

Understanding Society User Support - Support #745

Interviewer number query

03/14/2017 09:27 AM - Jonathan Burton

Status:	Closed	Start date:	03/14/2017
Priority:	Normal	% Done:	100%
Assignee:	Jonathan Burton		
Category:	Data inconsistency		

Description

Hi,

I received this (below) - I can look into it, but wanted to post it here in case it's come up before (I did a quick search and couldn't find anything).

Hello Jonathan,

Sorry to have to open this thread again, and please let me know if I should send this set of queries elsewhere, but we've hit a bit of a roadblock in our analysis.

In order to include data on interviewer characteristics in our analysis of interviewers' ratings of respondents' health, we had to merge the interviewer, household, and respondent level data. Merging was done in 3 parts. The interviewer data (xivdata.dta) was first merged with wave 8 household data based on intnum (h_hhsamp_ip.dta). Household data was then distributed to the individual level wave 8 data (h_indresp_ip.dta). Wave 7 individual level data (g_indresp_ip.dta) was merged with wave 8. Resulting in Wave 7, wave 8 and the interviewer level data at the individual level.

In these data, there are 2,378 respondents interviewed across various modes:

```
. tab h_indmode
```

```
mode this |
  individual |
    was given |
    final ind |
outcome in |      Freq.    Percent    Cum.
-----+-----
      proxy |         111     4.67     4.67
      capi |        1,439    60.51    65.18
      cati |          29     1.22    66.40
      cawi |         799    33.60   100.00
-----+-----
      Total |        2,378   100.00
```

For the capi respondents, 37 are missing interviewer identification numbers. For the telephone, 5 are missing interviewer ID. For the CAWI, there are 3 respondents that have an interviewer ID number (would expect no respondents in this mode to have an interviewer ID number).

When we go to examine the interviewers' characteristics in the respondent level file we created (to see, e.g., if the interviewer's gender is related to how they rate respondents' health), we have large amounts of missing data. We have valid data for 682 respondents in terms of intsex intyearofbirth intyearstarted intoparea intrace_dv intveteran. We would expect the number to be closer to 1400 given the number of CAPI respondents and available interviewer IDs.

So my questions are:

- 1) Should we merge another way in order to get the interviewer characteristics to the respondent level data? Maybe there is a mismatch with intnum somewhere in this set of procedures?
- 2) Maybe the xivdata.dta file we used (which we downloaded in May 2016) did not have information for some of the wave 8 interviewers yet? The xivdata file does not have large amounts of missing data, but maybe the cases are incomplete? For example, I see that when we merge xivdata with the wave 8 household data (h_hhsamp_ip.dta), there are 1,738 households, 1,200 households with a valid intnum, and only 531 households with valid values of interviewer characteristics. So there problem starts to appear when linking xivdata to the household data, and does not appear to be linking this merged data to the respondent level data (h_indresp_ip.dta).

Any thoughts you have on how to proceed would be most appreciated!

History

#1 - 03/14/2017 03:56 PM - Jonathan Burton

Update:

I've had a look at the h_intnum variable on the h_indresp_ip data and this is "inapplicable" for 2267 out of 2378 cases, 9 "proxy" cases, and the rest are numbers, but not of the same format as intnum on the hhsamp file - these go from 113 to 82764.

#2 - 04/25/2017 10:58 AM - Victoria Nolan

- *Category set to Data inconsistency*
- *Status changed from New to Closed*
- *Assignee set to Jonathan Burton*
- *% Done changed from 0 to 100*
- *Private changed from Yes to No*

This will be addressed for the release of IP9 data.