

Understanding Society User Support - Support #1608

Question about weights for analysis of child and youth data

11/19/2021 03:54 PM - Marie Mueller

Status:	Resolved	Start date:	11/19/2021
Priority:	Normal	% Done:	100%
Assignee:			
Category:			
Description			
Dear Understanding Society support team,			
Today I attended a workshop on using weights in Understanding Society. I am analysing child (0-9) and youth (10-15) data. For the selection of weights I have followed instructions in the user guide and FAQs. However, in the workshop today I realised I may need to tailor my own weights. I wanted to check if this is actually necessary in my case.			
<u>Analysis of child data</u>			
I see there are no specific weights for children between 0 and 9 years. However, looking at the hierarchy of analysis, I suspect I can use the adult self-completion interview weight (_indsc). In the child analysis, I am not using blood sample, nurse visit, youth, or extra 5 minutes interview data. Therefore, the adult self-completion data are my lowest level of analysis (level 2). Can you confirm that the _indsc weight is appropriate in this case?			
<u>Analysis of youth data</u>			
In the youth analysis, the youth data are my lowest level of analysis (level 2). Again, I am not using blood sample or nurse visit data. However, I do use adult self-completion interview data (also level 2). Until today I assumed I would use the youth weight anyway (_ythsc), as it is the lowest in the table in the weighting FAQs document. However, in the workshop I heard that if I use data from different sources at the same level (in this case level 2), I may need to tailor my own weight. I was wondering if this is necessary in my case? My outcome of interest is the SDQ (from the youth data set) and data provided by adults (in this case parents) are only used as covariates in my models. One of these variables (the GHQ) has much missingness (around 30%). I use multiple imputation to impute these missing data. Do you think it is necessary to tailor my own weight for this analysis, or will the youth weight be just fine?			
Thank you very much in advance for your help!			
Best wishes, Marie			

History

#1 - 11/19/2021 03:57 PM - Understanding Society User Support Team

- Status changed from New to In Progress
- Assignee set to Olena Kaminska
- % 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. 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 - 11/19/2021 04:24 PM - Olena Kaminska

Marie,

Thank you. It sounds like the best approach would be for you to create a tailored weight. But you could also start your analysis with a suboptimal weight. Before I suggest which one it is, could you specify which information you use and which waves you use it from? I understand you use information from youth q-re and youth parents' self-completion. Is this correct? Does the information come from the same wave?

Thank you,
Olena

#3 - 11/19/2021 04:46 PM - Marie Mueller

Dear Olena,

Thank you for your rapid response. Yes, for the **youth** analysis, I am using data from the youth questionnaire (SDQ, happiness, self-esteem, sex, age, ethnicity, number of natural parents), data from the adult main interview and self-completion questionnaire (education, GHQ), and data from the household questionnaire (housing tenure).

(For the **child** analysis, I can assume the adult self-completion weight is the most appropriate?)

Thank you very much!

Best wishes,
Marie

#4 - 11/19/2021 04:48 PM - Marie Mueller

I forgot to mention: I use data from waves 1-8, and youth and adult data are matched for each wave (i.e., youth wave 1 + adult wave 1, youth wave 2 + adult wave 2, ...).

#5 - 11/23/2021 12:04 PM - Olena Kaminska

Marie,

Thank you. Do you mean that you pool all youth data from all waves? And for each wave you want to combine youth questionnaire with a parent questionnaire from full and self-completion response from the same wave?

Are you studying youth (and adult information is inferred as attributes of youth)? So, is your analysis at youth level?

When you say you use data for waves 1-8 - do you conduct longitudinal analysis (looking if something at wave 1 affected something at wave 8) or are you trying to combine different waves just to increase sample size (this may be problematic as you will have same people multiple times), or are you just studying each wave separately?

Thanks,
Olena

#6 - 11/23/2021 12:26 PM - Marie Mueller

Hi Olena,

I use youth data from waves 1-8. I pool these data into one data set. I do not run a longitudinal analysis because youth data do not seem to be suitable for longitudinal analysis (i.e., there are no longitudinal weights for youth, and enumeration weights contain too many zeros). This was the conclusion drawn from a different issue in this forum. I pool data of eight waves but run a cross-sectional analysis. I stratify my analysis by age, so in each analysis every individual has only one observation (i.e., there will be no multiple observations). And yes, for each wave I combine youth with parent data, and information from parents are only used as covariates/confounders. Main outcomes and exposures are all for youth, so my analysis is at youth-level. This is why I assumed the youth cross-sectional weight is appropriate. The only problem is with one parent variable, the GHQ, which has 30% missing values. As my sample sizes are small already (due to a focus on London), I need to impute this variable (so I do not lose too many observations in my model). What do you think is the appropriate sub-optimal weight? And what problem would tailoring my own weight solve? Right now I do not actually see any problem with using the youth weight, but I may be missing something.

Thank you!

Marie

#7 - 11/23/2021 12:52 PM - Olena Kaminska

Marie,

First of all, you could use youth in a longitudinal analysis, just keeping in mind that a few years before they are children and a few year after they are adults - hence information would be coming from different instruments. This is why we don't provide youth longitudinal weights - the combination of instruments is very large. Instead we recommend creating tailored weight specific to the combination of instruments in your analysis.

It is ok to pool cross-sectional data, but this is usually done for analysis of events, not people (e.g. getting or losing a job). In each wave an event is a new occurrence.

There are issues with pooling our cross-sectional data for studying people as observations are not independent. How do you ensure that a person is only present once. I am asking because depending on how you do this your pooled data may or may not be representative of a theoretical population. So the weight will depend on this.

Thanks,
Olena

#8 - 11/23/2021 01:10 PM - Marie Mueller

Hi Olena,

Essentially, I am interested in the effect of exposure X on outcome Y. Exposure X is linked data (via LSOA and postcode identifiers). Outcome Y is taken from the youth data set. As I focus on London, sample sizes at each wave are small (say between 400 and 800). To increase the sample size, the initial idea was to pool data of all individuals across all waves into one analysis. However, there was no suitable weight for this analysis, due to the lack of a longitudinal youth weight and too many zeros in the enumeration weight at the last time point (which had been suggested as a sub-optimal weight for this analysis). Instead, I do the same (i.e., pool data of eight waves) but use corresponding cross-sectional weights from waves 1-8. To ensure I only have one observation per individual, I stratify my analysis by age: I pool youth data from eight waves but then run separate analyses for each age group (10, 11, 12, 13, 14, 15 years). This ensures that in each of these separate analyses every individual has only one observation. To summarise my sub-sample selection:

youth (across waves) --> youth in London (across waves) --> 10-year-olds in London (across waves) [and this for each age group]

Thank you!

Marie

#9 - 11/24/2021 09:07 AM - Understanding Society User Support Team

- % Done changed from 10 to 80

#10 - 11/30/2021 10:23 PM - Marie Mueller

Hi Olena,

Can I assume that the youth cross-sectional weight is the sub-optimal (or indeed the optimal) weight for my analysis, or do you need more information from me to be able to confirm this? As my focus is on youth, I don't really see how a different weight (e.g., an adult weight) would be the better choice. I also don't know how exactly tailoring my own weight would help (i.e. what 'problem' it would solve).

Thank you very much in advance for your help!

Best wishes,
Marie

#11 - 12/01/2021 12:15 PM - Olena Kaminska

Marie,

Yes, youth weights will be suboptimal weights for your analysis (assuming analysis within one wave, and not longitudinally). A tailored weight will correct for additional nonresponse of parent self-completion interview (you will have some youth that completed but no parent self-completion interview - these youth will be out of your analysis due to missingness of covariates). If it is only one parent variable that you use (or a handful) you could consider imputation to correct their missingness. This may be especially useful approach if adults have completed full interview but not self-completion - you should have rich covariates for this imputation model.

Best,
Olena

#12 - 12/01/2021 03:01 PM - Marie Mueller

Hi Olena,

Yes, for each wave I only link data of the same wave. After creating 'complete' data sets for each wave, I pool the data of the eight waves and reshape from wide to long format in Stata. Then I run my models on these long data. However, to ensure that I do not have multiple observations in my models, I run separate models for each age group.

Yes, I impute missing values on some of my covariates, using an imputation model including all exposures, outcomes, and covariates.

Thank you very much for all your help.

Best wishes,
Marie

#13 - 01/14/2022 11:59 AM - Understanding Society User Support Team

- Status changed from In Progress to Feedback

#14 - 01/18/2022 11:43 AM - Olena Kaminska

Marie,

Sounds like using youth weights should be a good option for you.

Best,
Olena

#15 - 01/27/2022 02:06 PM - Understanding Society User Support Team

- Status changed from Feedback to Resolved
- Assignee deleted (Olena Kaminska)
- % Done changed from 80 to 100