Improving efficiency and accuracy in data management for naturalistic driving studies

Rusan Chen

Georgetown University

Overview

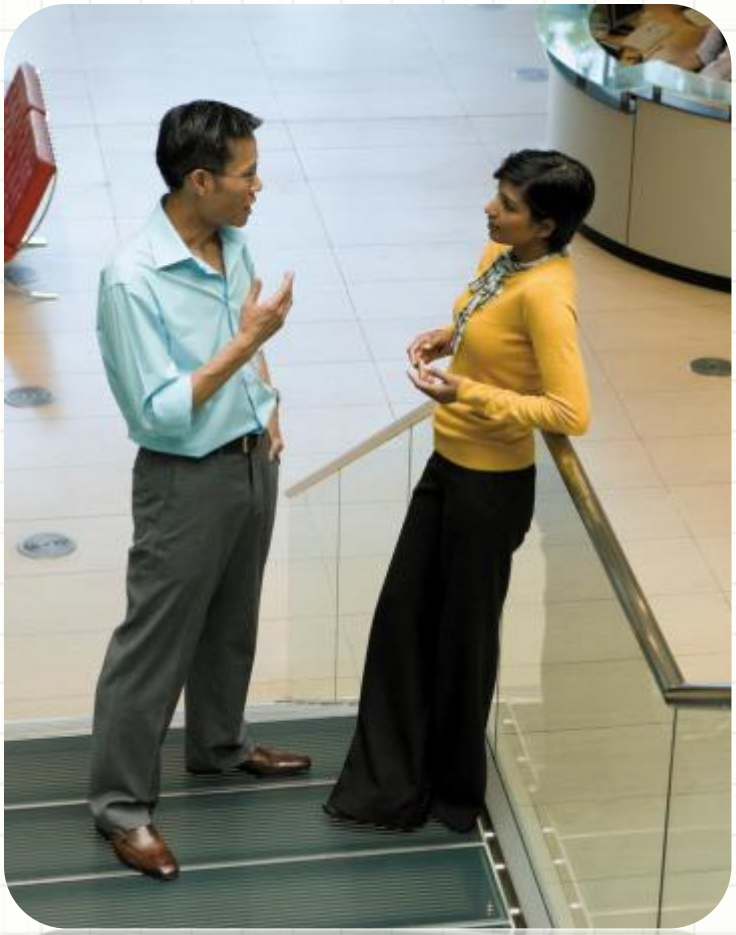
1. Naturalistic driving studies involve complicated, dynamic datasets
2. Efficient data management is essential for the analysis results being replicable
3. Based on my experience working on the 40-car Naturalistic Driving Study

Sound familiar?

- *You have multiple versions for the same file and don't know which is which.*
- *You cannot find an important file and think you may have deleted it.*
- *There are two versions of the 'latest' draft for a paper, with the same name 'final.doc'*

Efficient workflow requires proper

- Organizing
- Documenting
- Automating
- Archiving



Organizing

- \Work and \Post directories are critical
- Once a file is posted, it is never changed!

Example:

C:\40Car

\ADS

\Work

\Post

40carAnalysis.doc

Organizing folders

- \Post
  - \2009
  - \012710 survey questionnaire analysis
  - \013110 personality related to risky driving
  - \031110 predicting C/NC from g-force
  - \032710 SAS Glimmix
  - \033010 risky friends interaction
  - \052110 speeding analysis
  - \052410 perception of risk as mediators
  - \060610 high vs low risky drivers

Documenting

It is always better to document today than tomorrow

What to document?

- *Date*
- *Purpose*
- *Data sources*
- *How to form new composite scores*
- *Steps of analysis*
- *Where to save the results*
- *To whom you sent the results*

Automating

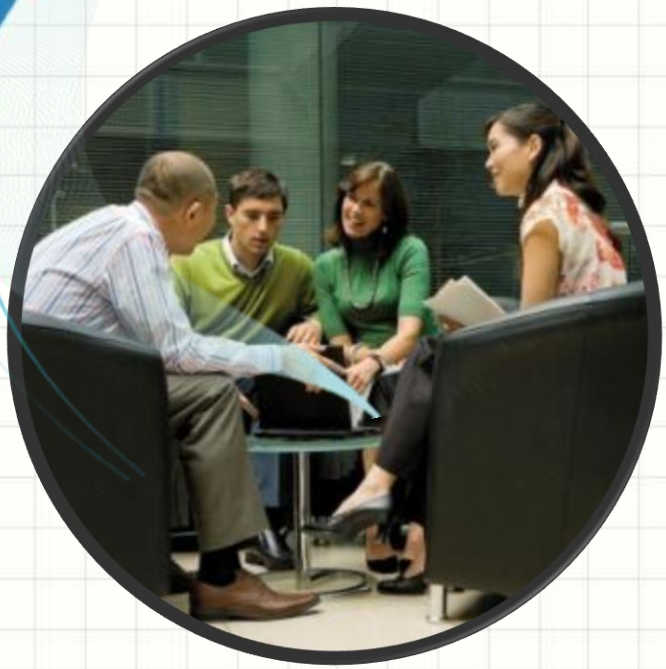
- ***Data management involves doing the same task multiple times.***
- ***Automating these tasks can save time and prevent errors***

What to automate? (using macros and loops)

- *To update, merge, and subset datasets*
- *to create and label new variables*
- *To check outliers*
- *To define and report missing values*
- *To fit a sequence of similar models*
- *To save analysis results*

Archiving: to protect your files

short-term	mid-term	long-term
mirror	backup	archive



Thank you!

Reference

- Long, JS (2009) The workflow of data analysis using Stata. Stata Press, TX: College Station