5-17-2021

Estimation of household smoke-exposure risk using Demographic and Health Survey (DHS) data

Mazbahul G. Ahamad

Fahian Tanin

MethodsX 8 (2021) 101390

Contents lists available at ScienceDirect

MethodsX

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

Method Article

Estimation of household smoke-exposure risk using Demographic and Health Survey (DHS) data

Mazbahul G Ahamad[a,*], Fahian Tanin[b]

[a] University of Nebraska–Lincoln, Lincoln, NE 68583, USA
[b] Independent Researcher, Sylhet 3100, Bangladesh

A B S T R A C T

We introduce Stata and R codes to estimate the household smoke-exposure risk (SER) variable using cooking fuel- and cooking place-related information obtained from country-level demographic and health survey (DHS) data. Two categories of cooking fuels (smoke-producing and non-smoke producing fuels) and two categories of cooking places (indoor and outdoor) are used to estimate the household SER. Finally, household SER is classified into four levels of risk: high (cooking indoor using smoke-producing fuels), medium (cooking outdoor using smoke-producing fuels), low (cooking indoor using non-smoke-producing fuels), and very low (cooking outdoor using non-smoke-producing fuels). An example of a household SER calculation using the DHS data and codes is provided for clarification. The available DHS data of over 90 countries contain both cooking fuel- and cooking place-related information, so the method of estimating household SER would be the same for these countries.

- Household-level cooking fuel and cooking place data can be used to estimate household SER.
- This paper illustrates an estimation technique for household smoke-exposure risk (SER) using demographic and health survey (DHS) data.
- This method can be used to estimate household SER data for any country on the DHS country list.

A R T I C L E I N F O
Method name: Method of household smoke-exposure risk using demographic and health survey (DHS) data
Keywords: Cooking place, Cooking fuel, Smoke producing fuel, Smoke exposure risk, Indoor air quality, Demographic and health survey, DHS data
Article history: Received 5 April 2021; Accepted 15 May 2021; Available online 17 May 2021

* Corresponding author.
E-mail address: mahamad2@unl.edu (M.G. Ahamad).

https://doi.org/10.1016/j.mex.2021.101390

Specifications table

Subject Area:	Environmental Science
More specific subject area:	Environmental health Public health
Method name:	Method of household smoke-exposure risk using demographic and health survey (DHS) data
Name and reference of original method:	Household smoke-exposure risks associated with cooking fuels and cooking places in Tanzania: a cross-sectional analysis of demographic and health survey data [1]. Original reference of the method article is available at: https://www.mdpi.com/1660-4601/18/5/2534
Resource availability:	The demographic and health survey (DHS) data of over 90 countries worldwide are publicly available at: https://dhsprogram.com/data/available-datasets.cfm

Method details

Background

Household air pollution from cooking fuels and cooking places is a global health concern and a leading cause of different air pollution-related diseases. Different types of household cooking fuels (i.e., smoke-producing and non-smoke-producing cooking fuels) and cooking places (i.e., indoor and outdoor cooking places) create different levels of household smoke-exposure risks (SERs), which can be estimated using the demographic and health survey (DHS) data of a country [1]. In this article, we presented explicit and reproducible estimation and classification techniques of household SER using Stata (www.stata.com) and R (www.r-project.org) codes.

Data

Using country-level DHS data, SER can be estimated to understand the different levels of household SERs linked to households' cooking fuels and cooking places. Two variables of household-level data (indicated as HR file of DHS), namely, `hv226 (type of cooking fuel)` and `hv241 (food cooked in the house/separate building/outdoors)`, are required to estimate the household SER. The country-level DHS data of different countries are publicly available at https://dhsprogram.com/data/available-datasets.cfm and can be accessed freely upon data request approval (https://dhsprogram.com/data/new-user-registration.cfm).

Conceptual approach of household SER variable estimation

Household SER is an ordered variable that denotes the four types of SERs associated with the use of different types of household cooking fuels and cooking places. In the study, we used 2015–16 Tanzania Demographic and Health Survey (TDHS) data to illustrate the estimation method. To construct the household SER, we followed the following steps:

Step 1: We classified the different types of cooking fuel (`hv226`) into two major categories, namely, "smoke-producing fuels" and "non-smoke-producing fuels." Kerosene, charcoal, wood, straw, shrubs, grass, animal dung, and others were categorized as smoke-producing cooking fuel. Bottled gas and electricity were grouped as non-smoke-producing cooking fuel. Different countries might have different types of fuels, which need to be adjusted based on the available data.

Step 2: The second variable (`hv241`) represents the households' cooking places. "Households that cooked inside their houses or in a separate building were categorized as "indoor cooking" households, whereas households that reported using outdoor cooking places were categorized as "outdoor cooking" households" [1]. Different countries might have different types of cooking places, which need to be modified based on the available data.

Table 1
Classification of household smoke-exposure risk (SER).

Cooking Fuel (cf)			Cooking Place (cp)			Level of Household SER
DHS Variable	Value Level(s)*	cf Type	DHS Variable	Value Level(s)*	cp Type	
hv226	1, 12	Smoke-producing cf	hv241	1, 2	Indoor cp	High SER
	1, 12	Smoke-producing cf		3	Outdoor cp	Medium SER
	5, 7, 8, 9, 11	Non-smoke-producing cf		1, 2	Indoor cp	Low SER
	5, 7, 8, 9, 11	Non-smoke-producing cf		3	Outdoor cp	Very low SER

Note: The 2015—16 TDHS data consist only of the *indicated fuel types and places, which might be different in other countries.

Step 3: The combination of the type of cooking fuel (hv226) and the location of the cooking place (hv241) produced an ordinal variable, which is categorized into four ordinal levels of household SERs: high (cooking indoor using smoke-producing fuels), medium (cooking outdoor using smoke-producing cooking fuels), low (cooking indoor using non-smoke-producing cooking fuels), and very low (cooking outdoor using non-smoke-producing cooking fuels) (Table 1). In this estimation, it is assumed that smoke-producing cooking fuels are commonly more hazardous and unhealthier to humans than non-smoke-producing cooking fuels. It is also assumed that indoor cooking is more hazardous and unhealthier than outdoor cooking because of the high possibility for smoke exposure in a confined space [1].

Step-by-step approach in household SER estimation

This section presents both Stata and R codes for estimating household SER using 2015—16 TDHS data. Different countries might have different types of cooking fuels and cooking places, so the respective variables should be accordingly defined by the user.

Stata codes for estimating household (HH) SER variable

```
  **step 1: define variable: cooking fuel (cf)
  label list HV226
  tab hv226
  recode hv226 (1 12 = 0) (5 7 8 9 11 = 1) (95 96 = .), gen(cf)
  label define CF 0 "Non-smoke-producing Fuels" 1 "Smoke-producing
Fuels"
  label values cf CF
  label variable cf "HH's Cooking Fuels"
  tab cf
  **step 2: define variable: cooking place (cp)
  label list HV241
  tab hv241
  recode hv241 (3 = 0) (1 2 = 1) (6 = .), gen(cp)
```

```
label define CP 0 "Outdoor Cooking" 1 "Indoor Cooking"
label values cp CP
label variable cp "HH's Cooking Places"
tab cp
**step 3: define variable: smoke-exposure risk (ser)
gen ser = .
replace ser=1 if cf==1 & cp==1
replace ser=2 if cf==1 & cp==0
replace ser=3 if cf==0 & cp==1
replace ser=4 if cf==0 & cp==0
label define SER 1 "High SER" 2 "Medium SER" 3 "Low SER" 4 "Very Low
SER"
label values ser SER
label list SER
label variable ser "HH's SER Levels"
tab ser
```

The following R codes can be used to construct the household SER variable similar to the Stata codes using the same demographic and health survey data.

R codes for estimating household (HH) SER variable

```
## step 1: define variable: cooking fuel (cf)
table(newtdhs$hv226)
newtdhs$cf[newtdhs$hv226=="1"] <- 0
newtdhs$cf[newtdhs$hv226=="12"] <- 0
newtdhs$cf[newtdhs$hv226=="5"] <- 1
newtdhs$cf[newtdhs$hv226=="7"] <- 1
newtdhs$cf[newtdhs$hv226=="8"] <- 1
newtdhs$cf[newtdhs$hv226=="9"] <- 1
newtdhs$cf[newtdhs$hv226=="11"] <- 1
newtdhs$cf[newtdhs$hv226=="95"] <- "NA"
newtdhs$cf[newtdhs$hv226=="96"] <- "NA"
newtdhs$cf_label <- ordered(newtdhs$cf,
   levels = c(0,1),
   labels = c("Non-smoke-producing Fuels", "Smoke-producing Fuels"))
## step 2: define variable: cooking place (cp)
table(newtdhs$hv241)
newtdhs$cp[newtdhs$hv241=="3"] <- 0
newtdhs$cp[newtdhs$hv241=="1"] <- 1
newtdhs$cp[newtdhs$hv241=="2"] <- 1
newtdhs$cp[newtdhs$hv241=="6"] <- "NA"
newtdhs$cp_label <- ordered(newtdhs$cp,
   levels = c(0,1),
   labels = c("Outdoor Cooking", "Indoor Cooking"))
## step 3: define variable: smoke exposure risk (ser)
newtdhs$ser[newtdhs$cf=="1" & newtdhs$cp=="1"] <- 1
newtdhs$ser[newtdhs$cf=="1" & newtdhs$cp=="0"] <- 2
newtdhs$ser[newtdhs$cf=="0" & newtdhs$cp=="1"] <- 3
newtdhs$ser[newtdhs$cf=="0" & newtdhs$cp=="0"] <- 4
table(newtdhs$ser)
newtdhs$ser_label <- ordered(newtdhs$ser,
   levels = c(1,2,3,4),
   labels = c("High SER"), "Medium SER", "Low SER", "Very Low SER")
```

Illustration of methods

We executed all the Stata commands sequentially presented in the previous section to estimate the household SER using 2015–16 Tanzania DHS data following the method described in Ahamad et al. [1]. The presented R codes will produce similar results to those obtained using the R software . First, we need to load the respective household (HR) file of the DHS data.

In Step 1, we need to find the available cooking fuel information captured by the `hv226 (type of cooking fuel)` variable. The following code illustrates the complete list (`label list HV226`) and different types of cooking fuels (`tab hv226`) used by the households in Tanzania. The following code extracts the value labels of `hv226`.

```
. label list HV226
HV226:
           1 electricity
           2 lpg
           3 natural gas
           4 biogas
           5 kerosene
           6 coal, lignite
           7 charcoal
           8 wood
           9 straw/shrubs/grass
          10 agricultural crop
          11 animal dung
          12 bottled gas
          95 no food cooked in house
          96 other
```

As shown below, Tanzanian households used nine different types of cooking fuels.

```
. tab hv226
```

type of cooking fuel	Freq.	Percent	Cum.
electricity	61	0.49	0.49
kerosene	187	1.49	1.97
charcoal	3,143	25.02	26.99
wood	8,743	69.59	96.59
straw/shrubs/grass	8	0.06	96.65
animal dung	1	0.01	96.66
bottled gas	285	2.27	98.93
no food cooked in house	131	1.04	99.97
other	4	0.03	100.00
Total	12,563	100.00	

Next, we need to define two major types of cooking fuels (non-smoke-producing fuels = 0 and smoke-producing fuels = 1) by using the `recode` command to produce a new variable `cf`, as shown below. The user can also modify the recoding strategies to define `cf` depending on the cooking fuels available for a specific country. It is important to note that the percent values of the following figures are unweighted, and appropriate weight needs to be used during formal statistical analysis.

Households that did not cook food in the house or used other cooking places were not studied in our analysis and were recoded as missing values for exclusion. The `cf` is estimated as:

```
recode hv226 (1 12 = 0) (5 7 8 9 11 = 1) (95 96 = .), gen(cf)
label define CF 0 "Non-smoke-producing Fuels" 1 "Smoke-producing Fuels"
. tab cf
```

HH's Cooking Fuels	Freq.	Percent	Cum.
Non-smoke-producing Fuels	346	2.78	2.78
Smoke-producing Fuels	12,082	97.22	100.00
Total	12,428	100.00	

In Step 2, similarly, we need to find the available cooking places captured by `hv241 (food cooked in the house/separate building/outdoors)` variable. As shown below, the following code illustrates the different types of cooking places in Tanzania.

```
. label list HV241
HV241:
           1 in the house
           2 in a separate building
           3 outdoors
           6 other
```

Now, we need to define two major types of cooking places (outdoor cooking = 0 and indoor cooking = 1) by using the `recode` command to produce a new variable `cp`. The user can also modify the recoding strategies to define `cp` depending on the cooking places available for a specific country. The `cp` is computed as:

```
recode hv241 (3 = 0) (1 2 = 1) (6 = .), gen(cp)
label define CP 0 "Outdoor Cooking" 1 "Indoor Cooking"
. tab cp
```

HH's Cooking Places	Freq.	Percent	Cum.
Outdoor Cooking	2,330	18.75	18.75
Indoor Cooking	10,099	81.25	100.00
Total	12,429	100.00	

Finally, in Step 3, the combination of two types of cooking fuels (non-smoke-producing fuels = 0 and smoke-producing fuels = 1 of `cf`) and location of the cooking places (outdoor cooking = 0 and indoor cooking = 1 of `cp`) produced four levels of household SER (`ser`), namely, high SER (smoke-producing cooking fuels = 1 and indoor cooking = 1), medium SER (smoke-producing cooking fuels = 1 and outdoor cooking = 0), low SER (non-smoke-producing cooking fuels = 0 and indoor cooking = 1), and very low SER (non-smoke-producing cooking fuels = 0 and outdoor cooking = 0). To compute `ser`, we need to use the following codes:

```
gen ser = .
replace ser=1 if cf==1 & cp==1
replace ser=2 if cf==1 & cp==0
```

```
  replace ser=3 if cf==0 & cp==1
  replace ser=4 if cf==0 & cp==0
  label define SER 1 "High SER" 2 "Medium SER" 3 "Low SER" 4 "Very Low
SER"
  The output of ser is estimated as:
  . tab ser
```

HH's SER Levels	Freq.	Percent	Cum.
High SER	9,766	78.60	78.60
Medium SER	2,313	18.62	97.22
Low SER	330	2.66	99.87
Very Low SER	16	0.13	100.00
Total	12,425	100.00	

This result provides four levels of SER based on cooking fuels (`hv226`) and cooking places (`hv241`) variables.

Conclusion

In this method article, we presented both Stata and R codes for household SER estimation using country-level demographic and health survey data, following Ahamad et al. (2021) [1]. We further provided an example of the household SER estimation method using 2015–16 TDHS data for additional clarification and hands-on illustration of the SER estimation technique.

Funding

None.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

We thank DHS Program of ICF International, USA for authorizing 2015–16 TDHS data. We appreciate comments and suggestions from participants of the CUGH's 2020 Global Health Conference, APHA's 2020 Virtual Annual Meeting and Expo, and anonymous reviewers. M. G. Ahamad acknowledges support from a postdoctoral travel grant from the University of Nebraska–Lincoln.

References

[1] M Ahamad, F Tanin, N. Shrestha, Household smoke-exposure risks associated with cooking fuels and cooking places in Tanzania: a cross-sectional analysis of demographic and health survey data, Int. J. Environ. Res. Public Health 18 (2021) 2534.