

Understanding Society User Support - Support #2255

Request for Feedback on Weighting Strategy for Longitudinal Event History Analysis

05/20/2025 12:28 PM - Irene Frageri

<div>Status:Feedback</div> <div>Priority:Normal</div> <div>Assignee:Understanding Society User Support Team</div> <div>Category:Weights</div>	<div>Start date:05/20/2025</div> <div>% Done:50%</div>
<div>Description</div> <div>Dear Understanding Society Support Team,</div> <div>I am writing to seek your advice on the weighting strategy I am using for a longitudinal Event History analysis based on the Understanding Society data. I have consulted the available documentation and discussed with other researchers, but given the specific structure of my data and research design, I would appreciate your expert opinion.</div> <div>My data setup:</div> <div>I have constructed a long-format monthly panel dataset, where each respondent appears in multiple rows. I follow individuals from their entry into the sample until one of the following:</div> <div><div>-they experience the event (first birth),</div><div>-they exit the reproductive age window, or</div><div>-they drop out of the panel (non-intermittent response only).</div></div> <div>As a result, different respondents exit the analysis at different waves. I observed that many weights are zero, which I assume is because the respondent is not part of the OSM (original sample members).</div> <div>My weighting strategy:</div> <div>I use the longitudinal individual weights (<code>_indscus_lw</code>) from each wave (b, c, ..., n).</div> <div>For each wave, I compute the mean weight across all individuals with non-missing value. I use this to rescale the weights: <code>prefix_longitudinalweight = prefix_indscus_lw / mean(prefix_indscus_lw)</code>. This ensures the rescaled weights have a mean of 1. This ensures that the rescaled weights have a mean of 1. Since my dataset is in monthly long format, each individual appears multiple times — once for each month they are observed — and their original weight is repeated across those rows. However, I think that because I compute a mean, this repetition does not affect the validity of the rescaling, as each individual’s weight contributes proportionally to the average.</div> <div>I calculate the total rescaled weight for each wave by summing the rescaled weights: <code>prefix_totalweight = sum(prefix_longitudinalweight)</code>. I then generate a constant variable per wave containing this total for all individuals.</div> <div>I compute an average total weight across all waves: <code>average_longitudinal</code>.</div> <div>I calculate a scaling factor for each wave: <code>prefix_scale = average_longitudinal / prefix_totalweight</code>.</div> <div>I apply the scaling factor to the rescaled weight: <code>prefix_weight_rescaled = prefix_scale * prefix_longitudinalweight</code>.</div> <div>Finally, for each respondent, I assign their weight based on the last wave in which they are observed (prior to their event or censoring).</div> <div>For respondents who are only observed in wave 1, I use the weight from wave 2.</div> <div>I would be grateful if you could let me know whether this approach is methodologically sound, particularly in the context of a monthly, long-format Event History analysis with varying exit points across individuals.</div> <div>Thank you in advance for your time and support. It is truly appreciated.</div> <div>Best wishes,</div>	

History

#1 - 05/21/2025 12:31 PM - Olena Kaminska

Irene,

Thank you for your question. I can tell you that your current method is wrong as you can't use the last observed weight. Weights only represent a population taken from a particular point in time, and can't be taken from different points in time for different people (unless this is directly related to a particular event and not attrition).

To help you choose the best weight I need to know what you want to represent. Are you representing events? Also, are you using survival analysis (the weight for this may be different than for other analysis).

Hope this helps,
Olena

#2 - 05/21/2025 01:51 PM - Irene Frageri

Dear Olena,

thank you very much for your help.

I am conducting a discrete-time event history analysis. Basically it is a logistic regression in which the dependent variable is the occurrence of an event, in my case the conception of a child. I have a long format dataset in which conception=0 for every observation of the subject, except for the last one, in which it is =1 in case the subject has a child (and =0 if the person exits the observation with censoring). Therefore, to answer to your question, I am trying to represent a sample of individuals in their reproductive age (for now, childless individuals but maybe later I will extend to others as well). And regarding survival analysis, yes I think the methodology I am using should fall under the umbrella of survival analysis.

I would also add another point to you which I forgot to mention yesterday: by now I was also using the psu and strata of last available weight.. I guess this is also not correct. could you also advice me on that, please?

Thank you very much again, your help is truly appreciated.

Best regards,

Irene

#3 - 05/22/2025 03:10 PM - Olena Kaminska

- Assignee changed from Olena Kaminska to Alita Nandi

Irene,

Psu and strata can be used from any point of time - they don't change, so you can as well take them from the last wave. For a weight, it sounds like you are looking at longitudinal analysis, so you will need either an issue weight or longitudinal weight.

If your analysis deals with censoring (similar to survival) you may want to use it to correct for attrition instead of relying on weights for that part.

Double check whether this is the case. If so, you can use an `_lw` or an issue (`_li`) weight from any (first wave) of your analysis, but from the same wave for everyone. No additional scaling needed.

There may be alternative ways of setting up your data and therefore weights, and I am forwarding this question to the appropriate team members who can give you a better advice on this.

Hope this helps,
Olena

#4 - 05/22/2025 04:33 PM - Irene Frageri

Dear Olena,

once again, thank you very much for your help and support. I think that yes, my analysis methodology should account for the censoring, because the individual is not present anymore after censoring (so I have a certain amount of observation-rows up till event of censoring). Is this that you are saying, right? That depending whether my methodology accounts for censoring, I might not need to use the weights to account for censoring.

But I am a bit confused: when I need to choose the longitudinal weights, I have both to identify the entry time and the exit time of my observed sample. In this sense, I don't understand when you say "you can use an `_lw` or an issue (`_li`) weight from any (first wave) of your analysis, but from the same wave for everyone". So say my first wave of analysis is wave1.. what should I put for the last wave then? I believe if I put wave 14 I won't have a value for most of the people, which are those individual that have attrition and therefore exit the panel before wave14.

I hope I explain myself well enough.

Thank you also for asking the help of Alita. Looking forward to reach your suggestions.

A big thanks to you both.

Irene

#5 - 06/10/2025 02:38 PM - Understanding Society User Support Team

- File Worksheet ex6 R.pdf added
- File Worksheet ex 6 Stata.pdf added
- Status changed from New to Feedback
- Assignee changed from Alita Nandi to Understanding Society User Support Team
- % Done changed from 0 to 50
- Private changed from Yes to No

Dear Irene,

We thought that you might find useful the exercise on weighting from our Introduction to Understanding Society workshop, particularly section 6.9 Longitudinal weights (attached in R and Stata). Please let us know if that helped to clarify the issue.

Best wishes,
UKHLS User Support team

Files			
Worksheet ex6 R.pdf	462 KB	06/10/2025	Understanding Society User Support Team
Worksheet ex 6 Stata.pdf	400 KB	06/10/2025	Understanding Society User Support Team