

Understanding Society User Support - Support #155

Number of biological children in the household (nmatch)

06/06/2013 12:44 AM - Ben Wilson

Status:	Closed	Start date:	06/06/2013
Priority:	Normal	% Done:	100%
Assignee:	Ben Wilson		
Category:	Data analysis		

Description

Hi there, apologies if this is a known issue (I did search the FAQ & all open and closed issues), but I think I have found an issue with the variable `a_nmatch` for about 350 cases. I may be wrong, but it seems that this variable does not correctly calculate the total number of resident biological children for some household members. If this is a new issue, then I guess it is probably best for me to send syntax to explain and/or so you can replicate (I can't attach anything to this web form). Of course, please let me know if you want this (email: b.m.wilson@lse.ac.uk).

To give a little more info of what I did to discover this, a brief summary is: In order to obtain ages of all resident biological children, I did some matching, separately by sex (of parents), using a combination of the various id variables to match all resident biological children (taken from indall and isolated using `mnpno/fnpno`) to all parents (taken from indall and using their `pidp` to match to the `mpid/fpid` of resident biological children). I then derive my own version of `nmatch` by counting non-missing values of the ages of resident biological children. My final dataset includes all household members (from indall) and I can see that `nmatch` is miscounting biological children (assuming that the `mnpno/fnpno/mpid/fpid` variables are correct).

History

#1 - 06/06/2013 08:57 AM - Redmine Admin

- Status changed from New to In Progress

- % Done changed from 0 to 50

Ben,
Please see [#141](#)

Jakob

#2 - 06/10/2013 01:25 PM - Ben Wilson

Jakob

I've spent quite a bit of time understanding the differences between the various fertility variables, so am fairly happy with most of [#141](#). The relevant parts seem to be the statement that (a) there were "quite a number of errors in collecting the relationship grid data", and (b) that there are corrected variables at the end of the dataset.

On (a), are you saying that we know `a_nmatch` is incorrect? And what about the fertility history file; are we happy that this contains all the natural children resident in each parent's household (it appears that this is not the case).

On (b), I can only see one variable relating to counts of children (`a_nchild_dv`), which does not measure resident biological children only. Did I miss something?

Thanks for your help!
Ben

#3 - 06/10/2013 01:40 PM - Redmine Admin

[#141](#) (note 5) is the best advice we can give at the moment. The relationship pointers mentioned can be used to distinguish different types of parent-child relationships (adoptive, foster, biological).

The inconsistencies in `a_nmatch` and `a_relationship` are flagged up in the online documentation system. We will also be reviewing the script computed variables for the next revision.

Jakob

#4 - 06/20/2013 06:57 PM - Ben Wilson

Jakob

Thanks for your reply.

I have done some more work on this, and read everything I could find in the documentation. Unfortunately, there is still a substantial issue with the

fertility history data, which is not discussed anywhere (that I can see). I appreciate that you may not be able to (or have time to) help further (sorry for asking if so), but I'm wondering if you (or someone else) can give any more guidance relating to the following questions:

Q1: For resident biological children, should we trust (a) counts based on the fertility histories (natchild), or (b) counts based on information from the enumeration (indall). They give very different answers.

Q2: Is there any view on how we should correct errors in the fertility history (natchild) file? For example, each child in that file should only appear once per adult, but there are a considerable number of duplicates - (i.e. duplicate lchno's for a given mum or dad). The below code gives an example of this showing men and women who have more than one child with person number (lchno) equal to 3.

```
use "$dir1\a_natchild", clear
renpfix a_
sort hidp pno
bys hidp pno: egen test=total(lchno==3)
ta test, m
keep if test>=2
br
```

Q3: If you cannot provide any guidance with either of these, do you have any idea when this might change? (you mention a review for the next release)

Grateful for any help
Thanks
Ben

#5 - 07/05/2013 10:41 AM - Gundi Knies

Hi Ben
as you know from the study documentation and user guide there were some problems with the collection of the household grid. The issues you describe all relate to this. There was a higher chance for members of larger households to have coding errors in their relationships to each other. In addition, as you can see in the questionnaire (check the sequence of questions and the universe!), the household and relationship grid are reported by one member of the household only. There is of course some scope for reporting error in this as this person may not actually know for all the children in the household whether they are a particular person's biological or step child, or whether the children are each others' half-, step- or biological siblings.
The household reference person's account will be used to derive flags for each member of the household which decide over which questions are addressed to them in their individual interview. Now, if the household grid suggested that somebody has 3 natural children but they were in fact not. It could be that the interviewer has then picked a random child from the suggested list of children in the households three times to get out of the loop over # of (suggested) natural co-resident children. The respondent would then appear to have three natural children who all have the same _lchno. To make things yet a little more complicated, the interviewer would also have gone back to the household grid to correct the information initially provided in the grid. However, _nmatch etc would not be re-calculated unless the interviewer went through the whole household grid again (again: the question sequence and careful reading of the universe instructions will give you hints!). This is an iterative process where some errors may and others may not have been picked up in the interviews, and it is difficult to work out which information to trust more.

You will have to sit down and work out where each piece of information is likely to come from, and possibly also consider who was interviewed first. Personally, I would tend to trust reports from the respondents' themselves more than reports from another member of the household. You may also want to consider that some natural children may in fact be step children and their information may therefore be stored in a different data file. I would suspect that some of the duplicates will have -8 on the birthweight (or other variables) while others don't; this could also help you identify which children were misclassified in the household grid and the interviewer just filled in -8 to get out of the inapplicable loop. It may also help to use the edited relationship variable as a reference, if only to see which households had known problems in the reported relationships.

Gundi

#6 - 07/11/2013 11:46 AM - Redmine Admin

- Status changed from *In Progress* to *Closed*
- % Done changed from 50 to 100