

Understanding Society User Support - Support #1458

Weights for Covid longitudinal sample

12/07/2020 11:21 AM - Manuel Serrano

Status:	Resolved	Start date:	12/07/2020
Priority:	Normal	% Done:	100%
Assignee:			
Category:	Weights		
Description			
Hi,			
I am doing a longitudinal analysis of a balanced sample from Wave 9 to Wave Covid 4. So far I am using the cross-sectional weights for wave 4 (cd_betaindin_xw) that according to the users Guide account for non-response on Covid 4 as compared to Wave 9. Additionally, in order to account for non response in Wave 10 and in Waves Covid 1, 2, 3, I am multiplying these weights by the inverse of the probability of response in those waves, conditional on a set of variables (following the point 15 in the weighting FAQs). I just wanted to make sure that my weighting strategy is correct or whether there are other more optimal weighting strategies that involve longitudinal weights for Covid waves.			
Thanks in advance			
Manuel			

History

#1 - 12/07/2020 11:50 AM - Olena Kaminska

Manuel,

Yes, this sounds correct. Although it has one unnecessary assumption of no household composition change between waves 9 and 10. Alternatively, you could just use wave 9 longitudinal weight as a base weight and model response from there to your model. Remember to take out those who died and moved out of the country between waves 9 and 10 from your response model.

Hope this helps,
Olena

#2 - 12/07/2020 11:54 AM - Manuel Serrano

Dear Olena, thanks very much for the help.

Should I take out of my response model also those who died and moved out of the country in Waves Covid 1, 2 and 3?

Thanks in advance

Manuel

#3 - 12/07/2020 12:28 PM - Olena Kaminska

Manuel,

Yes, you should.

Thanks,
Olena

#4 - 12/07/2020 05:47 PM - Alita Nandi

- Status changed from New to Feedback
- Assignee changed from Olena Kaminska to Manuel Serrano
- % Done changed from 0 to 90
- Private changed from Yes to No

#5 - 12/09/2020 10:11 AM - Theocharis Kromydas

Manuel Serrano wrote in [#note-2](#):

Dear Olena, thanks very much for the help.

Should I take out of my response model also those who died and moved out of the country in Waves Covid 1, 2 and 3?

Thanks in advance

Manuel

This is a very interesting discussion as I am more or less trying to do the same thing with just using wave 9 onwards (including the four Covid waves). So, with both strategies I guess the weight will be valid only to those that have participated to all waves since the starting wave (in my case Wave 9) and have no missing values on our variables of interest (outcome and predictors). So, since the longitudinal weight produced at the end would have the same value for all waves for each participant I assume we can only do a complete case analysis so I guess we would need to exclude those who stopped responding during the time period we are interested in, irrespective of the reason. Specifically for me, and since weights in Covid waves are based on Wave 9 weights and account for non-response, if I get the Wave 9 longitudinal weight and multiply it with `cd_betaindin_xw (9,10,Covid1, Covid2, Covid3, Covid4)` and assign this product to each individual who have participated to all waves, I would then produce a longitudinal weight which is valid to be used from wave 9 onwards. Am I right here?

#6 - 12/09/2020 10:29 AM - Theocharis Kromydas

Also there is a difference between using Wave 10 or the sub-sample of Wave 10 and Wave 11, which was integrated to the latest Covid release and includes data only for those who have also participated in at least one of the Covid waves. This data includes no weights at all, so I guess this means that those in this sub-sample with complete data to all other Waves of interest (Wave 9 and all Covid waves) can get the longitudinal weight the way I have described before with no additional adjustment for non-response?

#7 - 01/07/2021 02:43 PM - Alita Nandi

If you are using Wave 9 + all Covid waves, and conducting a longitudinal analysis, you should use the longitudinal weight from the last Covid wave, `ce_betaindin_lw`. This weight is based on those who had responded in Wave 9, and then their wave on wave response weights are multiplied. Please see the user guide for details.

"Also there is a difference between using Wave 10 or the sub-sample of Wave 10 and Wave 11, which was integrated to the latest Covid release and includes data only for those who have also participated in at least one of the Covid waves. This data includes no weights at all, so I guess this means that those in this sub-sample with complete data to all other Waves of interest (Wave 9 and all Covid waves) can get the longitudinal weight the way I have described before with no additional adjustment for non-response?"

The data that was released with the Covid data only includes respondents of Wave 10 year 2 and Wave 11 year 1, which mostly covers the interviews in 2019. So, this sample is not identical to the Wave 10 data released with the main survey. Yes you are right that no weights were included in this file. If you use the weights provided with the Covid data, you will have to produce weights that additionally account for non-random non-response between that sample (W9+Covid waves) and responding in 2019.

#8 - 01/13/2021 10:38 AM - Manuel Serrano

Dear Alita and Olena

Thanks for alerting about the release of the new longitudinal weights.

I have been reading the user guide version 5.1 carefully and I still have doubts about which weights to use in my analysis. To sum up, in my analysis I am using the balanced longitudinal sample of individuals who responded to waves 9, 10, Covid 1, Covid 2, Covid 3 and Covid 4 for the mental health variables (GHQ). Following the advice in this post and the guide I can think of several options for my weights:

Option 1. The weights I have been using so far, following Olena's previous advice and the weighting FAQs are based on the cross-sectional weights of wave 9 ("`i_indscui_xw`", since I am using the GHQ variable from self-completion questionnaire). Then, to account for non-response between waves I have used inverse probability weight (IPWs). These were created by estimating the probability of responding in all 6 waves in the mental health variables (i.e.: taking part of the balanced sample) as a function of observable variables at baseline wave (i.e.: wave 9): age, sex, education level, labour market status, self-reported health status, smoking status, access to internet, household income (OECD-equivalent), and region. IPWs were then formed by the inverse of the predicted probability of responding in the balanced sample. Lastly, I multiplied these weights by the above mention cross-sectional weights from the wave 9 ("`i_indscui_xw`")

Option 2. Following Alita's new post and using "`cd_betaindin_lw`" as my baseline weights. However, as far as I understood these weights would not take into account non-response in wave 10, nor the fact that I am using variables from the self-completion questionnaire. From these weights, I would create a non-response model with respect to my final longitudinal sample similar to the one in option 1, in order to take into account non-response in Wave 10 and the self-completion questionnaires. However, according to the user guide v5.1, the new longitudinal weights ("`cd_betaindin_lw`") were created with chains of models of response conditional on response to all previous waves. In such model, non-response at wave 10 and non-response on the self-completion questionnaire would actually go at the beginning of the chain of models of response, nor at the end as I am doing here, right? This leads me to think about Option 3.

Option 3. Create my chain of models of response, using as baseline the cross-sectional weight in wave 9 ("`i_indscui_xw`"). This would include the same variables as in the response model of Option 1 (and any extra variables that you could suggest), and I would do then a response model for each wave, creating interim weights until getting the final weight, as explained in the user guide v5.1.

I would appreciate any insights on the issue since I am quite confused about which weights I should use.

Thanks in advance,
Manuel

#9 - 01/13/2021 01:04 PM - Alita Nandi

- Assignee changed from Manuel Serrano to Olena Kaminska

#10 - 01/15/2021 01:15 PM - Olena Kaminska

Manuel,

Your option 1 is the best: it goes back to the most recent of our weights and then models nonresponse in one step to your model. It has advantage of using sc weight as well.

Your option 3 is possible, but includes extra, possibly unnecessary steps. Modelling wave on wave may be more important over longer time period as predictors change over time - and you can use more recent information relevant to panel members in your nonresponse models. But in your situation I don't think it would improve models substantially.

Options 1 an 3 would be preferable over option 2.

Hope this helps,

Olena

#11 - 01/21/2021 01:43 PM - Alita Nandi

- Assignee changed from Olena Kaminska to Manuel Serrano

#12 - 10/12/2021 05:00 PM - Understanding Society User Support Team

- Status changed from Feedback to Resolved

- Assignee deleted (Manuel Serrano)

- % Done changed from 90 to 100