

Understanding Society User Support - Support #848

Clinical Depression H_COND variables

09/04/2017 04:55 PM - Luca Bernardi

Status:	Closed	Start date:	09/04/2017
Priority:	Immediate	% Done:	100%
Assignee:	Olena Kaminska		
Category:	Weights		
Description Dear Support group, I am measuring clinical depression and I would kindly need your advice on a couple of questions. I apologise sincerely for putting immediate priority on this, but your answer might also have implications for a paper I am co-authoring within the Understanding Society EU Referendum project and we have a deadline shortly for submitting the paper. As I am interested in objective depression, I was using the questions H_COND17 and H_CONDS17 to create a measure of depression. What I was doing is to assign value 1 to respondents who replied that they still have depression in H_CONDS17=Yes (as I am interested in the effects of depression, I do not care much if the person was diagnosed with depression at some point in his/her life - i.e. H_COND17=Yes - but rather it is important that the person is depressed at the time of the interview). I assign value 0 if the respondent mentioned that he/she has never been diagnosed with depression in H_COND17=No. So far I was using data from waves 1, and 3 to 6 as I noticed that these two variables are available in all waves but wave 2 (https://www.understandingsociety.ac.uk/documentation/mainstage/dataset-documentation/wave/2/datafile/b_indresp), where instead a slightly different question is asked: H_CONDN17. In turn, this question is not available in all waves and sometimes is asked together with the previous two questions (e.g., https://www.understandingsociety.ac.uk/documentation/mainstage/dataset-documentation/wave/4/datafile/d_indresp). My questions thus are the following. Do you please know what is the reason of such a variation and, more importantly, can I "maximise" my number of depressives by creating a measure of depression that combines both sets of questions (i.e., H_COND17 and H_CONDS17, and H_CONDN17) and makes use of all available waves (i.e. 1 to 6)? My idea was to do the following: gen depression = . replace depression = 1 if hconds17==1 hcondn17==1 replace depression = 0 if hcond17==0 hcondn17==0 However, I wonder how problematic can be mixing questions that are not available in all waves, as this is certainly a point that reviewers will raise. I would really appreciate your thoughts on this. Many thanks and best wishes, Luca			

History

#1 - 09/04/2017 05:06 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
- Private changed from Yes to No

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 12:15 PM - Stephanie Auty

Dear Luca,

In Wave 1 respondents were asked to report whether they'd ever been diagnosed with clinical depression and whether at that time they still had the condition (w_hcond17 and w_hconds17). At each subsequent interview (i.e., wave), they were only asked about any new diagnoses (w_hcondn17, the 'n' in the name meaning new) – meaning that, perhaps importantly for you, diagnoses from previous waves were not followed up to ascertain whether they still had the condition. The variables w_hcond17 and w_hconds17 have been collected since Wave 3 but this is for new entrants to the study: at subsequent interviews they are only asked about new diagnoses too.

You may be interested in the GHQ score of mental health that is collected in every wave, although this is a subjective measure and possibly not ideal for your purposes (although it could be argued that someone's assessment of whether they still have a previously diagnosed condition is somewhat subjective too). This module was part of the self-completion paper questionnaire at waves 1 and 2, as you can download here <https://www.understandingsociety.ac.uk/documentation/mainstage/questionnaires>, and you can see it in the online documentation of the scaghq module from wave 3 onwards:

https://www.understandingsociety.ac.uk/documentation/mainstage/dataset-documentation/wave/3/questionnaire-module/scaghq_w3

You may also be interested to know that we are currently developing the questionnaire to collect information about previous diagnoses, including whether the respondent reports still having such conditions. This information should be available from Wave 10 onwards. In some types of analyses the presence of post-diagnosis symptoms/conditions is not important, it is the diagnosis that's relevant (it's objective, it signals the likely presence of an underlying condition even if symptoms are either not currently present or managed via medication, etc.). However this is not sufficient for all research questions and we are revising our data collection strategy.

Best wishes,
Stephanie Auty - Understanding Society User Support Officer

#3 - 09/07/2017 12:15 PM - Stephanie Auty

- Status changed from In Progress to Feedback
- Assignee changed from Stephanie Auty to Luca Bernardi
- % Done changed from 10 to 80

#4 - 09/12/2017 03:13 PM - Stephanie Auty

----- Forwarded message -----

From: Bernardi, Luca <lb350@leicester.ac.uk>

Date: Fri, Sep 8, 2017 at 9:50 AM

Subject: Re: [Understanding Society User Support - Support [#848](#)] Clinical Depression H_COND variables

To: "info@understandingsociety.ac.uk" <info@understandingsociety.ac.uk>

Dear Stephanie,

Many thanks for your super quick reply and for better explaining the situation with the variables of interest.

Best wishes,
Luca

#5 - 10/09/2017 12:21 PM - Stephanie Auty

From: Bernardi, Luca [mailto:lb350@leicester.ac.uk]

Sent: 08 September 2017 10:05

To: Auty, Stephanie

Subject: Understanding Society Issue [#848](#)

Dear Stephanie,

Thank you again for getting back to me so quickly.

Sorry for contacting you directly to your email address, but I wanted to make sure I understand what it is sensible to do and what is not. Given that the hcond17, hconds17 and hcondn17 questions are not asked in all waves contemporaneously, does my proposal of using information from all variables make sense? Note that in this particular project I am not doing a panel study but I am simply pooling the data so I am not interested in tracking respondents over time.

I recall that what I had in mind is to create a dummy variable for depression equals 1 if the person replied positively to the hconds17 and hcondn17 questions and equals 0 if the person replied negatively to the hcond17 and hcondn17 questions.

Otherwise, what I was previously doing is to use data only from the hcond17 and hconds17 questions, in which case Wave 2 was not included in the analysis. Hence, what I was doing is to create a dummy variable for depression equals 1 if the person replied positively to the hconds17 question and equals 0 if the person replied negatively to the hcond17 question.

As you know the data much better than I do, I would really appreciate if you could let me know if both solutions are feasible or only the latter is.

Thanks and best wishes,
Luca

#6 - 10/09/2017 12:22 PM - Stephanie Auty

Dear Luca,

These are valid questions. It would be helpful if you could post your reply on the User Forum (

<https://www.understandingsociety.ac.uk/support/projects/support>) as

(i) we manage and track these requests and make sure someone from the team answers within 10 working days

(ii) we send these requests to the person best suited to answer that particular question

(iii) and most importantly other users can benefit from the answers. Similarly, you can search on the User Forum to see if someone has raised this issue before.

I will follow your reply on the forum with my answer as below, so that other users can benefit.

The variables you are discussing measure different things, and we don't know the respondents' current status. We can't tell you what assumptions to make as it depends on your research question. Our remit at the User Forum is to answer queries related to Understanding Society data and provide general advice about how to manage the data. Given the number of users we have I'm afraid we cannot advise on individual users' analysis specifically.

The new entrants at each wave are not part of the original sample. They provide a household context for the OSMs and also increase the sample size, which compensates somewhat for attrition. So if you didn't include wave 2 new entrants it would not impact on how representative your analysis is.

If you would like to discuss weighting based on this then I can assign this issue to Peter Lynn after we have posted our replies on the forum.

There are two other variables you could look at, w_hcondno1-10 and w_hcondns1-10 from wave 2 onwards (not all waves have 10 of these variables). These measure whether the respondent still has their new diagnosis as asked in w_hcondn, but which variable is related to depression varies depending on which conditions the respondent has.

Best wishes,
Stephanie

#7 - 10/09/2017 12:23 PM - Stephanie Auty

From: Bernardi, Luca (Dr.) [mailto:lb388@leicester.ac.uk]

Sent: 04 October 2017 14:23

To: Auty, Stephanie

Subject: Re: Understanding Society Issue [#848](#)

Dear Stephanie,

I am replying from this account because I have just changed email. Many thanks for your message. Please feel free to post our exchange on the forum if this can be of help for other researchers and sorry for not having done this myself immediately. So if I understand correctly your point on the representativeness of the analysis, I should be ok on this side even without including Wave 2, am I right? If so, that is good news. I appreciate that you cannot offer suggestions that depend on the research question, but since the depressive sample declines considerably after Wave 1, I would be grateful if you could please advise me on whether it still makes sense to use data from subsequent waves or if I am good enough with only Wave 1.

I think forwarding the message to Peter Lynn is a great idea and thanks for this, because it would be very helpful for my analysis to know whether to use a weight would be more appropriate than not weighting at all, given the low N of depressive respondents compared to non-sufferers. I look forward to Peter's reply and yours.

Best wishes,
Luca

#8 - 10/16/2017 04:28 PM - Stephanie Auty

- Category changed from Questionnaire content to Weights

- Assignee changed from Luca Bernardi to Peter Lynn

- % Done changed from 80 to 60

#9 - 11/09/2017 04:19 PM - Olena Kaminska

Luca,

I will try to answer some of your questions in terms of weighting. But please if my response is not sufficient do not hesitate to rephrase and ask again.

In terms of representation you are fine if

1. you just use wave 1 - this is completely fine - just use the cross-sectional weight relevant to your analysis;

2. if you pool your data over the waves but exclude wave 2 - this is again completely fine from the perspective of population representation. You should use cross-sectional weight relevant to your analysis from each wave in your analysis. Because with time the sample size varies (due to attrition and refreshments) - you should add a scaling factor so that each wave contributes to the analysis similar to other waves. Read this note on how to create the scaling factor: <https://www.understandingsociety.ac.uk/support/issues/494> . Also please remember to indicate clustering within respondents - otherwise the statistical program will think that you have much higher sample than you actually do.

Hope this helps,
Olena

#10 - 11/10/2017 04:36 PM - Peter Lynn

- Assignee changed from Peter Lynn to Luca Bernardi
- % Done changed from 60 to 70

#11 - 11/28/2017 12:20 PM - Stephanie Auty

- Status changed from Feedback to In Progress
- Assignee changed from Luca Bernardi to Olena Kaminska

From: Bernardi, Luca (Dr.) [mailto:lb388@leicester.ac.uk]

Sent: 09 November 2017 19:48

To: usersupport@understandingsociety.ac.uk

Subject: Re: [Understanding Society User Support - Support [#848](#)] Clinical Depression H_COND variables

Dear Olena,

Many thanks for your reply and the useful suggestions. I was wondering whether you could please tell me more about the advantages and disadvantages of only using wave 1 instead of pooling the waves or, vice versa, if I would instead be better-off using the available information from all waves. Relatedly, if I only use Wave 1 reviewers might fairly ask why I have not used information from the other available waves. I am not much worried about the validity of my results, for as far as I can see they still hold using either wave 1 or pooling all waves. Yet some clarification on this would be great for me to be able to take an informed decision on which data to use and to fully justify the choice in the paper.

Thank you and best wishes,
Luca

#12 - 12/18/2017 11:28 AM - Stephanie Auty

- Status changed from In Progress to Resolved
- % Done changed from 70 to 100

This issue has been resolved outside of the forum.

#13 - 01/09/2018 01:46 PM - Stephanie Auty

- Status changed from Resolved to Closed