Understanding Society User Support - Support #228

Wave 3 equivalent of b_indscbh_xw

12/12/2013 10:42 AM - David Bayliss

Status:	Closed	**Start date:**	12/12/2013
Priority:	Normal	**% Done:**	100%
Assignee:			
Category:	Weights		

Description

Good morning,
I am conducting analysis of a several waves of data spanning BHPS and USoc, using only the BHPS cohort. I am including cases with missing waves and so I am using cross-sectional weights for each wave. For USoc wave 2 I used b_indscbh_xw as the cross-sectional weight for the self-completion form for adults. In USoc wave 3 this variable is not available, and is instead replaced by a combined weight c_indscub_xw. My question is whether this 'combined' weight is suitable for a cross-sectional analysis of only the BHPS cohort, or whether it is intended to be used when analysing the combined BHPS, GPS, EMB samples together?
If it cannot be used to analyse only the BHPS cohort, please could you advise me which weight, if any, is appropriate.
Best wishes,
David Bayliss

History

#1 - 12/12/2013 12:53 PM - David Bayliss

Sorry - this is not an IP3 question (mistook this for wave 3).

#2 - 12/16/2013 06:20 PM - Olena Kaminska

David,

You are right that the wave 3 cross-sectional weight is a combined BHPS + GPS + EMB weight, so using it only for BHPS will give biased results.

There are two options for you:
1. Simple but crude: take longitudinal weight for BHPS for wave 3. Longitudinal weight also represents cross-sectional population (with tiny differences due to recent immigrants longitudinal and cross-sectional weights should give similar results);
2. Take the cross-sectional weight but scale it (and all other waves) such that each wave has the same weighted sample size. The explanation is below.

The aim of pooling data from different waves is often to represent events (e.g. number of events in the last 20 years). This works fine if each wave has the same number of people. As you know even BHPS does not have the same number of people (there is a boost in 1999 for example). While each single wave (once weighted using cross-sectional weight) represents the population in that year, the waves that have higher number of people will contribute to your estimates more than the waves with smaller number of people. Even before wave 3, if one uses pooled BHPS data to study events in GB over the last 20 years the years before 1999 would be underrepresented and therefore events after 1999 would make a larger contribution on your estimate.

It is easy to correct for this. 1) first calculate the weighted sample size for each wave (total of weight variable will give you this - note the weight variable should have mean of one); 2) take the average of weighted sample sizes across the waves you use; 3) divide the average by the weighted sample size for each year to get the scaling factor; 4) multiply the scaling factor for each wave by its cross-sectional weight. Use this product as a new weight for pooled data. This will ensure that each wave has the same weighted sample size and therefore each year has the same importance in your estimate. For example if one wave has weighted sample size of 1000 and another has 2000, then the average is 1500, the scaling factor for wave 1 is 1500/1000=1.5; for wave 2 is 1500/2000=0.75. The new weighted sample size (you could check this) will be the same in both waves (1500).

Treat the new BHPS + GPS + EMB sample in the same way - the scaling factor will be small for this wave and the scaling factor for BHPS waves will be over 1. But after correction your analysis will have higher precision (then if you were to not use GPS and EMB data) and will correctly and evenly represent all years. Finally, this method also corrects for differences in sample size due to non-response as well. In other words it should be used with pooled data even when there aren't sample boosts.

Hope this helps,
Olena

#3 - 12/17/2013 02:57 PM - David Bayliss

Hi Olena,

Thank you very much for your response. I do need to produce some statistics for which your advice will be necessary. However, my main analysis is longitudinal, making use of information about how the same people have changed over different waves (going back to the mid-2000s). I use multilevel and structural equation models which are well adapted to deal with non-balanced data. Prior to the wave 3 release I used cross-sectional weights because the longitudinal BHPS and USoc weights are only provided for cases which were sampled at each and every wave (and obviously many

people do not have a continuous record for the whole 20+ years). Because multilevel and SEM can deal with non-balanced data, using the cross-sectional weights allows me to retain people who may have missed some years so they can contribute to cross-sectional estimates where they did responded. However based on what you have said, I don't think there is a suitable weight for this purpose. Are you aware of anyone having tried a model based approach, or the feasibility of such an approach for such a complex survey design?

Best wishes,
David

#4 - 12/18/2013 09:31 AM - Redmine Admin

- % Done changed from 0 to 50

#5 - 01/08/2014 12:47 PM - Olena Kaminska

David,

Cross-sectional weights are not appropriate for your situation - this is because some people that have cross-sectional weights by design are not followed and therefore will not have longitudinal information (they are called TSMs - temporary sample members - i.e. those who live with someone who if followed). Because of the way the cross-sectional weight is calculated (through a weight-share method) the weight for those people who remain in your longitudinal analysis will have wrong weights. Your analysis in this situation is likely to underrepresent people who move in with others (young renters? young couples? etc.)

There is an easy way out for your situation though. I understand that your analysis controls for missingness - that's wonderful. You should think about it in the following way: 1) there is design and nonresponse that occurs before the first wave that you use in your analysis (e.g. wave j if you start with the year of 2000) - this you should take into account through a weight; 2) there is nonresponse (attrition) that occurs afterwards, sometimes nonmonotone attrition (e.g. missing only one wave) - this will be taken into account as part of the model.

Therefore use the appropriate weight from wave j of BHPS (or the first wave from which you start your analysis), and as long as the model controls / adjusts for the following missingness due to attrition - you are fine.

Hope this helps,
Olena

#6 - 01/15/2014 11:00 AM - Redmine Admin

- Status changed from New to Closed

- % Done changed from 50 to 100