
Subject: Merging Women's and Men's File
Posted by [parul](#) on Tue, 16 Aug 2022 05:28:27 GMT
[View Forum Message](#) <> [Reply to Message](#)

Dear DHS User Forum,

Dataset: National Family Health Survey (IAIR7ADT, IAKR7ADT, and IAPR7ADT)
Year: 2019-21
Country: India

For my research objective, I want to identify pairs of siblings (all siblings in one household), mother-child pairs (including all possible children a mother has), father-child pairs (including all possible children a father has), mother-father pairs, and parent-child pairs (including all possible children parent have). Thus, I need information on mothers, fathers, and all their kids in one row.

I have been using the women's file, as it already has all kids a mother has in one row. Thus, only information from men remains to be merged. However, the men's file did not include information on height and weight. So, I merged the person's file and the men's file.

Now I am trying to merge the (men's+person's) file to the women's file. However, I am unable to do so. Although there is an already existing couple's file available in the dataset, it does not contain any information that I need. As one man can have more than one wife, I am using m:1 merging. I have attached the codes that I have been using.

I have already searched the existing queries posted in the DHS User Forum. However, none of them addresses the issue I have. I would request you to please have a look at my query and please suggest a way out. If you need any additional information on the issue, please write back.

Regards,
Parul

File Attachments

1) [DHS.do](#), downloaded 143 times

Subject: Re: Merging Women's and Men's File
Posted by [Janet-DHS](#) on Thu, 01 Sep 2022 20:21:43 GMT
[View Forum Message](#) <> [Reply to Message](#)

Following is a response from DHS Research & Data Analysis Director, Tom Pullum:

Note that the 7A files have been replaced with 7B versions. I recommend that you use the 7B versions. However, that's not really relevant to your question.

What's most relevant is that in these surveys men were subsampled. Only 15% of men who appear in the PR file are in the MR file. The subsampling was done by selecting 30% of clusters and, within the selected clusters, including men in alternate households (in the listing of households in the cluster).

If you want to identify couples, you are restricted to those in the CR file. The criterion for a couple in the CR file is that the woman names the man AND the man names the woman. You could add a few more couples by replacing AND with OR but very few. The reason why the CR file is smaller than you would expect is that the MR file is also small, because of the subsampling.

If you look at just the measurement of height in the PR file, say with "summarize ha3 hb3" you will find 764,455 women and only 124,445 men. Anthropometry for men was only in the 15% subsample of households.

I think your best option for constructing child-parent pairs is to work with the PR file and the variable hv111-hv114, which apply to all children age 0-17. hv112 and hv114 give the line number of the mother and father, respectively, if they are alive and in the household.
