
Subject: Merging KR with PR for Ethiopia 2011
Posted by [Yue_Chen](#) on Thu, 24 May 2018 14:56:58 GMT
[View Forum Message](#) <> [Reply to Message](#)

Dear DHS Experts,

I am trying to replicate Table 10.7 Prevalence of diarrhoea in Ethiopia's 2011 report. However, I am having troubles merging PR dataset with KR. I have tried different ways but none of them works. I copied my codes as below.

```
use "/Ethiopia/2011/ETPR61DT/ETPR61FL.DTA", clear
keep hv001 hv002 hvidx hv201 hv205 hv225
rename hv001 v001
rename hv002 v002
rename hvidx b16
sort v001 v002 b16
save "/Ethiopia/2011/ETKR61DT/ETkKR61FLdiarrhoea.dta", replace
```

```
use "/Ethiopia/2011/ETKR61DT/ETKR61FL.DTA", clear
keep v001 v002 b16 v005 v008 b3 b5 h11 h11b v024 v025
sort v001 v002 b16
merge v001 v002 b16 using "/Ethiopia/2011/ETPR61DT/ETPR61FLdiarrhoea.dta"
```

Stata reports that the variables don't uniquely identify observations. If that's the case, which variables should I use instead? Thanks in advance!

Subject: Re: Merging KR with PR for Ethiopia 2011
Posted by [Bridgette-DHS](#) on Thu, 14 Jun 2018 16:02:17 GMT
[View Forum Message](#) <> [Reply to Message](#)

Please reference the response to a similar post: 15192

Subject: Re: Merging KR with PR for Ethiopia 2011
Posted by [kingx025](#) on Mon, 09 Jul 2018 03:04:08 GMT
[View Forum Message](#) <> [Reply to Message](#)

If I am understanding correctly, you want to merge the KR and PR files for Ethiopia 2011 to replicate the categories in Table 10.7 of the final report, with information on diarrhea from the KR files and household characteristics taken from the PR files. A simpler approach, in my opinion, is simply to merge the household record for the child's mother to the KR file. IPUMS-DHS has

already done this for you; for the KR (child) and BR (birth) files, we have merged the mother's record from the IR file and the mother's household record. Thus, you can get all the variables you need if you choose "children" as your unit of analysis and select the following variables in IPUMS-DHS:

Child's age in months: HW1 (KIDAGEMO)

https://www.idhsdata.org/idhs-action/variables/KIDAGEMO#code_s_section

Child's sex: B4 (KIDSEX) https://www.idhsdata.org/idhs-action/variables/KIDSEX#codes_section

HH source of drinking water: HV201 (DRINKWTR):

https://www.idhsdata.org/idhs-action/variables/DRINKWTR#code_s_section

Type of HH toilet facility: HV205 (TOILETTYE):

https://www.idhsdata.org/idhs-action/variables/TOILETTYE#codes_section

Toilet facility shared: HV225 (TOILETSHAREYN);

https://www.idhsdata.org/idhs-action/variables/TOILETSHAREYN#codes_section

Urban-rural residence: V025 (Urban)

https://www.idhsdata.org/idhs-action/variables/URBAN#codes_section

Region of residence: V101_ET2011 (GEO_ET2011):

https://www.idhsdata.org/idhs-action/variables/GEO_ET2011#codes_section

Mother's education: V106 (EDUCLVL):

https://www.idhsdata.org/idhs-action/variables/EDUCLVL#codes_section

Household wealth quintile: V190 (WEALTHQ):

https://www.idhsdata.org/idhs-action/variables/WEALTHQ#codes_section

Child had recent diarrhea: H11 (DIARRECENT):

https://www.idhsdata.org/idhs-action/variables/DIARRECENT#codes_section

Blood in child's stools: H11B (DIABLOODST)

https://www.idhsdata.org/idhs-action/variables/DIABLOODST#codes_section

Weight: V005 (PERWEIGHT) https://www.idhsdata.org/idhs-action/variables/PERWEIGHT#codes_section

When you merge the PR and KR data, you have to drop the cases of children not included in the KR file, anyway. The only possible problem that I see is that some of the background variables (on urban residence, region of residence, and wealth quintiles) use the variable from the IR file, rather than the variable from the linked household file, so you may get some results with "non-resident" as a code in your results, which are not in the table. We're planning on adding the household-based version of those variables to the IPUMS-DHS data for women, children, and births by Fall.

You can make a customized data file from IPUMS-DHS (www.idhsdata.org) by logging in with your DHS user name and password, selecting the above variables and ET2011 as a sample for a customized dataset, selecting your preferred data format, and downloading the customized (free) data file to your computer for analysis. You will have to do some recoding of the variable on type of toilet facility, but that would be the case with the original DHS files as well.

I hope this helps.

Miriam King

Subject: Re: Merging KR with PR for Malawi 2015-16
Posted by [Felixo](#) on Mon, 18 Mar 2019 12:26:39 GMT
[View Forum Message](#) <> [Reply to Message](#)

Dear Expert,

I am working on children malnutrition using Mdhs dataset.

I have been able to merge KR with PR files using PR file as the base file.

This is achieved by

1. convert variables v001, v002 and B16 to HV001, HV002 and HVIDX respectively, in the KR file.
2. sort cases using the key identifier variables HV001, HV002 and HVIDX
3. open the PR file and sort by HV001, HV002 and HVIDX
4. select cases based on $HVIDX \geq 2$ & $HVIDX \leq 17$ & $HV103 = 1$ & $0 \leq HC1 \leq 59$
5. Compute weight variable using $HV005/1000000$; and weigh the reduced dataset PR

With the above, I could obtain the prevalence of the stunting (37.1%), wasting (2.7%) and underweight (11.5%) as it were in the Malawi DHS report.

However, the sample size increases by 20 (from 6,033 to 6,053). Please advise appropriately on what I could do next to avoid this increase.

Thank you for your prompt intervention.

Subject: Re: Merging KR with PR for Malawi 2015-16
Posted by [kingx025](#) on Mon, 18 Mar 2019 17:28:26 GMT
[View Forum Message](#) <> [Reply to Message](#)

I don't know the reason behind the extra 20 cases you encountered after merging; perhaps someone from The DHS Program staff will have an idea about that.

When you pick up an extra 20 cases out of more than 6,000, I really doubt that the validity of your results would be affected. Why do you feel it is essential to get exactly the same number of cases?

Subject: Re: Merging KR with PR for Malawi 2015-16
Posted by [Bridgette-DHS](#) on Thu, 28 Mar 2019 15:41:51 GMT
[View Forum Message](#) <> [Reply to Message](#)

Following is a response from Senior DHS Stata Specialist, Tom Pullum:

Why are you doing this merge? The variables you need are already in the KR file, and you do not need to merge with the PR file to get them. The variables hc70, hc71, and hc72 in the PR file are the same as hw70, hw71, and hw72 in the KR file (or the BR file).

Table 11.1 on page 170 of the main report for this survey gives the % stunted as 37.1%, based on 5707 children. This number comes directly from the PR file, using weights hv005 and restricting

to de facto residence (hv103=1).

After you open the PR file, enter these lines:

```
gen stunted=0
replace stunted=1 if hc70<-200
replace stunted=. if hc70<-600 | hc70>600
tab stunted if hv103==1 [iweight=hv005/1000000]
```

You will get 37.13% stunted, with 5706.83 weighted cases. That is, you match the report this way.

The KR file includes the children of women who are de facto members of the household. I don't know why you have a difference between 6033 and 6053 because I don't know where the 6033 and 6053 themselves come from.