Understanding Society User Support - Support #2115

Help with weights for longitudinal analysis

06/04/2024 10:36 AM - Amy Taylor

Status:	Feedback	**Start date:**	06/04/2024
Priority:	Normal	**% Done:**	80%
Assignee:	Olena Kaminska		
Category:	Weights		

Description

I was wondering if you could advise on the correct weights that I should use for an analysis. I have identified arthritis cases from waves 1-12 and am running multilevel models in Stata. I am including anyone with data at any waves so I am not sure that the longitudinal weights are appropriate. Am I right in thinking these only account for individuals with data at all waves? If so, would you be able to advise how I calculated appropriate weights?

Should I be using PSU as a level in the multilevel model and is this sufficient for accounting for clustering by household? I understand PSU is at the postcode level.

I am also worried about the effect of mortality/drop out on analyses. Is this accounted for when using the sampling weights?

Many thanks

History

#1 - 06/07/2024 03:21 PM - Understanding Society User Support Team

- Category set to Weights

- Status changed from New to In Progress

- Assignee changed from Understanding Society User Support Team to Olena Kaminska

- Private changed from Yes to No

#2 - 06/11/2024 12:43 PM - Olena Kaminska

Amy,

Thank you for your question. Read questions 14-16 here, which should help you choose the weight you need:
https://www.understandingsociety.ac.uk/wp-content/uploads/working-papers/2024-01.pdf

Yes, psu should be a level in your multilevel model.

One advice I have for you is to carefully define the population that you want to study outside of UKHLS. Meaning the population should not have in its definition 'observed in any wave of UKHLS'. It may be people who currently have arthritis, who ever have in their lives etc. Think what happens if a person with arthritis in present in all waves - you shouldn't use the same person multiple times for person-level analysis, but you can use multiple waves for event / changing state-level of analysis. My suggestion is to devote to this a bit of time and thought before you start working with the data.

Let me know if I can help,
Olena

#3 - 06/11/2024 02:09 PM - Amy Taylor

Thanks Olena,

The population of interest is anyone with a new diagnosis of arthritis at some point between 2009 and 2022. The analysis is of changes in physical symptoms over time (including both before and after diagnosis) so we are using all available waves of data for each person. But this means that as long as people have data at at least two waves they can be included in the analysis. I'm struggling with the weights because someone might contribute at wave 2 and wave 5, another person may contribute at all waves, someone else at waves 10,11,12.

Thanks

Amy

#4 - 06/12/2024 12:37 PM - Olena Kaminska

Amy,

What is your level of analysis? Is it people? And diagnosis etc. are their attributes? Or are you analysing actual conditions within people? Do you always start one wave before diagnosis?

Thanks,
Olena

#5 - 06/12/2024 12:56 PM - Amy Taylor

Hi Olena,

Analysis is at the person level. We are looking at GHQ score over time in relation to arthritis diagnosis so anyone who is diagnosed with arthritis at any point in waves 1-12 is included. But any individual could have between 1 and 11 waves before they are diagnosed. Our time zero is wave of diagnosis but this is a different wave for each person.

#6 - 06/12/2024 03:48 PM - Olena Kaminska

Amy,

Thank you. You have two options. You could either use one of our _lw weights which would be from the last wave in your analysis. This also would be limited to the sample that starts when you start (so likely not all samples will be in your analysis). I suggest you run a quick analysis using this set with our weight. Maybe what you are looking for will be significant and you won't need to do anything else.
The best option would be to create your own tailored weight https://www.understandingsociety.ac.uk/help/training/creating-tailored-weights/
Your base weight can be the _lw weight from the wave when the person is diagnosed. You should pool the data with the time 0 and weight at time 0, which would be your base weight. Base weight is part of tailored weight calculation.

Hope this helps,
Olena

#7 - 07/16/2024 04:34 PM - Understanding Society User Support Team

- Status changed from In Progress to Feedback

- % Done changed from 0 to 80