Understanding Society User Support - Support #1789

Missing responses for racel_dv at each wave

10/19/2022 11:43 AM - Albert Ward

Status:	Resolved		**Start date:**	10/19/2022
Priority:	Normal		**% Done:**	100%
Assignee:				
Category:				

Description

Hi,

I'm using the crosswave variable racel_dv: when I match it into UKHLS indresp waves that I am interested in, I notice that there are a surprising number of missing responses (usually around a 1000). Why is this? I would think that ethnicity would be available for every respondent.

Thanks,
Albert

History

#1 - 10/19/2022 01:55 PM - Understanding Society User Support Team

- Status changed from New to Feedback

- % Done changed from 0 to 50

- Private changed from Yes to No

Every adult respondent is asked questions about their ethnic group the first time they are interviewed. Those who are interviewed by telephone are asked this information through a series of questions (racelt*). This information collected in different waves is combined into the racel_dv. So, if someone answered by telephone they would have missing racel but a valid value in racel_dv. Does this explain these 1000 cases?

Best wishes,
Alita

#2 - 10/19/2022 05:27 PM - Albert Ward

Hi Alita,

Not really unfortunately, maybe I'm not understanding something - I'm only talking about the racel_dv variable: when I look at it in the indresp files I see that there are usually quite a lot of missing values. For instance, in wave J, there are 567 missing values. I was just wondering why this was.

Albert

#3 - 10/20/2022 08:22 AM - Understanding Society User Support Team

I see.

racel_dv is combined from the information on racel (or racelt*) collected from across all the waves, i.e., the information is picked up from whichever wave it was reported. The first time someone completes an adult questionnaire they are asked this question (ff_everint~=1 & ff_ivlolw~=1). However, there are two groups that get left out - (i) rising 16s, i.e., those who completed a youth questionnaire prior to the first adult interview (N=143) (ii) whose first adult interview is a proxy interview (N=283), (iii) continuing BHPS sample members who hadn't reported racel (N=97) (iv) those who refused or didn't know (N=28). That leaves 16 for whom this information is missing due to other reasons.

We also provide another variable, j_ethn_dv, which imputes missing racel information using these sources (i) yprace (ethnic group question answered during the youth interview) (ii) ethnic group of parents (iii) ethnic group answered as part of the enumeration grid in Waves 1 & 6. An accompanying variable ethn_dv_source explains what is the source of this information.

There is a trade off between racel_dv and ethn_dv. The latter has fewer missing values but includes ethnic group information that is not self-reported.

Best wishes,
Alita

#4 - 10/20/2022 08:55 AM - Albert Ward

Ah I see: thanks for your answer, that's incredibly helpful! I wasn't aware of ethn_dv. I think I'm looking for most completeness in my data, so I think I'll choose to use that, but I understand the trade-off.

Thanks again,

Albert

#5 - 10/20/2022 10:38 AM - Understanding Society User Support Team

- Status changed from Feedback to Resolved

- % Done changed from 50 to 100

Albert

#5 - 10/20/2022 10:38 AM - Understanding Society User Support Team

- Status changed from Feedback to Resolved

- % Done changed from 50 to 100