
Subject: merging HR and KR files and weightng 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

Subject: Re: merging HR and KR files and weightng for analysis
Posted by [Janet-DHS](#) on Mon, 12 Jun 2023 13:20:38 GMT
[View Forum Message](#) <> [Reply to Message](#)

Following is a response from DHS staff member, Tom Pullum:

You are on the right track. Below I will post a modification of how you were doing the merge. You can save the temp file as you did--I just prefer to save it differently.

I recalculate age in single years at the time of the birth, using the cmc (century month code) for when the child was born minus the cmc for when the mother was born, dividing by 12, and converting an integer . v013 is age in five-year intervals and at the time of the survey, not the time of the birth. You can recode to intervals of age as you wish.

The sample weight variable to use is v005. This is the weight for the mother and it is used for her children. It is approximately proportional to hv005 but is adjusted for women's nonresponse.

The cases in the IR file are women. The cases in the KR file are children. The numbers of cases in the two files are different. The number of mothers is less than the number of women but more than the number of children. So the numbers you gave are ok.

Here are Stata lines that will do the merge. Let us know if you have other questions.

```
* specify a workspace
cd e:\DHS\DHS_data\scratch
```

```
use hv001 hv002 hv226 hv241 hv252 using "...RWHR81FL.DTA", clear
```

```
rename hv001 cluster
rename hv002 hh
sort cluster hh
save RWHRtemp.dta, replace
```

```
use "...RWKR81FL.DTA", clear
rename v001 cluster
rename v002 hh
gen mo_age_at_birth=int((b3-v011)/12)
sort cluster hh
merge m:1 cluster hh using RWHRtemp.dta
tab _merge
```

```
* _merge=2 for households with no children in the KR file
keep if _merge==3
```

Subject: Re: merging HR and KR files and weightng for analysis
Posted by [ashonci](#) on Wed, 14 Jun 2023 06:42:41 GMT
[View Forum Message](#) <> [Reply to Message](#)

Thank you for your reply! It was very helpful.

I have one more question.

For the variable of mother's smoking status (table 10.1 in the 2019-2020 DHS Rwanda final report) , I decided to use "v463z" only among v463a~z. Is it correct?

If not, can you advise me which recode I should use and how I can regroup or recode if needed?

Thanks!!

Subject: Re: merging HR and KR files and weightng for analysis
Posted by [Janet-DHS](#) on Tue, 20 Jun 2023 15:25:49 GMT
[View Forum Message](#) <> [Reply to Message](#)

Following is a response from DHS staff member, Tom Pullum:

Yes, v463z summarizes all the other v463 variables. Just be sure to get the coding right. 0 means smokes or uses tobacco; 1 means DOES NOT smoke or use tobacco. This kind of reverse coding always reminds me of an old song you can hear here:
https://www.youtube.com/watch?v=_hF05ik5TFQ,

Subject: Re: merging HR and KR files and weightng for analysis
Posted by [ashonci](#) on Mon, 26 Jun 2023 06:43:19 GMT
[View Forum Message](#) <> [Reply to Message](#)

Thank you for your witty reply.

I have one more question. By several literatures, twin babies are more likely to be born with low birth weight compared to single babies. I would like to drop the cases of twin babies. I checked a variable of "b0" and there were 241 twin babies by the analysis. I want to double-check whether it is correct before I drop them.

Thank you!

Subject: Re: merging HR and KR files and weightng for analysis
Posted by [Janet-DHS](#) on Mon, 03 Jul 2023 14:40:29 GMT
[View Forum Message](#) <> [Reply to Message](#)

Following is a response from DHS staff member, Tom Pullum:

In most surveys, the mortality of children from multiple births is much higher than the mortality of singletons. There are probably several reasons, not just low birthweight. If you want to calculate overall mortality rates, then all live births should be included, but if you are looking into causes of mortality, you could drop the multiple births--just be sure you say that was done. This is not a matter of being correct or not, it's just something that is often done.
