

## Understanding Society User Support - Support #1610

### Weighting method for IP4 refreshment sample

11/23/2021 08:54 AM - Jun Wang

<b>Status:</b>	Resolved	<b>Start date:</b>	11/23/2021
<b>Priority:</b>	High	<b>% Done:</b>	100%
<b>Assignee:</b>			
<b>Category:</b>			
<b>Description</b>			
Hi,			
I'm user from Peking University, my research interests include survey sampling and data integration.			
Recently, our team is considering about adding refreshment sample to our longitudinal survey in next wave while keep PSUs the same as they were in Wave 1. I'm writing to inquire about some details about design weights calculation procedures for refreshment sample(IP4 for example).			
I have read the user guide and technical report the website provide, but there are some questions:			
1. as the post-sectors were selected in IP1 and refreshment sample units in IP4 were sampled in those IP1 post-sectors, how to calculate inclusion probabilities for post-sectors in IP4? are inclusion probabilities for those IP1 post-sectors the same for IP4 or updated by new sampling frame?			
2. as IP4 addresses in refreshment sample are selected using systematic randomly amongst addresses not already selected for the IP1, how to calculate inclusion probabilities for those addresses in refreshment sample?			
3. if a household member A in IP3 moved to a IP4 selected address, and A is selected into IP4 refreshment sample, how to calculate design weights for A?			

#### History

##### #1 - 11/24/2021 09:05 AM - Understanding Society User Support Team

- Status changed from New to In Progress
- Assignee set to Olena Kaminska
- % Done changed from 0 to 10

Many thanks for your enquiry. The Understanding Society team is looking into it and we will get back to you as soon as we can. We aim to respond to simple queries within 48 hours and more complex issues within 7 working days.

Best wishes,  
Understanding Society User Support Team

##### #2 - 11/24/2021 09:06 AM - Understanding Society User Support Team

- Private changed from Yes to No

##### #3 - 11/24/2021 11:20 AM - Olena Kaminska

Dear Jun Wang,

Thank you for your question.

1. Selection probabilities for refreshment sample are independent of the selection probabilities for the first sample. You will know them from your design;
2. The combined probabilities are the sum of the two;
3. For the rare possible case the actual probability does not depend on whether the person appears twice or not - everyone's probability is the sum of the two probabilities.

If you could let us know which study you are working on, it would help if we can include this for our impact case.

Best wishes,  
Olena

##### #4 - 11/26/2021 08:00 AM - Jun Wang

Dear Olena,

I'm working on CHARLS project. Thanks for your reply, but there is still some questions here:

According to the sample design of Innovation Panel, the inclusion probability for a respondent in Wave 1 can be written as  $\times(\text{inclusion probability for address})\times(\text{inclusion probability for a household})\times(\text{adjustment for multiple dwellings and household})$ . As [Innovation Panel, Wave 1-13: User Manual] [Sample Design part] wrote, 120 PSUs (post-sectors) are all kept fixed after they were selected as SSUs in Wave 1 sample. Refreshment sample addresses in Wave 4 were selected in those original PSUs (post-sectors) excluding addresses have already been selected as SSUs in Wave 1.

First question is, as 120 PSUs in Wave 4 refreshment sample are original PSUs in Wave 1, how to calculate inclusion probabilities for those PSUs (post-sector) for Wave 4 refreshment sample? Are those inclusion probabilities for PSUs in Wave 4 refreshment sample the same as they were in Wave 1? Or they are recalculated by new PSUs frame in Wave 4?

Second question is, for Wave 4 refreshment sampling, those addresses have already been selected as SSUs in Wave 1 in each selected PSU (post-sector) were excluded from addresses frame first. Addresses for refreshment sample in Wave 4 were selected from the remaining addresses in each original PSU. Let B be the remaining addresses sampling frame for PSU<sub>i</sub>, let A be the sampling frame the combining of B and those selected addresses in Wave 1 for PSU<sub>i</sub>. Are inclusion probabilities for refreshment sample addresses in PSU<sub>i</sub> calculated by frame B? or by frame A?

Best wishes

Jun

**#5 - 11/26/2021 08:03 AM - Jun Wang**

Dear Olena,

I'm working on CHARLS project. Thanks for your reply, but there is still some questions here:

According to the sample design of Innovation Panel, the inclusion probability for a respondent in Wave 1 can be written as  $\times(\text{inclusion probability for address})\times(\text{inclusion probability for a household})\times(\text{adjustment for multiple dwellings and household})$ . As [Innovation Panel, Wave 1-13: User Manual] [Sample Design part] wrote, 120 PSUs (post-sectors) are all kept fixed after they were selected as SSUs in Wave 1 sample. Refreshment sample addresses in Wave 4 were selected in those original PSUs (post-sectors) excluding addresses have already been selected as SSUs in Wave 1.

First question is, as 120 PSUs in Wave 4 refreshment sample are original PSUs in Wave 1, how to calculate inclusion probabilities for those PSUs (post-sector) for Wave 4 refreshment sample? Are those inclusion probabilities for PSUs in Wave 4 refreshment sample the same as they were in Wave 1? Or they are recalculated by new PSUs frame in Wave 4?

Second question is, for Wave 4 refreshment sampling, those addresses have already been selected as SSUs in Wave 1 in each selected PSU (post-sector) were excluded from addresses frame first. Addresses for refreshment sample in Wave 4 were selected from the remaining addresses in each original PSU. Let B be the remaining addresses sampling frame for PSU<sub>i</sub>, let A be the sampling frame the combining of B and those selected addresses in Wave 1 for PSU<sub>i</sub>. Are inclusion probabilities for refreshment sample addresses in PSU<sub>i</sub> calculated by frame B? or by frame A?

Best wishes

Jun

**#6 - 11/26/2021 12:39 PM - Olena Kaminska**

Jun,

Let's talk about a theoretical example.

In IP1 let's say we sample 10 people out of each PSU. If each PSU has 100 people, the selection probability is  $10/100=0.1$ .

In IP4 let's say we again sample 10 people from the each same PSU (we exclude previous people, although this isn't theoretically necessary). For refreshment sample the selection probability is  $10/100=0.1$ .

What you should be interested in though if you want to use original sample alongside refreshment is the joint probability. In simple terms we now have 20 people selected (either through the original sample or refreshment sample) from each PSU. So the total selection probability (for both samples) is  $20/100=0.2$ .

Does this help?

Olena

**#7 - 11/26/2021 01:50 PM - Jun Wang**

Hi, Olena, thanks very much, the second question is solved.

For the first question, suppose there are 100 PSUs in population, we select 10 PSUs in 100 in Wave 1 by PPS, the size of population at the time of Wave 1 is 10000, the size of one selected PSU is 100, then the selection probability for this PSU can be approximated by  $100*10/10000=0.1$ ; if we select 10 people in this PSU, then the selection probability for each people in this PSU is  $10/100=0.1$ , this is right. Then for this people, the selection probability for he/she can be calculated by  $(100*10/10000)*(10/100)=0.01$ , which is the product of (selection probability of PSU)  $\times$  (selection probability of a people given PSU), selection probability here is defined as: the probability of a people in population is included to the sample.

but at the time Wave 4, the population size increase to 20000, the size of this selected PSU keep the same as it was in Wave 1 (that is 100), if we select this PSU to our sample by PPS at Wave 4, then the selection probability for this PSU should  $10*100/20000=0.05$  at Wave 4. However, at Wave 4, this PSU is not selected randomly, it was selected to the sample at Wave 1. as selection probability (for a people in population) equals (selection

probability of PSU) × selection probability of a people given PSU), the main issue here is how to calculate (selection probability of PSU) at Wave 4, should it be equal to  $100 \times 10 / 10000 = 0.1$  (wave 1) or  $10 \times 100 / 20000 = 0.05$  (wave 4)?

Best wishes  
Jun

**#8 - 11/26/2021 02:03 PM - Olena Kaminska**

Jun,

Yes, there may be a population increase (by newborns or immigration). It depends on your longitudinal plans for renewing your sample. For us we get newborns through mums but we don't get immigrants. Immigrants between wave 1 and 4 to the country have 0 chance to be selected at wave 1 and the usual probability at wave 4. Similarly, you would need to think for each category of people whose population grows separately.

Does this help?  
Olena

**#9 - 11/26/2021 02:04 PM - Olena Kaminska**

Jun,

Also, if people move across PSUs you are in the territory of a cross-classified design, which would need to be taken into account when analysing the data.

Hope this helps,  
Olena

**#10 - 11/26/2021 06:12 PM - Understanding Society User Support Team**

- Status changed from *In Progress* to *Feedback*
- % Done changed from 10 to 80

**#11 - 11/29/2021 08:14 AM - Jun Wang**

Olena Kaminska wrote in [#note-3](#):

Dear Jun Wang,

Thank you for your question.

1. Selection probabilities for refreshment sample are independent of the selection probabilities for the first sample. You will know them from your design;
2. The combined probabilities are the sum of the two;
3. For the rare possible case the actual probability does not depend on whether the person appears twice or not - everyone's probability is the sum of the two probabilities.

If you could let us know which study you are working on, it would help if we can include this for our impact case.

Best wishes,  
Olena

Hi, Olena,

The second you provided <2. The combined probabilities are the sum of the two> was the reply to question:

**2. as IP4 addresses in refreshment sample are selected using systematic randomly amongst addresses not already selected for the IP1, how to calculate inclusion probabilities for those addresses in refreshment sample?**

Refreshment sample is quite different with on-going longitudinal sample. Addresses selected at Wave 1 were followed at Wave 4, but new addresses were selected at Wave 4 for refreshment sample within each PSU selected at Wave 1.

According to **Innovation panel sample design**, addresses for refreshment sample at Wave 4 are selected using systematic randomly sampling excluding those addresses have already been selected at Wave1 given a selected PSU. However, for each PSU, those addresses selected at Wave 4 are not representative of all addresses within this PSU, because those addresses have been selected at Wave 1 were excluded from the sampling frame.

so for second question, I want to know, are the selection probabilities for those addresses selected at Wave 4 for refreshment sample calculated by full addresses frame information or by frame information that exclude those selected addresses have already been selected at Wave1? Can you provide some details or documents about how to calculate this inclusion probability?

We are considering add refreshment sample to our on-going sample, and now are reading technical documents on sample designs and weighting processes for refreshment sample of different longitudinal surveys, like ELSA. If you can provide some details or documents about weighting processes of Innovation Panel, you can sent it to my email [jun.wang@pku.edu.cn](mailto:jun.wang@pku.edu.cn). Thanks very much for your help.

This is the Website of CHARLS: <http://charls.pku.edu.cn/index/en.html>.

Best,  
Jun

**#12 - 11/29/2021 08:28 AM - Jun Wang**

Olena Kaminska wrote in [#note-8](#):

Jun,

Yes, there may be a population increase (by newborns or immigration). It depends on your longitudinal plans for renewing your sample. For us we get newborns through mums but we don't get immigrants. Immigrants between wave 1 and 4 to the country have 0 chance to be selected at wave 1 and the usual probability at wave 4. Similarly, you would need to think for each category of people whose population grows separately.

Does this help?  
Olena

hi,Olena,

UK understanding society:\*Innovation panel\* add a refreshment sample at Wave 4, but PSUs (post-sectors) were the same PSUs in Wave 1. inclusion probabilities for PSUs at Wave 1 can be calculated by  $(\text{number of selected PSUs}) \times (\text{size of PSU}) / (\text{total size of all population PSUs})$ . Even though PSUs at Wave 4 were selected at Wave 1, are their selection probabilities at Wave 4 the same as they were at Wave 1? or re-calculated by the PSUs size information at Wave 4?

Best Wishes,  
Jun

**#13 - 11/30/2021 11:55 AM - Olena Kaminska**

Jun,

Assuming no change in population size of PSUs over time, the selection probability at the PSU stage is the same in IP1 and IP4.

Hope this helps,  
Olena

**#14 - 12/22/2021 09:55 AM - Jun Wang**

Olena,

Problems are solved, thanks for your help.

Best Wishes  
Jun

**#15 - 01/27/2022 02:05 PM - Understanding Society User Support Team**

- Status changed from *Feedback* to *Resolved*

- Assignee deleted (*Olena Kaminska*)

- % Done changed from 80 to 100