Science of the Total Environment 676 (2019) 811–822

Contents lists available at ScienceDirect

Science of the Total Environment

journal homepage: www.elsevier.com/locate/scitotenv

Machine-learned modeling of $PM_{2.5}$ exposures in rural Lao PDR

L.D. Hill [a,*], A. Pillarisetti [a], S. Delapena [b], C. Garland [b], D. Pennise [b], A. Pelletreau [c], P. Koetting [c], T. Motmans [c,1], K. Vongnakhone [c], C. Khammavong [c], M.R. Boatman [d], J. Balmes [a,e], A. Hubbard [f], K.R. Smith [a]

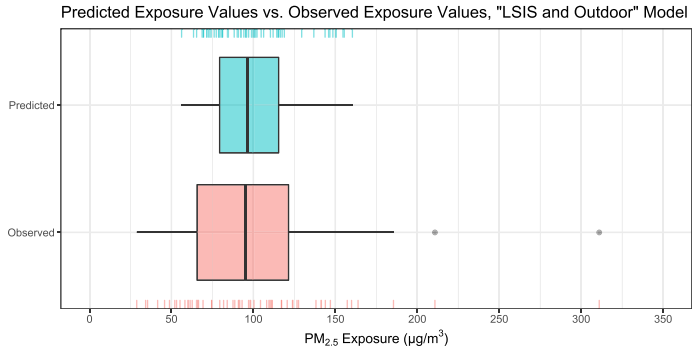
[a] Division of Environmental Health Sciences, School of Public Health, University of California, Berkeley, 2121 Berkeley Way #5302, Berkeley, CA 94720, USA
[b] Berkeley Air Monitoring Group, Inc., 1900 Addison St #350, Berkeley, CA 94704, USA
[c] Lao Institute for Renewable Energy, Ban Watnak Lao-Thai Friendship Road, Sisattanak District, Vientiane, Lao People's Democratic Republic
[d] Geo-Sys (Lao) Co., Ltd, 136/9, Hom 7, Sokphaluang Village, Sisattanak District, Vientiane, Lao People's Democratic Republic
[e] Department of Medicine, University of California, San Francisco, 505 Parnassus Ave, San Francisco, CA 94143, USA
[f] Division of Epidemiology and Biostatistics, School of Public Health, University of California, Berkeley, 2121 Berkeley Way #5302, Berkeley, CA 94720, USA

HIGHLIGHTS

- $PM_{2.5}$ exposures of rural Lao women cooking with wood are within the global range.
- Ensemble modeling and machine learning can improve personal exposure estimation.
- Kitchen exposure factor poorly predicted exposures for traditional stove users.

GRAPHICAL ABSTRACT

Graphical abstract showing predicted exposure values alongside observed exposure values for a specific model examined in the analysis.



ARTICLE INFO

Article history:
Received 17 January 2019
Received in revised form 30 March 2019
Accepted 17 April 2019
Available online 18 April 2019

Editor: Lidia Morawska

Keywords:
Air quality
Cookstoves
Prediction
Environmental health
Biostatistics'

ABSTRACT

This study presents a machine-learning-enhanced method of modeling $PM_{2.5}$ personal exposures in a data-scarce, rural, solid fuel use context. Data collected during a cookstove (Africa Clean Energy (ACE)-1 solar-battery-powered stove) intervention program in rural Lao PDR are presented and leveraged to explore advanced techniques for predicting personal exposures to particulate matter with aerodynamic diameter smaller than 2.5 μm ($PM_{2.5}$). Mean 48-h $PM_{2.5}$ exposure concentrations for female cooks were measured for the pre- and post-intervention periods (the "Before" and "After" periods, respectively) as 123 μg/m^3 and 81 μg/m^3. Mean 48-h $PM_{2.5}$ kitchen air pollution ("KAP") concentrations were measured at 462 μg/m^3 Before and 124 μg/m^3 After. Application of machine learning and ensemble modeling demonstrated cross-validated personal exposure predictions that were modest at the individual level but reasonably strong at the group level, with the best models producing an observed vs. predicted r^2 between 0.26 and 0.31 ($r^2 = 0.49$ when using a smaller, un-imputed dataset) and mean Before estimates of 119–120 μg/m^3 and After estimates of 86–88 μg/m^3. This offered improvement over one typical method of predicting exposure – using a kitchen exposure factor (the ratio of exposure to KAP)— which demonstrated an $r^2 \sim 0.03$ and poorly estimated group average values. The results of these analyses highlight areas of methodological improvement for future exposure assessments of household air pollution and

* Corresponding author at: Ramboll USA, 201 California Street, Suite 1200, San Francisco, CA 94111, USA.
E-mail address: drew.hill@berkeley.edu (L.D. Hill).
[1] Author wishes to note they are currently at BASE – Basel Agency for Sustainable Energy, Elisabethenstrasse 22, 4051 Basel, Switzerland.

https://doi.org/10.1016/j.scitotenv.2019.04.258

provide evidence for researchers to explore the advantages of further incorporating machine learning methods into similar research across wider geographic and cultural contexts.

1. Introduction

Approximately 2.8 billion people meet most of their cooking needs with solid fuels such as wood or coal (Bonjour et al., 2013), creating much human exposure to household air pollution ("HAP") from particulate matter (PM) with aerodynamic diameter of 2.5 μm or smaller ($PM_{2.5}$). HAP exposures can vary due to heterogeneity of pollution within the home, including vertical stratification (Johnson et al., 2011; Kandpal et al., 1995) and changes in dilution relative to distance from the stove or windows. Additional factors influencing exposure include the amount of time spent outside of the household, $PM_{2.5}$ sources unaccounted for in measurements, pollution at a neighbor's home during a visit, or high outdoor ambient pollution concentrations due to traffic, trash burning, or other sources.

Personal exposures that directly affect health are, thus, more than a function of fuel type alone, and even the most sophisticated indoor concentration models are often incapable of assessing the human-environment interactions (Steinle et al., 2013) that affect actual exposures. Errors in exposure estimation caused by relying on proxies like fuel type may lead to significant bias in the estimation of related health burdens, especially at points along the exposure-response relationship where risk is thought to be highly non-linear, i.e. be changing rapidly with exposure (Burnett et al., 2014; Cohen et al., 2017; Smith et al., 2014; World Bank and Institute for Health Metrics and Evaluation, 2016). The general paucity of HAP exposure data stems from the difficulty, invasiveness, and resource-intensive nature of current exposure measurement techniques (Balakrishnan et al., 2011, 2014b; Clark et al., 2013). Gold standard methods for measuring exposure have required outfitting individuals with delicate, expensive pumps that are sometimes also bulky, heavy, and noisy and require sophisticated laboratory backup. Alternate approaches that rely on somewhat easier to collect data suggest that variability in both indoor concentrations and personal exposures to HAP may be explained in part by household and behavioral indicators like fuel type, stove type, kitchen structure, and cooking duration (Balakrishnan et al., 2002, 2013).

We investigate the utility of kitchen air pollution and other environmental measurements and easier-to-assess household and survey-based behavioral indicators to predict 48-h average $PM_{2.5}$ exposures in rural Lao women, utilizing machine learning, and specifically "super learning" – the production of a single "super" learner by combining a set of candidate learners like random forests or neural networks, as weighted by their predictive utility (van der Laan et al., 2007). We focus specifically on survey questions similar to or drawn directly from the Demographic and Health Survey (DHS) questionnaires. DHS is an internationally administered, structured, and standardized survey designed to collect accurate and representative data on demographics and health (ICF International, 2011) that is deployed in over 90 countries. Better use of existing DHS indicators to predict HAP exposures may prove useful for future disease burden research, quantifications of the amount of illness avoided through interventions (e.g. Averted Disability Adjusted Life Years, or aDALYs) (Anenberg et al., 2017; Smith et al., 2015), and, in general, cookstove monitoring and evaluation programs. Finally, we provide additional data from a Southeast Asian stove intervention program, adding to a growing global pool of $PM_{2.5}$ exposure and kitchen-level concentration data and providing measured concentrations that may augment existing global databases (Shupler et al., 2018a).

2. Methods

Data collection occurred during a separate biomass stove intervention program in Xonboury District, Savannakhet Province, Lao People's Democratic Republic (Lao PDR), where approximately 95% of households cook with solid fuels on traditional appliances (Health Effects Institute, 2018). The program has been described elsewhere (Hill et al., 2015). Briefly, it sought to evaluate the displacement of traditional stoves with the African Clean Energy (ACE)-1 battery-powered[2] blower stove in three villages, hereafter identified as villages A, B, and C, in no particular order. During the program, a total of 72 households (24 from each village) were enrolled. Selection criteria included the use of wood as the primary cooking fuel and that the main cook be 18 years of age or older and not pregnant. All primary cooks were identified as female. Households were encouraged to use their new stove during the post-intervention study period, but no penalty was imposed if they did not comply.

The study population described in this paper is a subset derived from the above population receiving the ACE-1 intervention. Participants from half of the households in each village ($n = 36$) were enrolled in a personal exposure assessment on a convenience basis. Meteorology and ambient $PM_{2.5}$ concentrations, measured in the separate study (Hill et al., 2015) were utilized in the analysis of data in this study.

The sampling scheme took the form of a before-and-after intervention study. Each household participated in a baseline survey, a round of air pollution monitoring and follow-up surveys approximately 2–4 weeks prior to receiving the ACE-1 stove ("Before"). Subsequently, each household participated in another round of air pollution monitoring and follow-up surveys weeks after receiving the ACE-1 stove ("After"). The time between stove dissemination and measurement was intended to allow cooks to gain familiarity with the new stove. Dissemination occurred over the course of a week in January 2015. Distributors made follow-up visits to villages and households to address problems and queries.

Each round of air pollution sampling occurred in segments of approximately 3-weeks, with about one week in each village. The first round spanned December 2014 to January 2015 and the second round occurred between January and February 2015. In this paper, we describe the following measurements: 2 days of simultaneous gravimetric measurement of KAP and personal $PM_{2.5}$ exposure; a post-monitoring survey to provide data specific to each 2-day set of measurements; and ambient air pollution and meteorology measurements at a central site in each village. This detailed sampling scheme is summarized in Fig. 1.

Protocols used to collect the data analyzed in the study were approved by the University of California, Berkeley Committee for Protection of Human Subjects, protocol number 2014-06-6457. Lao Institute for Renewable Energy (LIRE) formally sought permission for the study through Ministry of Energy and Mines – IREP – who informed the provincial and district authorities. District authorities supported the implementation.

2.1. Air pollution measurements

Gravimetric air pollution measurements of KAP and personal $PM_{2.5}$ exposures were taken before and after stove dissemination. Sampling was performed using Triplex cyclones (SCC 1.062, Mesa Labs, Butler, NJ, USA) with a 2.5 μm size cut at 1.5 l per minute attached to AirChek

[2] The ACE-1 comes with a small solar panel to recharge the internal battery of the stove.

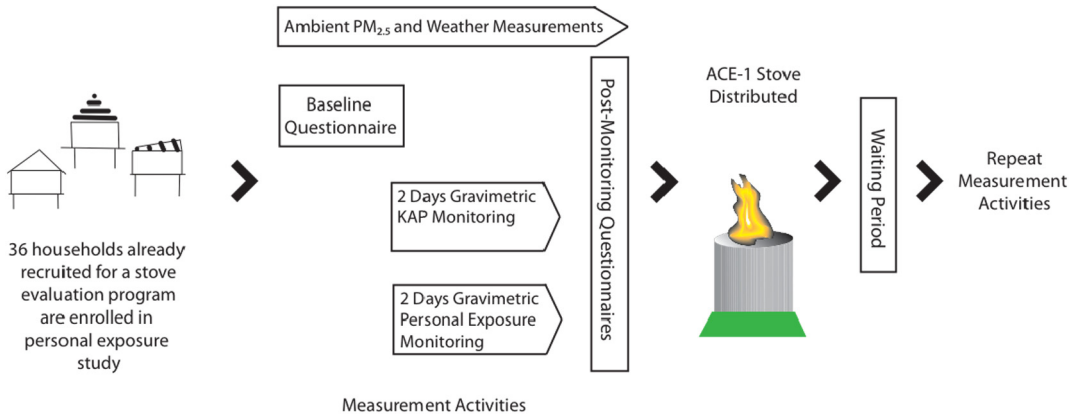


Fig. 1. Key aspects of the sampling scheme.

XR5000 pumps (SKC, Inc., Eighty Four, PA, USA). Cyclones were fit with 37 mm Polytetrafluoroethylene (PTFE) filters with support ring and 2 μm pores (Pall Corporation, Port Washington, NY, USA). Gravimetric KAP and personal $PM_{2.5}$ exposure samples were collected for two consecutive approximately 24-h periods (about 48 h total).

KAP monitors were placed on the kitchen wall about 1.5 m from the floor and about 1 m from the edge of the main cooking stove. Where possible, monitors were placed about 1 m away from major ventilation sources such as windows, eaves, and doors. Field teams installed pushpins during the Before sampling period to facilitate consistency in monitor placement with the After period. Duplicate KAP filter measurements were taken for quality assurance and quality control ($n_{original} = 16$).[3]

Measurements of personal exposure and gravimetric KAP concentrations were taken concurrently. A vest was designed to hold personal exposure monitoring equipment with the gravimetric sampling inlet approximately in the breathing zone while limiting discomfort (Fig. 2). Participants were requested to place the garment next to their sleeping area (e.g. bed) at night and in a nearby location while bathing.

Sample run times were determined primarily by start and stop times logged by fieldworkers. Discrepancies were assessed and resolved manually using context, like the internal timers on the pumps. Flow rates were measured at the start and end of each gravimetric sample using standard methods. Paper based forms were double-entered and discrepancies were resolved manually.

Ambient 24-h $PM_{2.5}$ measurements were collected using a MiniVol $PM_{2.5}$ Sampler (Air Metrics, Springfield, OR, USA). 47 mm PTFE filters with built-in support ring and 2 μm pores (Pall Corporation, Port Washington, NY, USA) were used for sampling at a flow rate of approximately 5 l per minute. The MiniVol was placed in a relatively central location in each village where it would be safe and would not be disturbed. Placement rotated through villages with KAP and exposure sampling.

Measurements of outdoor temperature (°C), relative humidity (%), barometric pressure (mb), wind speed (km/h), wind direction, and precipitation (mm) were collected using a Vantage Vue 6250 wireless weather station with Integrated Sensor Suite 6357 and a 6351 console (Davis Instruments, Hayward, CA, USA) co-located with the MiniVol in each village. Meteorology readings were manually entered into a field form once in mid-morning ("morning") and in the late morning or early afternoon ("afternoon").

2.2. Determination of mass concentrations

Filters from KAP and personal monitoring were weighed before and after sampling on a Mettler Toledo XP2U (Mettler Toledo, Columbus, OH, USA). Field blanks were collected for both 37 mm ($n_{Before} = 15$, $n_{After} = 12$) and 47 mm systems ($n_{Before} = 3$, $n_{After} = 2$). During the study, an equipment malfunction in the weighing facility invalidated pre-sample filter weight measurements for a large fraction of the 37 mm samples (about 20%). Samples were recovered using a validated method (Garland et al., 2018). Briefly, post-sample filters were sonicated in a solvent bath to extract PM mass from sampling. Filters were air dried, conditioned, and reweighed. The post-extraction mass was used as a proxy for the filter's pre-sample weight. Final mass deposition from sampling was calculated as the difference between the post-sample (pre-extraction) filter mass and post-extraction filter mass, and adjusted using the average post-extraction field blank mass from each sampling period. The method has been shown to recover pre-sample filter weights with high reliability ($r^2 > 0.99$). For consistency, all gravimetric samples were treated with this method. The method discussed by Garland et al. (2018) identifies and excludes samples below a limit of detection of 60 μg of mass deposition – we employed data produced by an earlier version of their model that did not exclude values below the LoD (<5% of non-blank filters and about 12% of all filters)[4] but which still demonstrated high reliability ($r^2 > 0.99$). Filters from ambient PM measurement were not affected by this equipment malfunction and thus not treated for extraction.

2.3. Questionnaires

Three separate questionnaires were administered during the personal exposure study: a baseline questionnaire, a post-KAP monitoring questionnaire (not used in this analysis), and a post-personal exposure monitoring questionnaire (Hill et al., 2015). All questionnaires were drafted in English and translated into Lao with input from colleagues with experience in the field, rural Lao context, or both. All were administered in Lao by trained surveyors familiar with local customs using Microsoft Excel-based tools and portable computers. Questionnaires were reviewed for missing and implausible answers and manually adjusted to available audio recordings of survey administration where possible.[5]

The post-monitoring questionnaires were administered upon equipment removal. The post-KAP monitoring questionnaire asked specifically about the approximately 4 days of KAP measurement (which included about 48 h of gravimetric measurement), and the post-personal exposure monitoring questionnaire asked only about the approximately 48 h of simultaneous gravimetric exposure and KAP measurement.

All post-personal exposure monitoring questions were included because they pertain directly to the air pollution measurement periods of interest. Redundant questions were removed. Topics considered from

[3] 2 of the 16 samples intended as duplicates were analyzed as actual area samples to replace invalid area samples.

[4] Mass deposition assessed to be <60 μg after application of the extraction method.

[5] In some instances of review, especially at later dates, recordings were not available.

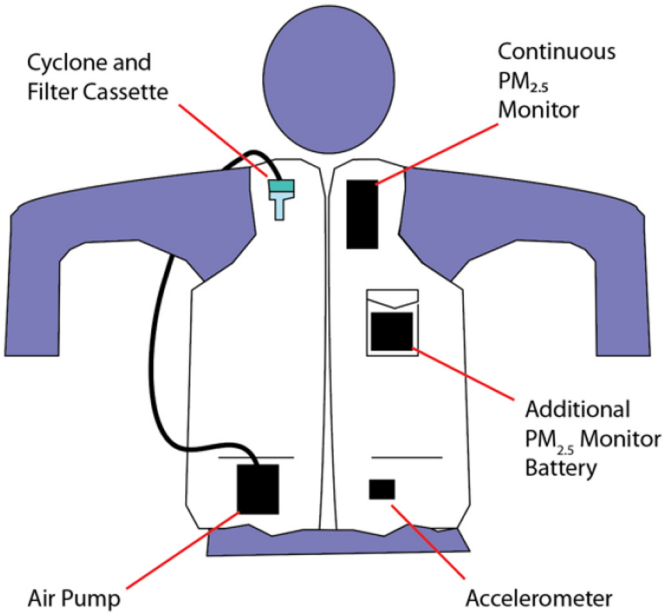


Fig. 2. An example of the sampling garments used.

the baseline questionnaire included household characteristics and demographics, architectural characteristics of the primary house and kitchen, respondent age and smoking status, attitudes toward cooking smoke and its impacts on health, wood fuel preparation, electricity access, and electric appliance use. Baseline questionnaire variables were combined for each household's Before and After post-monitoring datasets.

A primary goal of the analysis was to understand the utility of DHS-type survey questions in the prediction of personal $PM_{2.5}$ exposure concentrations in the wood-fuel cooking context. In Lao PDR, DHS questions have been administered in the form of the Lao Social Indicator Survey (LSIS) (Ministry of Health and Lao Statistics Bureau of the Ministry of Planning and Investment, 2012). LSIS questions for which reasonable overlap with the baseline or post-monitoring questionnaires existed or could be produced during data processing covered topics including household size by age, education level of household members, drinking water source, sanitation, ethnic identity, architectural characteristics of the main house, primary fuel type and cooking location, household asset and financial status, and land and livestock ownership.[6]

Several existing variables were modified and new variables created by combining individual responses or questions. Cooking and grilling exposure activity scores were created from post-monitoring questionnaire responses about cooking and grilling behavior. This is described in more detail in the Supplement.

2.4. Statistical analysis

Analyses were performed primarily using the R statistical programming language (R Core Team, 2016). Data were analyzed using two methods: one using a user-specified library of algorithms (SuperLearner; Polley et al., 2016) and one using so-called "bagging" of regression trees (randomForest; Liaw and Wiener, 2002). SuperLearner implementations used a library containing the following learners: random forest, cForest, extreme gradient boosting, neural networks with a single hidden layer, support vector machines, and 10-fold cross-validated (CV) generalized linear modeling with regularization.[7] Though both are flexible methods for prediction, SuperLearner has the advantage on more parsimonious fits if the underlying true model is relatively smooth. In addition, as long as the library of learners is extensive enough, the SuperLearner provides something akin to an upper bound on how any approach would do given the available data, so a low CV-based r^2 resulting from SuperLearner is unlikely to be from model misspecification and more likely to be because of the limitations of the predictors.

The general goals of the analysis were to understand how well 48-h average exposures and 24-h average exposures could be predicted using all variables collected, only a targeted subset of the variables collected (for several targeted subsets), and the traditional kitchen exposure factor (KEF; the ratio of a person's exposure to their KAP concentration) as well as which categories of information lead to relatively better performance when included in a predictive model. Special emphasis was placed on data similar to those collected routinely around the globe as part of the DHS program.

Most, but not all, households contributed two sets of data to the models (Before and After measurements); values from the same household were kept together (by fold) during cross-validation. Because not all households contributed to both the Before and After datasets (i.e. not all Before and After data were paired), performing paired (dependent) statistical tests in many of the comparisons in this paper would require a sample size reduction which may prove detrimental to the primary predictive analyses of this paper. For this reason, we often perform testing that assumes independence when comparing variables between the Before and After periods. Doing so may increase bias and affect the likelihood of Type I error – possibly increasing the likelihood of Type I error in *t*-tests (e.g. Guo and Yuan, 2017). For this reason, reported statistical significances in variables Before vs. After the intervention should be used as general indications of periodic difference and with caution. Missing personal $PM_{2.5}$ exposure concentration values determined the final sample size for analysis: 33 After samples and 27 Before samples from 34 unique participants (60 total samples; 19 in village A, 22 in village B, and 19 in village C). Variable importance in the prediction of 48-h exposure was assessed using the randomForest package (Liaw and Wiener, 2002) (based upon the incremental decrement in fit by removing each candidate predictor). Model performance was evaluated using the coefficient of determination (r^2) of observed exposure values regressed on predicted exposure values and by comparing the similarity of observed and predicted exposure values as stratified by sampling period (Before vs. After). The latter was chosen to reflect real-world use in the context of monitoring and evaluating an exposure intervention program.

The kitchen exposure factor is sometimes used along with KAP concentration data to quantify personal $PM_{2.5}$ exposure. The predictive value of KEF was explored in this dataset using 10-fold CV: a mean 48-h KEF was calculated on approximately 90% of sample data and multiplied across individual 48-h KAP concentration measurements on the remaining ~ 10% of data to produce estimates of exposure for those participants. This was performed ten times with each iteration leaving out a unique ~ 10% of the data until cross validated exposure predictions were calculated for all participants. This procedure was performed on all KAP and exposure data and again on the dataset stratified by sampling period (5 folds per stratum). Before and After samples from the same household were kept together during CV fold creation.

The ability to predict 48-h average $PM_{2.5}$ exposure was explored using 10-fold CV SuperLearner methods on three primary variable sets. The first training set, the "Full" set, included all original and imputed (described in the Supplement) independent variables with the exception of village, as it is largely non-transportable, and sampling period, as it is a direct aspect of the pre-determined model performance metric. We can use such an approach to examine the ability of predicting based upon future data that is a combination of actual and imputed measures. Stove type, as a near-perfect proxy for sampling period, was also excluded. An iteration of this model was also produced

[6] LSIS questions HL6, HH11, ED3, ED4, WS1, WS8, WS9, HC1C, HC3, HC4, HC5, HC6, HC7, HC8, HC9, HC10, HC12, HC14, and HC15.

[7] Learners as identified by SuperLearner: SL.randomForest, SL.cforest, SL.xgboost, SL.nnet, SL.svm, SL.glmnet. All were used with default settings.

using the un-imputed dataset.[8] The second training set excluded all KAP and meteorology variables from the Full set ("Full Without KAP" set). The third set included only KAP, meteorology, and ambient concentration data from the Full set ("KAP Without Surveys" set).

The Full set was also explored with CV SuperLearner using each of the first- and second-day 24-h average exposure measurements as the outcomes of interest. When predicting individual 24-h average outcomes, gravimetric KAP and meteorology data were subset to reflect only the specific day of measurement.

A group of five training datasets that focused on LSIS-type questionnaire data was also explored with CV SuperLearner to better understand the predictive power of DHS indicators and the adjuvant power of other measurements. These included the following:

- a set of only the LSIS-type questionnaire variables ("LSIS Only")
- a set of LSIS-type questions, ambient $PM_{2.5}$ concentrations, and all meteorology variables ("LSIS and Outdoor")
- a set of LSIS-type variables and self-reported wood fuel use (kg) from the post-monitoring questionnaires ("LSIS and Wood Use")
- a set of LSIS-type variables and self-reported heating variables from post-monitoring questionnaires ("LSIS and Heating")
- a set of LSIS-type variables and the previously described cooking and grilling exposure activity scores ("LSIS and Exposure Scores")

$PM_{2.5}$ concentrations observed during the Before and After periods were examined, though the overall analysis does not hypothesize or explore the causal link between ACE-1 use and exposure reduction.

3. Results

3.1. Air pollution and meteorological measurements

Summary statistics for the complete pollution concentration samples ($n = 60$) are shown in Table 1 and visualized in Fig. 3. The 48-h average personal $PM_{2.5}$ exposure concentration in the Before period was significantly different from that of the After period (reduced from 123 μg/m^3 to 81 μg/m^3, $p < 0.001$) as was the 48-h average KAP concentration (from 462 μg/m3 to 124 μg/m^3, $p < 0.001$). Day 1 average exposures were marginally different from Day 2 averages ($p = 0.07$), with a stronger and significant difference ($p < 0.01$) in the Before period. In the Before period, the Day 2 average value was lower than the Day 1 average value; this was reversed in the After period. Average KAP concentrations were not significantly different between Days 1 and 2. Paired ($n = 52$) and mixed paired-unpaired ($n = 60$) 48-h samples were significantly different between the Before and After periods for both personal exposure and KAP ($p < 0.001$). The same was true for KEF ($p < 0.01$).

Budget constraints and technical difficulties resulted in too few valid samples to produce daily ambient $PM_{2.5}$ concentration estimates. Thus, valid 24-h measurements were aggregated to produce a single estimate of average ambient $PM_{2.5}$ concentration for each village during each sampling period (Table 2). Each sample was assigned an ambient concentration based on the village in which the participant lived and the sampling period. A significant difference in average ambient concentrations between sampling periods was not observed ($p = 0.74$). Temperature and wind speed varied significantly ($p < 0.05$) between Before and After periods; barometric pressure varied marginally significantly ($p < 0.10$) between periods (Table 3).

3.2. Questionnaires

A total of 246 variables from the baseline and post-monitoring questionnaires were available for analysis (not including variables for household ID and village). Missing values were imputed, and an indicator variable was created for each imputation. A more complete description of the questionnaire dataset, imputation techniques, and construction of the final dataset used for subsequent analysis is included in the Supplement. Table 4 summarizes select variables of the survey-based data.

3.3. Statistical analyses and prediction models

Summary statistics for models produced using all 48-h exposure datasets are shown in Table 5. Fig. 4 shows observed exposure values plotted against values predicted using CV SuperLearner for each of the three primary datasets and against values predicted using the KEF method.

Among the three primary datasets, the model produced using the "Full Without KAP" dataset estimated exposures from both Before and After periods with the least reliability. Little difference was observed between the predictive power of the "Full" and "Full Without Surveys" datasets, which both produced models that accurately predicted Before and After concentrations. Both models demonstrated modest strength in predicting individual exposure values, with $r^2 \sim 0.25$ and estimated means within a reasonable range of observed values. The model produced using the Full un-imputed dataset performed similarly well to the model produced on the Full imputed dataset.

KEF proved a poor predictor of exposure, especially in the Before sampling period. Stratification by sampling period improved results slightly, with the best results observed in the After period predictions. 48-h exposures were also modeled from 48-h average KAP concentrations with SuperLearner[9] as a complementary analysis of what a KEF-type model would look like with the benefits of super learning and machine learning. This improved performance slightly ($r^2 = 0.09$), producing mean Before and After exposure estimates of 108.0 μg/m^3 ($p = 0.18$) and 86.4 μg/m^3 ($p = 0.30$), respectively.

The model produced on LSIS-only data did not successfully predict mean concentrations for either study period. Addition of ambient $PM_{2.5}$ and meteorology data dramatically improved the LSIS dataset's predictive power, resulting in the most predictive LSIS-based model. The addition of heating variables, wood use variables, and exposure score variables also improved performance slightly.

Variable importance analyses uncovered useful data groupings (Table 6). Relative importance is demonstrated in terms of the percent increase in mean squared error (MSE) caused by removing each variable during Random Forest selection. Meteorology and KAP data made the greatest contributions to the Full data model. The sets of 10 most important variables were very similar between the Full and Full Without Surveys models. The model produced on the Full set without meteorology or KAP data was less robust, and produced relatively poor estimates of sample period mean exposures. Generally, LSIS-type variables related to household size, ethnicity, and cooking location made the greatest contributions to LSIS model performance. Among all models, a select few variables produced a considerably larger impact than the others. For example, relative importance in the Full dataset drops considerably after Day 1 24-h KAP concentration, the fourth most important variable.

[8] The un-imputed "Full" dataset consisted of only 37 samples ($n_{Before} = 13$, $n_{After} = 24$) and 278 covariates, compared to the 60 samples ($n_{Before} = 27$, $n_{After} = 33$) and 305 covariates included in the post-imputation dataset. Differences in covariates arose from the changes to data processing outputs caused by data missingness. In general, the types of covariates included in the un-imputed and post-imputation datasets were similar. A description of the imputed and final datasets are in the Supplement to this article.

[9] Excluding learners reliant on forests and regularization (i.e. SL.randomForest, SL.cforest, and SL.glmnet), which did not function with only one predictor variable.

Table 1
Kitchen air pollution and personal $PM_{2.5}$ exposure concentrations, by period and sample day.

		Kitchen air pollution concentration				Personal $PM_{2.5}$ exposure concentration				KEF	
	n[c]	Mean (μg/m^3)	+/− 95% CI [b]	GM	GSD	Mean (μg/m^3)	+/− 95% CI [b]	GM	GSD	Mean	SD
Before	*27*										
Day 1		499 [a]	182	376	2.1	139 [a]	28	124	1.6	0.45 [a]	0.43
Day 2		470 [a]	163	350	2.1	107 [a]	19	98	1.5	0.39 [a]	0.36
48-h Avg.		462 [a]	144	370	2.0	123 [a]	22	113	1.5	0.42 [a]	0.42
After	*33*										
Day 1		131 [a]	30	113	1.7	78	12	71	1.6	0.70	0.32
Day 2		116	15	109	1.5	83	14	69	2.2	0.80 [a]	0.50
48-h Avg.		124 [a]	20	114	1.5	81	11	75	1.5	0.72	0.29

Note: GM = Geometric Mean, GSD = Geometric Standard Deviation, KEF = Kitchen Exposure Factor.
[a] Values comprising this arithmetic mean are distributed log-normally ($p < 0.05$).
[b] For non-transformed data.
[c] Note the difference in Before and After sample sizes. Some data did not have both Before and After samples (i.e. were unpaired).

Observed Day 1 and Day 2 24-h average exposure values plotted against the values predicted using CV SuperLearner on the Full dataset are shown in Fig. 5. Models created with either Day 1 or Day 2 measurements accurately predicted sample period means ($p > 0.15$). The model trained on Day 1 measurements produced a stronger fit ($r^2 = 0.30$) than the model trained on Day 2 measurements ($r^2 = 0.07$)

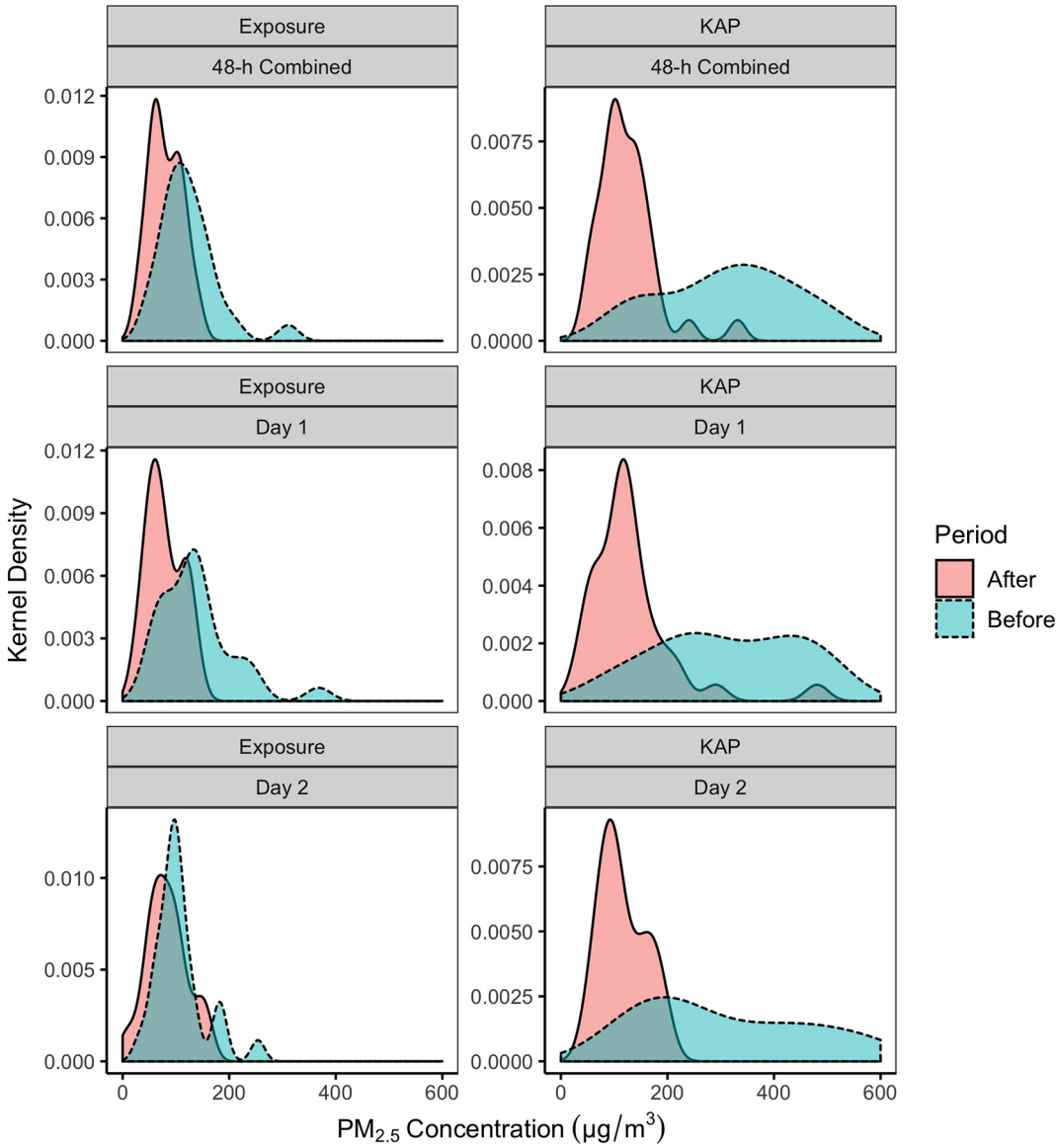


Fig. 3. Kernel density plots of distributions of personal $PM_{2.5}$ exposure (left) and Kitchen $PM_{2.5}$ Air Pollution concentrations (KAP; right), by sampling period (blue fill with dashed-line and red fill with solid-line) and sampling time (from top to bottom). Concentrations are limited to 600 μg/m^3 to allow easier interpretation; 6 points from the 48-h Combined KAP, 6 points from the Day 1 KAP, and 5 points from the Day 2 KAP plots were excluded.

Table 2
Summary statistics for outdoor ambient $PM_{2.5}$ concentrations, by sampling period and village[a].

	n	Mean (μg/m^3)	Max (μg/m^3)	Min (μg/m^3)	SD (μg/m^3)
Before	*7*	52	73	26	16
Village A	*2*	49	50	47	–
Village B	*3*	57	66	43	12
Village C	*2*	49	73	26	–
After	*14*	53	147	15	34
Village A	*4*	57	82	15	30
Village B	*5*	70	147	38	44
Village C	*5*	34	57	20	15

The difference between Before and After concentrations was not statistically significant ($p = 0.74$).
[a] Data collected Before and After the intervention cannot be claimed as independent. Because the data as analyzed included a mix of paired and un-paired data, independent sample testing methods were applied. This may increase bias and the likelihood of Type I error (Guo and Yuan, 2017).

or any of the models produced using 48-h average exposures as the outcome of interest (highest 48-h average $r^2 = 0.27$) and un-imputed data.

4. Discussion

This work provides one of the first exposure estimates among rural Lao females cooking with solid fuel and adds to the small number of studies on HAP in the region (e.g. Huang et al., 2013; Mengersen et al., 2011; Morawska et al., 2011). The personal $PM_{2.5}$ exposure measurements were within the range of those experienced by cooks using solid fuels globally. One study (Shupler et al., 2018b) estimated KAP and female exposure concentrations for rural Lao homes cooking with wood at 380 μg/m^3 and 155 μg/m^3, respectively during the "wet (summer)" season, and at 517 μg/m^3 and 211 μg/m^3 respectively during the "dry (winter)" season. A recent review of the literature (Balakrishnan et al., 2014b) estimated the mean of 24-h average exposures for women cooking with solid fuels around the world at 267 μg/m^3 (SD: 297 μg/m^3). The lower estimates of exposures experienced by the Lao females in our study may be due to the hybrid indoor-outdoor nature of cooking in the region. Dwellings and kitchens in Lao homes were observed to be highly ventilated; rooms often have large eaves, windows, and walls that do not reach the roof (Fig. 6).

4.1. Concentrations observed before and after the ACE-1 intervention

KAP concentrations and personal exposures were higher during the Before (pre-ACE-1 dissemination) than during the After (post-ACE-1 dissemination) periods ($p < 0.001$). The magnitude of this difference as depicted in Table 1 should be treated with caution, as the testing methods used may introduce bias and affect Type I error, as discussed in the Methods section. This is also true of the below-discussed tests for significance between the Before and After periods.

Factors other than the ACE-1 intervention may have impacted the magnitude of observed exposure reductions. Meteorological parameters changed between the Before and After periods (Table 3). Mean temperature was lower during the Before period ($p < 0.05$) while mean wind speed was greater ($p < 0.05$). Such changes may decrease physical comfort and encourage heating with biomass, increasing exposures among participants. In fact, use of heating during sampling was significantly higher ($p < 0.01$) during the Before period than during the After period. The effect may amplify the observed decrease in exposures. Future work should explore study designs that allow for delineation of the exposure effects of seasonality. Meteorological as well as behavioral and cultural factors should be considered – e.g. harvest periods, seasonal biomass burning, changes in ingredients and cooking types, religious holidays, and celebrations with pyrotechnic or major cooking components.

The amount of self-reported fuel used and time spent in the kitchen dropped significantly between the Before and After periods ($p < 0.01$ and $p < 0.05$, respectively). Time spent cooking at home and grilling at home also fell significantly ($p < 0.01$ and $p < 0.05$, respectively). These reductions may have been due to possible improvements in cooking efficiency afforded by the new stove.

It should be noted that follow-up time and time between the intervention and After measurements (approximately 2–4 weeks) was relatively short. A longer or more intensive stove transition period may have improved usage rates and efficiency, while a longer period of follow-up may have incurred increases in hardware failure or declines in usage.

4.2. Predicting $PM_{2.5}$ exposures with traditional and new methods

A primary aim of this paper is to explore the utility of advanced statistical techniques in predicting the difficult-to-measure personal $PM_{2.5}$ exposure concentrations within the rural solid-fuel cooking context. Exposure data were collected only on females with self-reported ages of or above the age of 18 years, and so may not be generalizable to males or children. Models using machine learning and super learning produced reliable estimates of mean exposures in both the Before and After study periods. The best models calculated individual 48-h $PM_{2.5}$ exposures with reasonable strength (r^2: 0.26–0.31 [and 0.49 for a model performed on the smaller un-imputed set]), and predicted period-specific mean exposure values that were within 10% of and not significantly different from measured values.

Direct cross-validated prediction modeling for HAP *exposures* is sparse in the literature. One recent study (Shupler et al., 2018b) used Bayesian hierarchical methods, leave-one-out model validation, and variables like fuel and stove type, season, and location from 44 studies

Table 3
Summary statistics for key meteorology data, by sampling period and village[a].

	Mean barometric pressure (mb)[b]	Mean temperature (°C)[c]	Mean RH (%)	Mean wind speed (km/h)[c]	Most-observed wind direction[d]	Mean afternoon precipitation (mm)
Before	1017.1	24.3	56	6.1	SW	0.3
Village A	1018.7	23.3	53	8.5	SE	0.0
Village B	1017.3	23.3	54	6.5	NW	0.8
Village C	1015.2	26.3	61	3.2	SW	0.0
After	1015.9	25.9	54	3.9	NW	0.2
Village A	1015.4	27.8	53	4.2	SE	0.0
Village B	1015.0	24.5	57	3.2	NW	0.7
Village C	1017.3	25.3	51	4.3	NW	0.0

Note: with the exception of "Mean afternoon precipitation," "mean" values are taken as the average among all morning and afternoon measurements.
[a] Data collected Before and After the intervention cannot be claimed as "independent." Because the data as analyzed included a mix of paired and un-paired data, independent sample testing methods were applied. This may increase bias and the likelihood of Type I error (Guo and Yuan, 2017).
[b] Difference between Before and After concentrations marginally significant using un-paired student's *t*-test ($p < 0.10$).
[c] Difference between Before and After concentrations statistically significant using un-paired student's t-test ($p < 0.05$).
[d] Difference between Before and After values not tested for significance.

Table 4
Summary of selected questionnaire responses.

	All study data		Dataset as analyzed[a]		
	Mean (SD or %)		Mean (SD or %)		
	Before	*After*	*Before*	*After*	p-value[c]
n	*36*	*36*	*27*	*33*	–
Age of Cook (years; reported during Baseline Questionnaire)	35 (11)		34 (11)	35 (11)	0.95
Female Head of Household	4 (11%)		4 (15%)	4 (12%)	1.00
Self Reported Spending (KIP/month)	316,000 (710,000)		201,000 (343,000)	314,000 (745,000)	0.44
Electricity Access	36 (100%)		27 (100%)	33 (100%)	–
Primary Biomass Stove is ACE 1	0	36 (100%) [b]	0	33 (100%)[b]	<0.01*
Primary Biomass Stove is Open Fire	30 (83%)	0 [b]	24 (89%)	0 [b]	<0.01*
Reported Wood Use for Cooking (kg)	10.5 (5.1)	2.4 (1.2)	10.2 (4.5)	2.4 (1.3)	<0.01*
Kitchen Size (m^3)	40 (33)		41 (32)	42 (34)	0.91
Cooking Location at Home					
In House (Separate Room)	1 (3%)	0	1 (4%)	0	0.44
In House (Elsewhere)	30 (83%)	33 (92%)	22 (81%)	30 (91%)	
In Separate Building	5 (14%)	3 (8%)	4 (15%)	3 (9%)	
Time Spent Cooking per Day at Home (minutes)	543 (211)	364 (146)	542 (223)	368 (147)	<0.01*
Total Cooking Exposure Score	6896 (3441)	4356 (3498)	7423 (3643)	4353 (3516)	<0.01*
Time Spent Grilling per Day at Home (minutes)	86 (113)	29 (59)	87 (123)	31 (61)	0.04*
Total Grilling Exposure Score	3744 (6537)	1414 (5375)	4284 (7315)	1542 (5603)	0.12
Smoking Occurred in House	28 (78%)	27 (75%)	23 (85%)	25 (76%)	0.52
Smoking Occurred in Kitchen	5 (14%)	3 (8%)	4 (15%)	3 (9%)	0.69
Ever Used a Heat Source During Sampling	23 (64%)	10 (28%)	20 (74%)	9 (27%)	<0.01*
Time Activity During 48-h Sampling (hours)					
Kitchen	8.9 (3.3)	7.4 (3.2)	9.3 (3.5)	7.5 (3.2)	0.04*
Inside Home, but Not in Kitchen	21.6 (5.2)	27.7 (5.2)	22.4 (5.4)	27.8 (5.4)	<0.01*
Inside, at a Job	0.1 (0.5)	0	0.1 (0.6)	0	0.33
Inside, Elsewhere	2.4 (3.1)	1.1 (2.0)	2.0 (3.1)	0.9 (1.4)	0.08
Outdoors, at a Job Site	6.6 (5.2)	6.2 (6.8)	6.5 (5.2)	6.7 (6.8)	0.88
Outdoors, Elsewhere	8.4 (4.3)	5.6 (5.7)	7.7 (4.1)	5.1 (5.3)	0.04*

[a] Twelve samples were dropped prior to analysis for a lack of outcome data as described in the supplement. Tests for significant differences between Before and After responses were performed only for the dataset as analyzed.
[b] Field notes indicate that at least one participant's ACE-1 stove was non-operational during one of the two days of After period sampling.
[c] Data collected Before and After the intervention cannot be claimed as "independent." Because the data as analyzed included a mix of paired and un-paired data, independent sample testing methods were applied. This may increase bias and the likelihood of Type I error (Guo and Yuan, 2017); reported statistical significance should be used as a general indicator and with caution.
* Significant difference; see note "c".

across 13 countries to predict household concentration and, using an indicator variable for Female Exposure, to also predict $PM_{2.5}$ exposures in women in 106 countries (best model Bayesian $r^2 = 0.57$). The sample size and geographic and cultural diversity covered by the dataset used in Shupler et al. (2018b) go far beyond those of the data analyzed in our paper. Still, our model produced values (120.1 μg/m^3 for the Full model) considerably closer in magnitude to the measured mean 48-h average personal exposure concentration (123.2 μg/m^3) in rural Lao women using wood cookfuels on non-improved cookstoves than those estimated (as annual average 24-h concentrations) by the Bayesian techniques of Shupler et al. (2018b; 183 μg/m^3). The Bayesian model heavily over-predicted annual 24-h average exposure concentrations of Lao women using an improved biomass cookstove (ICS, e.g. the ACE-1) at 449 μg/m^3 in relation to the mean of our measured 48-h average values (80.8 μg/m^3). Our Full model produced estimates much closer at 88.1 μg/m^3.

Table 5
48-h model results - observed vs. predicted.

		Before sampling period			After sampling period		
Observed		123.2	54.5		80.8	30.5	
Model	CV r^2	Mean	SD	p-value[a]	Mean	SD	*p*-value[a]
Full	0.26	120.1	17.9	0.76	88.1	16.4	0.15
Full, Un-imputed Dataset	0.49	117.6	10.8	0.46	86.8	15.3	0.19
Full without KAP	0.01	105.5	11.6	0.12	91.7	6.9	0.07
Full Without Surveys	0.27	119.0	23.4	0.68	87.1	22.7	0.21
KEF Only	0.03	240.8	206.1	0.02	67.4	58.2	0.00
KEF Only - Stratified	0.02	199.5	170.5	0.04	89.0	40.3	0.28
LSIS Only	0.00	97.5	6.0	0.02	97.7	6.5	0.01
LSIS and Outdoor	0.31	119.7	26.0	0.72	85.8	16.4	0.30
LSIS and Wood Use	0.05	102.7	14.5	0.06	91.9	9.1	0.06
LSIS and Heating	0.04	108.1	17.7	0.16	91.7	10.7	0.09
LSIS and Exposure Scores	0.01	103.6	8.8	0.08	98.2	10.9	0.00

[a] Against observed.

Our analysis holds an advantage over that of Shupler et al. (2018b), as it employs data from the specific rural Lao population of interest. However, given the small relative sample size of our analysis and the diversity of the covariate set employed by Shupler et al. (2018b), the relative strength of accuracy observed in our models is likely assisted by factors other than population specificity. We believe our model benefits substantially from the ability of machine learning techniques like Random Forest to uncover and harness complex variable interrelationships to produce more accurate predictions. Future work should explore this notion more directly by comparing the results of machine learning techniques to the types of Bayesian analysis employed by Shupler et al. (2018b) using the same underlying datasets.

A more recent study (Yuchi et al., 2019) predicted *indoor* $PM_{2.5}$ concentrations using 10-fold cross validation of models produced with multiple linear regression, random forests regression, and a blend of the two during an air cleaner randomized controlled trial in Ulaanbaatar, Mongolia – a city heavily impacted by solid fuel use (Hill et al., 2017). Individual models predicted indoor concentrations, each with an r^2 around 0.50, similar to the prediction efforts in India. However, the use of a blended model produced an r^2 of 0.815, demonstrating that ensemble methods that include machine learning techniques may produce dramatic improvement in indoor concentration prediction models.

The best models in our analysis calculated 48-h $PM_{2.5}$ exposures with correlations (r^2) between 0.26 and 0.31 ($r^2 = 0.49$ using an un-imputed dataset). In many contexts, these correlations may be

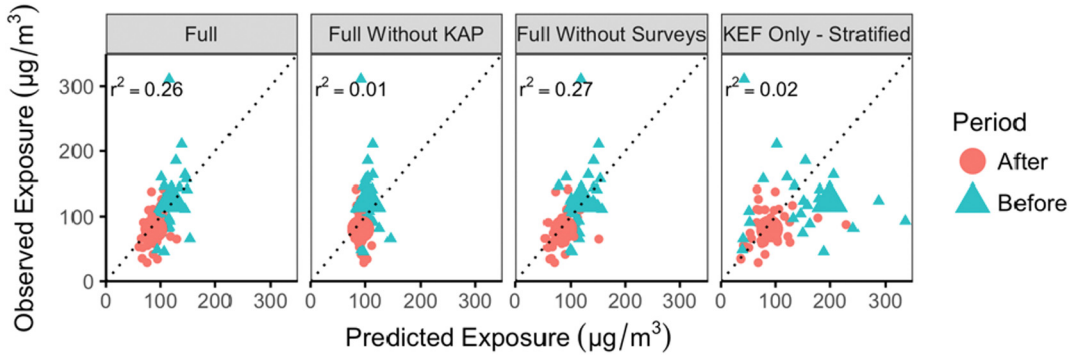


Fig. 4. Observed 48-h average exposures plotted against predicted values for the three primary models and the KEF with stratification model. Individual concentrations are depicted by smaller points, and mean concentrations, by larger points. All data are stratified by sampling period. The coefficient of determination (r^2) for the regression of observed values on predicted values is shown in the top left corner of each panel. X and Y axes were limited to 350 μg/m^3 to improve interpretability of the figures; these bounds were exceeded by 4 predicted values (max X = 861.3 μg/m^3) in the KEF Only – Stratified panel.

considered modest and are smaller than those produced by Shupler et al. (2018b) and Yuchi et al. (2019). In the context of predicting population-level exposures, however, they are of reasonable magnitude. For example, a cross-validated model derived from measurements of a small number of household-level, DHS-type variables in 617 households from 24 villages in India (Balakrishnan et al., 2013) demonstrated an r^2 of 0.31, and was ultimately employed in the GBD 2010 to estimate global HAP concentrations.

With cross validation, the machine learning and super learning methods applied to larger covariate sets strongly outperformed the more conventional method of linear KEF-based exposure assessment. KEF-based predictions overestimated mean Before exposures by 60–95%, but produced estimates of the After period mean that were similar to the measured value. Application of super learning and machine learning did improve the predictive performance of KAP (a mathematical variant of KEF when used as the sole predictor), but did not produce results on par with the best larger datasets. Researchers and practitioners should approach the predictive use of KEF with caution, at least in contexts similar to Lao PDR.

Poor generalizability of the bivariate relationship between indoor area concentration and exposure is supported by work performed elsewhere. A study of homes cooking with wood in Guatemala showed 48-h average $PM_{2.5}$ exposure estimates for mothers that were typically 70% lower than 48-h average KAP measurements in homes cooking with open fires but only 35% lower in homes using chimney stoves (Northcross et al., 2010). Despite average exposure values that were *lower* than average kitchen concentrations, 33% of measured exposure concentrations were *greater* than corresponding KAP levels. Women cooking with solid fuels over traditional open fires in Mexico (Armendáriz-Arnez et al., 2008) were shown to have mean 24-h average $PM_{2.5}$ exposures over 75% lower than mean 48-h average KAP concentrations. This relationship changed after the introduction of a Patsari chimney stove. In women cooking solely with the Patsari, this difference dropped to about 70%; in women cooking with a Patsari who also maintained an open fire outside, it dropped to about 55%; and in women cooking with a Patsari who also maintained an open fire in the kitchen it dropped to about 40%. A study of cookstove users in rural China (Baumgartner et al., 2011) demonstrated a significant relationship between individual 24-h kitchen concentrations and exposures in adults, but found no such connection in children and did not test predictive performance using CV. These regional examples are supported by a more-global analysis (Shupler et al., 2018b), which calculated female exposure to kitchen concentration ratios spanning a range of 0.33–0.74 across Global Burden of Disease regions.

The relative performance of the models examined in this analysis provides insight into which types of covariates may be of most predictive utility. Models trained on outdoor environmental data and KAP-related data were most accurate. Variable impact analysis identified morning wind speed, morning relative humidity, and KAP concentrations as producing the greatest impacts on model error in the Full Dataset. However, a direct causal link between these variables and exposures should not be automatically assumed. It is possible that these relationships are due to related, more-direct exposure factors such as outdoor ambient concentrations, which were only generally assessed in this analysis and which can be interrelated with HAP concentrations (Balakrishnan et al., 2014a) and influence concentrations in microenvironments in which a considerable portion of a person's time is spent. Future work should more comprehensively assess and incorporate ambient concentrations into their modeling approaches. Future modeling efforts should also explore the impacts of these variables on model strength across global regions of varying climate, housing type, and general culture.

Table 6
10 most important variables for prediction as estimated for select datasets.

Full dataset		Full without KAP		LSIS and heating	
Variable	% MSE increase	Variable	% MSE increase	Variable	% MSE increase
Morning Wind Speed - Day 1	407	Combined Cooking Exposure Score	147	Indoor Heating - Yes	319
Morning Wind Speed - Day 2	257	Wood Fuel Use (kg)	52	Ethnicity - Lao	87
Morning %RH - Day 1	187	Home Cooking Exposure Score - Morning	48	Any Heating	31
24-h KAP Concentration - Day 1	150	Indoor Heating Type - None[a]	45	Total Household Size	29
48-h KAP Concentration	73	Ambient $PM_{2.5}$ Concentration	34	Number of Household Members aged 5–14 Years	19
Combined Cooking Exposure Score	41	Indoor Heating - Yes[a]	34	Ethnicity - Katang	14
Afternoon Pressure - Day 1	32	Number of Pigs Owned	28	Kitchen Type - Cooks In House (Elsewhere)	13
Afternoon Wind Speed - Day 1	31	Home Cooking Exposure Score - Afternoon	25	Number of Pigs Owned	8
Morning Wind Direction, Blowing West - Day 2	29	Combined Grilling Exposure Score	24	Highest Education in Home - Lower Secondary	8
Afternoon Temp - Day 2	29	Owns Radio	14	Kitchen Type - Cooks In Separate Building	6

[a] Perfectly correlated with each other.

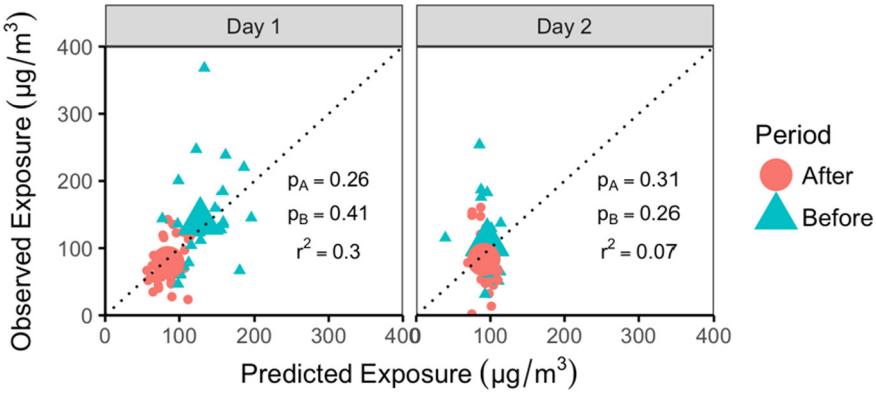


Fig. 5. Observed 24-h average exposures plotted against predicted values produced with CV SuperLearner and the Full dataset. Individual concentrations are depicted by smaller points, and mean concentrations, by larger points. All data are stratified by sampling period. The coefficient of determination (r^2) for the regression of observed values on predicted values, the *p*-value for a paired student's *t*-test for significance between observed and predicted Before values (p_B), and the same for After values (p_A) are shown in the bottom right corner of each panel.

Models including questionnaire data, while outperformed by models that used other information, demonstrated some predictive power. Of notable utility were heating indicators – heating is likely an important determinant of personal exposure in regions were combustion-based heating exists (Baumgartner et al., 2011). The predictive power shown by questionnaire data in general suggests that accurate prediction of group mean exposures in the context of rural solid-fuel cooking might not require lengthy surveys. Moreover, there is reason to believe that certain survey-related queries would be of use when predicting exposures in populations with household characteristics and behavior patterns that are more heterogeneous than those of the Lao study population. Previous studies have been able to explain a good deal of statistical variation in 24-h and 8-h average HAP exposures with data obtainable by questionnaire (Baumgartner et al., 2011; Clark et al., 2010).

LSIS data produced the least well-performing super learning models. Poor performance of the LSIS data model in this specific study population may be because the LSIS dataset does not include variables with much differentiation between seasons or cooking appliance. Globally administered DHS surveys like LSIS may be improved by adding a small number of questions about stove type, fuel usage, and heating behaviors. It is also likely that DHS indicators provide more predictive benefit in regions in which household characteristics, like kitchen and stove type, are more variable or seasonal effects, like heating, are less variable. In support of the usefulness of DHS indicators, the GBD 2010 model was able to reliably predict KAP concentrations from DHS-like data (Balakrishnan et al., 2013).

The single best candidate learner on the Full dataset was random forest, with an average root mean squared error (RMSE) of 40.0 μg/m^3, followed by cForest at 43.2 μg/m^3. Generalized linear modeling with regularization performed about as well as cForest with an RMSE of 44.6 μg/m^3. Researchers interested in HAP exposure prediction should explore decision trees and regression with regularization, and may benefit from the use of a super learner, which, in this analysis, did not improve upon the best individual candidate learner (RMSE: 40.1 μg/m^3). The neural network package, our implementation of which was relatively simplistic, did not outperform most other candidates (RMSE: 48.1 μg/m^3). However, future analyses should investigate the utility of more complex and more nuanced neural networks.

The findings described in this paper provide guidance to future HAP studies, especially those measuring and modeling changes in exposure due to an intervention. Based on our experience, we have

Fig. 6. A kitchen with ventilation features typical of those found in the area of study (Credit: Philipp Koetting).

produced a short list of lessons learned and suggestions for future research:

1. KAP and KEF alone should not be relied upon as unbiased predictors of exposure without verification.
2. Exposure-related parameters should be dealt with carefully. In some instances, confounders may be measured and adjusted for or employed as predictors.
3. Researchers across the globe should include DHS survey questions in their questionnaires to allow for stronger evaluation of the utility of DHS data in estimating HAP concentrations and related exposures across a wider global context.
4. Covariates with high "importance" in HAP exposure prediction models are not necessarily causal factors for the associated changes in exposure, and may not be externally generalizable. More studies in a wider variety of populations are needed.
5. The development of a broadly applicable and easy-to-calculate cooking exposure score should be explored, perhaps for use by the DHS program. Our random forests analyses, while specific to our rural Lao study population, suggest that an exposure score could prove useful in the prediction of exposures in the wider solid-fuel cooking context.
6. Structured, timely, and repeated quality control measures – like reviewing participant responses with interviewers and discussing unanticipated points of confusion at the end of each sample day or week – should be used during periods of questionnaire administration.

The prediction methods we used have several strengths. Exposure, the metric of interest to health investigators, was directly predicted. The use of machine learning enabled exploration of hidden and complex relationships precluded by common linear and log-linear regression approaches. Furthermore, ensemble methods had the potential to improve predictive power, and were enabled by the exploration of datasets with dozens to hundreds of covariates, allowing a broad investigation of predictive power. Despite the inclusion of more covariates than samples, cross-validation methods limited error from overfitting.

The limitations of this work highlight areas where further research is needed. Our sample size of 60 exposure measurements from a single district in Lao PDR is unlikely to provide the heterogeneity required to produce valid inferences or models with wide external validity. Additionally, collection of the large number of survey indicators and measurements described is resource intensive.

Evidence suggests the final gravimetric analysis approach used in this study is reliable (Garland et al., 2018). It should not be overlooked, however, that our analysis may have incurred some error or bias from the gravimetric equipment malfunction and related remedies discussed in the Methods section. Additional assessments using more traditional gravimetric analysis techniques may prove useful in confirming the data and our findings.

The World Health Organization recently undertook a collaborative process of revising existing survey indicators related to household energy use (World Health Organization, 2018a, 2018b). This includes an expanded set of questions focusing on both primary and secondary use of fuels and stoves for cooking, heating, and lighting. It is likely that this set of more nuanced survey indicators – combined with other traditional LSIS-style questions – may provide more robust household energy indicators and improve the predictive power and precision of models that rely upon them.

The impact of participant compliance on predictive power was not analyzed. Future work should assess compliance, perhaps through the use of accelerometers on monitors. This will also help investigators better understand how their measured exposures relate to true exposures. Approaches with smaller, more targeted variable sets – perhaps informed by the initial findings of our analysis or similar future work performed across a wider geographic and cultural domain – may produce models that are more manageable and less onerous for researchers and, importantly, participants. Finally, comprehensive longitudinal measurements of both predictors and exposures were not collected in each sampling period. Expanded collection and more in-depth analysis of repeated or long-term measurements would better capture between- and within-individual variability. This has been shown to considerably improve the reliability of predictions of HAP-related exposures (McCracken et al., 2009).

Acknowledgements

The success of this research was fostered by the generosity and backing of local leaders in Lao PDR, Lao Government officials like Mr. Seumkham Thoummavongsa at the Ministry of Energy and Mines, and, of course, the kind study participants who allowed us into their homes, kitchens, and lives. Berkeley Air Monitoring Group, Inc., including Kirstie Jagoe, put a great deal of effort into the field preparation, team training, design, initial field work, and data quality assurance involved in the study from which this research evolved. We owe a debt of gratitude to the field teams who collected the data used. These include the following colleagues from Lao Institute for Renewable Energy (LIRE) who were managed by co-author A. Pelletreau: Keolamphanh Bouttasing, Laura Magni, Phetdavone Nammaxay, Souksengdao Phetsaphangthong, Soraia Sadid, and Lamduan Thammavong. We thank Geo-Sys colleagues Vichitta Linthavong and Line Ellen Ankjaer Nielsen who worked to revisit Lao questionnaire audio recordings. In addition, we are thankful for the support and guidance of Ken Newcombe, Rutu Dave, Natsuko Toba, Maria Teresa Hernandez, Ellen Eisen, and Mark Nicas and for the support of SNV Vientiane, the stove disseminator.

This work also benefits from the development of a collaborative report on the stove intervention project presented by many of the study co-authors to the Ministry of Health and Inter-Ministerial Clean Stove Initiative of the Lao People's Democratic Republic in 2015. The current study contributes substantial additional analysis to the original report, including the implementation of super learning and machine learning to model personal exposures.

A version of this paper appeared as part of the UC Berkeley PhD dissertation, 2017, by Lawson Andrew Hill, which is available from the University of California's online repository, "eScholarship," and ProQuest.

Partial funding came through Berkeley Air Monitoring Group, Inc. from the World Bank, United States (Contract Number 7171980), under the Asia Sustainable and Alternative Energy Program (ASTAE), as well as the National Science Foundation Integrative Graduate Education and Research Traineeship under Grant No. DGE-1144885. NSF traineeship support was generously provided through the Berkeley Center for Green Chemistry. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation. Partial financial and technical support by the Energy Sector Management Assistance Program (ESMAP) is gratefully acknowledged. ESMAP—a global knowledge and technical assistance program administered by the World Bank—assists low- and middle-income countries to increase their know-how and institutional capacity to achieve environmentally sustainable energy solutions for poverty reduction and economic growth. ESMAP is funded by Australia, Austria, Canada, Denmark, the European Commission, Finland, France, Germany, Iceland, Italy, Japan, Lithuania, Luxemburg, the Netherlands, Norway, the Rockefeller Foundation, Sweden, Switzerland, the United Kingdom, and the World Bank.

Conflicts of interest

No conflicts of interest are known.

Appendix A. Supplement. Additional details on employed methods and findings

Supplementary information for this article can be found online at https://doi.org/10.1016/j.scitotenv.2019.04.258.

References

Anenberg, S., Kinney, P., Newcombe, K., Talyan, V., Goyal, A., Hewlett, O., 2017. Gold Standard Methodology to Estimate and Verify Averted Mortality and Disability Adjusted Life Years (ADALYs) from Cleaner Household Air. pp. 1–52.

Armendáriz-Arnez, C., Edwards, R.D., Johnson, M., Zuk, M., Rojas, L., Jiménez, R.D., et al., 2008. Reduction in personal exposures to particulate matter and carbon monoxide as a result of the installation of a Patsari improved cook stove in Michoacan Mexico. Indoor Air 18, 93–105. https://doi.org/10.1111/j.1600-0668.2007.00509.x.

Balakrishnan, K., Parikh, J., Sankar, S., Padmavathi, R., Srividya, K., Venugopal, V., et al., 2002. Daily average exposures to respirable particulate matter from combustion of biomass fuels in rural households of southern India. Environ. Health Perspect. 110, 1069–1075. https://doi.org/10.1289/ehp.021101069.

Balakrishnan, K., Thangavel, G., Ghosh, S., Sambandam, S., Mukhopadhyay, K., Adair-Rohani, H., et al., 2011. The Global Household Air Pollution Database 2011 (Version 3.0).

Balakrishnan, K., Ghosh, S., Ganguli, B., Sambandam, S., Bruce, N., Barnes, D.F., et al., 2013. State and national household concentrations of PM2.5 from solid cookfuel use: results from measurements and modeling in India for estimation of the global burden of disease. Environ. Health 12, 77. https://doi.org/10.1186/1476-069X-12-77.

Balakrishnan, K., Cohen, A., Smith, K.R., 2014a. Perspectives | editorial addressing the burden of disease attributable to air pollution in India: the need to integrate across household. Environ. Health Perspect. 122, A6–A7. https://doi.org/10.1289/ehp.1307822.

Balakrishnan, K., Mehta, S., Ghosh, S., Johnson, M., Brauer, M., Zhang, J., et al., 2014b. Review 5: Population Levels of Household Air Pollution and Exposures. WHO Indoor Air Qual. Guidel. Househ. Fuel Combust. World Health Organization, Geneva, Switzerland.

Baumgartner, J., Schauer, J.J., Ezzati, M., Lu, L., Cheng, C., Patz, J., et al., 2011. Patterns and predictors of personal exposure to indoor air pollution from biomass combustion among women and children in rural China. Indoor Air 21, 479–488. https://doi.org/10.1111/j.1600-0668.2011.00730.x.

Bonjour, S., Adair-Rohani, H., Wolf, J., Bruce, N.G., Mehta, S., Prüss-Ustün, A., et al., 2013. Solid fuel use for household cooking: country and regional estimates for 1980-2010. Environ. Health Perspect. 121, 784–790. https://doi.org/10.1289/ehp.1205987.

Burnett, R.T., Pope, C.A., Ezzati, M., Olives, C., Lim, S., Mehta, S., et al., 2014. An integrated risk function for estimating the global burden of disease attributable to ambient fine particulate matter exposure. Environ. Health Perspect. 122, 397–403. https://doi.org/10.1289/ehp.1307049.

Clark, M.L., Reynolds, S.J., Burch, J.B., Conway, S., Bachand, A.M., Peel, J.L., 2010. Indoor air pollution, cookstove quality, and housing characteristics in two Honduran communities. Environ. Res. 110, 12–18. https://doi.org/10.1016/j.envres.2009.10.008.

Clark, M.L., Peel, J.L., Balakrishnan, K., Breysse, P.N., Chillrud, S.N., Naeher, L.P., et al., 2013. Health and household air pollution from solid fuel use: the need for improved exposure assessment. Environ. Health Perspect. 121, 1120–1128.

Cohen, A.J., Brauer, M., Burnett, R., Anderson, H.R., Frostad, J., Estep, K., et al., 2017. Estimates and 25-year trends of the global burden of disease attributable to ambient air pollution: an analysis of data from the global burden of diseases study 2015. Lancet 389, 1907–1918. https://doi.org/10.1016/S0140-6736(17)30505-6.

Garland, C., Delapena, S., Pennise, D., 2018. An alternative technique for determining gravimetric particle mass deposition on filter substrate: the particle extraction method. Open J. Air Pollut. 07, 309–321. https://doi.org/10.4236/ojap.2018.74016.

Guo, B., Yuan, Y., 2017. A comparative review of methods for comparing means using partially paired data. Stat. Methods Med. Res. 26, 1323–1340. https://doi.org/10.1177/0962280215577111.

Health Effects Institute, 2018. State of Global Air 2018. https://www.stateofglobalair.org/data/#/air/map, Accessed date: 20 October 2018.

Hill, L., Pillarisetti, A., Smith, K., Delapena, S., Garland, C., Jagoe, K., et al., 2015. Air Pollution and Impact Analysis of a Pilot Stove Intervention, Report to the Ministry of Health and Inter-Ministerial Clean Stove Initiative of the Lao People's Democratic Republic. UC Berkeley Household Energy, Health and Climate Group and the Berkeley Air Monitoring Group, Berkeley, California.

Hill, L.D., Edwards, R., Turner, J.R., Argo, Y.D., Olkhanud, P.B., Odsuren, M., et al., 2017. Health assessment of future PM2.5 exposures from indoor, outdoor, and secondhand tobacco smoke concentrations under alternative policy pathways in Ulaanbaatar, Mongolia. PLoS One https://doi.org/10.1371/journal.pone.0186834.

Huang, K., Fu, J.S., Hsu, N.C., Gao, Y., Dong, X., Tsay, S.-C., et al., 2013. Impact assessment of biomass burning on air quality in southeast and East Asia during BASE-ASIA. Atmos. Environ. 78, 291–302. https://doi.org/10.1016/j.atmosenv.2012.03.048.

ICF International, 2011. Demographic and Health Surveys Methodology - Questionnaires: Household, woman's, and man's.

Johnson, M., Pennise, D., Lam, N., Brant, S., Charron, D., Gray, C., et al., 2011. Modeling indoor air pollution concentrations from stove emissions using a Monte Carlo single-box model. Atmos. Environ. 45, 3237–3243.

Kandpal, J.B., Maheshwari, R.C., Kandpal, T.C., 1995. Indoor air pollution from domestic cookstoves using coal, kerosene and LPG. Energy Convers. Manag. 36, 1067–1072. https://doi.org/10.1016/0196-8904(94)00087-G.

Liaw, A., Wiener, M., 2002. Classification and regression by randomForest. R News 2, 18–22.

McCracken, J.P., Schwartz, J., Bruce, N., Mittleman, M., Ryan, L.M., Smith, K.R., 2009. Combining individual- and group-level exposure information. Epidemiology 20, 127–136. https://doi.org/10.1097/EDE.0b013e31818ef327.

Mengersen, K., Morawska, L., Wang, H., Murphy, N., Tayphasavanh, F., Darasavong, K., et al., 2011. Association between indoor air pollution measurements and respiratory health in women and children in Lao PDR. Indoor Air 21, 25–35. https://doi.org/10.1111/j.1600-0668.2010.00679.x.

Ministry of Health, Lao Statistics Bureau of the Ministry of Planning and Investment, 2012. Lao PDR Social Indicator Survey (MICS-DHS) 2011–2012. Lao PDR, Vientiane.

Morawska, L., Mengersen, K., Wang, H., Tayphasavanh, F., Darasavong, K., Holmes, N.S., 2011. Pollutant concentrations within households in Lao PDR and association with housing characteristics and occupants' activities. Indoor Air 45, 882–889. https://doi.org/10.1021/es102294v.

Northcross, A., Chowdhury, Z., McCracken, J., Canuz, E., Smith, K.R., 2010. Estimating personal PM2.5 exposures using CO measurements in Guatemalan households cooking with wood fuel. J. Environ. Monit. 12, 873–878. https://doi.org/10.1039/b916068j.

Polley, E.C., LeDell, E., Kennedy, C., van der Laan, M.J., 2016. SuperLearner: Super Learner Prediction.

R Core Team, 2016. R: A Language and Environment for Statistical Computing.

Shupler, M., Balakrishnan, K., Ghosh, S., Thangavel, G., Stroud-Drinkwater, S., Adair-Rohani, H., et al., 2018a. Global household air pollution database: kitchen concentrations and personal exposures of particulate matter and carbon monoxide. Data Br 21, 1292–1295. https://doi.org/10.1016/j.dib.2018.10.120.

Shupler, M., Godwin, W., Frostad, J., Gustafson, P., Arku, R.E., Brauer, M., 2018b. Global estimation of exposure to fine particulate matter (PM2.5) from household air pollution. Environ. Int. 120, 354–363. https://doi.org/10.1016/j.envint.2018.08.026.

Smith, K., Bruce, N., Balakrishnan, K., Adair-Rohani, H., Balmes, J., Dherani, M., et al., 2014. Millions dead: how do we know and what does it mean? Methods used in the comparative risk assessment of household air pollution. Am Rev Public Heal 35, 185–206. https://doi.org/10.1146/annurev-publhealth-032013-182356.

Smith, K., Pillarisetti, A., Hill, L.D., Charron, D., Delapena, S., Garland, C., et al., 2015. Proposed Methodology: Quantification of a Saleable Health Product (aDALYs) from Household Cooking Interventions. Berkeley, CA. .

Steinle, S., Reis, S., Sabel, C.E., 2013. Quantifying human exposure to air pollution —moving from static monitoring to spatio-temporally resolved personal exposure assessment. Sci. Total Environ. 443, 184–193. https://doi.org/10.1016/j.scitotenv.2012.10.098.

van der Laan, M.J., Polley, E.C., Hubbard, A.E., 2007. Super learner. Stat. Appl. Genet. Mol. Biol. 6, 25. https://doi.org/10.2202/1544-6115.1309.

World Bank, Institute for Health Metrics & Evaluation, 2016. The Cost of Air Pollution: Strengthening the Economic Case for Action. International Bank for Reconstruction and Development/The World Bank, Washington, DC.

World Health Organization, 2018a. Survey harmonization process. http://www.who.int/airpollution/household/survey-harmonization/en/, Accessed date: 20 October 2018.

World Health Organization, 2018b. New harmonized household energy survey questions finalized. http://www.who.int/airpollution/household/harmonized-survey/en/, Accessed date: 20 October 2018.

Yuchi, W., Gombojav, E., Boldbaatar, B., Galsuren, J., Enkhmaa, S., Beejin, B., et al., 2019. Evaluation of random forest regression and multiple linear regression for predicting indoor fine particulate matter concentrations in a highly polluted city. Environ. Pollut. 245, 746–753. https://doi.org/10.1016/j.envpol.2018.11.034.