM_{2.5}$ in the study location and year to calculate the excess particulate matter for individuals using solid fuel. The final model coefficients are included below:

Table 5: HAP mapping model and coefficients

Variable	**Beta, log (95% UI)**	**Beta, exponentiated (95% UI)**
Intercept	5.34 (5.16–5.52)	208.51 (174.16–249.64)
Fuel type • Clean (ref) • Crop • Coal	 3.15 (3.06–3.25) 1.66 (1.57–1.73)	 23.34 (21.33–25.79) 5.26 (4.81–5.64)

• Dung • Wood	2.35 (2.22–2.48) 1.99 (1.94–2.04)	10.49 (9.21–11.94) 7.32 (6.96–7.69)
Measure group • Indoor (ref) • Female • Male • Child	 -0.37 (-0.42 to -0.32) -0.27 (-0.36 to -0.18) -1.09 (-1.19 to -1.00)	 0.69 (0.66–0.73) 0.76 (0.70–0.84) 0.34 (0.30–0.37)
24-hour measurement	-0.68 (-0.83 to -0.54)	0.51 (0.44–0.58)
LDI	-2.93e-4 (-4.94e-4 to -8.37e-5)	1.00 (1.00–1.00)

To derive final predicted $PM_{2.5}$ exposure values due to solid fuel usage, instead of using direct model outputs for males and children, we scaled $PM_{2.5}$ exposure values for females to the other two groups. There are few studies of personal monitoring in men and children, so we derived ratios of female-male and female-child exposures using studies that reported PM exposure values for females and one or both of the other groups. To calculate these ratios, we first subtracted off the outdoor value from each PM measurement (using GBD 2019 ambient $PM_{2.5}$ predictions as above for $PM_{2.5}$ studies and the studies' published values for PM_4 and PM_{10} studies) and then calculated ratios weighted by sample size.

Table 6: HAP mapping personal monitoring input observations

Study	Location	Year	Pollutant	Female N	Female PM	Group	N	PM	Outdoor
Balakrishnan et al, 2004	Andhra Pradesh, Rural	2004	PM4	591	352	male	503	187	94
Gao X et al, 2009.	Tibet	2009	PM2.5	52	127	male	85	111	78
Dasgupta et al, 2006	Bangladesh	2006	PM10	944	209	male	944	166	50
Devkumar et al, 2014	Nepal	2014	PM2.5	405	169	male	429	167	167
Balakrishnan et al, 2004	Andhra Pradesh, Rural	2004	PM4	591	352	child	56	262	94
Dionisio et al, 2008.	Republic of the Gambia	2008	PM2.5	13	275	child	13	219	147
Dasgupta et al, 2006	Bangladesh	2006	PM10	944	209	child	944	199	50
Gurley et al, 2013	Bangladesh	2013	PM2.5			child	37	308	
Shupler et al, 2020	Sub-Saharan Africa	2018	PM2.5	37	153	male	20	120	26.05
Shupler et al, 2020	India	2018	PM2.5	11	150	male	5	178	42.3
Shupler et al, 2020	India	2018	PM2.5	63	89	male	48	82	42.3
Shupler et al, 2020	South Asia	2018	PM2.5	5	148	male	3	147	64
Shupler et al, 2020	South Asia	2018	PM2.5	27	148	male	17	90	64
Shupler et al, 2020	South Asia	2018	PM2.5	5	147	male	2	73	64
Shupler et al, 2020	South Asia	2018	PM2.5	15	183	male	6	135	64
Shupler et al, 2020	Latin America and Caribbean	2018	PM2.5	24	39	male	12	40	27.2

Shupler et al, 2020	China	2018	PM2.5	36	71	male	35	61	58.9
Shupler et al, 2020	China	2018	PM2.5	23	94	male	21	93	58.9
Shupler et al, 2020	China	2018	PM2.5	55	45	male	47	44	58.9
Shupler et al, 2020	China	2018	PM2.5	4	64	male	3	37	58.9

The final ratios, updated with information from the 2020 PURE-AIR study, were 0.85 (95% UI 0.67–1.09) for children and 0.64 (0.52–0.79) for males compared to 0.85 (0.56–1.31) for children and 0.64 (0.45–0.91) for males in GBD 2019. These results were used to scale the $PM_{2.5}$ mapping model fuel-type-specific predictions for these age and sex groups to calculate relative risks from the $PM_{2.5}$ risk curves.

HAP population-attributable fractions (PAFs) are calculated jointly with those for ambient particulate matter pollution. Details of PAF calculation, relative risks, and evidence scores for all outcomes besides cataract are provided in the Ambient Particulate Matter Pollution appendix.

References

1. World Health Organization (WHO). WHO Household Energy Database 1960-2017. Geneva, Switzerland: World Health Organization (WHO), 2019.
2. Smith KR, Bruce N, Balakrishnan K, Adair-Rohani H, Balmes J, Chafe Z, *et al*. Millions Dead: How Do We Know and What Does It Mean? Methods Used in the Comparative Risk Assessment of Household Air Pollution. Annu Rev Public Health. 2014; **35**(1):185–206.
3. Tanchangya J, Geater AF. Use of traditional cooking fuels and the risk of young adult cataract in rural Bangladesh: a hospital-based case-control study. BMC Ophthalmology. 2011; **11**.
4. Shupler M, Balakrishnan K, Ghosh S, *et al*. Global household air pollution database: Kitchen concentrations and personal exposures of particulate matter and carbon monoxide. Data in Brief 2018; **21**: 1292–5.
5. Shupler M, Hystad P, Birch A, Miller-Lionberg D, Jeronimo M, Arku R, *et al*. Household and personal air pollution exposure measurements from 120 communities in eight countries: results from the PURE-AIR study. Lancet Planet Health. 2020; **4**(10): E451-E462.
6. Shupler M, Godwin W, Frostad J, Gustafson P, Arku RE, Brauer M. Global estimation of exposure to fine particulate matter ($PM_{2.5}$) from household air pollution. Environment International 2018; **120**: 354–63.