NATIONAL DATA ARCHIVE ON
NDACAN CHILD ABUSE AND NEGLECT

How to merge LONGSCAN data files when the files contain multiple records per participant

Based on SPSS version 24

This document is intended to assist LONGSCAN data users with merging two data files which each contain multiple records per participant. This file format is also known as a "long" or "stacked" data file. This document will show data users how to navigate the SPSS menus to conduct a merge and also how to use syntax to conduct the merge.

Launch SPSS and open both of your two data files of interest that each have multiple records per participant. For this instruction sheet, we will merge the Child Behavior Checklist item level data (CBCL1201.sav.sav) with the scored data for the same measure (CBCS1201.sav). We are using these two data files because they are related, one is the item-level data and the other contains the scores, so we know that all child ID-visit pairs appearing in one file will appear in the other file. The instructions in this document are suited for merging data files where there are matching records between the files and also for merges where one of the data files contains cases for which there is no match in the other file. In the set-up described in this document, all cases will be preserved in the resulting merged data file.

You must sort each data file by the key IDS which in the case of the files in this example, from LONGSCAN, is the LONGSCAN ID variable called "ID" and the visit number variable called VISIT." Click on "Data" from the menu bar at the top of the main SPSS window and select "Sort cases…" from the Data menu options, as shown in Figure 1.

Figure 1. Screenshot of the "Data" menu options showing "Sort Cases…" highlighted.
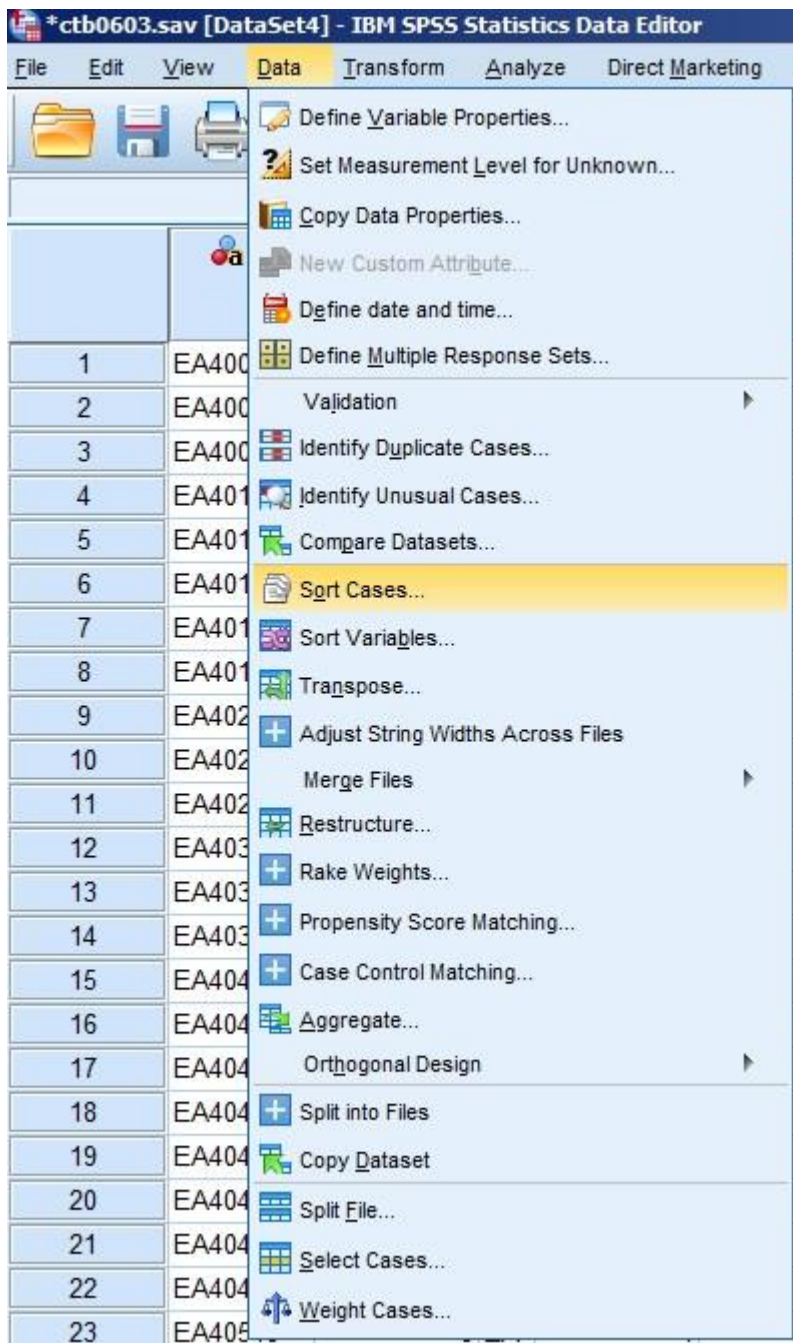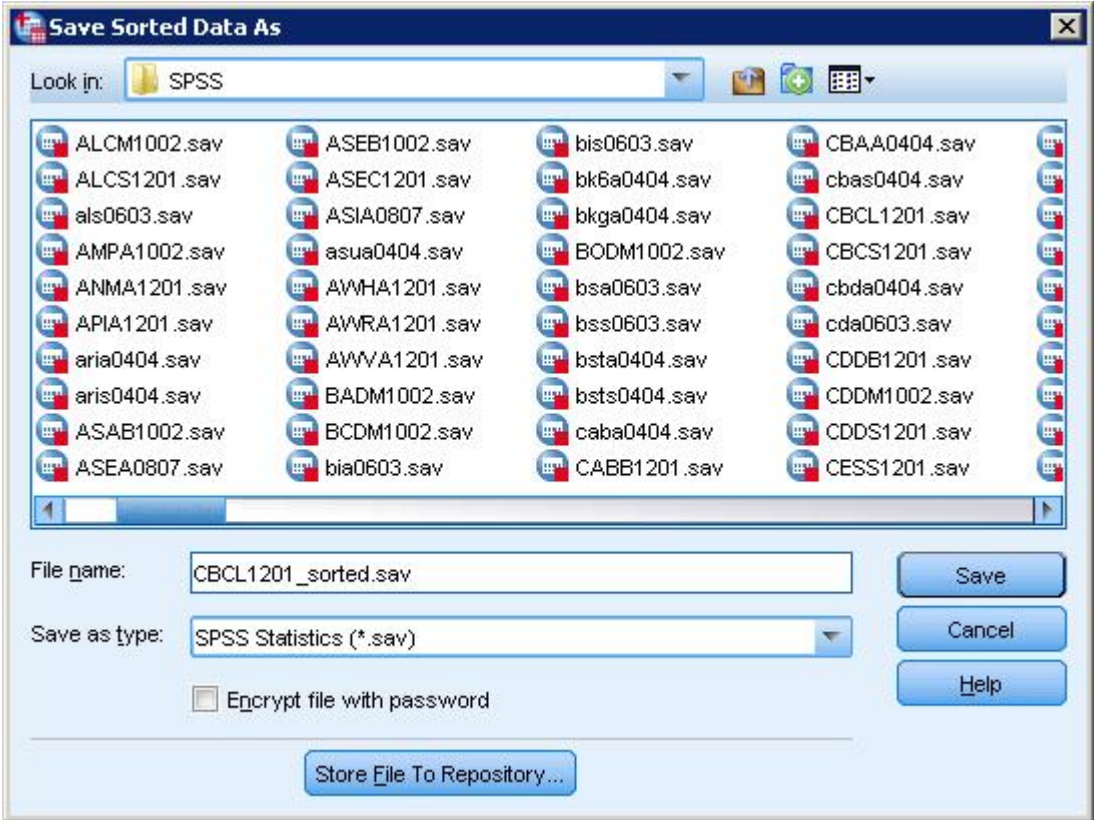
In the "Sort Cases" window that appears, select the "LONGSCAN SUBJECT ID (ID)" variable from the list on the left of the window and use the arrow button to click the variable into the "Sort by" box on the right side of the window. Do the same for the "VISIT NUMBER (VISIT)" variable. We need to sort by both because they will both be used as our key ID variables in the merge due to a unique record being defined by an ID-VISIT pair. Be sure that the "Ascending" radio button is active to sort the file in ascending order by ID. Click in the checkbox next to the words "Save file with sorted data" as shown in Figure 2 to save the resulting sorted data file. Click on the "File…" button beneath the "Save file with sorted data" text.

*Figure 2. Screenshot of the "Sort Cases" window.*
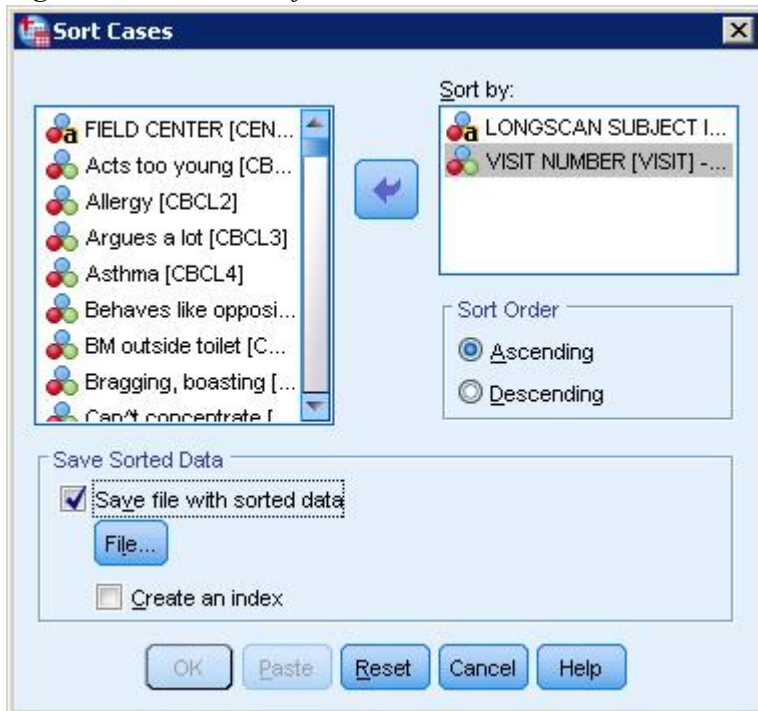
Once the "Save Sorted Data As" window is open, specify a file location and name to save the sorted data file (see Figure 3). Click the "Save" button on the lower right portion of the window to save the sorted data file once you are finish.

Figure 3. Screenshot of the "Save Sorted Data As" window.

After clicking "Save" in the "Save Sorted Data As" window, the "Sort Cases" window will reappear as shown in Figure 4. If the information appearing in the window looks correct, then click "OK" to sort the cases and save the file as specified.

Figure 4. Screenshot of the "Sort Cases" window.
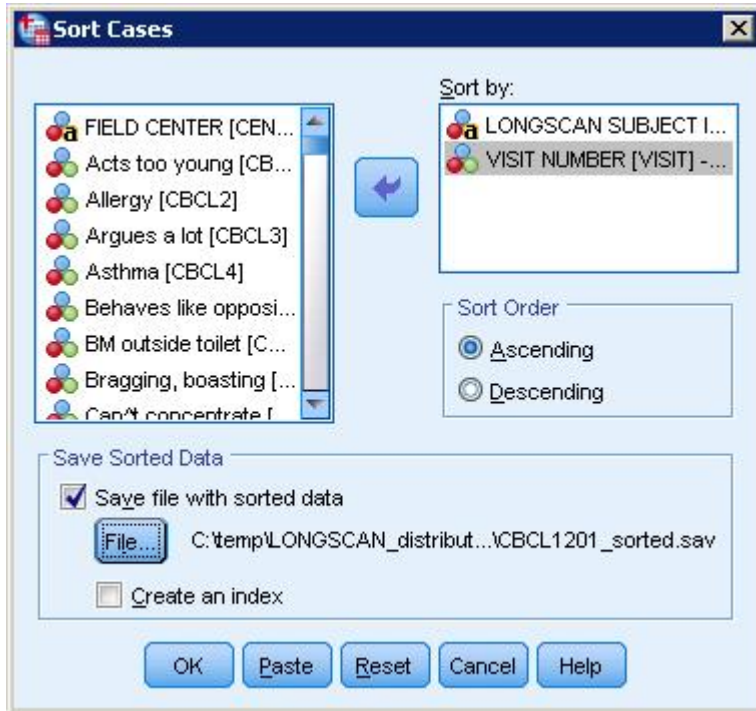
Repeat all instructions from the beginning of the document through this point for the other long/stacked format data file you want to use in the merge. In the case of the example being used for this instruction sheet, we would repeat the steps for the CBCS1201.sav data file. After both files have been sorted by ID and saved, proceed to the next step which is the merge. If you want variables from the item level data to appear first in the merged data file, be sure to open the item level sort data file and conduct the next steps from inside of that SPSS window.

Click on the "Data" menu in the SPSS main window and select "Merge Files" from the list of options in the Data menu and then select "Add variables..." Figure 5 shows the "Data" menu list with the "Merge Files" option selected. We are adding variables because we know two things about these data files: First, they are related and the only way to produce scores in the scored file the participant must have a record in the item level data, so records will not be added during the merge. Second, each of the files only contain one time point of data collection, which is Visit 8.

Figure 5. Screenshot of the "Data" menu list showing "Merge Files" highlighted.
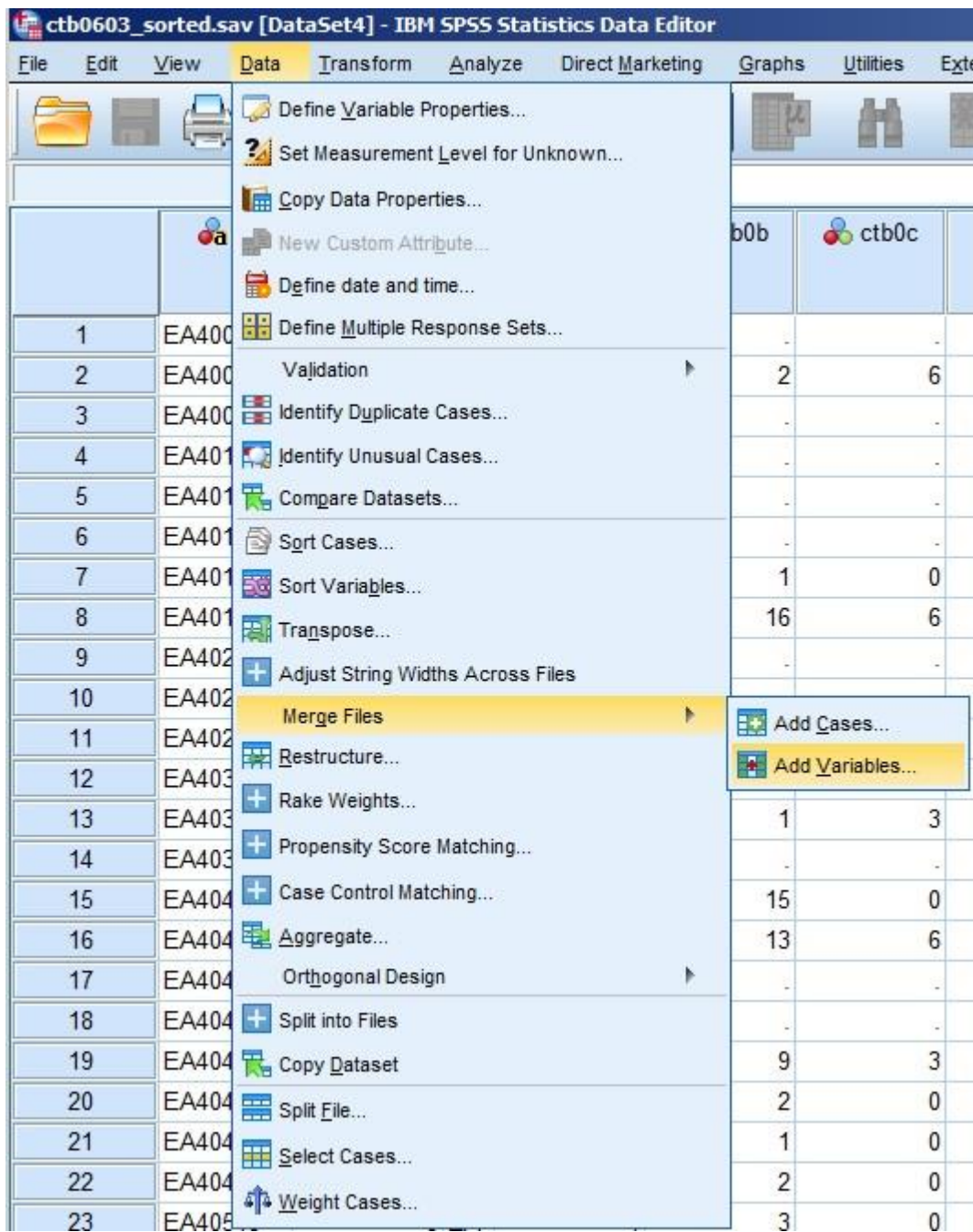
If both data files are already open, then you can activate the radio button in the "Add Variables to CBCS1201_sorted.sav" (or insert your item-level data file here) which says "An open dataset." The open scored data file should appear in the list inside of the box beneath the "An open dataset" text. Click on the name of the data file to select it. Be sure that it is your sorted data file. Click the "Continue" button along the bottom of the window (see Figure 6).

Figure 6. Screenshot of the "Add Variables to CBCS1201_sorted.sav" window.
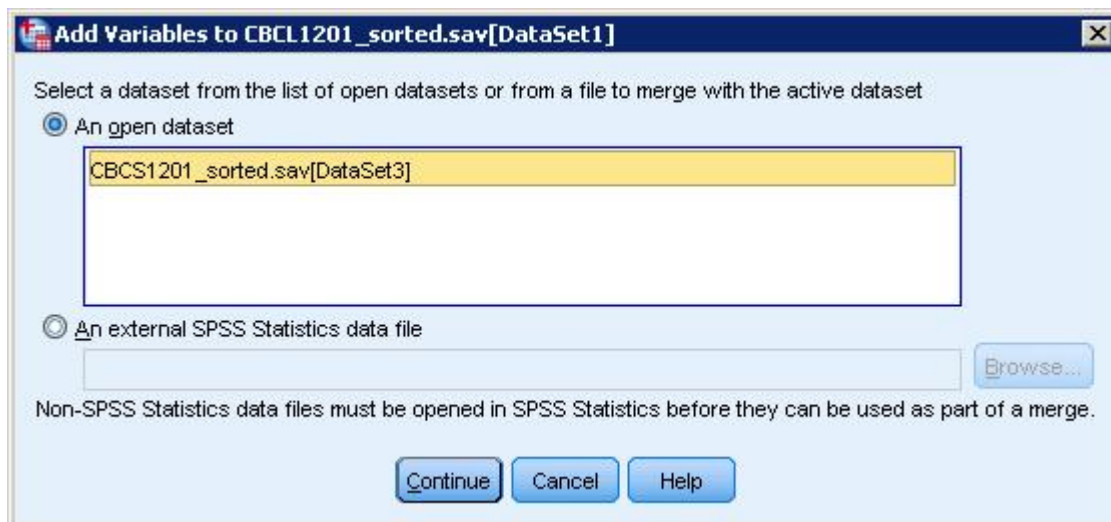
The next window that will open is called the "Add Variables from Datasets" window (see Figure 7). Check mark the "Match cases on key variables" check box. Check mark the "Cases are sorted in order of key variables in both datasets" box. Activate the radio button for "Both files provide cases" option in the outlined box beneath the "Cases are sorted…" check box. Activating the "Both files provide cases" option means that if there were cases that existed in one file but not the other, the case would be preserved in the resulting merged data file. From the "Excluded Variables" list, select the "id" variable and then use the arrow button to the left of the "Key Variables" box to click "id" into that box. We can leave "Center" and "Visit" variables in the "Excluded Variables" box since those variables are duplicated in the other data file. Click the "Ok" button.

Figure 7. Screenshot of the "add Variables from Datasets" window.
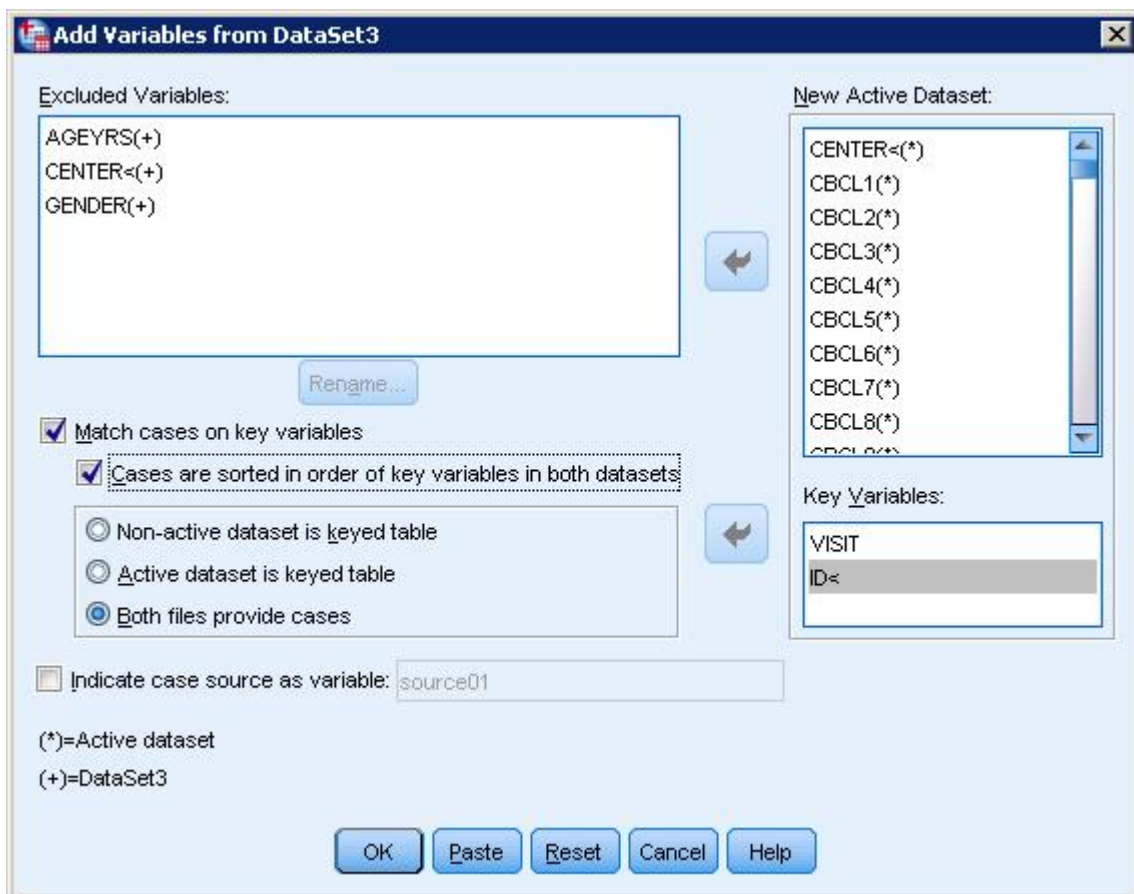
Once you click the "Ok" button, a warning will pop up like the one below in Figure 8. If you are certain that you sorted both data files in the same order by the key id, then click the "Ok" button.

Figure 8. Screenshot of the warning window.

After clicking the "Ok" button on the warning window, your files will be merged and the merged dataset will be open in front of you. Be sure to examine the data to make sure that the merge was performed correctly. One quick way to be sure that your merge was performed correctly is to add the count of variables from each data file together and then subtract the duplicate variables that were in the "excluded variables" section of Figure 7 which was the "Add Variables from Datasets" window and also subtract key ID variable(s), since they were used to match up the records, and therefore, only one copy of the key IDs are included in the resulting merged data file. Checking to see that all cases made it into the file is a littler trickier. You can do a very basic spot check using the count of records from the largest of the two data files to see if it is close to matching the record count in the resulting merged data file.

Data users who are comfortable using syntax to perform statistical procedures can modify and use the syntax below to conduct the merge. The syntax below leverages the MATCH FILES command in SPSS to merge the data files. Just like when we went through the menus to conduct the merge, we will need to sort each data file in ascending order prior to performing the merge. We will also employ the use of dataset nicknames (Dataset Name).

Note of caution: When copying and pasting syntax from a pdf document into the SPSS Syntax Editor window, characters, such as, single quotes may not paste into the Syntax Editor as being recognizable by SPSS. They may appear to look like a single quote in the Editor window, but they will not allow the program to run without errors. To fix it, delete all single quotes appearing the editor and add them back in from inside of the Syntax Editor.

*Assigning the nickname t0 to the CBCL1201 data file.

GET FILE = 'C:\temp\LONGSCAN\Data\SPSS\cbcl1201.sav'.

DATASET NAME t0.

EXECUTE.

DATASET ACTIVATE t0.

EXECUTE.

SORT CASES ID Visit (A).

EXECUTE.

SAVE OUTFILE = 'C:\temp\LONGSCAN\Data\SPSS\cbcl1201_sorted.sav'.

EXECUTE.

GET FILE = 'C:\temp\LONGSCAN\Data\SPSS\CBCS1201.sav'.

EXECUTE.

DATASET NAME t1.

EXECUTE.

DATASET ACTIVATE t1.

EXECUTE.

SORT CASES ID Visit (A).

EXECUTE.

SAVE OUTFILE = 'C:\temp\LONGSCAN\Data\SPSS\CBCS1201_sorted.sav'.

EXECUTE.

DATASET ACTIVATE t0.

EXECUTE.

MATCH FILES FILE = 'C:\temp\LONGSCAN\Data\SPSS\cbcl1201_sorted.sav'

/FILE = 'C:\temp\LONGSCAN\Data\SPSS\cbcs1201_sorted.sav'

/BY ID visit.

SAVE OUTFILE = 'C:\temp\LONGSCAN\Data\SPSS\cbcl_cbcs_merged.sav'.

EXECUTE.

Please send questions to NDACANsupport@cornell.edu.