

## Understanding Society User Support - Support #723

### Combining USOC/BHPS and zero weights

02/14/2017 05:16 PM - David Hussey

<b>Status:</b>	Resolved	<b>Start date:</b>	02/14/2017
<b>Priority:</b>	High	<b>% Done:</b>	100%
<b>Assignee:</b>	Peter Lynn		
<b>Category:</b>	Weights		
<b>Description</b>			
<p>I have a couple more very small queries following on from our analysis; these are mainly for re-assurance that we've done things correctly.</p> <p>1) With regard to the Youth data, is it ok to combine BHPS and USOC? We couldn't find a combined weight but assumed this was still ok given that it's done for adults. Therefore, when I created my weights (for w2/4 and w2/6), I started off by combining the weights thus (after checking the overlap):</p> <ul style="list-style-type: none"><li>• select cases with a non-zero weight for USOC or BGPS. sele if not (b_ythscus_xw =0 and b_ythscbh_xw =0).</li><li>• calculate weight (wt) as either USOC weight or BHPS weight. compute wt=b_ythscus_xw. compute samptype=2. if wt=0 samptype=1. if wt=0 wt=b_ythscbh_xw. Val labels samptype 1 "BHPS" 2 "USOC".</li></ul> <p>I then re-scaled the weights to average 1 within samptype and went on to model response to wave 4/6 contingent on wave 2, using interactions with samptype (USOC/BHPS) to take account of possible differences in response process between the two surveys. If it's not ok to combine USOC and BHPS then my weights should still be fine given the way I've done them but it'd be helpful to know whether or not it's ok to do so.</p> <p>2) In the adult data, there seem to be many zero weights, more than we expected. We'd like to know if these are mainly related to students living in halls and other institutional addresses? E.g. among wave 5 fully productive interviews, the following syntax:</p> <pre>TEMP. SELECT IF e_outcome = 11. FRE e_indinub_xw.</pre> <p>yields 4,481 cases with 0 value for e_indinub_xw (cross-sectional adult main interview weight for USoc &amp; BHPS samples).</p> <p>Many thanks for your help.</p>			

#### History

##### #1 - 02/14/2017 05:23 PM - Victoria Nolan

- Status changed from New to In Progress
- Assignee changed from Olena Kaminska to David Hussey
- % Done changed from 0 to 10
- Private changed from Yes to No

Assigned to Peter.

##### #2 - 02/14/2017 10:58 PM - Peter Lynn

- Target version set to X M
- % Done changed from 10 to 80

David,

Re. point 1), this all looks broadly fine to me. The only caveat is that the two sets of weights for the different samples are designed to represent slightly different populations. The BHPS weights represent the UK population excluding any households consisting solely of people who entered the

country since 2001 or are descended from such immigrants, while the UKHLS weights represent the UK population excluding any households consisting solely of people who entered the country since 2009 or are descended from such immigrants. In principle, one should up-weight the UKHLS post-2001 immigrant households to compensate for their absence in BHPS (this is exactly what we do in the "ub" weights), but I'm not sure how much difference this would make in practice.

Re. point 2), zero weights will arise whenever a household does not contain any adult who has been enumerated at every wave since the relevant start point (wave 2 in the case of "ub" weights; wave 1 for "us" weights). This is because the xw weights are derived from the lw weights and you need at least one person to have a non-zero lw weight in order for it to be shared to the other household members.

HTH,

Peter

**#3 - 02/27/2017 09:46 AM - Victoria Nolan**

- Status changed from In Progress to Closed

- % Done changed from 80 to 100

**#4 - 02/27/2017 02:32 PM - David Hussey**

Many thanks for the reply. I have a couple of follow-up questions re point (1):

1) May I just clarify that there isn't a ub weight for the Youth data?

2) The respective profiles of w2 youth participants weighted by the two weights - b\_ythscus\_xw and b\_ythscbh\_xw - are rather different, which is partly why I queried the method I used to combine them. E.g. the aren't any 10 year olds in the BHPS sample (just 11-15 year olds), the proportion in owner occupied tenures is 8 ppts higher in the BHPS sample compared to the USOC sample and the proportion in urban areas is around 6 ppts lower. Given that the two are meant to be representing very similar populations the differences look rather large. Is there an explanation? Should we be concerned about this or is it ignorable?

**#5 - 02/27/2017 02:34 PM - Victoria Nolan**

- Status changed from Closed to In Progress

- % Done changed from 100 to 80

**#6 - 03/06/2017 05:14 PM - Peter Lynn**

David,

1. n\_ythscub\_xw exists from wave 3 onwards, but not for wave 2.

2. I don't know the answer and I am concerned! I'm looking into it and will get back to you. There may be an error in the weight for the BHPS sample.

Peter

**#7 - 04/27/2017 01:53 PM - Victoria Nolan**

- Assignee changed from David Hussey to Peter Lynn

**#8 - 10/19/2017 11:33 AM - Olena Kaminska**

With regard to number 1: you can combine the weights on your own and your thinking is correct. Your weight will be close to ours but imperfect in underrepresenting recent immigrants.

With regard to number 2: thank you for pointing this error to us. We were aware of this and this was corrected for all other weights but for the cross-sectional youth weight for w2. This is now corrected and the updated weight will be released with w7 release. Thanks again.

**#9 - 10/19/2017 11:33 AM - Olena Kaminska**

- % Done changed from 80 to 100

**#10 - 02/12/2018 12:29 PM - Alita Nandi**

- Status changed from In Progress to Resolved