Understanding Society User Support - Support #850

Non-overlapping data query (waves a to c)

09/07/2017 04:06 PM - Emma Bridger

Status:	Closed		**Start date:**	09/07/2017
Priority:	Normal		**% Done:**	100%
Assignee:	Emma Bridger			
Category:	Questionnaire content			

Description

Hello Understanding Society team,

I am working on a project in which (amongst other variables) we are using parental occupation and parental education variables from waves a & b (specifically a_paedqf, a_maedqf, b_paedqf, b_maedqf, a_pasoc10_cc, a_masoc10_cc) to predict mental health outcomes at wave c (including c_sf12mcs_dv, c_scghq1_dv and c_big5n_dv). I have summed education data from across waves a and b to derive separate retrospective measures for mother's and father's education (i.e. ma_educ = a_maedqf + b_maedqf and pa_educ = a_paedqf + b_paedqf) and as a consequence have Ns of around 40,000 for mother's and father's education. The sample sizes for a_pasoc10_cc and a_masoc10_cc are smaller but still >25,000.

After cleaning the wave c data, I have usable mental health data for around 40,500 cohort members, which is also very healthy.

My problem arises, however, when I try to run the models and it would appear that there is a considerable mismatch/nonoverlap between the sample for parental education/occupation (at a&b) and mental health at wave c. For example, although I have usable data on mother's education for 40,541, this drops to around 27,500 who also appear to have usable wave c mental health data (similar mismatches occur for other mental health variables). It isn't clear to me where the nearly 13,000 data points have disappeared to.

Is there an obvious way to account for this mismatch that I have overlooked and does this map onto your records of the data? Are these missing data represented elsewhere? Was a large part of the sample who completed the main questionnaire at wave c distinct from those who completed wave a&b?

Many thanks for any help you can provide,
Emma Bridger

History

#1 - 09/07/2017 04:25 PM - Stephanie Auty

- Category set to Questionnaire content

- Status changed from New to In Progress

- Assignee set to Stephanie Auty

- % Done changed from 0 to 10

Many thanks for your enquiry. The Understanding Society team is looking into it and we will get back to you as soon as we can.

Best wishes,
Stephanie Auty - Understanding Society User Support Officer

#2 - 09/07/2017 04:25 PM - Stephanie Auty

- Private changed from Yes to No

#3 - 09/14/2017 05:11 PM - Stephanie Auty

- Status changed from In Progress to Feedback

- Assignee changed from Stephanie Auty to Emma Bridger

- % Done changed from 10 to 60

Dear Emma,

First I want to confirm that I understand how you are combining the data from waves 1 and 2. a_paedqf and a_maedqf were originally intended to be asked of the whole sample, but due to timing was only asked of the first 6 months' samples in wave 1. Therefore, b_paedqf, b_maedqf were asked of the OSMs in the remaining 18 months' samples in wave 2 so that we would have answers to this question for the full original sample, excluding those who did not take part at wave 2. There may be an occasion where someone was asked in both waves due to a routing error so in this case you will

need to decide which data to use. So, the respondents with applicable data from waves 1 and 2 should be appended and these pairs of variables combined. Is that what you have done?

Next, regarding your sample size, the parents' education questions were only asked of original sample members, so any new entrants in waves 2 and 3 will not have been asked. This will include rising 16s and people who joined the OSMs' households. Some OSMs from the 7-24 months' samples may have missed wave 2 and completed wave 3, so these would also have missing data for the education questions. So I can confirm that your sample sizes look correct from our investigations, and there is no other data available.

It's not obvious from the online documentation, but if you look at the questionnaire (https://www.understandingsociety.ac.uk/questionnaires/ms01_questionnaire.html) you can see that the questions a_pasoc10_cc and a_masoc10_cc are filtered, based on a_paju and a_maju, whether each parent was working when the respondent was 14. If they weren't working, a_pasoc10_cc and/or a_masoc10_cc would be coded as inapplicable. You may want to code this differently depending on your analysis.

An option for you to increase your sample size would be to look at rising 16s who completed the mental health questions at wave 3 and investigate the education and job codes of their parents where they have also responded to the survey.

Best wishes,
Stephanie Auty - Understanding Society User Support Officer

#4 - 10/02/2017 01:45 PM - Stephanie Auty

- Status changed from Feedback to Resolved

- % Done changed from 60 to 100

#5 - 10/16/2017 04:28 PM - Stephanie Auty

- Status changed from Resolved to Closed