
Subject: Re: Merging KR file with PR file India DHS 2019-21 (NFHS-5)

Posted by [Bridgette-DHS](#) on Mon, 14 Aug 2023 14:55:52 GMT

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

Following is a response from Senior DHS staff member, Tom Pullum:

We apologize for the delay in this response, due to travel. I have slightly revised your code, as follows:

* specify workspace

```
cd e:\DHS\DHS_data\scratch
```

```
use "...IAPR7EFL.DTA", clear
keep hv001 hv002 hvidx hv024
gen cluster=hv001
gen hh=hv002
gen chline=hvidx
gen state=hv024
sort state cluster hh chline
save IAPR_temp.dta, replace
```

```
use "...IAKR7EFL.DTA", clear
keep v001 v002 b16 v024
keep if b16>0 & b16<.
gen cluster=v001
gen hh=v002
gen chline=b16
gen state=v024
sort state cluster hh chline
```

* At this point a merge will fail, because of a few duplicated values of b16 in the KR file.

* Identify the duplicates.

```
egen repeated_b16=seq(),by(cluster hh chline)
tab repeated_b16
```

* There are 7 households in which the same value of b16 appears twice

* Remove the second case in duplicates. Note: it is possible that in some households it is the

* first, not the second, that should be removed. Values of b16 could be edited.

```
list if repeated_b16>1, table clean
```

```
drop if repeated_b16>1
```

```
drop repeated_b16
```

* sort again

```
sort state cluster hh chline
```

* Note that this is a 1:1 merge

```
merge 1:1 state cluster hh chline using IAPR_temp.dta
```

tab _merge

- * There is one case with _merge=1. This is a child in the KR file with a valid value of b16 but
- * the value does not appear in the PR file. Should not happen; drop this case but could edit b16.

list if _merge==1, table clean
drop if _merge==1

There are 3 issues. First, the KR file includes 7 households in which the same value of b16 is repeated for different children. I simply drop the repeats. Ideally, someone could probably figure out a better solution for those 7 cases but there are nearly 222,000 children and the effect is negligible. Second, the KR file includes one child who has a valid value of b16 but that line number is not found in the household in the PR file. Again, I just drop the case but there are other options. Third, you do this as a m:m merge but it should be 1:1. The children with _merge=3 will be matched in the two files.
