
Subject: Mismatch in the partner's age given in the IR file vs MR file

Posted by [Navya](#) on Thu, 11 Jul 2024 10:07:52 GMT

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

Hi,

I want to add a few husband/partner characteristics from the MR file to the IR file for the survey waves 2005-06, 2015-16, and 2019-21. For this, I first created a men's file with the indicators I needed, then I duplicated observations based on the number of unique wives' line numbers given for each respondent in the men's files, after which I created a single indicator for the wife's line number (replacing mv034_x with a single mv034 variable), renamed the variables and did a 1:1 merge. I removed all the line numbers for which the wife was not in the household (since the household number could not be matched). Here is the code I used to create the men's dataset for 2019-21:

```
** identification - husband
```

```
rename mv000 v000
```

```
rename mv001 v001
```

```
rename mv002 v002
```

```
rename mv003 v034
```

```
** only keep partners who are currently married
```

```
keep if mv501 == 1
```

```
sort v001 v002 v034
```

```
** missing if wife is not in household (cannot match household code with individual file if wife is from a different household)
```

```
forvalues i = 1(1)8 {
```

```
  replace mv034_`i' = . if mv034_`i' == 0
```

```
}
```

```
** identification - wife
```

```
** drop if all wife line numbers are missing (cannot match with no line number)
```

```
drop if mv034_1 == . & mv034_2 == . & mv034_3 == . & mv034_4 == . & mv034_5 == . &
```

```
mv034_6 == . & mv034_7 == . & mv034_8 == .
```

```
save temp.dta, replace
```

```
** append datasets based on line numbers to create woman-level dataset
```

```
forvalues i = 1(1)8 {
```

```
  use temp, clear
```

```
  rename mv034_`i' v003
```

```
  drop mv034_*
```

```
  order v003, after(v002)
```

```
  drop if v003 == .
```

```
save temp`i`.dta, replace
}

** append datasets
use temp1, clear

append using temp2
append using temp3
append using temp4
append using temp5
append using temp6
append using temp7
append using temp8

sort v000 v001 v002 v003 v034

save hus_2020.dta, replace
```

I then used the following code to merge with the IR file:

```
use "2020/IAIR7EDT/IAIR7EFL.dta", clear

** merge with husband data
merge 1:1 v001 v002 v003 v034 using hus_2020
drop if _merge == 2
```

Since the number of observations for each male respondent matched the number of wives listed, the 1:1 merge using the MR file (with the IR file as the master data) on the cluster number ((m)v001 in both files) household number ((m)v002 in both files), respondent's line number (v003 in the IR file, and renamed from mv034_x to v003 in the MR file), and line number of husband (v034 in the IR file, and renamed from mv003 to v034 in the MR file) worked. However, I was trying to match the partner's age given in the IR file (v730) with the merged age variable from the MR file (mv012) to check if the merge was correct, and I was not able to do so. The age for around 22.7% of the partners in the MR file that were successfully matched with the IR file (for which `_merge == 3`) did not match with the partner's age given in the IR file. I am also getting a mismatch in husband's employment data, since husband/partner's occupation (v705 in IR file) is showing that all husbands in the IR file are working, but the merged indicator from the MR file (mv717) shows that 2.84% of the husbands are not working. I was wondering if I'm doing something wrong here.

Thank you!

Subject: Re: Mismatch in the partner's age given in the IR file vs MR file

Posted by [Bridgette-DHS](#) on Thu, 11 Jul 2024 15:40:34 GMT

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

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

I believe you did the merge correctly, but it would have been simpler to use the CR files for these surveys. They already have matched the partners using the same criteria. I opened the CR file for the NFHS-5 (IACR7EFL.dta) and then entered "gen delta=v730-mv012". I entered "tab delta" to get the unweighted distribution of delta. delta is 0 for 77% of the cases; +/-1 for 12%, and +/- 2 for 4%. The remaining cases are spread over a very wide range, from -37 to +66!

The differences indicate reporting errors, and they reflect on the quality of all the age data. Reports of age in many surveys, especially in South Asia, show high levels of heaping at multiples of 5, and that alone can shift a true age by one or two years. Very large differences are difficult to explain, but fortunately there are very few of them. The great majority of these pairs of ages are within 2 years of each other.

Subject: Re: Mismatch in the partner's age given in the IR file vs MR file

Posted by [Navya](#) on Fri, 12 Jul 2024 05:34:25 GMT

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

Dear Tom,

Thank you for your reply, this is very helpful. For indicators which show large differences, like husband/partner's occupation (v705 in the IR file vs. mv717 in the MR file giving male respondent occupation) where v705 shows that all men in couples are working, but mv717 shows that 2.84% of the men in couples are not working, which indicator should I be using for my analysis?

Thank you so much for your help!

Subject: Re: Mismatch in the partner's age given in the IR file vs MR file

Posted by [Bridgette-DHS](#) on Fri, 12 Jul 2024 12:46:15 GMT

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

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

The "mv" variables come from the men's questionnaire and describe the responses given by the men themselves. We would consider them to be more accurate. Similarly, women's responses about themselves (such as their age) would be considered to be more accurate than responses from their partner or from the household respondent (in the household survey). Sometimes there is an interest in the partner's perception, but self-reports are usually (maybe not always!) more accurate.
