

Which weights to use when combining the mainstage and Covid-19 waves of UKHLS

02/20/2024 02:23 PM - James Laurence

Status:	Feedback	Start date:	02/20/2024
Priority:	Normal	% Done:	100%
Assignee:			
Category:	Weights		
Description			
<p>Hi there,</p> <p>I have been reading through the forum support on weighting when using the Covid-19 data and Mainstage data combined, but I was hoping to just get some clarification on a couple of points.</p> <p>I am conducting a study which looks at trends over time in adult mental health across UKHLS Mainstage surveys 2, 3, 4, 5, 6, 7, 8 and 9, and the Covid-19 survey 1, 2, 3, 4, 5, 6, 7, 8 (only the web survey). I end at Mainstage survey wave 9 (survey period Jan 2017-May 2019) as that appears to be the last Mainstage wave that doesn't cover the pandemic period. My data is in long format and I'm using Stata.</p> <p>However, I am unsure of which weights to use precisely for my aims.</p> <p>The first thing I would like to do is to treat the data as pooled cross-sectional data to look at how mental health changes over time (between wave 2 of the mainstage and wave 8 of the Covid survey). So, I'd like to see the level of mental health in the UK at each wave of the UKHLS. Ideally, I would like to model all the waves together. Am I right in thinking that to do so I would need to create a new 'weight' variable, which is the self-completion cross-sectional weight for each wave? I'll explain below...</p> <p>So, for wave 2 of the Mainstage survey, the new 'weight' variable would have the value of <code>b_indscub_xw</code> For wave 3 it would have the value of: <code>c_indscub_xw</code> For wave 4 it would have the value of: <code>d_indscub_xw</code> For wave 5 it would have the value of: <code>e_indscub_xw</code> For wave 6, it seems there are two self-completion, cross-sectional weights (<code>_ub</code> and <code>_ui</code>): would it be <code>f_indscub_xw</code> or <code>f_indscui_xw</code>? For wave 7 it would have the value of: <code>g_indscui_xw</code> (as there is no <code>_ub</code> version) For wave 8 it would have the value of: <code>h_indscui_xw</code> (again, as there is no <code>_ub</code> version) For wave 9 it would have the value of: <code>i_indscui_xw</code> (again, as there is no <code>_ub</code> version)</p> <p>Then, turning to filling in the COVID-19 survey values of the new 'weight' variable it would be: For wave 1 of the COVID-19 survey it would have the value of: <code>ca_betaindin_xw</code> For wave 2 of the COVID-19 survey it would have the value of: <code>cb_betaindin_xw</code> ... For wave 8 of the COVID-19 survey it would have the value of: <code>ch_betaindin_xw</code> For wave 9 of the COVID-19 survey it would have the value of: <code>ci_betaindin_xw</code></p> <p>So, the Stata code would look something like this to give you an idea of what I mean:</p> <pre>gen weight = b_indscub_xw if mainstage_wave==2 replace weight = c_indscub_xw if mainstage_wave==3 replace weight = d_indscub_xw if mainstage_wave==4 ... replace weight = i_indscui_xw if mainstage_wave==9 replace weight = ca_betaindin_xw if cv19survey_wave==1 ... replace weight = ch_betaindin_xw if cv19survey_wave==8 replace weight = ci_betaindin_xw if cv19survey_wave==9</pre> <p>So, it would be one new 'weight' variable, where each wave within each pidp had a weight value which corresponds to the cross-sectional weight for that wave.</p> <p>- Is that the correct approach to take in to treat the data as repeated cross-section data and look at levels of mental health in each wave? - Am I handling the COVID-19 weights correctly, and can I combine the Mainstage (waves 1-9) and Covid-19 surveys (waves 1-9) in this way? - I'm not sure I fully understand the switch between <code>_ub</code> (waves 1-6 mainstage) and <code>_ui</code> (waves 7-9 mainstage). I can only use <code>_ub</code> up</p>			

to wave 5 and only _ui from waves 7 to 9. Is it correct to take the approach I've outlined above, looking at _ui in some waves and _wub in others?

- Also, for wave 6, which self-completion cross-sectional weights should I use? The _ui or _ub?

I hope this makes sense and please do let me know if you require any further clarifications.

Best wishes,

James

History

#1 - 02/20/2024 03:05 PM - Understanding Society User Support Team

- Category set to *Weights*

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

#2 - 02/21/2024 12:10 PM - Olena Kaminska

James,

Yes, your approach is generally correct and the syntax makes sense.

For wave 6 use ui weight (although check if your questions were asked to IEMB sample in memorig variable as some questions were not asked to everyone in wave 6). At wave 6 we had an immigrant and ethnic minority boost, so from wave 6 the coverage is updated with recent immigrants - hence ui weight, which now covers the population with recent immigrants.

If Covid does not have a separate weight for web part, you will have to create this yourself. You could create it from the general Covid weight in one step. For this follow our online training course:

<https://www.understandingsociety.ac.uk/help/training/creating-tailored-weights/>

Hope this helps,
Olena

#3 - 02/21/2024 01:22 PM - James Laurence

Hi Olena,

That's really helpful. Thanks so much.

Best wishes,

James

#4 - 02/23/2024 10:16 AM - Understanding Society User Support Team

- Status changed from *New* to *Feedback*

- Assignee deleted (*Olena Kaminska*)

- % Done changed from *0* to *100*

- Private changed from *Yes* to *No*