# Understanding Society User Support - Support #692

## merging _indresp and _child across all waves

01/03/2017 05:37 PM - Nico Ochmann

| | | | | |
|---|---|---|---|---|
| **Status:** | Closed | | **Start date:** | 01/03/2017 |
| **Priority:** | Normal | | **% Done:** | 100% |
| **Assignee:** | Nico Ochmann | | | |
| **Category:** | Data analysis | | | |

**Description**

Dear Ladies and Gentlemen,

I hope this question is USoc-specific enough to post it here. With Stata, I would like to merge the  files _indresp and _child across all six waves. In the end, I would like to have the age and sex of the children, the biological father, his wage etc. in one panel data set (long format). So far, I have created two separate panel data files, the master file with numerous selected variables from _indresp and the using file from _child with a number of selected variables. So I have these two panel sets and I want to merge them. I went ahead and ran

merge 1:1 pidp wave using ........, gen(....). In the master set, I have roughtly 290 000 observations and in the using set 9 000. The number of matches were zero, though. My guess is that I cannot use pidp as common variable for both sets, but I am not sure which one to use. If you could give me a hint as to how to set this merging procedure up, I would highly appreciate that.

**History**

**#1 - 01/11/2017 11:43 AM - Alita Nandi**

*- Status changed from New to In Progress*

*- Assignee changed from Alita Nandi to Nico Ochmann*

*- % Done changed from 0 to 90*

Dear Nico,

The pidp in w_child file is the pidp of the child (< 16 year olds). The pidp in w_indresp is the pidp of the adult respondents (16+ year olds). So, if you merge these two files on pidp they will never match. What you need to do is identify the PIDP of the parents of the children in w_child: w_fnspid w_mnspid in w_indall. Then rename these variables to PIDP and then merge that with w_indresp using PIDP. Note you will have to do this separately for the father and mother.

Best wishes,
Alita

**#2 - 01/11/2017 11:44 AM - Alita Nandi**

*- Private changed from Yes to No*

**#3 - 01/11/2017 08:27 PM - Nico Ochmann**

Thank you very much once again!

Just as a note, it seems to me that I need to merge these two files using PIDP and wave.

Cheers!

Nico

**#4 - 01/16/2017 10:46 AM - Alita Nandi**

You can either
(i) do the merge exercise separately for each wave, save the final file as a wave specific file and then append them, or
(ii) produce a long file for each type of file and then merge them together - using pidp and wave

**#5 - 01/23/2017 01:00 PM - Nico Ochmann**

Dear Dr Nandi,

I do have a short follow up question before you close this one. In the cild file I have 90,000 obs. I end up with  only 60,000 matches when I merge the files.
I cannot see why I would have such a large number of mismatches. If you have any ideas, I would appreciate a short comment.
Thanks a lot!

Nico

```
Result                           # of obs.
    ---------------------------------------
    not matched                      289,433
         from master                 259,370  (_wemerge==1)
         from using                   30,063  (_wemerge==2)

matched                              60,664  (_wemerge==3)
    ---------------------------------------
```

global Stata11_se "F:\UnderstandingSocietyData\Data\stata\stata11_se"
global data "P:\Mergeddata_master2016_religion"
global dofiles "P:\Dofiles_master2016_religion"
global logfiles "P:\Logfiles_master2016_religion"

_____

foreach w in a b c d e f{

```
use `w'_ukborn pidp fpid mpid `w'_hidp `w'_pno `w'_istrtdaty `w'_plbornc `w'_jbhrs `w'_qfhigh `w'_qualoc `w'_s
ex `w'_dvage `w'_marstat `w'_fimngrs_dv `w'_fimnlabgrs_dv `w'_jbstat `w'_plbornc_all `w'_pacob_all `w'_macob_a
ll ///
         `w'_oprlg1 `w'_oprlg0 ///
         `w'_paygu_dv `w'_oprlg `w'_yr2uk4  `w'_birthy `w'_feend `w'_scend `w'_jbterm1 `w'_jbsize `w'_jbsect `w
'_jbsectpub `w'_racel `w'_jbbgy `w'_sf1 `w'_jbmngr `w'_lnprnt `w'_jbsemp ///
         `w'_gor_dv `w'_urban_dv `w'_jshrs ///
         using "$Stata11_se/`w'_indresp_protect", clear

gen wave = strpos("abcdef","`w'")

renpfix `w'_

save $data/`w'wave, replace
     }
foreach w in a b c d e {

append using $data/`w'wave.dta
     }
sort wave pidp
rename pidp PIDP
rename dvage age
save $data/abcdef_long, replace
```

foreach w in a b c d e f{
erase $data/`w'wave.dta
}

foreach w in a b c d e f{

```
use pidp `w'_hidp `w'_sex `w'_dvage `w'_hgbiom `w'_hgbiof `w'_adresp15 `w'_birthy `w'_birthm  fpid mpid  `w'_f
npno `w'_fnspid `w'_mnpno `w'_mnspid ///
         using "$Stata11_se/`w'_child_protect", clear

gen wave = strpos("abcdef","`w'")

renpfix `w'_

save $data/`w'wave, replace
     }
foreach w in a b c d e {

append using $data/`w'wave.dta
     }
sort wave pidp
rename fnspid PIDP
rename dvage childage
rename sex childgender
save $data/abcdef_long_2, replace
```

foreach w in  a b c d e f{
erase $data/`w'wave.dta
}

use $data/abcdef_long, clear

merge 1:m PIDP wave using $data/abcdef_long_2, generate (_wemerge)

**#6 - 01/30/2017 06:38 PM - Alita Nandi**

A quick look at your syntax suggests that you have only matched with fathers and so are missing all children living in households without fathers.

Best wishes,
Alita

**#7 - 02/13/2017 03:32 PM - Victoria Nolan**

*- Status changed from In Progress to Closed*

*- % Done changed from 90 to 100*