

## Understanding Society User Support - Support #1864

### Using survey weights in longitudinal analysis

02/23/2023 03:34 PM - Tanya Braune

<b>Status:</b>	Resolved	<b>Start date:</b>	02/23/2023
<b>Priority:</b>	Normal	<b>% Done:</b>	100%
<b>Assignee:</b>	Olena Kaminska		
<b>Category:</b>	Weights		
<b>Description</b> <p>I am using Waves 7, 9 and 11 to conduct longitudinal analyses on changes in fruit and vegetable intake with age. I am using a multilevel (mixed effects) model where the measurement wave is level 1, individual (pidp) is level 2 and psu is level 3. I have looked through all the documentation and the other support cases (particularly Support <a href="#">#1572</a>) but I have not found a solution on how to apply the survey weights to my model. I was advised to use the 'g_indscui_lw' weight for this analysis. The online documentation advises to use the 'survey' package in R but I am struggling to properly apply this to my multilevel model as the resources provided do not go into detail about this.</p> <p>Any recommendations or advice you have on this would be greatly appreciated.</p> <p>Many thanks</p> <p>Tanya</p>			

#### History

##### #1 - 02/24/2023 08:44 AM - Understanding Society User Support Team

- Status changed from New to In Progress
- Assignee set to Olena Kaminska
- % Done changed from 0 to 10

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.

We aim to respond to simple queries within 48 hours and more complex issues within 7 working days.

Best wishes,  
Understanding Society User Support Team

##### #2 - 02/24/2023 08:44 AM - Understanding Society User Support Team

- Private changed from Yes to No

##### #3 - 02/24/2023 08:48 AM - Understanding Society User Support Team

- Category set to Weights

##### #4 - 03/02/2023 03:41 PM - Olena Kaminska

See whether this information may help you:

#### 18. Pooled analysis

You can pool the data however you want. There are three most important points to keep in mind:

1. Always take into account clustering within PSUs with UKHLS data. Taking into account clustering within a person (in case you have multiple entries per person) is optional and could be used in addition to clustering within PSUs. This implies that you don't need to use multilevel models while pooling – you could use the standard svy command if this suits your purpose.
2. When pooling information from multiple waves, especially BHPS waves and UKHLS waves you need to apply additional scaling to weights in order for each wave to contribute a similar level as all others. See question 19 in this document for how to implement it.
3. Define your population carefully. Unlike unpooled analysis, where population definition is straightforward, we find that many users get confused with the population definition in the pooled analysis. A few examples follow presenting the population definition and the data structure:
  - Events, e.g. hospitalization occurrences (staying in a hospital for over 24 hours) observed in GB between 1991 and 2009 and UK between 2009 and 2020. In this situation hospitalization variable would be created and data is pooled from all waves and all people observed in each wave between 1991 and 2020. Note, you are studying events, not people, in this situation.
  - Event triggered situations, e.g. happiness upon marriage observed in UK between 2009 and 2020. If you study the state after marriage – you could pool all the observations after marriage in the data from all the time points. Your data will consist of all marriages and relevant observations following from all waves between 2009 and 2020. You are studying happiness following marriage, i.e. a state following an event (not people).
  - A subgroup defined by a time point, e.g. 11 year olds living in UK between 2009 and 2020. You could pool information from 11 year olds from each wave and analyse them together in one model, which gives you more statistical power. In this situation you will have one observation per person as a person is 11 only once per lifetime (and wave).
  - A subgroup defined by an event where event may happen multiple times, e.g. first year students studying in UK between 2009 and 2020. You could pool first year students from all the years we have in the study. Note, some people may have multiple occurrences of being a first year student.

It then depends on your definition. If you want to study number of books read in a year by the first year students it may be appropriate to count all the multiple occurrences per person. In this situation you don't study people really but 'event triggered states'.

- Time variant state or characteristic, e.g. wellbeing observed in UK between 2009 and 2020. While wellbeing changes over time and it may be more appropriate to study it using a classic longitudinal analysing, there are situations, especially when studying very small subgroups, where pooling may add statistical power. In essence you are studying wellbeing states observed over a specific time period, (again not people). For this you just pool all the information on wellbeing from all the relevant waves.
- It does not make sense to study time-invariant states (e.g. eye colour) with pooled analysis. If you happen to do it, your effective sample size will not be any higher than in an unpooled analysis. So, technically there won't be any gain from pooling, and it would be easier and clearer to avoid it.

Pooling can be cross-sectional or longitudinal. Theoretically, you will be combining 'separate samples of events / states' each of which will have the corresponding weight.

If you are just interested in events (and what happened at the same time / wave) you are looking at pooling cross-sectional information. For this create a new weights variable `new_weight`, and give it a value of the cross-sectional weight from each wave (e.g. `new_weight=a_indinus_xw` if `wave==1`; `new_weight=b_indinub_xw` if `wave==2` etc.)

Alternatively you may be interested in what happens before and / or after a particular event (e.g. studying work pattern for 3 years before birth of a first child and 3 years after for new mothers). In this situation you need to choose a longitudinal weight from the last wave in your analysis for each combination of waves (e.g. for birth at wave 3 where we observe waves 1-6, the weight will be `f_indinus_lw`; for birth at wave 4 with information in the model of waves 2-7 –it will be `g_indinub_lw` etc.).

#### **#5 - 03/08/2023 12:23 PM - Understanding Society User Support Team**

- *Status changed from In Progress to Feedback*

- *% Done changed from 10 to 80*

#### **#6 - 11/30/2023 12:52 PM - Understanding Society User Support Team**

- *Status changed from Feedback to Resolved*

- *% Done changed from 80 to 100*