

# Standard Operating Procedures

**Subject:** Error Detection

**Date:** 15Jun2006

**Purpose:** Set the minimal standards for error checking in data files handled by the Data Management and Analysis Center (DMAC).

**Scope:** These standards apply to computerized data files managed by the Data Management and Analysis Center.

**Responsibility:** Designing a system to check for possible errors in the computerized data files for a project and running the programs to perform the error detection is the responsibility of the programmer assigned to that project.

**Procedures for Error Detection:** Every time observations are added or values are changed in a permanent dataset, the dataset should be thoroughly checked for possible errors. The error detection process is typically accomplished with an error checking program which scrutinizes the values for variables in each observation for possible errors.

Some of the types checks for possible errors are listed below:

- Duplicate key values checks. Do the values for the designated key variables for the dataset create a unique identification for each observation?
- Range checks. Is the value for a continuous numeric variable within the range of expected or believable values?
- Valid values checks. Is the value for a categorical or ordinal variable on the list of acceptable values?
- Consistency checks within observation. Are the values for variable in an observation in agreement with each other? For example, date of birth for the parent must occur a reasonable number of years before date of birth for the child.
- Consistency checks between observations. Are values for variable(s) in different observations consistent with each other? For example the date of the observation for visit 1 should occur before the date of the observation for visit 2.
- Check for missing values. Are there missing values that may be researched and recovered?
- Check for missing observations. Is an expected observation missing? For example, is there are observations for visit 1 and visit 3, should there be an observation for visit 2?

Possible errors should be reported in a fashion that can be used by those responsible for resolving them. Useful information is the identifiers for the observation(s) in which the problem occurs, the variable(s) that hold the problem values, the value(s) in question, and

a description of the problem. It is also useful to provide a space for the corrected value(s) and comments by those resolving the problem.

Typically it is not the programmer's responsibility to resolve suspected errors. The report of possible errors is generally sent to the project staff who originally gathered the data. They may then authorize corrections or verify that the suspect value is indeed correct. If the suspect value is not an error, then the error detection program may need to be refined so that that value will not show up the next time the program is run. The error correction process will be covered in a separate SOP.