

## Understanding Society User Support - Support #1680

### Merging: Many to one & sorting

04/08/2022 05:46 PM - Irina Kolegova

<b>Status:</b>	Resolved	<b>Start date:</b>	04/08/2022
<b>Priority:</b>	Urgent	<b>% Done:</b>	100%
<b>Assignee:</b>	Irina Kolegova		
<b>Category:</b>	Data linkage and consents		
<b>Description</b>			
Hello,			
I wanted to merge youth datasets for Wave 10 (UKHLS), Wave 4 (COVID) and Wave 8 (COVID).			
1) Should I use "many to one" merging, "one to one", or "one to many"?			
2) Do I need to sort datasets by pidp (and pidp_c) before merging them?			
3) How can I unite those variables that repeat themselves in every wave (sex, ethnicity etc.)?			
Thank you			

### History

#### #1 - 04/12/2022 11:33 AM - Understanding Society User Support Team

- Private changed from Yes to No

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 - 04/12/2022 11:33 AM - Understanding Society User Support Team

- Status changed from New to In Progress

#### #3 - 04/12/2022 05:18 PM - Understanding Society User Support Team

- Status changed from In Progress to Feedback

- % Done changed from 0 to 50

Dear Irina,

1) all these files are individual level files in which pidp is a unique identifier, so this is 1:1 merge,  
2) if you are using Stata version 12 or older then you need to sort first, if newer versions of Stata then you do not need to sort,  
3) in wide format you by definition get extra variables preceded by a wave prefix (a\_ - wave 1, b\_ - wave 2 and so on) to accommodate additional time points. However, for time invariant variables (e.g. country of birth, sex, ethnicity) the value of these will be the same. One workaround is to omit these variables when merging and add them after the merge from the xwavedat file (for more information about xwavedat see <https://www.understandingsociety.ac.uk/documentation/mainstage/user-guides/main-survey-user-guide/list-of-data-files-and-their-descriptions>). Additionally, for sex specifically, in each youth datafile there is W\_ypsex which contains the answers to the YPSEX question from the youth questionnaire. Although this would be rare, W\_ypsex may vary between waves. We leave the choice which of these to use, d\_sex or d\_ypsex to users.

Best wishes,  
Understanding Society User Support Team

#### #4 - 06/06/2022 08:25 PM - Understanding Society User Support Team

- Status changed from Feedback to Resolved

- % Done changed from 50 to 100