



## **Joining two data files from the RPG-2 dataset, using Stata.**

This document is intended to assist data users with joining/merging two data files from RPG-2, dataset #218. The instructions are generally applicable for use with other data files with similar characteristics.

The two data files used in this example are the `rpg_esl_person_level` and the `rpg_esl_case_level`. These were selected because they feature the joining of two datasets where the unit of observation differs, resulting in different record counts for each file.

The `rpg_esl_person_level` dataset contains information about each person who participated in the RPG-2 data collection, including substance abusing adults and their children. As shown in the User's Guide Table 1, you will note that the variables that identify a unique record for the `rpg_esl_person_level` data file are: `Grantee_ID`, `Case_ID`, and `Ind_ID`. In this file, because each case has multiple people and the unit of observation is the person, the `Case_ID` repeats across records for each person represented in that case. The variables that identify a unique record in the `rpg_esl_case_level` data file are: `Grantee_ID` and `Case_ID`. Since `Grantee_ID` and `Case_ID` are unique to both data files, we will merge/join the records based on those unique identifiers.

In this example, we are interested in preserving all records from the `rpg_esl_person_level` data file that contains 8,626 records pertaining to each individual who participated in the RPG-2 study. Due to the `rpg_esl_case_level` data file containing information about the "case," there are only 2,887 records. To summarize, each **case** represents multiple **individuals** (substance abusing adults and their children). When we merge/join these two datasets together we want the information about the case properly associated with records for the individuals. This is known as a many-to-one relationship. The `Joinby` procedure paired with the "unmatched (master)" sub command will create a joined file consisting of all records from the `rpg_esl_person_level` and the matching records from the `rpg_esl_case_level` file.

The instructions are broken out into sections and should be followed in the order in which they appear. Data users should copy each section of syntax into the Do-file Editor window of Stata, omitting section headings and other explanatory text. You will need to update all file paths to point to the location of the data file on your own computer and provide locations to where the modified data files should be saved.

The instructions were written based on Version 14 of the Stata software package.

### **Section One: Sort the `rpg_esl_person_level` data file**

This section will assist you in sorting the `rpg_esl_person_level` data file. The `Joinby` procedure requires sorting each data file by the key IDs that will be used during the join/merge. Remember:

variable names are case sensitive in Stata so be sure to confirm that the variables listed in the syntax match the data file exactly.

```
clear  
  
use "c:\temp\change-the-file-path-to-  
location\rpg_esl_person_level-data.dta"  
  
sort Grantee_ID CASE_ID  
  
save "c:\temp\change-the-file-path-to-  
location\rpg_esl_person_level-data-sorted.dta", replace
```

## Section Two: Sort the rpg\_esl\_case\_level data file

This section will assist you in sorting the rpg\_esl\_case\_level data file, which is a necessary step prior to conducting the Joinby procedure in Section Three.

```
clear  
  
use "c:\temp\change-the-file-path-to-  
location\rpg_esl_case_level.dta"  
  
sort Grantee_ID CASE_ID  
  
save "c:\temp\change-the-file-path-to-  
location\rpg_esl_case_level-sorted.dta", replace
```

## Section Three: Merge/join the rpg\_esl\_person\_level data file with the rpg\_esl\_case\_level data file.

In this section, you will perform the data join/merge of the rpg\_esl\_person\_level data with the rpg\_esl\_case\_level data file.

```
clear  
  
use "c:\temp\change-the-file-path-to-  
Location\rpg_esl_person_level-sorted.dta"  
  
joinby Grantee_ID CASE_ID using "c:\temp\change-the-file-path-  
to-location\rpg_esl_case_level-sorted.dta", unmatched(master)  
  
drop _merge  
  
save "c:\temp\change-the-file-path-to-location\joined-  
rpg_esl_case_level-and-person_level.dta", replace
```

In this example, you should have a combined data file that contains 8,626 records and 67 variables. You should refer to each respective data file's codebook for information pertaining to the variables in the newly joined/merged dataset.

This document was prepared by NDACAN Analyst, Holly M. Larrabee. Email questions to [NDACANsupport@cornell.edu](mailto:NDACANsupport@cornell.edu).