
Subject: Re: Merging Individual and HH Member Recode Files
Posted by [Bridgette-DHS](#) on Tue, 16 Aug 2016 16:13:30 GMT
[View Forum Message](#) <> [Reply to Message](#)

Following is a response from Senior DHS Stata Specialist, Tom Pullum:

I'm sorry that the previous responses did not resolve your problem. The Peru Continuous Survey is now operating almost completely independently of DHS.

It is clear that hv002a is not what you need. It is an empty variable, and why it is even included in the data is a mystery. I still have no idea what is the name of the variable that identifies sub-households.

As a kind of last resort, I have checked the household id variable, hhid. This is a character string that is usually made up from hv001 and hv002. In this survey, however, there are two additional characters at the end. Those last two characters appear to be the missing variable. It is possible to extract them as a separate variable.

If you do the following lines, you will extract that variable, which I call "final2":

```
* use PEPR6IFL.dta, clear
list hhid hv001 hv002 hvidx if _n<=50, table clean
* the last two characters of hhid must be extracted
gen str2 final2=substr(hhid,8,2)
destring final2, replace
tab final2
```

This gives you the variable you need in the HR or PR file.

To get it in the IR or KR file, you need to extract from caseid, as follows:

```
* use PEPR6IFL.dta, clear
list caseid v001 v002 v003 if _n<=50, table clean
* the last two characters of hhid must be extracted
gen str2 final2=substr(caseid,11,2)
destring final2, replace
tab final2
```

The substring command goes to columns 8-9 for hhid and 11-12 for caseid. You can identify those positions by trial and error.

After you have constructed "final2" or whatever you want to call it in every file, you would include it in the sort and merge and I think everything will be fine.

There are other ways to do this, even simpler, but possibly confusing, so I will just recommend this strategy.
