

Data Collection Basics

A Mayo Resource

The Department of Health Sciences Research has developed a website to introduce investigators to the basics of data collection. The site also includes common mistakes and an introduction to collecting data properly using REDCap.

Here's an overview of what you'll find at this site:

<i>Page</i>	<i>Content</i>	<i>One Example... Check the website for more!</i>
Introduction to data collection		
• Clinical Data Management	Overview	
• Terminology	Terms used throughout the site	Key: 1 or more variables that make an observation unique within a table.
• Variable types	Descriptions and examples of different variables types	The analysis methods change depending on the variable type
○ Numeric variables	Definition, examples, collection suggestions, and software pitfalls	Numeric variables can become text variables when symbols such as < or > are included with the numerical value.
○ Categorical variables	Definition, examples, collection suggestions, and software pitfalls	REDCap checkboxes can be particularly problematic.
○ Date variables	Definition, examples, collection suggestions, and software pitfalls	If using Excel to collect data, take extra precautions when entering values because Excel may automatically reformat or change the value.
○ Free text variables	Definition, examples, collection suggestions, and software pitfalls	Because they can take a long time to review, the cost to analyze these can be high
• Variable details	Concepts, data dictionary	It is important to understand and document a number of items about the variables being collected before data collection begins.
• Missing data	Concepts	If it is important to make a distinction between any of these, make sure a unique code is assigned to each type of missing.
• Steps for building a case report form (CRF)	Multi-step process	At a minimum, the following people should review each version of the CRF: <ul style="list-style-type: none"> • Person abstracting/collecting the data • Person entering the data • Principle investigator • Statistician
• Endorsed electronic data collection systems	Brief description of the 3 systems	Excel and Access have been common software tools, so why are they not on this list?
○ Choosing the right data collection	What to consider	Choosing a data collection system is not based on the number of subjects you plan on including in your study. Rather, it is based on various requirements of your

Page	Content	One Example... Check the website for more!
system		study. A "small" study with 100 patients and extensive follow-up will not have the same needs as a "large" study with 3,000 patients and only 2 very basic forms.
Building effective REDCap databases		
• Classic vs. longitudinal design in REDCap	Definitions, useful study information, classic & longitudinal designs	If you plan to analyze your own data and hope to use a longitudinal design... It could be time consuming to redo this data manipulation, which you will need to redo each time you pull the data.
• Field types and validation in REDCap	12 field types, examples of which field types and validation to use	Avoid Open Response fields as much as possible. If the full list of possible options is not known, then use a Closed Response with an option for "Other" and a spot to record the value for "Other."
• Checkbox issues within REDCap	Checkboxes, example of problem with checkboxes, fixing the checkbox problem	Recall that checkboxes in REDCap by default only have 2 states -- checked and unchecked -- there is no option for missing. This behavior by REDCap is causing the problem.
• Building a matrix of fields in REDCap	Description, example, data from matrix of fields, comparing checkboxes and matrix of fields data	The data for the Checkboxes are filled in no matter what. The Branching Logic has no effect on the data.... the Branching Logic within the Matrix of Fields has result in the data behaving as expected.
• Miscellaneous tips	Variable names and field labels, calculated field type, user rights	
Contacting a statistician		Take advantage of the guidance a statistician can provide in study design and data collection planning.

- The site is located at <http://bsiweb.mayo.edu/node/781400> (must be logged in to the Mayo network)
- To access on a smartphone: use a QR reader app such as NeoReader to take a picture of this:

