

Understanding Society User Support - Support #1372

Question about age of respondents based on age_dv, and racel_dv (Re-created as original was deleted)

06/30/2020 11:42 AM - Alita Nandi

Status:	Resolved	Start date:	06/30/2020
Priority:	Normal	% Done:	100%
Assignee:			
Category:	Data inconsistency		
Description			
Hello Alita,			
<p>I am wondering why the age_dv variable is not consistent with birthy when the latter is read side by side with the interview start year (istrtdat). I randomly looked at missing cells for age_dv vs those with exactly similar pidp, i.e. same person/individual over time. There are instances that age_dv = istrtdat - birthy. I checked the variables for birth month and birth date but these are missing. Why are there discrepancies like an individual is interviewed in consecutive waves but the year is unchanged, so his age stays the same; or there are cases like individuals refuse to reveal his birth year in a previous year but this information is filled in later waves, that first year of interview still represents a missing cell for birthy and age_dv? Kindly advise whether it makes sense to base age_dv on birthy and istrtdat. Thanks so much</p> <p>I am wondering whether race_bh could be combined with racel_dv for BHPS waves 1-12, as racel_bh covers BHPS wave 13-18. Does it make sense to assume that this is a time-invariant characteristic for individuals? I noticed that this is only asked once for new entrants. Can then this be carried over as non-missing info for succeeding waves, or this is simply counted once for each respondent? Note that I did use racel_bh to be combined with racel_dv when generating a separate new variable for use in Stata, so that original ethnicity variables are preserved.</p>			

History

#1 - 06/30/2020 11:51 AM - Alita Nandi

- Private changed from Yes to No

#2 - 06/30/2020 12:17 PM - Alita Nandi

- Status changed from New to In Progress

- % Done changed from 0 to 30

Hello Abigail,

I am passing on your query about age_dv to the data team and will get back to you with their response.

racel_dv in the file xwavedat combines responses to **race** and **racel** reported in BHPS and racel* reported in UKHLS. These questions are asked only once, first time someone is interviewed, and that information is used to create racel_dv.

Best wishes,

Alita

On behalf of Understanding Society User Support Team

#3 - 06/30/2020 04:08 PM - Alita Nandi

Hi Abigail,

Have you read the variable note and the input variables mentioned on the variable page?

https://www.understandingsociety.ac.uk/documentation/mainstage/dataset-documentation/variable/age_dv

Best wishes,

Alita

#4 - 07/02/2020 02:16 PM - Abigail Dumalus

Hello Alita,

I have read this variable note. But one of the input variables, dob_dv, is based on birthy. That's why I am asking whether age_dv = istrtdat - birthy. In most cases, dob_dv is missing, whilst birthy has non-missing entries. I would assume that birthy is the source variable for dob_dv, a derived variable. Another input variable for age_dv is intdat_dv. Similar to dob_dv, intdat_dv has many missing cases/cells, but istrtdat has non-missing information to which I can refer. So, if I refer to istrtdat, birthy, and age_dv, there are many instances wherein the difference between the first two do not result in

the values indicated for age_dv. I tried looking at the variables dob_m_dv and dob_dv, but most are missing or inapplicable entries. What should I do if there are discrepancies under age_dv when looking at individuals being interviewed over many waves and their ages "evolve" through time do not make logical sense?

#5 - 07/03/2020 02:01 PM - Alita Nandi

- % Done changed from 30 to 60

For the 9 wave of UKHLS data:

The variable w_birthy was checked for consistency across the waves. After resolving the inconsistencies, the variable doby_dv was created.

The interview date, w_istrtdat? is only available for adult interviews. This information was imputed for children and adult non-respondents from their household interview dates and recorded in w_intdaty_dv. See the variable note for this variable for details on how it was created.

https://www.understandingsociety.ac.uk/documentation/mainstage/dataset-documentation/variable/intdaty_dv

Then using dob?_dv and intdat?_dv the variable w_age_dv was created. Here ? represents y m d - year, month & day. The year of birth variables are available in levels of data, while month of birth is available in SL version & Secure access version, while day of birth is only available in Secure version of the data. See

https://www.understandingsociety.ac.uk/documentation/mainstage/dataset-documentation/variable/age_dv

The imputation flag for this variable is w_age_if

For the 18 BHPS waves, the variable bw_age_dv was computed using the interview date variables (bw_intdat?) and date of birth variables (bw_birth?). But unlike for the UKHLS waves, bw_birth? were not checked for consistency across the waves. So, you will find a handful of cases where a respondent's age has decreased over the waves. If you want to discuss how to resolve these cases please email us at usersupport@understandingsociety.ac.uk and we will forward your query to Gundi.

#6 - 08/14/2020 05:21 PM - Alita Nandi

- % Done changed from 60 to 90

#7 - 09/03/2020 01:02 PM - Alita Nandi

- Status changed from In Progress to Feedback

#8 - 10/13/2021 11:34 AM - Understanding Society User Support Team

- Status changed from Feedback to Resolved

- Assignee deleted (Alita Nandi)

- % Done changed from 90 to 100