

Understanding Society User Support - Support #2267

Question on Merging and Weighting with R - Understanding Society Calendar Year 2022

07/30/2025 11:47 AM - Balsam Gharib

Status:	Feedback	Start date:	07/30/2025
Priority:	High	% Done:	80%
Assignee:	Understanding Society User Support Team		
Category:	Weights		
Description			
Hello,			
I am conducting a comparative study on household conditions in London and the South West, using the calendar year 2022 dataset available via the UK Data Service (Open Access version). I would like to double-check that I have correctly implemented the merging and survey weighting procedures to ensure a representative sample.			
My unit of analysis is the individual, but I also need to incorporate household income to so I attempted to merge the indresp and hhresp files using the below:			
<pre>merged_data <- merge(individual_data, household_data, by = "lmn_hidp")</pre>			
I then constructed the survey design object in R using the survey package as follows:			
<pre>design <- svydesign(id = ~lmn_psu, #this is to account for clustering strata = ~lmn_strata, #stratification weights = ~lmn_inding2_xw, #the only cross sectional weight I found for the main individual interview data = mydata, nest = TRUE)</pre>			
I would be grateful if you could confirm:			
Is this the correct approach for merging and weighting when conducting individual-level analysis that includes household-level variables?			
Is the use of lmn_inding2_xw appropriate for generating representative estimates for calendar year 2022?			
Can I assume that the results produced using svytable() or svymean() with this design object are representative of the UK population for 2022?			
As an example, I am using the following line to get the weighted sample distribution across regions (with regional_breakdown being a recode of lmn_gor_dv):			
<pre>svytable(~regional_breakdown, design)</pre>			
I appreciate any feedback you can provide. Thank you in advance!			

History

#1 - 07/30/2025 12:05 PM - Balsam Gharib

Hello,

I am conducting a comparative study on household conditions in London and the South West, using the calendar year 2022 dataset available via the UK Data Service (Open Access version). I would like to double-check that I have correctly implemented the merging and survey weighting procedures to ensure a representative sample.

My unit of analysis is the individual, but I also need to incorporate household income to so I attempted to merge the indresp and hhresp files in R using the below:

```
merged_data <- merge(individual_data, household_data, by = "lmn_hidp")
```

I then constructed the survey design object in R using the survey package as follows:

```
design <- svydesign(
id = ~lmn_psu, #this is to account for clustering
strata = ~lmn_strata, #stratification
weights = ~lmn_inding2_xw, #the only cross sectional weight I found for the main individual interview
data = mydata,
nest = TRUE
)
```

I would be grateful if you could confirm:

- Is this the correct approach for merging the files when conducting individual-level analysis that includes household-level variables?
- Is the use of lmn_inding2_xw appropriate for generating representative estimates for calendar year 2022? This weight is only for wave 14 even though the calendar year 2022 includes wave 13 and a few respondents from wave 12.
- Can I assume that the results produced using svytable() or svymean() with this design object are representative of the UK population for 2022?

As an example, I am using the following line to get the weighted sample distribution across regions (with regional_breakdown being a recode of lmn_gor_dv):

```
svytable(~regional_breakdown, design)
```

I appreciate any feedback you can provide. Thank you in advance!

#2 - 08/04/2025 02:41 PM - Understanding Society User Support Team

- Category set to *Weights*
- Status changed from *New* to *Feedback*
- % Done changed from 0 to 80

Hello,

The approach you described sounds correct. About cross-sectional weights - there are 3 _xw waves available in the calendar year 2022 indresp file: lmn_indpxg2_xw, lmn_inding2_xw, lmn_indscg2_xw. When you want to include proxies in the analysis use lmn_indpxg2_xw (other two exclude proxies altogether), if your analysis includes questions that come from the self-completion questionnaire and the main questionnaire use lmn_indscg2_xw. If you're using only questions from the main questionnaire with no self-completion questions use lmn_inding2_xw. In principle, the same rules as described below apply to picking the weights for the calendar year dataset:

<https://www.understandingsociety.ac.uk/documentation/mainstage/user-guides/main-survey-user-guide/selecting-the-correct-weight-for-your-analysis/>

Best wishes,
Piotr Marzec
UKHLS User Support

#3 - 08/04/2025 02:46 PM - Understanding Society User Support Team

- Private changed from *Yes* to *No*

#4 - 08/04/2025 02:53 PM - Balsam Gharib

Understanding Society User Support Team wrote in [#note-2](#):

Hello,

The approach you described sounds correct. About cross-sectional weights - there are 3 _xw waves available in the calendar year 2022 indresp file: lmn_indpxg2_xw, lmn_inding2_xw, lmn_indscg2_xw. When you want to include proxies in the analysis use lmn_indpxg2_xw (other two exclude proxies altogether), if your analysis includes questions that come from the self-completion questionnaire and the main questionnaire use lmn_indscg2_xw. If you're using only questions from the main questionnaire with no self-completion questions use lmn_inding2_xw. In principle, the same rules as described below apply to picking the weights for the calendar year dataset:

<https://www.understandingsociety.ac.uk/documentation/mainstage/user-guides/main-survey-user-guide/selecting-the-correct-weight-for-your-analysis/>

Best wishes,
Piotr Marzec
UKHLS User Support

Hello Piotr,

Thank you very much for your response. Yes, I am using the main questionnaire so I'll depend on lmn_inding2_xw. Can I now assume, that following this weighting process, the results I obtain from the analysis constitute a representative sample of the UK population for the calendar year 2022?

(I am bit new to working with weights and want to make sure I did not miss anything else)

#5 - 08/04/2025 05:24 PM - Understanding Society User Support Team

Hello,

Yes, the results will be representative of the UK population. See the User Guide, section 4.1 (https://doc.ukdataservice.ac.uk/doc/9333/mrdoc/pdf/9333_main_survey_calendar_year_user_guide_2022.pdf)

Best wishes,
Piotr