

---

Subject: Merging Nepal DHS data sets across years  
Posted by [UAB\\_user](#) on Wed, 01 Oct 2014 22:52:30 GMT  
[View Forum Message](#) <> [Reply to Message](#)

---

Hello,

For my current project I am merging all the Nepal DHS files (Household, (individual)Women, (individual)Men, Children, Births, Household member) for the years 1996, 2001, 2006, 2011. I know you have to merge all of the files within a year before you can concatenate each of the years together. However, I have noticed that several variables have the same name within the Nepal DHS files (Household, (individual)Women, (individual)Men, Children, Births, Household member) and some variables which measure the same thing are named differently in separate DHS files. I have also noticed that some of the variables change their names over the years.

Is there a way to merge all of the files together across different years or do I have to choose the variables I need before I merge the data sets together and limit the number of variables in my final dataset.

Thank you  
Derek

---

Subject: Re: Merging Nepal DHS data sets across years  
Posted by [Reduced-For\(u\)m](#) on Fri, 03 Oct 2014 02:33:52 GMT  
[View Forum Message](#) <> [Reply to Message](#)

---

At what level are you trying to "merge" all of these? I mean - what do you want the final unit of observation to be? In some of those recodes it is a person, and in some it is a household. Do you want one observation for each human being that was contacted in any way, and then have their household's information attached to them? Or do you want one observation per household, but need information on all the men, women and children in that household?

Depending on what you want your data to look like, you might be doing some merging and some appending. You should also just be wary of problems with sample selection, weighting, estimation, and inference (p-values) when you combine all of those recodes. I do not believe there is any set procedure for getting consistent estimates when joining together all the different survey types into one dataset, and there is certainly scope for a lot of accidental double-counting or running un-interpretable regressions or comparisons.

---

Subject: Re: Merging Nepal DHS data sets across years  
Posted by [UAB\\_user](#) on Fri, 03 Oct 2014 22:23:05 GMT  
[View Forum Message](#) <> [Reply to Message](#)

---

Hello,

Thank you for helping me with my question. I would like my unit of observation to be the individual (women) and have their household's information attached to them. I will definitely be conscious of accidental double-counting and making sure my estimates are correct. However, I know merging the files correctly is an important step to getting the correct values and I just want to make sure I am doing it correctly.

Currently, for the 2011 files I am merging the "household" file to the "memberhousehold" file using the variables "hv001" and "hv002". Then I merge the "individual women" file using "v001" and "v002". Lastly, I merge the birth file using "v001, v002, v003".

I follow this merging order for the years 96', 01', 06', and 11'. I then concatenate them together into one data set.

Is there a template or set of instructions I can follow (other than what is on the DHS website) to make sure I am merging the files correctly? SAS or STATA code on how to merge the files would be a great help.

Thank you very much!  
Derek

---

Subject: Re: Merging Nepal DHS data sets across years  
Posted by [Trevor-DHS](#) on Sat, 04 Oct 2014 03:58:56 GMT  
[View Forum Message](#) <> [Reply to Message](#)

---

Hi Derek,

I think you are adding some extra complexity that you don't need. The household members (PR) dataset already includes the household variables, so there is no need to merge those. Similarly, merging the birth history data and the women's data is unnecessary as the birth history data has the women's data included.

You say that you want the unit of analysis to be the individual women. In that case you should start with the IR file and the merge the PR file to it. That will give you all of the variables you need with women as the unit of analysis. You can do this using the following variables:

v001 = hv001  
v002 = hv002  
v003 = hvidx

Each of the surveys has different variables included in the datasets, so you will probably have to restrict the variables to a common set. Additionally you will need to check that all of the variables you select are coded in the consistent manner across years.

You also need to look at how to pool data across years. I suggest starting by looking at the following thread: <http://userforum.dhsprogram.com/index.php?t=msg&th=1189&>

amp;goto=2028&S=e803cfb9483c4984bc8aaa934d882b45#msg\_202 8

Finally, why do you want to use a pooled dataset?

---

Subject: Re: Merging Nepal DHS data sets across years

Posted by [UAB\\_user](#) on Fri, 10 Oct 2014 04:04:54 GMT

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

---

Thank you very much for the helpful info. It has made merging the different files much easier. I will definitely check to make sure the variables I use are coded in a consistent manner across years. Also, thank you for the link on how to pool data across years.

The purpose for pooling the different years together was to look at factors influencing infant mortality over several years of the Nepal DHS.

Thank you again

Derek

---