

# Understanding Society User Support - Support #1162

## Month of Sample Issue

03/15/2019 12:34 PM - Liam Wright

<b>Status:</b>	Resolved	<b>Start date:</b>	03/15/2019
<b>Priority:</b>	Normal	<b>% Done:</b>	100%
<b>Assignee:</b>			
<b>Category:</b>			
<b>Description</b>			
Hi,			
I'm interested in what specifically determined the month of sample issue in the General Population Sample. I looked in the user guide, but it was quite vague:			
<a href="https://www.understandingsociety.ac.uk/sites/default/files/downloads/documentation/mainstage/user-guides/mainstage-user-guide.pdf">https://www.understandingsociety.ac.uk/sites/default/files/downloads/documentation/mainstage/user-guides/mainstage-user-guide.pdf</a>			
Page 11 states: "From the sorted list, a systematic random sample of 2,640 sectors was selected, with probability proportional to the number of residential addresses in the sector. These sectors were then allocated systematically to 24 monthly samples, with 110 sectors in each monthly sample".			
I'm interested because the empstat modules in Waves 1 and 5 were routed by month of sample issue. I want to see whether respondent answers differ between Wave 1 or 5, and whether this could be feasibly put down to recall bias from longer recall periods.			
Thanks,			
Liam			

### History

#### #1 - 03/18/2019 11:48 AM - Stephanie Auty

- Category set to Survey design
- Status changed from New to In Progress
- Assignee set to Stephanie Auty
- % 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.

Best wishes,  
Stephanie Auty - Understanding Society User Support Officer

#### #2 - 03/18/2019 11:48 AM - Stephanie Auty

- Private changed from Yes to No

#### #3 - 03/19/2019 02:49 PM - Stephanie Auty

- Status changed from In Progress to Feedback
- Assignee changed from Stephanie Auty to Liam Wright
- % Done changed from 10 to 70

Dear Liam,

Households were assigned randomly to sample months, apart from BHPS and NI sample members. The BHPS sample members were all assigned to months in the first year, but they were not asked their employment history in UKHLS as they had already answered in BHPS. Northern Ireland households were also assigned to months in the first year.

The first six months' sample were asked these questions in Wave 1, but then the questionnaire had to be shortened as it was taking too long and so these questions were cut, and followed up with the last 18 months' sample and new entrants since then at Wave 5.

The recall period will depend more on when they started work than which wave they were asked, although there may be differences between

employment history at Wave 5 and the annual events information on employment those individuals had previously given which may be influenced by recall bias.

Best wishes,  
Stephanie Auty - Understanding Society User Support Officer

**#4 - 03/24/2019 01:42 PM - Liam Wright**

- File *Region by Life History Wave.PNG* added

- File *Working Paper.do* added

Hi Stephanie,

Thanks for the response.

I'm surprised by the answer as I find differences in (Wave 1) region of residence between those who eligible to answer life history in Wave 1 and those eligible to answer life history in Wave 5. Screenshot and code attached. (For comparability I limit to those in the `a_indresp` file with `a_sampst==1`, `a_ivfio==1` & `e_ivfio==1`). There's also differences in birth year and ethnicity.

I suspect this isn't a fluke, given the description in the user guide. Was there no geographical patterning to the `a_month` variable?

Thanks,

Liam

**#5 - 03/27/2019 11:56 AM - Olena Kaminska**

Liam,

Thank you for contacting us. Yes, your question is very interesting. The differences observed are very likely to be due to nonresponse and attrition that you don't currently take into account. I looked at the numbers in your table - the differences are not very large, but a nonresponse correction would adjust for the small differences that you observe.

Unfortunately we don't have specific weights for your analysis, but you can easily create them from our weights. As you are using GPS sample you can create a weight, for example:

```
weight=(e_indinus_lw/a_psnenus_xd)*a_psnengp_xd
```

We also suggest that you use strata and psu variables in svyset command.

Hope this helps,  
Olena

**#6 - 03/28/2019 09:35 AM - Peter Lynn**

- File *month and GOR.pdf* added

- % Done changed from 70 to 80

Liam,

Just to add to the above:

I can confirm that the allocation of GPS PSUs to sample month was random - specifically, systematic random after sorting by the stratification variables, as in the text you quote;

It seems to me that the slight imbalance you observe is caused by the inclusion of the Ethnic Minority Boost sample in your analysis. Though you refer in your message to the General Population Sample, your syntax does not exclude the EMBS. The EMBS is not evenly distributed over months, as sampling fractions increased for some groups in Year 2. And, of course, ethnic minorities are not evenly distributed over regions - hence the imbalance.

I ran the attached. First table is GPS-only: no significant difference in the distributions. Second includes EMBS and some differences can be seen (notably London). Third table shows that GPS-only differences are even smaller amongst full wave 1 sample, so attrition may make some minor contribution to the imbalance.

Peter

**#7 - 04/09/2019 04:10 PM - Liam Wright**

Hi Olena and Peter,

Thank you for these replies. These are very helpful.

Best wishes,

Liam

**#8 - 03/02/2021 03:52 PM - Understanding Society User Support Team**

- Status changed from *Feedback* to *Resolved*

- Assignee deleted (*Liam Wright*)

- % Done changed from 80 to 100

**Files**

---

Region by Life History Wave.PNG	32.4 KB	03/24/2019	Liam Wright
Working Paper.do	1.68 KB	03/24/2019	Liam Wright
month and GOR.pdf	9 KB	03/28/2019	Peter Lynn