# Data Entry and Verification

## Purpose

Data entry and verification are two separate but interrelated steps in data management. Timely completion of these steps ensures maximum data completeness and accuracy. Though ideal timing will vary by project, data type, and available resources, completion of entry and verification within one week of data collection is a standard best practice. Important considerations include: \* Identification of parties responsible for entry, verification and oversight \* Utilization of software appropriate to the project’s demands \* Designation of a secure storage location \* Development of work-flow detailing tracking procedures and frequency of completion

## Scope

This SOP is intended for use by any clinical research staff who may be involved in data entry and verification of a dataset

## Responsibilities

#### Data Manager

* To identify research assistant(s) for entry and verification tasks
* To detail how staff members will know data is ready for entry and verification
* To implement the method in which timely entry and verification will be accomplished and logged

#### Research Assistant

* Enter data electronically after collection

Spreadsheets and/or calender-based project management tools can aid in oversight of data entry and verification. It is also wise to designate a physical location where forms in need of entry/verification are temporarily stored that is distinct from the long-term storage of forms that have been entered and verified.

## Procedure

1. Data Entry

* Data Manager will give data to designated Research Assistant(s) to enter electronically in a timely manner after collection.
  + Though the project’s specific demands and resources dictate the appropriate timing, prioritizing data entry can aid in identifying potentially missing data.
  + In designating Research Assistant(s) for data entry, investigators and managers should consider whether or not the data will unblind staff to treatment assignment.

1. Data Verification

Data is verified in one of two ways:

* 1. Data is entered once by one study staff member, who marks it ‘Ready to be reviewed’ electronically. Subsequently, the entered data is reviewed by a second study staff member and compared to the original data. The second staff member will then make any necessary corrections and mark it as ‘Completed’ or ‘Reviewed’.
  2. In the second verification method, double data entry is utilized. Two staff members will separately enter identical data sets into the collection instruments, and a third staff member will then compare the two entries for discrepancies, make corrections, and then merge them into one complete, verified file.

## References

#### Software

The most appropriate software for entry and verification will vary by a project’s demands, resources, and scope. In selecting software, investigators and managers should consider the types and volume of data involved, the capabilities of staff responsible for entry/verification and oversight, and institutional/field standards. Though there are numerous software options, there are several common packages used by the Division of Sleep Medicine Epidemiology, often in conjunction with one another for the same project:

1. [Slice](https://sleepepi.partners.org/slice/)

* Developed within the Division of Sleep Medicine Epidemiology, Slice is a secure web application that can be customized to the specific needs of a project.

1. [REDCap (Research Electronic Data Capture)](https://redcap.partners.org/redcap/)

* Originally developed at Vanderbilt University (2004), REDCap is a secure web application designed for building and managing online research databases.
* Databases can be designed online with minimal to no computer science background.
  + Data can be created interchangeably using drop-down menus, numeric or text entry, and can be restricted to meet data type requirements.
  + Inclusion of a drop-down menu for “Incomplete,” “Entered,” and “Verified” can document entry and verification process.
* Access can be restricted and logged by user.
* Data can be exported and stored locally for archiving in formats compatible with common statistical packages (SAS,R,SPSS,Stata).

1. Offline software

* Spreadsheet (e.g. Excel), database (e.g. Access), and other common software users are likely to have previously encountered
* Direct entry into statistical programming or other software formats (SAS,R,etc.).

#### Storage

Because of privacy and security concerns, special consideration is needed for the storage of research data. Additional consideration must be made regarding blinded data to ensure the integrity of the research process. Investigators and managers should designate storage space for:

1. Live data

* For security and data integrity purposes, restrict user access to “live data” to users who enter and verify data or oversee the process.
* Data should be stored in password protected and encrypted locations.
* Unless study demands and resources are such that only unblinded staff can enter, verify, and oversee all data, create separate access to blinded and unblinded data.

1. Archived data

* Live data should be archived frequently to maintain data integrity.
  + The specific frequency will vary by project, but archiving weekly or monthly is a generally good practice.
  + Archived data should incorporate the date it was stored (e.g. as a suffix in the filename).
* Like live data, archived data should be restricted by user and password protected.
* Archived data should be accessible only to unblinded staff or should be compartmentalized into different locations with different access rights.

=============== [Back to README](README.md)