Understanding Society User Support - Support #296

zero value weights using c_indnsub_xw

08/21/2014 11:16 AM - Esther Curnock

Status:	Closed		Start date:	08/21/2014
Priority:	Normal		**% Done:**	100%
Assignee:				
Category:	Weights			

Description

Hi,

I have been carrying out cross-sectional analyses using the combined wave 2 & 3 nurse assessment dataset 'xindresp_ns', along with data from the main survey of the corresponding wave. Having read through the user guide I have been using the survey weight c_indnsub_xw, as I believe this is the correct one for the type of analyses I'm doing. However there seems to be a significant proportion (7.8%) of the combined nurse assessment dataset that have a weighting value of zero when this is applied, and I would be grateful if you could explain this so I can decide whether to continue using the weighting.

The user guide seems to indicate that c_indnsub_xw is equal to the longitudinal weight c_indnsub_lw for households with no TSM, and that c_indnsub_lw itself is calculated using a method that includes multiplication by the nurse inclusion weight b_indnsub_li. I think it might be this inclusion weight that leads to the zero weights for c_indnsub_xw. If so, does this mean that individuals with a zero weight are basically those that had nurse data collected despite falling outside the inclusion criteria such an assessment?

To try to understand the problem, I also looked at the separate nurse assessment datasets for each wave ('b_indresp_ns' and 'c_indresp_ns'). It seems that applying the cross-sectional weight 'b_indnsus_xw' to the wave 2 dataset results in only 0.7% of cases with weights of zero, whereas applying 'c_indnsbh_xw' to wave 3 dataset has 10.8% with weighting values of zero. I wasn't sure why there should be such a big difference between the two, but in any case, as the proportion of the combined dataset derived from the wave 2 GPS sample is much greater than those coming from the wave 3 BHPS group this didn't really explain the issue of high zero weights in the combined dataset.

Many thanks,
Esther

History

#1 - 09/29/2014 04:28 PM - Peter Lynn

Esther,

First, apologies for the delay in responding to this. It seems to me that the main factor contributing to zero values of c_indnsub_xw is non-response to the individual interview in one of waves 2 and 3, rather than failing the criteria for nurse assessment. c_indnsub_xw is, as you say, based on c_indnsub_lw, which will only have a non-zero value if the sample member has responded to the individual interview in both waves 2 and 3, as well as the nurse assessment. If there are no persons in the household with a non-zero value of c_indnsub_lw, then the individual will get a zero value of c_indnsub_xw. The explanation would therefore seem to be that 7.8% of respondents to the nurse assessment were in a household (at wave 3) in which no-one had responded at both waves as well as to the nurse assessment. Specifically, it would appear to be the case that 13.1% of the BHPS nurse assessment respondents were in households where no-one was interviewed at wave 2 and 6.1% of the GPS nurse assessment respondents were in households where no-one responded at wave 3:

. ta ubxw_ind nursewave, col

```
c_indnsub_ |
xw |  Wave of nurse visit
indicator |   Wave b    Wave c |    Total
-----------+----------------------+----------
No weight |     953        662 |    1,615
|    6.09      13.10 |     7.80
-----------+----------------------+----------
xw wgt |   14,693      4,391 |   19,084
|   93.91      86.90 |    92.20
-----------+----------------------+----------
Total |   15,646      5,053 |   20,699
|  100.00     100.00 |   100.00
```

Incidentally, the explanation for the difference between the proportion of zeros in b_indnsus_xw and c_indnsbh_xw is because b_indnsus_xw is defined for each person who responded to the nurse assessment and to the individual interview at wave 2, whereas c_indnsbh_xw is only defined if you responded at waves 2 AND 3 (and the nurse assessment):

. ta wgtind2 nursewave, col

```
Nurse xw |  Wave of nurse visit
  weight  |    Wave b      Wave c  |     Total
----------+-----------------------+----------
No weight |       114         548  |       662
          |      0.73       10.85  |      3.20
----------+-----------------------+----------
   xw wgt |    15,532       4,505  |    20,037
          |     99.27       89.15  |     96.80
----------+-----------------------+----------
    Total |    15,646       5,053  |    20,699
          |    100.00      100.00  |    100.00
```

Peter

#2 - 09/30/2014 12:41 PM - Peter Lynn

-------- Original Message --------
Subject: RE: [Understanding Society User Support - Support #296] zero
value weights using c_indnsub_xw
Date: Tue, 30 Sep 2014 10:29:11 +0000
From: Esther Curnock <Esther.Curnock@glasgow.ac.uk>
To: info@understandingsociety.ac.uk <info@understandingsociety.ac.uk>

Hi Peter,

Many thanks for this clarification, though I admit the rationale for these choices relating to how the weights have been calculated remains unclear to me.

The conclusion I have come to is that if I want to make use of all available nurse assessment data (wave 2 and 3 combined), and tie this in with variables taken from the corresponding main survey data (wave 2 main survey for individuals where nurse assessment data was collected at wave 2, and wave 3 main survey for those where nurse assessment data was collected at wave 3), then there is no logical weight to use.

From what you have described, it does not appear to make sense to use c_indnsub_xw, and thereforeexclude 6.1% of GPS sample who have nurse assessment data & relevant wave 2 data, just because no one in the household went on to participate in wave 3 (when wave 3 data is irrelevant for the cross-sectional analyses), as well as exclude 13.1% of BHPS sample who have both nurse assessment data & relevant wave 3 data just because no one in the household participated in wave 2 (using c_indnsbh_xw has the same issue in this case). One option might be to use b_indnsus_xw, but this would only apply to the GPS sample.

Making use of every person's data available from the nurse assessments is needed to achieve the sample sizes required to achieve reasonable confidence intervals and avoid type II errors. I might be missing something here, but it seems there is no way to carry out cross-sectional analyses of the nurse assessment data whilst applying weights and making full use of available data at that specific wave.
This seems an odd situation as I imagine use of the data in this way would be a reasonably common requirement.

Esther

#3 - 09/30/2014 03:13 PM - Peter Lynn

Esther,

This is essentially a trade-off between bias and variance. As you say, using all the cases but no weighting will (appear to*) minimise the variance, whereas weighting is designed to reduce the bias (due to non-response), but comes at the price of sacrificing some sample.

The reason that some sample has to be sacrificed is that those cases are devoid of the relevant auxiliary data needed to make the weighting adjustment, due to not having responded at the other wave. This is not the only way to do weighting. We could have kept those cases in by using a much slimmer set of auxiliary variables (e.g. only sample frame/first wave variables), but that would have weakened the weighting's power to reduce bias. It would also complicate the task of producing the weights, as each weight would have to be re-calculated 'from scratch' at each wave, rather than starting with an existing (previous wave) weight and making a relatively simple adjustment.

Whether to use our weights, some other user-specified weights, or no weights, is ultimately a judgement call. Sometimes the best path is obvious and sometimes it isn't. Sorry if this seems a bit unhelpful, but I think it is an accurate reflection of the situation.

*In reality, you are underestimating the true variance of estimates if you use unweighted data and ignore non-response (i.e. treat your sample as if it were a SRS with 100% response). The variance in response probabilities is still present in the data, but the standard error estimation ignores this, whereas it would take it into account if weighting were used.

Peter

#4 - 09/30/2014 03:36 PM - Esther Curnock

Thanks Peter,

Realsing the other waves include relevant auxillary data for calculating the weights (which is not all available through the main survey of the wave corresponding to the nurse assessment data) helps make sense of the situation. Framing my decision-making as a trade off between bias and variance is also helpful.

Thansk for your input,

Esther

#5 - 10/16/2014 08:13 AM - Redmine Admin

- Category set to Weights

- Status changed from New to Closed

- Target version set to X Ns

- % Done changed from 0 to 100