

## Understanding Society User Support - Support #1370

### BHPS+UKHLS (26 waves) Repeated Time Values Error in Stata

06/25/2020 03:45 PM - Abigail Dumalus

<b>Status:</b>	Feedback	<b>Start date:</b>	06/25/2020
<b>Priority:</b>	Normal	<b>Due date:</b>	06/25/2020
<b>Assignee:</b>	Abigail Dumalus	<b>% Done:</b>	90%
<b>Category:</b>	Data analysis	<b>Estimated time:</b>	0.00 hour
<b>Target version:</b>			
<b>Description</b>			
Hello,			
I found that linking all 26 waves went without a hitch. Since I intend to do longitudinal analysis of life satisfaction from wave 1 to wave 26, I am trying to tell Stata this has a panel structure (xtset pidp wave). Unfortunately, I am getting error r(451) for "repeated time values within panel". I tried doing xtset separately for all 18 BHPS indresp waves, and it worked. I also applied xtset for all 8 UKHLS indresp waves, but I am getting error r(451). What am I missing here?			

#### History

##### #1 - 06/25/2020 05:23 PM - Gundi Knies

- Private changed from Yes to No
- % Done changed from 0 to 50
- Priority changed from High to Normal
- Assignee changed from Stephanie Auty to Abigail Dumalus
- Status changed from New to Feedback

Hi Abigail,

it seems like you the wave values 1-9 twice - once for the BHPS and once for the UKHLS. Maybe treat UKHLS wave 1 as wave=19, UKHLS WAVE 2 as wave =20 etc. Does this solve the issue?

Best,  
Gundi

##### #2 - 06/25/2020 06:33 PM - Abigail Dumalus

Hello Gundi,

I have been trying append all UKHLS waves, using the standard Stata syntax for looping through the waves of ukhls from Wave 2. I am still getting error r(451). Wave 2 = 19, since the command I use for wave is: gen wave=`waveno'+17, and so on.

I even added a label for each wave in turn. If I tabulate wave, the wave number starts from 19. After I append all 8 waves, I still get an error message after using the xtset command. What would you advice to get around this "repeated time values within panel"?

##### #3 - 06/25/2020 07:31 PM - Gundi Knies

Try not to append UKHLS wave 9 to the data file that already has wave 9 data - the number of cases for w9 is suspicious.

##### #4 - 06/25/2020 07:49 PM - Abigail Dumalus

Why is the number of cases for UKHLS wave 9 suspicious? Does this mean there should only be 25 waves in the longitudinal analysis? I am only appending UKHLS waves 2 to 9, so there shouldn't be a duplication of wave 9 data. The first 18 waves from the BHPS is still waiting to be appended if only i can sort out error r(451) from the UKHLS waves.

##### #5 - 06/25/2020 09:36 PM - Gundi Knies

the number of cases in wave 9 is suspicious because the sample size tends to reduce from wave to wave, due to non-response. In your listing, the sample size almost doubles from UKHLS wave 8 to 9. It is also a common error to include all wave prefixes in the append-loop following the loop that prepares the data for merge in long format. If your first loop ends on wave 9, and then the append loop goes from BHPS Wave 1 - UKHLS wave 9, wave 9 data will be there twice. the append loop should run until wave 8 only. Generally, try and use "duplicates report pidp wave" and inspect the cases.

Hope htis helps,  
Gundi

**#6 - 06/25/2020 10:11 PM - Abigail Dumalus**

Thanks for the advice, Gundi.

I will use the duplicates report command. Is this the main reason why the merging-individual-files-harmonised-bhps-ukhls-long-format Stata syntax on the USoc website only includes up to UKHLS wave 8? In doing so, the period coverage ranges from 1991 to 2018.

In terms of the other wave-specific files on the individual level (i.e., indall, indsamp, and income) that have been fully harmonised, does it make sense to link all each of these respectively, and then merge with the indresp longitudinal file in long format for analysis?

**#7 - 06/25/2020 10:31 PM - Gundi Knies**

Not sure why the code on the website only goes to wave 8; it is probably a co-incidence and I would not expect the code examples to be updated each time a new wave of data is released. The general principles of data processing and management apply across all waves!

As to merging all data files from all waves together in a long format data file, it is probably more efficient (and less error prone!) to load only the variables that you are interested in, and to do all the cleaning and coding in the file, before merging all elements together for analysis. But there are many ways to achieve the same thing.

Good luck!

Gundi

**#8 - 06/26/2020 01:03 AM - Alita Nandi**

- % Done changed from 50 to 80
- File UserForum\_1370.log added
- File UserForum\_1370.do added

Hi Abigail,

Following up on what Gundi has said, I took a look at the syntax file on our website and added Wave 9 and ran it. The 18+8 waves appended correctly. I am attaching the logfile and dofile for you to compare with your syntax file and output.

Best wishes,  
Alita  
On behalf of Understanding Society User Support Team

**#9 - 06/26/2020 01:31 AM - Abigail Dumalus**

Hello Alita,

I'm so glad to hear that you and Gundi are around to give advice!

I just have a follow up question about the longitudinal weighting variables for these 18+8 waves. How do I make sure that I am using the appropriate weights for the duration of these periods? I have read Module 8 of the training using Stata but I am still feeling lost and unsure. Is there an easier to digest the weighting reference?

**#10 - 06/26/2020 09:59 AM - Alita Nandi**

You're welcome Abigail!

From your description it seems that you are planning to use the BHPS sample over 26 waves to conduct longitudinal analysis of individual adult response data. So, the correct will be `i_indin91_lw`

Please read the weighting section in the user guide:  
<https://www.understandingsociety.ac.uk/documentation/mainstage/user-guides/main-survey-user-guide/>

As the questions you are using (life satisfaction) were asked as part of the self-completion module, additional non-response adjustment is required, but that weight is not available. You can either acknowledge this in your paper/report or produce additional non-response adjustment for self-completion. If you are interested in doing the latter, please take a look at Item 15 in the weighting FAQ:

[https://www.understandingsociety.ac.uk/sites/default/files/downloads/general/weighting\\_faqs.pdf](https://www.understandingsociety.ac.uk/sites/default/files/downloads/general/weighting_faqs.pdf)

If this does not answer your question please let us know and we will assign the issue to our weighting team.

Best wishes,  
Alita  
On behalf of the Understanding Society User Support Team

**#11 - 06/26/2020 12:37 PM - Abigail Dumalus**

Hello Alita,

Yes, I am especially interested in analysing life satisfaction and GHQ happiness. Thank you for sharing the relevant resources.

Since the indresp sample only includes adults, does it make sense to merge the youth sample (for the question/variable 'yphlf') with the 'lfsato' and 'sclfsato' variables? How complicated would the weighting considerations be?

**#12 - 06/26/2020 01:00 PM - Alita Nandi**

- % Done changed from 80 to 50

- Assignee changed from Abigail Dumalus to Olena Kaminska

I am assigning this to our Survey Statistician, Dr. Kaminska, to answer your question on appropriate weights for this particular analysis. Best wishes, Alita

**#13 - 06/26/2020 04:17 PM - Abigail Dumalus**

Thanks so much, Alita!

Hello Olena,

Does it make sense to merge the youth sample (for the question/variable 'yphlf') with the 'lfsato' and 'sclfsato' variables included in indresp longitudinal file? How complicated would the weighting considerations be? Another reason I wish to merge the youth sample with the indresp file is because I want to analyse how life satisfaction responses over 26 waves are changing among those adults with children below 16 years old vs. those who are childless. Now I am thinking whether it would be more comprehensive to include the child data file to measure household size. Please let me know whether these are possible. Many thanks in advance.

**#14 - 06/26/2020 04:43 PM - Olena Kaminska**

Abigali,

Yes, of course your analysis and research questions make sense, and indeed we do not have specific weights for your purpose, but a suboptimal weight can be enumeration longitudinal weight (W\_psnen??\_lw).

You could create your own tailored weight based on the above weight - let me know if you are interested.

Hope this helps,  
Olena

**#15 - 06/26/2020 05:22 PM - Abigail Dumalus**

Great to hear from you, Olena...

I am interested with creating my own tailored weight base on W\_psnen??\_lw. If I understand this correctly after reading the weighting FAQs document, it would be more appropriate to use psnen91\_lw for 26 waves: BHPS+UKHLS samples. But how is this different from using psnenub\_lw, representing combined BHPS+UKHLS longitudinal enumerated person weight? To clarify, this assumes that I am merging youth and child datafile with indresp, correct?

**#16 - 06/26/2020 05:35 PM - Olena Kaminska**

Abigail,

To understand the difference between '91' and 'ub' weight please see [https://www.understandingsociety.ac.uk/sites/default/files/downloads/general/weighting\\_faqs.pdf](https://www.understandingsociety.ac.uk/sites/default/files/downloads/general/weighting_faqs.pdf)  
Question 6 on zz\_

And yes, you can merge youth, child and adult information of any kind with these weights.

I can send you some instructions on how to create a tailored weight. Please email your request here and mention that you are after these instructions [usersupport@understandingsociety.ac.uk](mailto:usersupport@understandingsociety.ac.uk) .

Thank you,  
Olena

**#17 - 06/26/2020 05:54 PM - Abigail Dumalus**

Hi Olena,

I just sent an email request for creating a tailored weight.

About merging youth and child data files, I have been reading through past issues that relate to this. As far as I can comprehend, I need to link first all 22 wave files of the youth sample starting from BHPS wave 4, following the indresp long format syntax to come up with longitudinal youth file. Then, I would have to append the longitudinal youth file with the indresp sample using pidp. I am not so sure whether this exercise applies to the child data files, doesn't it?

**#18 - 06/27/2020 11:15 AM - Alita Nandi**

- % Done changed from 50 to 90

- Status changed from Feedback to Resolved

This issue is now continuing via email. So, this post is set to resolved.

**#19 - 06/30/2020 11:51 AM - Alita Nandi**

- Assignee changed from Olena Kaminska to Abigail Dumalus  
- Status changed from Resolved to In Progress

Hello Abigail,

I am sorry I deleted the issue 1371 that you had posted by mistake. As it is not possible for us to un-delete it, I have now created a new post (1372) with the same content as 1371. But I couldn't assign it to you as you were not the author. So, you may not have received an email about this issue. Here is the link to this post: <https://userswww.essex.ac.uk/support/issues/1372>

Best wishes,  
Alita

**#20 - 06/30/2020 04:34 PM - Alita Nandi**

I am responding to issue #17 above:

There are a few ways to do this. What you have suggested (first append all youth files into a long format file, do the same for indresp files, then append these two long format files) will work. But note you don't need to specify a variable for appending. If the two long format files are named youth\_allwaves and indresp\_allwaves then you just need to do the following to append them.

use youth\_allwaves, clear  
append using indresp\_allwaves

**#21 - 07/01/2020 01:18 PM - Abigail Dumalus**

Thanks so much, Alita.

I was trying to append the indall longitudinal file (all 26 waves) with the indresp+xwavedat. When I try to tell Stata that the data has a panel structure (xtset pid wave), the repeated time values error keep coming up. When I did this with the merged indresp+xwavedat longitudinal file, it was successful. I am wondering now whether does it make sense to append the indall data files without a specific common variable like pidp. I have been reading previous related issues and I have not found clarity on this yet. As far as I can understand it, indall files represent household grid data for all enumerated persons in household, including children and nonrespondents. Should I not be merging instead of appending? When I merged indall, the only mode that works is m:m and not m:1 (which worked when I merged indresp files with xwavedat). What am I missing here? Kindly advise. Thanks in advance.

**#22 - 07/02/2020 01:37 PM - Abigail Dumalus**

Hello Alita,

Kindly see above for my question about appending/merging indall files with indresp+xwavedat file. I have been unable to tell Stata that the data has a panel structure (xtset pidp wave). Thus, the repeated time values error is still persistent. Is there a way to avoid this and simply add those respondents not included in the indresp files?

**#23 - 07/03/2020 06:34 AM - Alita Nandi**

You said you ran the code "xtset pid wave" and got an error message. Did you use "pid" instead of "pidp" or this just a typo here? "pid" is only available for BHPS sample members and so will be missing for many (all UKHLS sample members and those who joined the households of BHPS sample members after they joined the UKHLS). As a result you will get this error message. The other possible reason for this error message is if you have appended the same file twice.

In terms of adding in indall files. The best way will be to merge indall with indresp & youth files within those loops, keeping the variables you need from indall and then dropping the extra observations in indall.

Best wishes,  
Alita

**#24 - 07/03/2020 11:12 AM - Abigail Dumalus**

Hello Alita,

It was a typo here. I used pidp. What I did was linking all indall files together for 26 waves (using the indresp syntax), and then I merged this longitudinal indall file with the longitudinal indresp file using pidp. But I still get that repeated time values error. I can't understand how I could've appended the same file twice when I only merge the longitudinal indall file once with the longitudinal indresp file. Is it because respondents from the indresp files are also included in the indall files?

**#25 - 07/03/2020 01:32 PM - Alita Nandi**

First check that that indall\_longfile and indresp\_longfile don't have repeated observations. Type

isid pidp wave

if there are no repeated observations this will run through, otherwise there will be an error message. Note this code will not produce any output if it is ok.

If that is not the case, then make sure you have merged indresp\_longfile and indall\_longfile using pidp wave and not appended them. If you append these 2 files you will get repeated observations because, as you have correctly pointed out, same individuals are in both files. Everyone who is in the households interviewed appear in the indall file. From among them, only those who are 16 and above are invited to participate in the adult interviews. If they agree then their responses are recorded in indresp. So, indall will include everyone in indresp + 0-15 year olds + 16 & above year olds who did not participate in the adult interview.

**#26 - 07/03/2020 01:47 PM - Abigail Dumalus**

I will try your advice and see how it goes. Thanks so much, Alita.

I meant to raise this before: just looking at indresp data, there are 15-year old respondents (even without merging the youth samples). So, I am a bit confused when you say that adult interviews are done among 16 & above year olds who participated. I checked the individual wave data files and still found these respondents aged 15 years old. Is this normal?

**#27 - 07/03/2020 02:17 PM - Alita Nandi**

Yes, this happens sometimes, but only a few such cases. In the 9 UKHLS waves there are 13 such cases. In BHPS the cut-off was based on age as of Dec of the year = bw\_age12. If you use that variable then there are 14 such cases over the 18 waves.

**#28 - 08/14/2020 05:28 PM - Alita Nandi**

- Status changed from In Progress to Feedback

**Files**

---

BHPS(waves 1-18)_UKHLS(waves 2-9) Error After xtset pidp wave.pdf	1.6 KB	06/25/2020	Abigail Dumalus
UserForum_1370.do	4.25 KB	06/26/2020	Alita Nandi
UserForum_1370.log	8.83 KB	06/26/2020	Alita Nandi