# File and Variable Naming: Proposed Guidelines

## **General Conventions:**

* FOLDER NAMES ARE ALL UPPER CASE
* all other names: all lower case
* no blank spaces, replace with:
  + hyphen (-) in file, folder names
  + underscore (\_) in variable names
* first character of name must be a letter
* no non-alphanumeric characters (e.g., &, /)
  + except for hyphens, underscores
* use three-letter acronyms/initials to replace longer strings, wherever possible
* all specifiers are at least three characters long, some are longer for clarity (e.g., “child” instead of “chi”)

## **Folder names**

#### Example of a good folder name:

INPUT-FILES

* all UPPER-CASE letters
* no blank spaces (replaced by hyphens)

## **File names**

#### Example of a good file name:

tod-data-stand-dsh.csv

* general format: [project]-[file type]-[data source]-[operator].csv
* all lower-case letters
* no blank spaces (replaced by hyphens)
* specifiers
  + project: three-letter acronym
  + file type: **data** (or items, plan, etc.)
  + data source: **stand** (or clin, trt, irr, etc.)
  + operator: three-letter initials

Discussion:

* What are some other possible file specifiers?
* Within each specifier type, what is the exhaustive list of categories?

## **Variable names**

### Case identifier

* id\_num
  + person\_id
  + site\_id

### Demographics

* age:age in months, expressed as a three-digit integer with leading zeroes (e.g., “005”)
  + dob: date of birth
  + doe: date of evaluation
  + age\_yr: age in whole years
  + age\_mo: month of current age year (0-11)
  + age\_range: the upper bound of each age range in months (three-digit integer with leading zeroes)
* gender
* educ
  + educ\_parent
  + educ\_self
* ethnic
* hispanic
* region

Discussion:

* Any other important demo variables?

### Items

##### Example of a good item name:

child\_dep\_i001

* general format: [form]\_[scale]\_i[item number]
* all lower-case letters
* no blank spaces (replaced by underscore)
* item number is preceded by “i” to signal that column contains item responses
* item number is always three digits with leading zeroes, to ensure that column names will sort properly
* if items are numbered separately by scale, they should be renumbered starting with 001. The scale specifier allows scale-wise numbering to be reapplied, if needed, during processing.

Discussion: How to distinguish items of the test under study from items of other measures administered concurrently?

## Scores

#### Example of a good score name:

dep\_raw

* general format: [three-letter score acronym]-[three-letter score type]

Discussion:

* What are other score types: tsc, raw, sts, pct, dif . . . ?
* How to distinguish scale scores from total scores or composite scores?