

## Understanding Society User Support - Support #430

### Sample size and weight within blood sample w2

10/09/2015 09:25 AM - Gaelle Albertus

<b>Status:</b>	Closed	<b>Start date:</b>	10/09/2015
<b>Priority:</b>	High	<b>% Done:</b>	100%
<b>Assignee:</b>	Olena Kaminska		
<b>Category:</b>	Weights		
<b>Description</b>			
Hi,			
I am working on Nurse Health Assessment database in order to use data related to biomarkers and health. I decided to use the wave 2 from the survey combined with the main sample (wave 2 - GPS sample).			
Therefore, I am contacting you because the user manual gives a number of subjects equals to 10,175 while when I import the database (with STATA), I only have 9,906 individuals for wave 2.			
In addition, I took a look at the variable called "b_nurout" in the b_indresp file from Nurse Assessment and I obtained 9 957 individuals belonging to "nurse visit conducted - blood sample sent to lab".			
How could you explain this difference ?			
Next, I thought about using the cross-sectional weight for GPS sample only (b_indnsus_xd) but I am wondering if I could use it even if I don't have the correct number of subjects. Is this weight is still accurate givent that I miss approximately 300 individuals ?			
Thank you for helping me.			
Kind regards,			
Gaelle			

#### History

##### #1 - 10/09/2015 02:44 PM - Olena Kaminska

Gaelle,

Thank you for your question. From your message I understand that you want to use information from three sources: nurse visit, blood and full adult interview. You also probably want to represent the UK (or some subgroup of it).

If so, you should use information from nurse visit and blood (from both waves) together with the full interview from either wave 2, wave 3 or both. For this, use blood cross-sectional weight called indbdub\_xw which can be found in xlabblood\_ns.dta dataset. This should give you 11955 people.

Note, b\_indnsus\_xd weight for GPS sample is a technical weight and should be used only by advanced users who would model nonresponse and selection probabilities themselves.

Let us know if you have further questions,

Olena

##### #2 - 10/12/2015 09:23 AM - Gaelle Albertus

Hi,

Thank you very much for your response.

Indeed, I would do a cross-sectional analysis of wave 2 data only, with the nurse visit and the blood sample combined to the full interview. I aim to obtain a *representative sample of UK population*.

Therefore, I will use the *indbdub\_xw* variable for weighting, just like you said.

Yet, I would like to understand because if I take a look on your [weight reference table page 12 in your User Guide of Nurse Health Assessment](#), you have a different weight for wave 2, for wave 3 and for wave 2 and 3 combined. I noticed this table refers only to the nurse visit and full interview data. So my question is why do we not have the same kind of weight with the blood sample? Thus, I don't understand why the weighting variable is the same for wave 2 and wave 3. For example, what weighting variable should I use if I work on wave 3? and both? Is it indbdub\_xw also?

My second problem is about the number of individuals. I obtain indeed 11,955 individuals (9,225 for wave 2 and 2,730 for wave 3) with the following commands:

```
> use "xlabblood_ns", clear
```

```
> tab wave if indbdub_xw!=0
```

But, this figure (for wave 2) does not correspond to the number of subjects given by the User Guide (N=10,175).

Many thanks for your help.

Kind regards,

Gaëlle

**#3 - 10/12/2015 09:25 AM - Redmine Admin**

- *Category set to Weights*
- *% Done changed from 0 to 70*

**#4 - 10/27/2015 10:23 AM - Olena Kaminska**

Dear Gaëlle,

Thank you for your questions. First, we indeed provide blood weight for combined sample of GPS and BHPS, and therefore for waves 2 and 3 (as these samples got blood collected at different waves). The weight `indbdub_xw` is appropriate for the following situations:

- if you are interested only in blood information;
- if you are interested in blood + nurse information;
- if you want to combine blood (with or without nurse information) with wave 2 full interview only;
- if you want to combine blood (with or without nurse information) with wave 3 full interview only;
- if you want to combine blood (with or without nurse information) with waves 2 and 3 full interviews;

Furthermore we create blood weights in each of the waves for longitudinal analysis (either starting at 1991 -`indbd91_lw` - or at wave 2 of UKHLS - `indbdub_lw`) which are for the use of blood (with or without nurse information) and full interviews starting in 1991 or w2 of UKHLS and onwards up to and including current wave.

The above should cover most of the situations that an analyst is interested in, we hope. In the rare situation we have users who can use information only from w2 or from w3 just because it is very important that some questions are asked before the nurse visit or blood tests. In this situation we recommend using w2 nurse weight or w3 nurse weight as suboptimal weights. There are w2 and w3 nurse weights separately because of the procedure used to calculate weights: these weights precede the combined weight for w2 and w3. For more details see the technical description of nurse weights.

To answer your second question, the user guide number for w2 (10,175) refers to number of people who agreed to give blood samples. Note that these samples were then collected with some success, sent to the lab, the blood was then stored, defrosted and the information was extracted some years later - at each stage there is a level of loss due to a number of varied factors. The usable information obtained was for 9225 people and these are used for weights as these are people you can include in your analysis.

Hope this answers your questions,  
Olena

**#5 - 11/09/2015 01:02 PM - Redmine Admin**

- *Status changed from New to Closed*
- *% Done changed from 70 to 100*

**#6 - 11/10/2015 10:37 AM - Gundi Knies**

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