
Subject: Re: Merging and appending data files
Posted by [Janet-DHS](#) on Thu, 21 Jul 2022 16:28:28 GMT
[View Forum Message](#) <> [Reply to Message](#)

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

The IR, MR, and KR files have individual women, men, or children as units. It would be virtually impossible to merge them with the HR file, which has households as units. You should use the PR file, which has individual household members as units, rather than the HR file.

You should do merges survey-by-survey and then append the merged files, rather than doing the appending first and then the merging.

The KR file includes children who are not in the PR file, and the PR file includes children who are not in the KR file. This is the trickiest merge and requires the use of b16 as well as v001 v002 v003.

I hardly ever use caseid or hhid in a merge. It is much easier to use the separate components of caseid, which are v001 v002 v003. There are survey-specific variations in the number of columns in caseid and hhid (which is made up of v001 v002) and in the number of columns assigned to the substrings for v001 v002 v003.

Some surveys in Francophone Africa have a sub-household code that must be used for merges.

Many surveys have survey-specific variables. Carrying them along will greatly increase the file size. Different surveys in the same country will have different coding for many variables, such as stratum, region, source of water, etc. I hope you are taking that into account and reducing the number of standard variables that you will keep.

My main advice would be that you do the merges for specific surveys and then append the surveys for each country. A massive file for all of the surveys done in Africa will be unwieldy.
