The Global Household Air Pollution Database 2018 (Version 5.0)

Kalpana Balakrishnan[1], Guruswamy Thangavel[1], Santu Ghosh[1], Matthew Shupler[2], Sankar Sambandam[1], Krishnendu Mukhopadhyay[1], Heather-Adair Rohani[3], Nigel Bruce[3], Kirk R. Smith[4]

[1]Department of Environmental Health Engineering, Sri Ramachandra University, Chennai, India ; [2]School of Population and Public Health, University of British Columba, Canada; [3]World Health Organization, Geneva; [4]School of Public Health, University of California, Berkeley, USA

Over two hundred studies over the last three decades have assessed household air pollution (HAP) exposures related to the use of solid fuels in developing countries. However, merging results from individual studies has been challenging because of differences in measurement protocols, types of summary measures reported, the types/ nature of household level determinants explored and quality control criteria used for sampling/analysis. To overcome this challenge a global indoor air pollution (IAP) database, was first prepared in 2003, that documented the results of these measurements from 70 studies from developing countries of Asia, Latin America and Africa [1] with a subsequent version that included 110 studies primarily from China[2]. The database, developed using Microsoft Access® 2000, allows researchers to extract and analyze findings within and across studies, and export the files to a statistical program for more in-depth analysis. Articles are abstracted in consistent and simple ways studies so that data across studies could be used to generate profiles that represent larger geographical regions within and outside a country.

Over the last several years, however there has been a continuous evolution of methods and protocols for assessing HAP. The quantum and quality of information collected has also become considerably more detailed. Results from large-scale studies (conducted over hundreds of households under multiple exposure configurations) that examine temporal, spatial, or multi-pollutant patterns, in addition to day-to-day or seasonal variability in concentrations and exposures have recently become available from many countries. In view of this progress and to inform the WHO HAP –IAQG (Household air pollution- Indoor Air quality Guidelines) process on the exposure evidence from HAP studies, the earlier version of global IAP database (cited above) was updated, to cover studies reported in peer-reviewed literature until 2011. The global IAP database has also been renamed as the Global HAP database to enable inclusion of a broader range of exposure assessment studies within this setting.

This version of the Global Household Air Pollution Database provides a concise overview of results from household air pollution measurements that have been extracted from 196 studies published between 1968 and 2016. This version of the database is based on data fields available in the earlier Global HAP database and provides expanded fields of information that reflect additional details available in more recent studies. As before, this working database includes publications that have been identified through systematic searches that report measurement results from rural and urban solid fuel using communities of low and middle-income countries. It is structured so that the independent researcher can extract and analyze findings within and across studies.

The revised database has been prepared by a team of investigators at the Department of Environmental Health Engineering at Sri Ramachandra University (SRU) in Chennai led Dr. Kalpana Balakrishnan, in collaboration with Matt Shupler at the School of Population and Public Health, University of British Columbia and The Global Alliance for Clean Cookstoves.

Search Methods

The following search terms were used with the PubMed (National Library of Medicine and National Institute of Health) and Science Direct (Elsevier) bibliographic search engines to identify publications for inclusion within the revised global HAP database.

MeSH Terms: [Air Pollution; Indoor]

Key words:

- Rural / Developing Countries/Less Developed countries/ Household/ Domestic/Village (Note: Developed country studies are currently not included in the global HAP Database.
- Solid fuels/ Biomass Combustion/ Bio Fuels/Household combustion/ Household energy
- IAP/Indoor Pollution/ Indoor Air Quality
- Kitchen/Cooking/Stove/Open fire/Improved Cook stove/Wood stove/Bio-mass stove
- Biomass / Wood / Dung / Coal / Agriculture residues/Crop residues/Kerosene/Clean Fuels/LPG
- Concentration / Exposure/Exposure Assessment/Personal
- Exposure/Environmental Exposure/ Particulate Matter/ PM/ Particles/ CO/Carbon monoxide/Volatile Organic Compounds/Sulphur Dioxide/Nitrogen Dioxide/Aromatic hydrocarbons

The search yielded over 8600 articles that were manually screened to identify suitability and currently, information from nearly 200 articles is available in the database.

Organisation of parameters available in the Database

The information is organized as an Excel Spread sheet with section headings. Section 1 contains general descriptives from studies to allow an assessment of type, number and range of measurement results available from individual studies. Section 2 provides detailed information on measurement results from studies that can be sorted based on individual categorical descriptives for each pollutant. With an increasing base of exposure –response studies now being published, a separate Section 3 describes health endpoints and exposure-response coefficients assessed in individual studies. Finally, Section 4 provides instrument and/or protocol details and quality control information pertaining to measurements.

Access Table 1 (Columns A through AO in Excel Spread sheet)

Table 1 has the following fields of information pertaining to general study descriptives

1. Serial Number (Also the study ID)
2. WHO Region (Coded as 1 to 5 to denote Americas, Africa, Eastern Mediterranean, South East Asian and Western Pacific regions respectively)
3. Author ,Title and Year of Publication
4. Study Label-Tag to provide a tag for information contained in the respective rows. The first row for each study contains only general descriptives. Subsequent rows are assigned for results from individual categories of measurements for each pollutant.
5. Location information including Country;Province or State;District/Region;Village/ Town/ Neighbourhood as well as a brief description of ambient air quality (if available)
6. Setting (i.e rural, urban , semi-rural or semi –urban)
7. Study Design.   While most earlier studies have been cross-sectional, many recent studies have used repeated measure, paired before-after and RCT designs. The design is specified either as reported by the study investigators or as assessed by the investigators during manual screening.
8. Specific study details reported
    a. Altitude (m) of study location(s)
    b. Season(s)
    c. Ambient Temperature range

d. Relative Humidity

e. Type(s) of Primary Cooking Fuel(s) and Secondary Cooking Fuel(s)

f. Type(s) of Heating Fuel(s) . Very few studies report results from settings using soild fuels for heating and fewer yet specify the type of heating fuel used.

g. Type(s) of Stove (s) specified as Open Fire/Traditional/Improved Cook-stove (ICS)/Kerosene/LPG/Electric stoves. Note that the terminology of 'ICS' has been retained to match stove classification reported in the earlier version of the database. Wherever available a model name has also been included. This terminology is likely to change as studies report results from stove technologies labelled using other criteria such as advanced combustion cookstoves, forced draft(forced convection) , natural draft (free convection) etc.

h. House Type including Construction Materials

i. Kitchen Location(s) specified as indoors , outdoors or both

j. Kitchen Type(s). The specification has been listed as reported by the study and was found to be variably specified across studies.

k. Ventilation Type(s) specified variably as reported by each study

l. Ventilation relevant measurements including Air Exchange Rate (m^3/h)   Room Volume (s) (m^3); Kitchen volume (m^3)

m. Average Family Size

n. Socioeconomic (SES) measures

o. Description of type of foods cooked in the study area

p. Stove use (hrs); Number of meals cooked per day

q. Fuel Quantity(kg/day); (kg/meal)

r. Number of households (or participants) covered by measurements

s. Number of ambient sites for ambient air pollution measurements within communities

t. Any additional inclusion/exclusion criteria identified by study investigators

u. Gender /Age groups of study participants

v. Pollutants Measured

w. Type of Measurements: Area (Measurements within or near households) , Personal Exposures or Ambient.

Access Table 2: Detailed Results from Measurements (Columns AP through EP in Excel Spreadsheet)

PM Measurement Results (Columns AP through BV in Excel Spreadsheet)

These fields present results of PM measurements across multiple categories of measurements, as reported in individual studies. Each categorical descriptive for a measurement (contained the column labelled PM Measurement Categorical Descriptives) includes labels used by the respective studies for setting (such as rural vs. urban), location within the study area; season; fuel type, kitchen type, stove type and other custom features to describe the conditions during measurement. Since the measurement categories vary widely across studies as compared to the type of measurement (such as area/ personal exposures or ambient) and the PM size fraction, these are contained in separate columns. It would thus be possible to sort first on the basis of size fraction followed by type of measurement and then across individual categories. Commonly used measures of descriptive statistics are contained in a series of columns to allow pooling across studies using specific measure(s). To accommodate an increasing base of studies reporting results from paired before-after assessments following the introduction of an intervention (usually an "improved cook-stove), a set of columns have been added to capture information on percent reductions accompanying individual interventions.

 Accordingly, fields reporting results of PM measurements include
1. PM Measurement Categorical Descriptives (with labels used by the respective studies for setting (such as rural vs. urban), location within the study area; season; fuel type, kitchen type, stove type and other custom features to describe the conditions during measurement as described above)
2. 'Has a subset measurement' indicates if the PM value provided in this row is further stratified by another categorical variable (e.g. fuel type, season) in the article. This is to help avoid repeating the same measurements when extracting the data
3. 'Is a subset of measurement' indicates if the PM value is a stratified value of another summary measure in the database. Also includes the categorical stratifying variable (e.g. fuel type, season)
4. PM size fraction/PM size distribution
5. PM Type of Measurement. Specified as Area (Kitchen); Area (Living/Bedroom); Area Outdoors); Ambient; Personal Exposures (Women/ Men/Children of specific age groups.

6. PM Measurement Method. Specified as Gravimetric or other direct read out Instrument Methods

7. PM Location of Sampler. Specified in terms of height and distance within a room often in relation to the location of the stove

8. PM Number of Measurements /PM Number of Repeat Measurements (if repeated measures study)

9. PM Sampler Percent of equipment failures to assess performance of technology

10. PM Averaging Period

11. PM Unit of Measurement

12. Descriptive Statistics including PM Arithmetic Mean; PM Std Deviation; PM 95% CI for Arithmetic Mean; PM Median; PM Geometric Mean;PM 95% CI for Geometric Mean; PM GSD; PM Range; PM Coefficient of Variation (%);PM 25th -75th percentile; PM IQR

13. Intervention type (for PM measurements)

14. Time elapsed between pre-intervention and post-intervention measurements

15. Parameters relevant for describing Intervention Impact including Post Intervention Percent change in PM Arithmetic; Mean 95% CI for Post Intervention Percent change in PM Arithmetic Mean; Post Intervention Percent change in PM Geometric Mean; 95% CI for Post Intervention Percent change in PM Geometric Mean; Post Intervention Percent change in PM Median; 95%CI for Post Intervention Percent change in PM Median

CO Measurement Results (Columns BW through CX in Excel Spreadsheet)

These fields present results of CO measurements across multiple categories of measurements, as reported in individual studies. Each categorical descriptive for a measurement (contained the column labelled CO Measurement Categorical Descriptives) includes labels used by the respective studies for setting (such as rural vs. urban), location within the study area; season; fuel type, kitchen type, stove type and other custom features to describe the conditions during measurement. Since the measurement categories vary widely across studies as compared to the type of measurement (such as area/ personal exposures or ambient), this is contained in a separate column. It would thus be possible to sort first on the basis of measurement and then across individual categories. Commonly used measures of descriptive statistics are contained in a series of columns to allow pooling across studies using specific measure(s). To accommodate an increasing base of studies reporting results from paired before-after assessments following the introduction of

an intervention (usually an "improved cook-stove), a set of columns have been added to capture information on percent reductions accompanying individual interventions.

Accordingly, fields columns reporting results of CO measurements include

1. CO Measurement Categorical Descriptives (with labels used by the respective studies for setting (such as rural vs. urban), location within the study area; season; fuel type, kitchen type, stove type and other custom features to describe the conditions during measurement (Rural/Urban; Location; Season; Fuel Type; Kitchen Type; Stove Type; Others)
2. CO Type of Measurement. Specified as Area (Kitchen); Area (Living/Bedroom); Area Outdoors); Ambient; Personal Exposures (Women/ Men/Children of specific age groups
3. CO Measurement Method
4. CO Location of Sampler
5. CO Number of Measurements/CO Number of Repeat Measurements (if repeated measures study)
6. CO Sampler Percent of equipment failures to assess performance of technology

7. CO Averaging Period
8. CO Unit Of Measurement
9. Descriptive Statistics including CO Arithmetic Mean; CO Std Deviation;CO 95% CI for Arithmetic Mean; CO Median;CO Geometric Mean;CO GSD;CO 95 % CI for Geometric Mean CO Coefficient of Variation(%);CO Range;CO-25th -75th percentile;CO IQR.
10. Intervention type (for CO measurements)
11. Time elapsed between pre-intervention and post-intervention measurements
12. Parameters relevant for describing Intervention Impact including CO Post Intervention Percent change in Arithmetic Mean;CO 95% CI Post Intervention Percent change in Arithmetic Mean;CO Post Intervention Percent change in Median;CO 95%CI Post Intervention Percent change in Median

Ratios Related to PM –CO Measurements (Columns CY to DE in Excel Spreadsheet)
Since many studies now report relationships between various PM/CO measures, these fields capture information related to commonly reported ratios in such studies.
Accordingly, fields reporting results from ratio calculations in PM/CO measurements contain

1. Ratios of Area Concentrations (PM) such as ratio of kitchen to living; kitchen to outdoors and indoors to ambient

2. Ratios of Area concentrations to Personal Exposures (PM)

3. Ratios of Area Concentrations (CO)

4. Ratios of Area concentrations to Exposures (CO) such as ratio of kitchen to living; kitchen to outdoors and indoors to ambient

5. PM/CO Ratios

6. PM/CO Correlation Coefficients

PM /CO Models (Columns CF to DN in Excel Spreadsheet)

Several recent studies have used household characteristics and area/personal measurements in multivariate regression models to identify significant predictors of household concentrations and/or personal exposures. An understanding of significant model variables across such studies is useful in identifying important household level determinants of such exposures. Accordingly fields reporting model results for PM/CO contain

1. Model variables (PM)

2. Significant Model Determinants (PM)

3. Model R^2 for PM

4. Model Variables (CO)

5. Significant Model Determinants(CO)

6. Model R^2 for CO

Air Toxics/Biomarker Measurement Results (Columns DO to EP in Excel Spreadsheet)

Since relatively few studies report results from air toxics or biomarker measurements, this is information presented as single category (instead of separate fields for a small number of measurements for a large number of individual pollutants). The organization of these fields is similar to what has been described for PM and CO. Accordingly columns reporting results from air toxics/biomarker measurements contain

1. Air Toxics /Biomarker Measurement Categorical Descriptives (Rural/Urban; Location; Season; Fuel Type; Kitchen Type;Stove Type;Others)

2. Air Toxics /Biomarker Type of Measurement

3. Air Toxics /Biomarker Measurement Method

4. Air Toxics /Biomarker Location of Sampler

5. Air Toxics /Biomarker Number of Measurements

6. Air Toxics /Biomarker Averaging Period

7. Air Toxics /Biomarker Units

8. Descriptive Statistics including Air Toxics /Biomarker Arithmetic Mean; Air Toxics /Biomarker Std Deviation; Air Toxics /Biomarker 95% CI for Arithmetic Mean; Air Toxics /Biomarker Median;Air Toxics /Biomarker Geometric Mean; Air Toxics /Biomarker GSD; Air Toxics /Biomarker 95 % CI for Geometric Mean; Air Toxics /Biomarker Coefficient of Variation(%); Air Toxics /Biomarker Range ;Air Toxics /Biomarker -25th -75th percentile; Air Toxics /Biomarker IQR

9. Intervention type (for Air Toxics /Biomarker measurements)

10. Parameters relevant for describing Intervention Impact including Post Intervention Percent change in Air Toxics /Biomarker Arithmetic Mean; 95% CI Post Intervention Percent change in PM Arithmetic Mean; Post Intervention Percent change in Air Toxics /Biomarker   Geometric Mean; 95% CI Post Intervention Percent change in Air Toxics /Biomarker   Geometric Mean; Post Intervention Percent change in Air Toxics /Biomarker Median; 95%CI Post Intervention Percent change in Air Toxics /Biomarker   Median

11. Organic Carbon /Black Carbon Ratio

Access Table 3: Results on Health Assessments included with Exposure Measurements (Columns EQ to EV in Excel Spreadsheet)

Since the primary purpose of the database is to provide exposure information, details of health endpoints are limited to their relationships with exposure measurements,

Accordingly fields reporting results of health assessments include

1. Health endpoints assessed

2. Epidemiological outcome measures of association used (e.g. odds ratio, population attributable risk)

3. Strength of epidemiological association

4. Descriptives associated/not associated with health outcome

5. Description of health endpoints assessed in relation to exposure measurements performed within the same study

Access Table 4: Measurement protocols and quality control information (Columns EW to EX in Excel Spreadsheet)

Measurement protocol details together with QA/QC details are being abstracted and currently not available in the database. These fields will be available in the next update of the database.

Access Table 5: Miscellaneous / Rarely Reported Information (Columns EY to FJ in Excel Spreadsheet)

1. Criteria for inclusion of predictor in multivariable health outcome model
2. Descriptives associated/not associated with using solid fuel vs 'improved cookstove'
3. Fuel Moisture Content (%)
4. Traditional Fuel Type Assessment (e.g. questionnaire, observation)
5. Cooking length type of assessment (e.g. Stove Use Monitors, fuel weighing)
6. Improved cookstove cost
7. Improved cookstove characteristics liked/disliked by participants
8. Change in daily cooking time (fuel) due to use of improved cookstove
9. Light scattering to Gravimetric Correction Factor
10. Light scattering/Gravimetric Correlation

References

1. Saksena , S., L. Thompson, and K.R. Smith, *The Indoor Air Pollution and Exposure Database: Household Pollution Levels in Developing Countries*. 2003, University of California, Berkeley; The World Health Organisation.
2. Sinton, J.E., et al., *Indoor air pollution database for China*. 2005, World Health Organization. Office of Global and Integrated Environmental Health.
3. Balakrishnan K., et al.. The Global Household Air Pollution Database 2011 (Version 3.0) http://www.who.int/indoorair/health_impacts/databases_iap/en/Accessed on 29/02/2016