

Understanding Society User Support - Support #1241

Problem merging indersp with children data

09/09/2019 07:52 PM - Marina Fernandez Reino

Status:	Resolved	Start date:	09/09/2019
Priority:	High	% Done:	90%
Assignee:	Marina Fernandez Reino		
Category:	Youth		

Description

Hi,

My objective is to construct a dataset that includes all household members, including children aged 0 to 15. I want to examine certain youth outcomes in migrant families. In order to do that, I need to know the country of birth of the natural/adoptive/step mother and natural/adoptive/step father living in the same household as the child. Note that, in this case, I am just going to focus on the co-resident parent(s) regardless of whether they are the natural parents of the child. I am aware that one of the natural parents of child might not reside with the child.

I'm going to detail the steps I followed so it is easier to understand the problem I have at the moment.

- Firstly I've merged the child and youth dataset. I saved all the variables of interest with capital letters except h_hidp and h_pno.
- Secondly, I've created couple-level variables in the indersp data file. These variables summarise the country of birth combinations of the partners within a couple, e.g. couples where both partners are UK born. I've also generated a variable identifying the country of birth of single adults.
- Thirdly, I've merged the new indersp dataset with the indall dataset.
- Fourthly, I've merged the resulting file from above with the children and youth dataset merged in step 1.

The resulting file includes all household members. I've generated a couple of variables to identify who is the responsible adult of children aged 0 to 15, so I know if the responsible adult is the mother other family member. The code reads as follows:

```
gen motherguardian=1 if (H_ADRESP15_DV==H_MNSPNO) & H_ADRESP15_DV!=. // Adoptive/Natural/Step mother is the
responsible adult
gen fatherguardian=1 if (H_ADRESP15_DV==H_FNSPNO) & H_ADRESP15_DV!=. // Adoptive/Natural/Step father is the
responsible adult
gen grandpguardian=1 if ((H_ADRESP15_DV==H_GRFPN0) | (H_ADRESP15_DV==H_GRMPNO)) & H_ADRESP15_DV!=.
// Grandparent is the responsible adult
gen parentguardian=1 if motherguardian==1 | fatherguardian==1
label var parentguardian "Guardian is nat/step/adopt parent"

gen guardian=1 if motherguardian==1 | fatherguardian==1
replace guardian=2 if grandpguardian==1
replace guardian=4 if H_ADRESP15_DV==8 | H_ADRESP15_DV==9
replace guardian=3 if H_ADRESP15_DV>=1 & H_ADRESP15_DV<=12 & guardian==.
lab define guardian ///
1 "resp.adult is mother or father" ///
2 " resp.adult is grandparent" ///
3 " resp.adult who is not parent/grandparent" ///
4 " resp.adult missing", replace
lab val guardian guardian
```

Then I want to identify the country of birth of children's responsible adults as well as their partners. I've already generated this variable in indersp, so I have this information. The problem I have is that there are some households with 2+ children where there is more than 1 responsible adult for those children. I've included below a fictitious example of such household (**file attached**). In this household there are 2 children (h_pno 4 and 5). Each children has different parents and hence different responsible adults. The mother of children 4 is UK born and her partner is also UK born. The mother of children 5 is Indian born and she is single.

I want to use a dataset only with children, but I need to have information about the country of birth of co-resident parents and their couples in such dataset. The problem is that I don't know how to attach the co-resident parents' country of birth to each child when there are more than 1 responsible adults in the household (like in the example below). This is mainly because the variable h_childpno in indersp is missing instead of indicating the pno of the child. This is actually surprising because I thought that h_childpno in indersp should take the value of the child the adult is responsible for.

History

#1 - 09/10/2019 06:57 PM - Gundi Knies

- Assignee changed from Gundi Knies to Marina Fernandez Reino
- % Done changed from 0 to 50
- Private changed from Yes to No

Hi Marina,

this is quite a complex data merge and, in general, I think it is best (as in: least error prone) to keep the different data components separate for as long as possible.

I would recommend to first get a file that contains all pidp's country of birth (COB). Also create look-up files between pidp and the pidp of significant others of interest (biological mother, biological father, spouse/partner). Attach the significant others' COB (and any other info you are interested in) to each look-up file. Then merge it all together so the info refers to each pidps' biological mother / biological father. Only then merge it all with the child or youth data.

So, in a little more detail:

1. Get your pidp-COB lookup from the XWAVEDAT file (or the bw_indresp or w_indresp data files).
2. Pointers to significant others such as the biological parent (*mnpid and *fnpid) are routinely provided in the indall data files. Create a long format data file that contains all pidp mnpid wave. Drop cases where mnpid equals -8 - no biological mother in household. Keep one observation per pidp (note that there will be some pidp who have more than one biological mother - obviously an error - that you need to correct). This is the pidp-biological mother lookup to which you can attach additional info about the mother. At this stage, there are only two variables: pidp mnpid.
3. To attach the mother's COB, temporarily rename mnpid to pidp (requires renaming pidp to something else first). Merge in the country of birth using pidp-COB lookup created in step 1. Keep only those with merge==3. When all required information about the mother has been added, undo the pidp and mnpid renames so pidp refers to the respondent and mnpid to their biological mother. Give the data relating to the mother a meaningful prefix. Save the data.
4. Repeat steps 2-3 for biological fathers.
5. To create the pidp-partner lookup wave by wave first because the identity of the partner may change over time (for valid reasons!). Use pidp and *ppid provided in the indall files. Drop cases where *ppid =-8 (no co-resident spouse or partner). Temporarily rename *ppid to pidp. Merge in the current partner's COB using the file pidp-COB lookup created in step 1. Kepp all cases in the pidp-partner lookup only. Undo the pidp and *ppid renames and give variables a meaningful prefix. Generate a wave identifier and remove wave prefixes. Save wave specific file and append to a long format. Then re-shape into wide format. The resulting data set is at the pidp level and has 1+ number of waves*2 variables: pidp ppid1 ppid2 .. ppid26 pp_cob1 pp_cob2 .. pp_cob26.
6. To attach the biological mothers's partners and their COB to the pidp-biological mother lookup file, rename pidp to mnpid. Merge with the data created in step 3. Drop cases who are only present in the pidp-partner lookup. Save.
7. To attach the biological father's partners and their COB to the pidp-biological father lookup file, rename pidp to fnpid. Merge with the data created in step 4. Drop cases who are only present in the pidp-partner lookup. Save.

Now you have a biological father file and a biological mother file at pidp-level and you can merge it to your youth data file using pidp.

Hope this helps!
Gundi

#2 - 09/12/2019 11:21 AM - Marina Fernandez Reino

Thanks for your detailed explanation, Gundi

#3 - 09/12/2019 05:20 PM - Marina Fernandez Reino

Hi Gundi,

I just wanted to make sure I am doing this OK.
These are the steps I've followed:

*Get your pidp-COB lookup from the XWAVEDAT file (or the bw_indresp or w_indresp data files). Pointers to significant others such as the biological parent (*mnpid and *fnpid) are routinely provided in the indall data files. Create a long format data file that contains all pidp mnpid wave.*

I've created a new data file which is the merged of the following files: xwavedat, indall and indresp. This file contains information of the responding and non-responding household members. I kept the variables pidp mnspid fnspid and a country of birth variable that I generated. Each row corresponds to one person, there are no duplicates. This file is saved as "pidpcob.dta"

Drop cases where mnpid equals -8 - no biological mother in household.

I created another data file where I only kept pidp mnspid and the country of birth variable. I dropped all individuals who are not living with their mother in the same household (drop if mnspid==8). I saved this file as "pidp_mnspid.dta".

```
drop if h_mnspid==8
rename pidp pidp_orig
rename indh_mnspid pidp
save "mnspid.dta", replace
```

Keep one observation per pidp (note that there will be some pidp who have more than one biological mother - obviously an error - that you need to

correct). This is the pidp-biological mother lookup to which you can attach additional info about the mother. At this stage, there are only two variables: pidp mnpid. To attach the mother's COB, temporarily rename mnpid to pidp (requires renaming pidp to something else first).

I don't see any duplicates. (I run duplicates report pidp)
Using the file "pidp_mnspid.dta", I rename the pidp identifier to pidp_original and I rename mnspid to pidp. I saved this data.

Merge in the country of birth using pidp-COB lookup created in step 1. Keep only those with merge==3. When all required information about the mother has been added, undo the pidp and mnpid renames so pidp refers to the respondent and mnpid to their biological mother. Give the data relating to the mother a meaningful prefix. Save the data.

```
use "pidp_mnspid.dta", replace
merge m:1 pidp using "pidpcob.dta" // Matched: 19,009 (mothers).
rename pidp h_mnspid
rename pidp_orig pidp
rename cob_det cob_detmnspid
rename cob cobmnspid
keep if _merge==3
```

Am I proceeding correctly?
Thanks for your help

#4 - 09/13/2019 10:07 AM - Gundi Knies

Hi Marina,

I am afraid something has gone wrong in your data merge as there really are respondents who have a different mother in different waves.

Here is how you can see it (following the procedure I suggested for spouses - step 5):

```
// Lookup file to mnspids
foreach x in a b c d e f g h {
use pidp `x'_mnspid using ${ukhls}/`x'_indall if `x'_mnspid!=-8, clear
gen wave=strpos("abcdefgh",`x'")
rename `x'_* *
save ${mydat}/pidp_mnspid_lookup_`x', replace
}
use ${mydat}/pidp_mnspid_lookup_a, clear
foreach x in b c d e f g h {
append using ${mydat}/pidp_mnspid_lookup_`x'
}
}
```

```
// Are there pidp who have different mnspids across time?
bys pidp (wave): gen index=_n
xtset pidp index
gen dif_mns_pidp=D.mnspid
fre dif_mns_pidp
bys pidp: egen mns_anydif=sum(dif_mns_pidp!=0 & dif_mns_pidp!=.)
lab var mns_anydif "Any change in mnspid over time?"
lab def mns_anydif 0"no", replace
lab val mns_anydif mns_anydif
fre mns_anydif
```

```
// Yes! And here is what the histories look like:
list pidp index wave mnspid dif mns_anydif if mns_anydif >0, sepby(pidp) noobs
```

```
// Save the PIDP-MNSPID-Lookup (long) to add the mothers' COB info to it.
drop index dif_mns_pidp
save ${mydat}/pidp_mnspid_lookup, replace
```

(add mothers' info). At some stage you want to re-shape this data to a wide format to merge with the youth file:

```
// Re-shape so there is just one row per PIDP
xtset pidp wave
reshape wide mnspid, i(pidp) j(wave)
describe
// PIDP- with mnspid for each wave (and any other info about mothers such as .... mns_cob1 etc.)
save ${mydat}/pidp_mnspid_lookup_wide, replace
```

Hope this helps,
Gundi

#5 - 09/13/2019 05:19 PM - Marina Fernandez Reino

Thanks, Gundi. I think the confusion is because you thought I was using all waves, but I am only using the last wave.
Marina

#6 - 09/16/2019 03:48 PM - Gundi Knies

Hi Marina,

I see - sorry for that. If it is just one wave your code seems to be fine. The h_indall file has 19,009 cases with mnspid~=8.

Best wishes,
Gundi

#7 - 09/18/2019 01:20 PM - Marina Fernandez Reino

Dear Gundi,

Thanks for your help. I have now a file with all the children below age 16 (1 child per row) and I have also the variables indicating the country of birth of their natural/step/adoptive mother and father. I have created a variable that combines both information about parents' country of birth. I have also rely on information about the parents' interview outcome and parents' de facto marital status to understand why sometimes the information about the mother/father country of birth is missing. The main problem I have now is that I get inconsistent results when I crosstab the variable that I've generated (about parents' country of birth) with h_fborn, which is a household-level variable indicating if there is somebody in the household that was born abroad. I find that, for example, some households reported that nobody is born abroad but then I know that the parents of 148 children living in those households were both born in a non-EU country. Are these discrepancies normal? Should I use h_fborn at all?

Thanks

Marina

#8 - 09/18/2019 06:39 PM - Gundi Knies

- % Done changed from 50 to 90

Hi Marina,

great to hear that you have managed to get the data merge together the way you intended!

As to the inconsistency with the variable h_fborn, as you can see in the questionnaire, this variables has been routed on information from the previous wave. As the household interview is typically recorded before the individual interviews, the information about the COB may not have been available then, or the individual may not have been a sample member previously, or not be an OSM. There are many potential sources of discrepancy. The fborn flag seems to be serving a particular purpose in routing questions and it may well not be capturing the construct you are after. Only you can decide whether it is a variable you want to / need to consider.

Here is the note defining the fborn flag:

Compute FBORN = Flag To Indicate Foreign Born Sample Using Ff_Fborn. Foreign Born OSMs As Of Wave 6 And All Household Members Co-Resident At The Current Wave. All Household Members Are Flagged Only If At Least One ADULT (Aged16+) Household Member Has Ff_Fborn = 1. Household Members Are Not Flagged If All ADULTS (Aged 16+) Have Ff_Fborn = 0.;

Best wishes,
Gundi

#9 - 07/15/2020 03:37 AM - Alita Nandi

- Status changed from New to Resolved

Files

question for forum.docx	15 KB	09/09/2019	Marina Fernandez Reino
-------------------------	-------	------------	------------------------