Understanding Society User Support - Support #2250

calculation of number of siblings

05/07/2025 11:02 AM - SHUQI LYU

Status:	Feedback		**Start date:**	05/07/2025
Priority:	Normal		**% Done:**	50%
Assignee:	Understanding Society User Support Team			
Category:	Data documentation			

Description

Hi, I plan to calculate "number of biological siblings" and "number of biological/step/adopt siblings" of each respondent.

I found these relevant variables in the UKHLS:
nnsib_dv (Number of R's natural siblings in hh) in w1-w14
nnssib_dv (Number of R's nat/step/adopt siblings in hh) in w1-w14
nrelsw12 (number of absent brothers or sisters) in w1
nrels2 (number of absent brothers or sisters) in w3,w5
nrels4 (number of absent brothers or sisters) in w7,w9,w11,w13,w14
bsbx_N   total no. of biological siblings
ssbx_N   total no. of step-siblings
asbx_N   total no. of adopted siblings

My questions are:
(1) I use nnsib_dv to calculate number of biological siblings in household, how do I calculate number of biological siblings not in hh?
(2) variable bsbx_N means siblings in hh or total number of siblings (in hh + not in hh)?
(3) I want to calculate number of biological/step/adopt siblings, should I use the sum of nnssib_dv + nrelsw12/nrels2/nrels4?

Thanks.

History

#1 - 05/08/2025 02:25 PM - Understanding Society User Support Team

- Category set to Data documentation

- Status changed from New to Feedback

- % Done changed from 0 to 50

- Private changed from Yes to No

Hello Shuqi,

Variables "nnsib_dv" and "nnssib_dv" capture the number of siblings within the household because they are derived from the egoalt file.

To capture family relationships for both co-resident and non-co-resident family members, I suggest using the information available in the family matrix file (xhhrel). It's important to note that the xhhrel file creates an individual-level, cross-wave file of all individuals ever enumerated in the study. This file contains familial relationship identifiers reported throughout the survey period for each sample member. Consequently, individuals who were never enumerated for any reason will not be included in the family matrix.

You can find more detailed information about the xhhrel file at these links:
https://www.understandingsociety.ac.uk/documentation/mainstage/user-guides/main-survey-user-guide/connecting-family-members-and-households-over-time/ and
https://www.understandingsociety.ac.uk/wp-content/uploads/documentation/user-guides/6614_main_survey_user_guide_family_matrix_xhhrel.pdf

I wouldn't recommend combining calculations using the variables nrelsw12, nrels2, and nrels4. These variables capture individuals who are absent at a specific wave in time. These individuals might return to the study later or become absent again in the future, making the combination with other variables less suitable for capturing stable family relationships across waves.

You might also find the data management syntax available here helpful:
https://www.understandingsociety.ac.uk/wp-content/uploads/documentation/main-survey/syntax/stata/stata-using-egoalt-file-create-household-composition-variables.do. This resource shows how to create household composition variables, including calculating the number of siblings in a household using the egoalt file.

I hope this information is helpful.

Best wishes,
Roberto Cavazos

#2 - 05/08/2025 05:21 PM - SHUQI LYU

Dear Roberto,

I want to calculate "number of biological/step/adopt siblings" respondents ever had.

Before I do the next step, I append all the indresp.dta files of wave 1,3,5,7,9,11,13,14.

```
/*way 1*/
use "indresp_8waves.dta",clear
keep pidp wave nnssib_dv nrels4 /*variable nrels2 in w3,w5, nrelsw12 in w1*/
gen nsibling_history= nnssib_dv+nrels4 /*number of siblings in each wave*/

/*Way 2*/
merge m:1 pidp using "xhhrel.dta",keepusing(bsbx_N ssbx_N asbx_N)
drop if _merge==2
drop _merge
gen nsibling_ukhls=bsbx_N+ssbx_N+asbx_N /*number of siblings in file xhhrel*/

/*difference*/
gen diff=nsibling_history-nsibling_ukhls
tab diff
```

I found that the number of sibilings calculated by these two ways are different. For most of the respondents, the value in xhhrel (nsibling_ukhls) is smaller than the value calculated in each wave (nsibling_history).

What are the reasons of this difference? Thanks.

Kind regards,
Shuqi Lyu

#3 - 05/09/2025 02:32 PM - Understanding Society User Support Team

Hello Shuqi,

The study's information is limited to individuals within enumerated households. For those who are absent, we don't have detailed information beyond these general types of questions.

Therefore, as previously mentioned, combining calculations using variables for absent family members might not yield stable family trajectories. However, the ultimate decision on how and which variables to use for their research rests with the users.

Additionally, please note that the variable nrels4 specifically measures the number of absent grandparents, not siblings.

I hope this information is helpful.

Best wishes,
Roberto Cavazos
Understanding Society User Support Team