

Understanding Society User Support - Support #453

Zero value weight with nurse data combined to blood sample and main sample (w2)

11/24/2015 08:15 AM - Gaelle Albertus

| | | | |
|---|----------------|--------------------|------------|
| Status: | Closed | Start date: | 11/24/2015 |
| Priority: | High | % Done: | 100% |
| Assignee: | Olena Kaminska | | |
| Category: | Weights | | |
| Description Hi, I am working on the nurse health assessment data combined to the blood sample data and the full main interview. I focus only on the wave 2. My aim is to describe the biomarkers and use them to construct a score representative of health condition according to gender and age. My problem is about the weighting strategy. I already posted a message on the forum (#430) and now I consider the indbdub_xw weight variable. <ul style="list-style-type: none">• Why is there people who have a weighting value of 0 ? I count 670 individuals in this case from n=9882. In addition, most of them do not have missing values on biomarkers so, why do we not use them ?• I have to delete individuals who have missing values (because I would use a PCA on my data). Is it still correct to use the indbdub_xw variable ? Thank you for your response. Gaelle | | | |

History

#1 - 11/24/2015 08:55 AM - Redmine Admin

- Category set to *Weights*
- Assignee set to *Olena Kaminska*
- Target version set to *M2*

#2 - 11/24/2015 11:20 AM - Olena Kaminska

Gaelle,

We indeed expect some people to have 0 weights in nurse cross-sectional dataset. These are mostly TSMs in the household where none of the OSMs has non-zero weight or OSMs that missed wave 2 (if they are part of GPS or EMB sample) or wave 2 or 3 (if they are part of BHPS sample). Nurse weight is conditional on enumeration weights in previous waves, and if the latter is zero the nurse weight is zero too. For further details please read the technical description of weights in the nurse user guide and in the mainstage user guide.

Yes, you should use the weight even if you have missing data on the variables of interest. But note that with the weights you correct for household and person nonresponse. Any nonresponse on items won't be correct and you may want to either keep it in mind or correct it additionally.

Hope this helps,
Olena

#3 - 11/26/2015 01:23 PM - Gaelle Albertus

Olena,

Thanks a lot for your response.

I read the User Guide and several questions on the forum (in particular [#127](#), [#137](#) and [#296](#)) and I noticed there are zero values because :

- they missed the previous wave (*new entrants : 34 cases among 670*)
- they are TSMs participants. I try to recognize them by using a_sampst (for TSMs from wave 1) and b_sampst variables: they correspond to the new entrants (*the previous 34 cases*).
- they missed wave 2 main interview (*none of them here even if they have a lot of missing values on w2*)

Thus, **I have 640 people with a weight of 0 and they do not satisfy the conditions above. I don't understand in which step of the calculation the weight (user guide health assessment w2 w3, section "Technical details") is it possible to obtain a weight of 0 ?**

I am quite sure I am missing something, could you point it out ?

I am still disturbed by a weight of 0 because if they have some non-missing values, why do we not take them into account ? By using a lower weight for example but not zero ?

I read "Non-EM persons in the EMB (who are TSMs) are given a design weight of 0 while non -EM persons in the GPS are given the household design weight".

Why do non-EM persons in the EMB have a design weight of 0 ? Design weights adjust unequal selection probabilities, so the less likely a person is to be included, the higher his design weight is, isn't it ?

Anyhow, **why we put a weight to 0 in general ?**

I don't know if I am really clear and I apologize for my confusion. I prefer correctly understanding the weight variable before using it in my analysis.

Thanks a lot for your help.

Gaelle

#4 - 12/07/2015 04:22 PM - Olena Kaminska

Dear Gaelle,

Indeed, you are right in thinking that generally for most of the studies all people who provide information receive a weight. This is particularly true for the studies that are designed to be one time off cross-sectional and aim to obtain all the information in one interview.

Understanding Society designed to measure different information, most of it over time. It is also designed to combine information over time and across instruments. The selection probabilities and response probabilities (which the weights reflect) depend on the complexity of the design. Note, we have some people who were selected in 1991, 1999, 2001 and 2007/2008. Many people also were born into the interviewed households and some join and leave at different points. There are complicated following rules that we implement as well as innovative methods to combine BHPS, EMB and GPS samples that we used. To fully understand how the weights are created you would need to read the sample design and following rules sections in addition to weight sections from BHPS documentation, UKHLS mainstage documentation and nurse / biomarker documentation.

The cross-sectional weight for blood is designed in mind of the users who are interested in combining it with the nurse information, and information from mainstage full interview (either wave 2, wave 3 or both). Such weight satisfies a number of scenarios. It is possible to create a separate weight for each scenario but this would lead to over 1000 weights which may be even more confusing for users.

For the data that you are looking at (and I am looking at combined blood weight indbdub_xw), I see 1303 weights with 0's for those who gave blood information.

Most of these can be explained by the major categories below:

348 are TSMs who live in a household where no OSM has a non-zero weight;

528 didn't respond in w3;

96 didn't give full interview at w2;

215 didn't give full interview in wave 1;

10 died by the time of w3;

16 moved out of scope by w3.

I hope this helps and gives you more confidence in our data,

Olena

#5 - 12/14/2015 07:45 AM - Gaelle Albertus

Hi Olena,

Thanks a lot for your helpful response.

I already read the User guide of both main stage and nurse health assessment in addition to the sample design written by P.Lynn.

I am not familiar with weighting, that's why I still have questions about it. Here, I only focus on the wave 2.

I understand that indbdub_xw weight variable mainly takes into account :

- GPS design weight (non equal selection probabilities)

- adjustment for non-response adult main interview w1 and w2

- adjustment for inclusion to nurse visit

- adjustment for non-participatin to nurse visit (propensity to respond)

So, a weight equals to 0 can be a consequence of these points above.

In other words, TSMs (in a household where no OSM has a non-zero weight), people who do not participate to the full interview, people who do not participate to wave 1 have a weight equals to 0.

I wonder if it is possible for a user like me to obtain these figures ? Are the variables available ?

I find the results below, on the merged table from xlabblood, b_indresp_ns and b_indresp:

1303 individuals have a zero weight.

681 for wave 2

Among the total of 681 people, 9 have not participated to the main interview w2.

tab b_elig if indbdub_xw==0

670 are "eligible - in nurse visit : productive"

tab b_elig b_sampst if indbdub_xw==0

Among the previous 670 individuals, only 34 are TSMs.

I also used b_ivfio, b_ff_ivlowl, b_ff_everint, b_newentrant ...
Finally, I merge my data with a_indall to see who does not match.

My objective is to use biomarkers and some other measures of health to create a score representative of wear-and-tear accross lifecourse.
In this way, I think I do not need to use weights (I do not necessary want to be representative).
However, I start to describe my population by age and in this case weighting is essential.

Thank you again for your help and the time you need to answer me. Understanding society is a very large survey and I aim to use correctly your data.
Have a nice week,

Gaelle

#6 - 12/15/2015 07:27 PM - Olena Kaminska

Dear Gaelle,

First, let me answer a question whether you need weights. You don't need weights if you want to talk about Understanding Society participants who provided nurse and blood data. Usually researchers are interested in talking about the population that our sample represents – in this situation you need to use weights (unless you use any other correction). You may see an issue when you look at age distribution, but note that our participates have different distribution on some important variables to the population like age, gender (especially within age groups), overall health, disabilities, ethnic minorities, education, among many others. The weights in analysis are rather important.

To answer your other question – yes it is possible and should be relatively easy for anyone to trace 0 weights. Among the categories that you tested you also need to look at:

~~refusal to give blood (look if the person is on the blood data);~~

giving full interview in w2 (look if a person is present at b_indresp.dta, but exclude proxy interviews – check ivfio variable)

- full interview in w3 (c_indresp.dta – same as above)

- full interview in w1 (for gps and emb parts)

- for bhps people – in all 19 waves of BHPS where they were eligible to respond (note, some joined in 1999, or 2001 – this needs to be taken into account)

- on a positive side, we add new entrances like those born since 1991, 1999 or 2001 or turning 16 since 1991, 1999, 2001 and 2007/ 2008 (depending on the sample of origin) – these get positive weight even if they were not 16 in other waves if they responded since.

- also on a positive side: anyone with 0 weight following the above rules but who has a household member with a positive weight will have a positive cross-sectional weight.

If you use the above rules you should come very close to the figure of 0 weights that we have.

Hope this helps,

Olena

#7 - 12/21/2015 09:51 AM - Gaelle Albertus

Thanks Olena, it helps a lot !

Thank you again for the time you take to answer me. I better understand the use of weight and why some people have a weight equals to 0.

I wish you a happy end of year.

Best regards,

Gaelle

#8 - 12/23/2015 10:37 AM - Redmine Admin

- Status changed from New to Closed

- % Done changed from 0 to 100