

---

Subject: Merging data sets for DHS 2011 for Nepal  
Posted by [pjoshi](#) on Mon, 06 Oct 2014 13:55:48 GMT  
[View Forum Message](#) <> [Reply to Message](#)

---

I would like to look into the differences in women's status in Nepal based on whether or not their husbands have migrated outside Nepal for work. I have the information on the migrants in the household data set and the women's information (indices for health, education etc.) in the IR data set. I want to merge these and match each individual woman with whether or not the household has a member who migrated abroad for work. I carried out the following steps to do this.

First, I converted the household data set from wide to long so that each migrant member is an observation using the following syntax.

```
reshape long SH22_ SH24_ SH26_ SH27_ SH28C_ SH28D_, i(HHID) j(line)
```

I then created a new dummy variable to separate migrant households (households with a member who has been living outside Nepal and working) from non-migrant households using the following syntax.

```
gen var1=2  
replace var1=1 if SH21==1 & SH27_==1 & SH28C_>=2
```

So, var1=1 for migrant households and var1=2 for non-migrant households

I now want to merge this household level data with individual data for women. So, I need to match each observation in IR data with var1 (whether the woman is from a migrant household or not). So, I sorted both the HR and IR data set by the identifying variables v001, v002 and v003. I then opened the HR dataset as the master data set and the used the following command to merge it to the individual dataset.

```
merge m:m v001 v002 v003 using "the file location for IR data set"
```

I did the many to many merge because for the same household there is multiple migrant members on the HR data set and multiple women in the IR data set. Does this make sense?

Thanks in advance for any suggestions you can make to help me out with this.

---

Subject: Re: Merging data sets for DHS 2011 for Nepal  
Posted by [Trevor-DHS](#) on Mon, 06 Oct 2014 16:31:14 GMT  
[View Forum Message](#) <> [Reply to Message](#)

---

It is hard to tell if this makes sense. There are a number of things that are not clear.

1) You said that you set var1=1 for migrant households and to 2 for non-migrant households, but your unit of analysis at this point is not migrant households, but migrants. You can have two people in the same household with different values. How are you combining this information together for one value for the household?

- 2) You said that you are using v001 v002 and v003 for matching, but these don't exist in the household data, so what are you matching them to? I'm assuming that v001 and v002 are being matched to hv001 and hv002, but I don't know what v003 is being matched to.
- 3) What do you want your final dataset to look like? Do you want migrants as the unit of analysis or women from the IR file as your unit of analysis?
- 4) Using a many to many match, if you have 4 people in your migrant dataset and 3 in your IR dataset, depending on what v001 v002 and v003 are matching to, you can have all possible combinations ( $4 \times 3 = 12$ ), but I don't think this makes sense.

If you want a variable for migrant households then you need to collapse the data from your migrants to data about households (you can use the collapse command with the right parameters to do this). You may need to pay attention to households that had no entries in the migration module.

I think you then want to match that data to the IR file, rather than the IR file to the migrant data, but that depends on what you want your final unit of analysis to be.

I hope this helps.

---

Subject: Re: Merging data sets for DHS 2011 for Nepal

Posted by [pjoshi](#) on Mon, 06 Oct 2014 17:18:58 GMT

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

---

Thanks so much for getting back to me. This information is definitely very helpful.

The main variable I need is a dummy variable at the individual level indicating whether a woman belongs to a household that has migrants working outside Nepal. So, my unit of analysis is women from IR file.

However, information on migration is only available in the household level data set (HR), so in this data set I categorized each household as migrant household or not by defining a new variable named var1.

Var1=1 if SH21==1 (i.e the household has a migrant member) and SH27\_==1 (the member migrated to work) and SH28C>=2 (the member moved outside Nepal).  
and var1=2 otherwise

Now, I want to add this variable (var1) to the individual data set, so I want to match women from each household with this new variable var1.

To merge the two data sets, I renamed hv001 and hv002 as v001 and v002 03 in the HR data set and used v001 and v002 as the identifying variables. I guess I didn't need hv003 for this purpose; thanks for correcting my mistake there.

---

Subject: Re: Merging data sets for DHS 2011 for Nepal

Posted by [Trevor-DHS](#) on Mon, 06 Oct 2014 17:38:42 GMT

I think there is a much easier way of doing this. You don't need to create a file with migrants as your unit of analysis. See the below code:

```
use "NPHR60FL.dta"
* Rename variable to drop 0 to allow for values below to work
rename sh27_0* sh27_*
rename sh28c_0* sh28c_*

* Initialize to code 2
gen var1=2
* Loop through all migrants and set to var1 if any meet the condition.
forvalues x = 1/16 {
  replace var1=1 if sh21==1 & sh27_`x'==1 & sh28c_`x'>=2
}

* rename variables for matching
rename hv001 v001
rename hv002 v002
keep v001 v002 var1
sort v001 v002
save "var1.dta"

* open IR file and merge
use "NPIR60FL.dta", clear
merge m:1 v001 v002 using "var1.dta"
```

---

Subject: Re: Merging data sets for DHS 2011 for Nepal  
Posted by [pjoshi](#) on Mon, 06 Oct 2014 19:11:40 GMT  
[View Forum Message](#) <> [Reply to Message](#)

---

Yes, this is exactly what I wanted. Thanks so much for your help!

---