

---

Subject: Re: Issue merging NFHS 2 women and household dataset

Posted by [Bridgette-DHS](#) on Thu, 01 Feb 2018 12:13:31 GMT

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

---

Another response from Tom Pullum:

The lines I gave yesterday will work for almost every DHS survey. However, I see that for this survey of India the id code includes another number. I saw this by comparing hhid in the PR file with caseid in the IR file. If you check you will find that hhid is a 12-character string and caseid is a 15-character string consisting of hhid in columns 1-12 and v003 in columns 13-15. I answered a related question on the forum yesterday regarding the Mali 2006 survey. In the case of Mali 2006 and some other surveys in West Africa, there is a sub-household code embedded in hhid and caseid. For this India survey, the extra code is the state, not the sub-household, but the strategy is basically the same--that is, to match caseid in the IR file with hhid hvidx in the PR file.

As an alternative, since state is given by v024, you could match v024 v001 v002 v003 in the IR file with hv024 hv001 hv002 hvidx in the PR file.

set more off

\* Prepare IR file for merge

use e:\DHS\DHS\_data\IR\_files\IAIR42FL.dta, clear

gen hhid=substr(caseid,1,12)

gen hvidx=v003

sort hhid hvidx

save e:\DHS\DHS\_data\scratch\IAIRtemp.dta, replace

\* Prepare PR file for merge

use e:\DHS\DHS\_data\PR\_files\IAPR42FL.dta, clear

sort hhid hvidx

\* Merge IR with PR

merge hhid hvidx using e:\DHS\DHS\_data\scratch\IAIRtemp.dta

tab \_merge