

---

Subject: merging HR and KR files and weighting for analysis

Posted by [ashonci](#) on Thu, 08 Jun 2023 06:49:31 GMT

[View Forum Message](#) <> [Reply to Message](#)

---

Hi, I am doing my MSc project using Rwanda 2019-2020 DHS data.

I am trying to analyze the association household air pollution from cooking (cooking fuel type and cooking place) and smoking inside (table 2.3 in the final report) and low birth weight (LBW) in Rwanda (table 10.1 in the final report). I also plan to examine the prevalence of LBW according to child factor (birth order of baby), maternal factor (age at birth, smoking status, education level) and socio-demographic factors (place of residence, province, wealth quintile) (table 10.1 in the final report). I will use HR files for household characteristics (primary exposures) and KR files for LBW (outcome) and co-variables.

1) I selected recode number for each variable. Could you check whether they are correct?

-Outcome: LBW (m19)

-Primary exposures: cooking fuel type (hv226) , cooking place (hv241), smoking inside (hv252)

-co-variables: birth order of baby (BORD), mother's age at birth (v013), smoking status (v463 a-z), education level (v106) and place of residence (v025), province (v024), wealth quintile (v190)

2) I chose to use KR file for the outcome and co-variables since table 10.1 in the final report represents the percent distribution of LIVE BIRTHS IN THE 5YEARS PRECEDING THE SURVEY.

However, I was wondering whether it is correct to use KR file or IR file. I wasn't sure because the total number of women are different. The total number for KR file is 8,092 (unweighted) and 14,634 for IR file. I was wondering the basis should be KR file or IR file. Can you give me an advice for it?

3) Based on the assumption that I need to use KR file, I tried to merge KR file with HR file following Guide to DHS Statistics DHS-7 (version 2). I followed the formulae below and changed the recode number. Could you check whether it correct?

\* open secondary file, e.g. household file, selecting just the variables needed  
use hhid hv001 hv002 hv226 hv241 hv252 using "RWHR81FL.dta", clear

\* rename, generate or clone variables to be used for matching

rename hv001 v001

rename hv002 v002

\* sort according to the ID variables

sort hhid

\* save temporary file of just the variables to merge in

tempfile secondary

save "`secondary'", replace

\* open primary file

\* e.g. Children's file

use "RWKR81FL.dta", clear

\* creating matching variables  
gen hhid = substr(caseid,1,12)

\* sort according to the ID variables needed for matching  
sort hhid

\* now merge the data from the secondary file to the primary file  
\* keep(master match) keeps all entries from the KR file  
merge m:1 hhid using "`secondary'", keep(master match) keepusing(hv226 hv241 hv252)

\* check the merge - should all be matched  
tab \_merge

4) After merging, I tried to do weighting for the data. Then I wasn't sure whether I need to account for weight for both of HR file and KR file or only KR file. Is it possible to do weighting for both HR file and KR file and use two 'svy' commands? I checked from a website that I need to weight for KR file only in this case, but I wasn't sure. Can you give me an advice for it? And which formulae do I need to use for weighting?

5) I tried to account for weight for KR file. However, it still showed unweighted number of 8,092 on the screen even though I finished weighting. I expected it to show weighted number of 8,924 as table 10.1 in the final report. Is it normal to show unweighted number? Please check the formulae I followed below,

```
gen wgt=v005/1000000  
svyset [pw=wgt], psu(v021) strata(v022) or svyset v021 [pw=wgt], strata(v022)  
svy: tab v106
```

It is my first time of using STATA for my project so, I confront many difficulties. Thank you for reading the long inquires. Hope to get your response as soon as possible!

Thank you