
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 ($4*3=12$), 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.
