GUI Data Workshop

Information Sheet 4c: How to match files - Infant Cohort - Wave 1 (at 9 months), Wave 2 (at 3 years) and Wave 2 (at 5 years)

This information sheet provides a guide on how to match the Infant Cohort Wave 1 (at 9 months), Wave 2 (at 3 years) and Wave 3 (at 5 years) files. Please note this information sheet is based on SPSS Version 22.

Convention A naming – Drop down menus[1]

To match the Wave 1 and Wave 2 datafile you must follow these steps:

1. Open the Wave 1 datafile – 'GUI Data_9MonthCohort.sav'

2. Click Data → Merge → Merge Files → Add variables

---

[1] See *Variable Naming Conventions And Longitudinal Data Dictionary For Wave 1 And Wave 2 Of The Infant Cohort Of Growing Up In Ireland* document for further details on variable naming conventions.
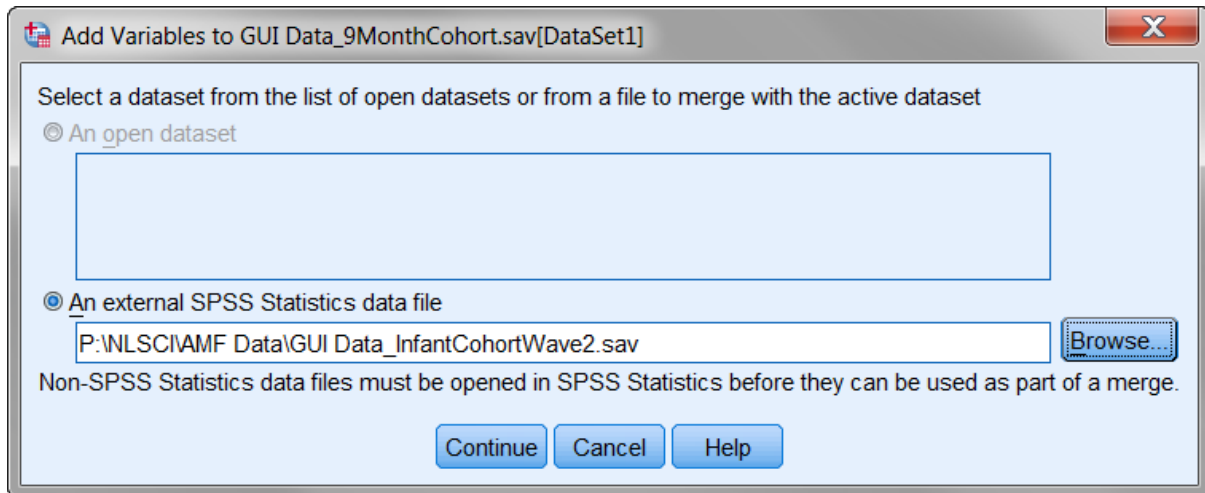
3. This will bring up the following dialog box. Click browse and select the location of where you have saved the Wave 2 datafile – 'GUI Data_ InfantCohortWave2.sav'. Click Continue.


4. In the following dialog box tick 'Match cases on key variables in sorted files', and 'Cases are sorted in order of key variables in both datasets'.

Using the arrow move the variable '*ID*' into the 'Key Variables' box.

Also tick 'Indicate case source as variable' and type a name for the variable into the box (the default variable name is '*source01*'. This has been changed to '*wave2*' in the example below). This will create a new variable which is coded 1 if the case is in Wave 2 and 0 if the case is not in Wave 2. Click 'ok'. There are 11,134 cases in Wave 1 and 9,793 cases in Wave 2.
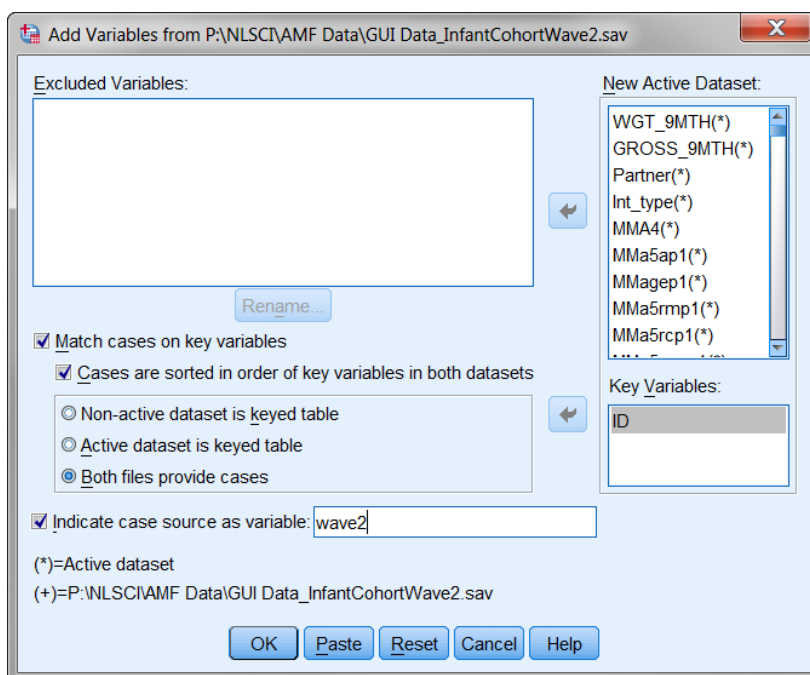
5. The following warning will appear. In order to match the files they both must be pre-sorted by the key variable used for matching – in this case '*ID*'. Click 'ok'.
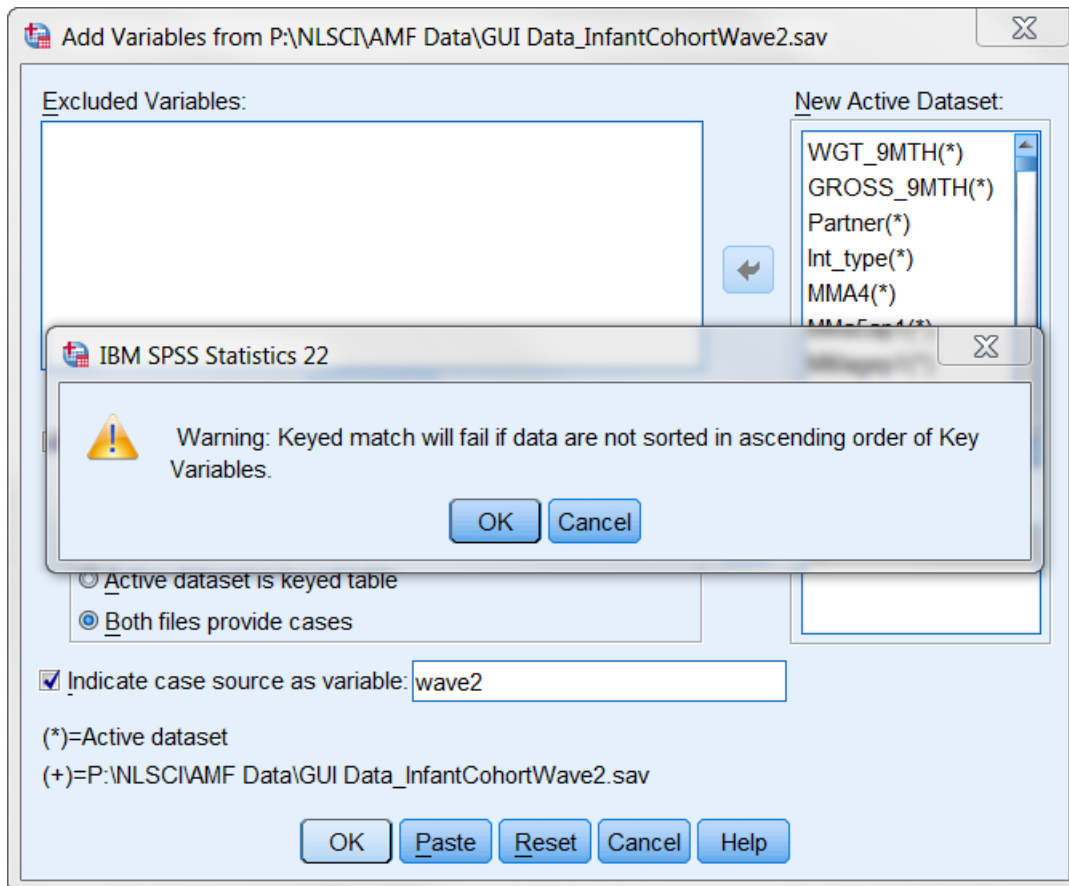
6. The matched file will contain 11,134 cases with all the variables from Wave 1 and all the variables from Wave 2. The Wave 2 variables will be system missing (sysmis) for any cases which are not included in Wave 2 (there are 9,793 valid cases in Wave 2). There will also be an identifier variable '*wave2*' which will flag which families participated in Wave 2.

7. To match the Wave 3 data repeat steps 2 – 6 which are summarized below.

- Click Data → Merge → Merge Files → Add variables
- In the dialog box, click browse and select the location of the Wave 3 datafile – 'GUI Data_ InfantCohortWave3.sav'. Click Continue.
- In the dialog box tick 'Match cases on key variables in sorted files', and 'Cases are sorted in order of key variables in both datasets'.
- Using the arrow move the variable '*ID*' into the 'Key Variables' box.

- Also tick 'Indicate case source as variable' and this time replace '*source01*' with '*wave3*'. Click 'ok'.

- Click 'ok' when the warning appears.

- The matched file will contain 11,134 cases with all the variables from Wave 1, Wave 2 and Wave 3. The Wave 3 variables will be system missing (sysmis) for any cases which are not included in Wave 3 (there are 9,001 valid cases in Wave 3). There will also be an identifier variable '*wave3*' which will flag which families participated in Wave 3.

8. This matched file should be saved under a new name (for e.g. 'GUI Data_InfantCohortWave1Wave2Wave3.sav'). <u>Simply saving the file will overwrite the original Wave 1 file.</u>

Convention A naming – SPSS syntax

Alternatively, the following SPSS syntax can be used to match the files together. (Note that the researcher will need to change the file handles to the relevant file path indicating where the data have been saved).

```
FILE HANDLE wave1 name = 'P:\NLSCI\AMF Data\GUI Data_9MonthCohort.sav'.
FILE HANDLE wave2 name = 'P:\NLSCI\AMF Data\GUI Data_InfantCohortWave2.sav'.
FILE HANDLE wave3 name = 'P:\NLSCI\AMF Data\GUI Data_InfantCohortWave3.sav'.

FILE HANDLE merged name = 'P:\NLSCI\AMF Data\GUI
Data_InfantCohortWave1Wave2Wave3.sav.sav'.

GET FILE wave1.
SORT CASES by id.

MATCH FILES FILE = * / FILE = wave2 / IN wave2 / BY id / MAP.
FREQUENCIES wave2.

MATCH FILES FILE = * / FILE = wave3 / IN wave3 / BY id / MAP.
FREQUENCIES wave3.

SAVE OUTFILE = merged.
```

Convention B naming – Drop down menus

Follow steps 1 – 8 above but using the following files:

Wave 1 – 'XGUI Data_9MonthCohort.sav'

Wave 2 – 'XGUI Data_ InfantCohortWave2.sav'

Wave 3 – 'XGUI Data_ InfantCohortWave3.sav'

And matching on the variable '*zid01*'.

Convention B naming – SPSS syntax

Alternatively, the following SPSS syntax can be used to match the files together. (Note that the researcher will need to change the file handles to the relevant file path indicating where the data have been saved).

```
FILE HANDLE wave1 name = 'P:\NLSCI\AMF Data\XGUI Data_9MonthCohort.sav'.
FILE HANDLE wave2 name = 'P:\NLSCI\AMF Data\XGUI Data_InfantCohortWave2.sav'.
FILE HANDLE wave3 name = 'P:\NLSCI\AMF Data\XGUI Data_InfantCohortWave3.sav'.

FILE HANDLE merged name = 'P:\NLSCI\AMF Data\GUI
Data_InfantCohortWave1Wave2Wave3.sav.sav'.

GET FILE wave1.
SORT CASES by zid01.

MATCH FILES FILE = * / FILE = wave2 / IN wave2 / BY zid01/ MAP.
FREQUENCIES wave2.

MATCH FILES FILE = * / FILE = wave3 / IN wave3 / BY zid01/ MAP.
FREQUENCIES wave3.

SAVE OUTFILE = merged.
```