

Understanding Society User Support - Support #1420

covid-19 observations not merging with wave 9 data

10/05/2020 11:24 PM - Cristina Sechel

Status:	Feedback	Start date:	10/05/2020
Priority:	Normal	% Done:	80%
Assignee:	Cristina Sechel		

Description

Hello

Some individuals from the April-June covid-19 data files do not match with wave 9 data (using the pidp variable). I don't understand why this is happening because as far as I understand the sample for the covid modules is based on those who have participated in waves 9-11 and I only use individuals who have a non-missing value for i_outcome variable in the covid-19 data for the matching.

best
Cristina

History

#1 - 10/06/2020 10:58 AM - Alita Nandi

- Status changed from New to Feedback
- Assignee set to Cristina Sechel
- % Done changed from 0 to 80
- Private changed from Yes to No

Hello Cristina,

Anyone 16+ years, who was in a household that participated in at least one of the last two waves of data collection as of April 2020 were invited to the Covid19 survey. So, even if someone did not participate in an adult interview in Wave 9 but did in Wave 10 or Wave 11 they could be included. Or if they didn't participate in any of the 3 adult interviews but were in a household which was enumerated, then they were eligible. For example, someone could have been 15 at the time of the last interview and turned 16 by April 2020 - they too would be eligible.

See User Guide Section 5: "The 'active' sample includes everyone in households who have participated in at least one of the last two waves of data collection..... In 'active' sample households, all household members who were aged 16+ in April 2020 were invited to the COVID-19 study, except for those who were adamant refusals or mentally or physically unable to make an informed decision to take part, and those with unknown postal address or addresses abroad."

Please let us know if this is not clear.

Best wishes,
Understanding Society User Support

#2 - 10/06/2020 11:36 AM - Cristina Sechel

Thanks Alita, I'm still a bit confused.

I understand that the covid sample also includes some individuals who were not interviewed in wave 9. But I am talking about matching only those individuals who have a full interview in wave 9 as identified by the i_outcome variable (so my matching is done only for those with i_outcome=11). According to the user guide "The interview outcomes will allow identification of the last wave a respondent was interviewed in the annual survey" so I would expect that all those with i_outcome=11 would match in with the i_indresp data file. Or is the i_outcome variable pertaining to whether the household (not the individual) had at least one full interview in wave 9 (which may or may not have been the person interviewed in the covid sample)? So say person A from household X is interviewed in wave 9 and person B from the same household is then interviewed in one of the covid modules, then person B would have i_outcome=11 even though they were never interviewed in wave 9?

And checking the age variable, there are plenty of older individuals (way past 16) who don't match.

best
Cristina

#3 - 10/06/2020 12:16 PM - Alita Nandi

Sorry I misunderstood your question. But I could not find the problem you have highlighted. Your understanding of i_outcome is correct - a value of 11 means that that individual had completed an adult individual interview in Wave 9.

I checked and found that everyone in ca_indresp_w with i_outcome=11 are also available in i_indresp. I used the following Stata syntax to check this. The merge showed that there were 15835 observations in both files, and 20220 only in i_indresp but not in ca_indresp_w. There were no cases that

were in ca_indresp_w but not in i_indresp.

```
// Stata syntax //  
use ca_indresp_w" clear  
keep if i_outcome==11  
merge 1:1 pidp using i_indresp
```

Best wishes,
Alita