

Understanding Society User Support - Support #585

Identifying individuals with whom an interview could not be completed

06/23/2016 06:25 PM - Sanne Velthuis

Status:	Closed	Start date:	06/23/2016
Priority:	Normal	% Done:	100%
Assignee:	Sanne Velthuis		
Category:	Data analysis		

Description

I have a question relating to respondents living in households which were contacted by an interviewer but with whom no interview was ultimately completed - either because the household as a whole could not be contacted or refused, or because the individual respondent refused to take part in an interview. I can see that the `_hhsamp` files include, for each household identifier (`_hidp`) the `_outcome` variable which indicates the outcome of the relevant interviewer's attempt to conduct interviews with the household members. I have merged this information (using `_hidp` and `_month`) into the `_hhresp` datafile which in turn I have merged with the `_indresp` datafile (using `_hidp`), and appended all five waves to create a panel. Theoretically, the resulting datafile should now include all households, and all individuals within these households) who were contacted by an interviewer at each wave, including households (and the individuals within these) with whom no interviews could be completed. I expected that the resulting datafile, when sorted on `_hidp` and wave, would look something like this:

```
_hidp wave _outcome pidp
1 1 complete 1
1 2 complete 1
1 3 no contact .
```

So this shows that household 1 was contacted in wave 1, that the interviewer was successful at completing the interviews, and that the household contained only one individual with the personal identifier (pidp) 1. It then shows that the same household was contacted at wave 2, again successfully completed an interview, and that the household still only contained the individual with pidp 1. Then it shows that in wave 3 household 1 was contacted again, but that no contact could be made, and as a result the interviewer was unable to identify whether pidp 1 still lived in this household and was unable to conduct an interview.

However, the data does not appear to be structured like this. Where the outcome of an attempt to contact a household was an inability to contact the household, a refusal of the household to take part, or any other unsuccessful outcome, the household identifier appears to be a unique identifier that does not occur in any other wave. As a result, it is impossible to tell whether a particular household with whom no interviews could be completed is the same as a household that was successfully interviewed the year before.

So my question is this: how does the allocation of household identifiers in the `_hhsamp` file work? I would have expected that all households selected for participation in wave 1 would have been given a household identifier (and upon completion of the first interview all individuals within these households a personal identifier), and that for wave 2 these households, listed under the same household identifiers, would have been allocated to an interviewer who would then try to contact them for their second interview. I would thus expect that all households who could not be contacted in wave 2 would show up in the `_hhsamp` file with the same identifier as they had at wave 1, with the specific outcome of the contact attempt next to that. But this does not appear to be the case. So how are household identifiers allocated? Is there a way to link a particular household that was interviewed in wave 1 to an unsuccessful attempt to interview this same household in wave 2?

History

#1 - 06/24/2016 03:17 AM - Alita Nandi

- Status changed from New to In Progress
- Assignee changed from Alita Nandi to Sanne Velthuis
- % Done changed from 0 to 90
- Private changed from Yes to No

Household Identifiers, `W_HIDP` are unique within a wave but not across wave. the individual identifier `PIDP` is unique both within and across waves. The survey does not impose a definition of longitudinal household and so there are no household identifiers that are unique across waves. For example, in wave 1, there are 3 people living in a household - a couple and their adult child. In the next wave, she moves out into her own accommodation and then in wave 2 these two belong to 2 different households and each have a different household identifier. Which of these two households in wave 2 should be considered to be the same as the wave 1 household?

If you decide on a definition of longitudinal household you will be able to link those households with the help of the `PIDP` of their members.

Best wishes,
Alita

#2 - 06/24/2016 09:06 AM - Sanne Velthuis

Dear Alita,

Thanks for your response. I understand that because of changes in household composition household identifiers are not unique across waves, and that it is possible to use my own definition of a longitudinal household to link households across waves based on the PIDP of their members - but what about those cases where a no household nor individual interviews could be conducted and so there is no information about which household members are present within the household? In this case there would be no PIDPs recorded under that household, and so no longitudinal linkage can be made based on the PIDPs of household members.

I'll explain to you what I wish to do. I would like to identify individuals living in households who are present in, say, wave 1, but not in wave 2. I would then like to see whether the interviewer that attempted to contact the household in which the individual lived (in wave 1) was the same as the interviewer who conducted the wave 1 interviews. To do this I would need to be able to identify these non-responding households across waves. For example, I have an individual with PIDP 12345, who in wave 1 lived in household 1000. In wave 2 individual 12345 is not in the _indresp datafile, indicating that they could not be interviewed for some reason. I would like to be able to find household 1000 in the _hhsamp datafile for wave 2, so I can see whether this household was contacted by an interviewer and what the outcome of this attempted contact was. If the outcome was that the household refused to participate, then that explains why PIDP 12345 is not present in wave 2 (of course without actually conducting a household interview it is impossible to know for sure that PIDP 12345 is still living in household 1000, but most of the time this will probably be the case), and I would then be able to see if the interviewer who attempted to contact household 1000 in wave 2 was the same person as before by comparing the interviewer IDs across waves 1 and 2. The reason I would like to do this is to construct a two-stage selection model that models selection into the sample at any given wave (after wave 1) as a function of whether there was a change in interviewer since the previous wave (plus some other variables).

Is there any way to achieve this?

Best wishes, Sanne

#3 - 07/11/2016 11:35 AM - Alita Nandi

Dear Sanne,

As HSDP are not consistent across waves you cannot use that to directly identify individuals in Wave 1 who are in non-responding households in Wave 2. But this can be done. You will need to use the W_INDSAMP files (available from onwards Wave 2). That file is an individual level. In this file, each row is uniquely identified by PIDP W_FINLOC. W_FINLOC =0 for the household where the person should have been found as per their last wave record but are not there. For example, if they have moved. In that case they are assigned a new household ID. Then W_FINLOC will be 1 for this new household where they have moved to. Note W_FINLOC=2 for those who are dead. So, you can use the PIDP to see whether that person was in a household in the current wave which was not interviewed (using the interview outcome variables W_IVFHO and W_IVFIO). You can also use the household IDs (attached to each person) to link to HHSAMP and CALLREC files which include INTNUM, interviewer ID. While HHSAMP is at the household level, CALLREC is at the call level. One household may have been contacted by more than one interviewer. It is not possible to link an individual to an interviewer.

There is also a file called XIVDATA which includes INTNUM and a few characteristics of the interviewer. If you wanted to account for interviewer characteristics, then use this file. It includes all interviewers ever connected with the survey and is updated each wave with information of new interviewers.

Please take a look at these files and see if that solves your problem. If not, please let us know.

Best wishes,
Alita

#4 - 07/25/2016 12:54 PM - Victoria Nolan

- Status changed from In Progress to Closed

- % Done changed from 90 to 100