

DATA COLLECTION PRACTICES



Claudia Mata, MPH
Samantha Wilson, MS

AGENDA

Getting Started

Data Dictionary

Variables

Building a Database

Tips & Tricks





GETTING STARTED

REDCap

Free with university account and CCTSI membership

- <https://redcap.ucdenver.edu/>

If you are a CHCO only employee, fill out a POI form:

- <https://research.cuanschutz.edu/comirb/home/request-poi>

Getting Started:

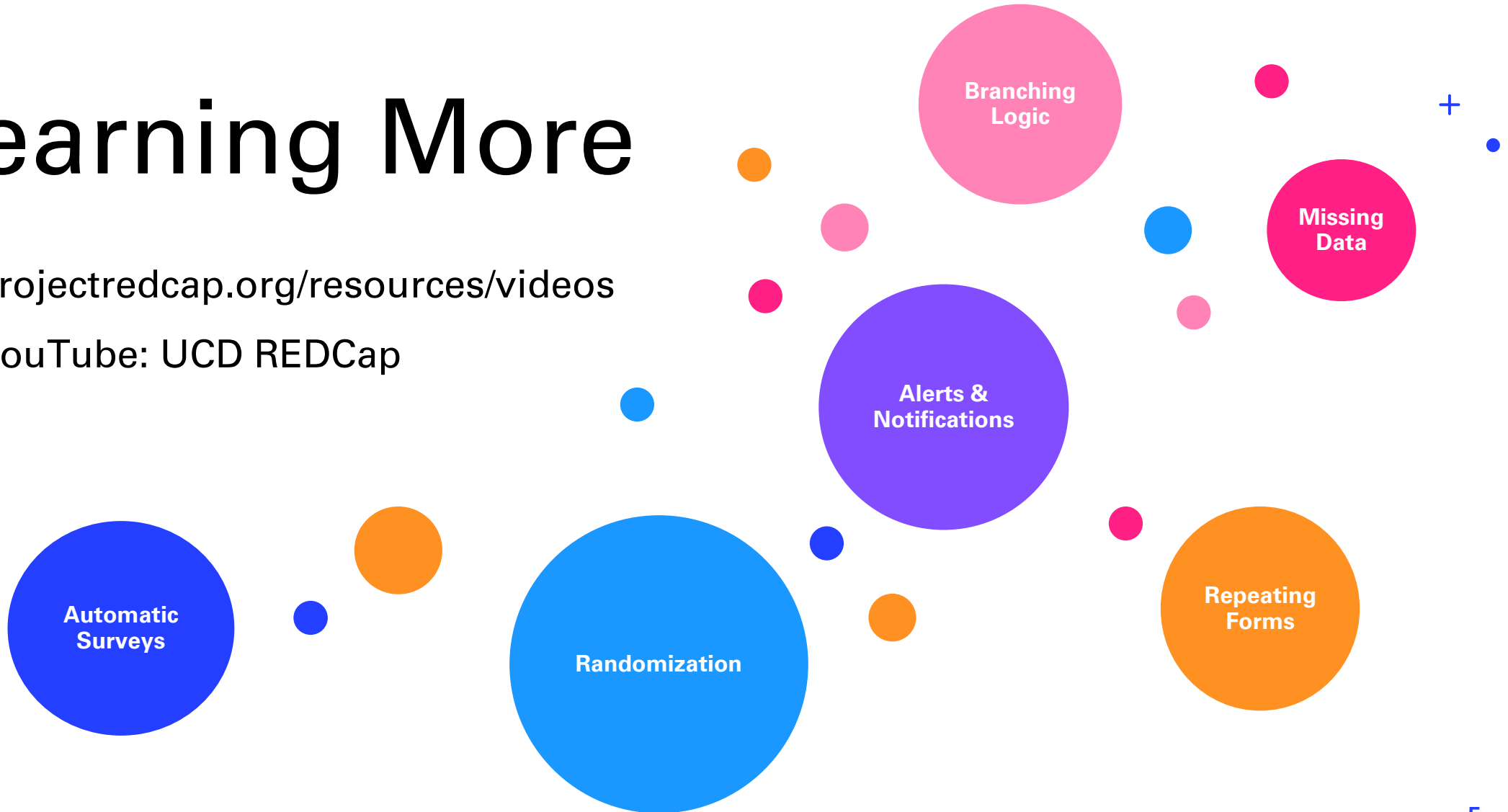
1. Complete the full online REDCap tutorial or attend an in-person tutorial
2. Online tutorial takes approx. 2 hours
3. View REDCap Resources for additional project design/setup help

<https://cctsi.cuanschutz.edu/resources/informatics/redcap-resources>



Learning More

- [Projectredcap.org/resources/videos](https://projectredcap.org/resources/videos)
- YouTube: UCD REDCap

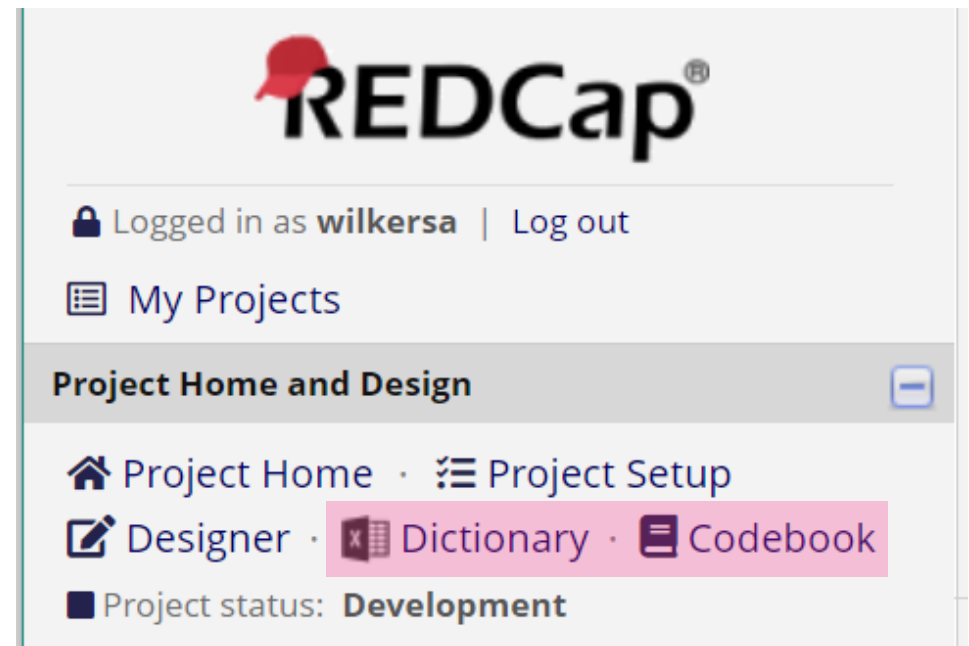













DATA DICTIONARY



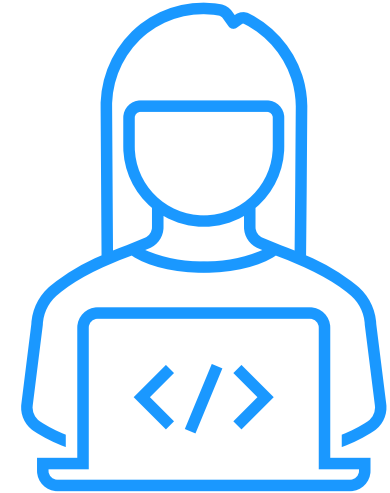
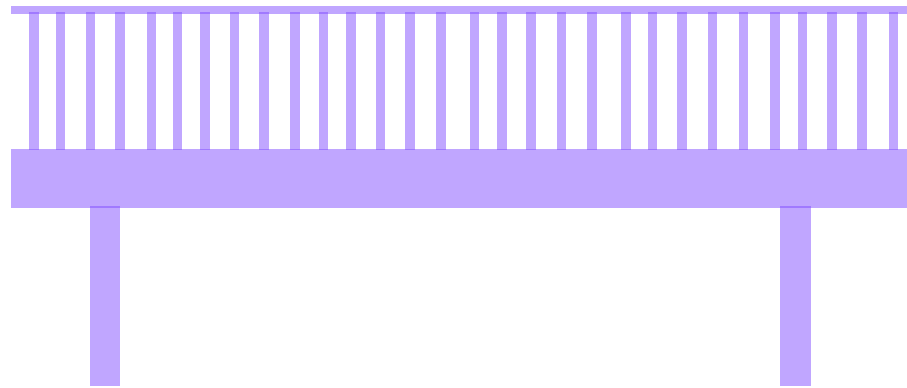
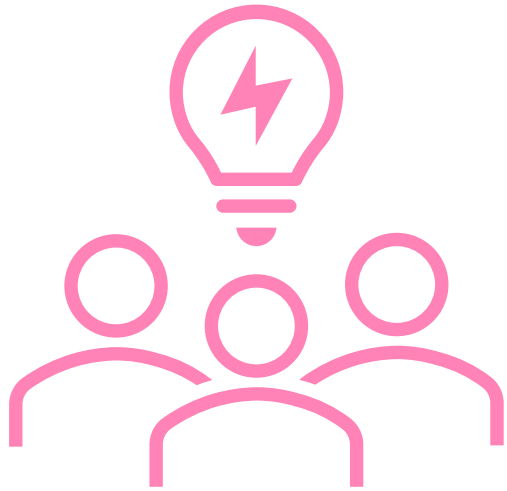
What is a Data Dictionary?

- A document, commonly .csv file, that contains your variables, labels/descriptions, choices, calculations, etc. for your database
- Download your data dictionary as excel file from the “Dictionary” tab
- View your data dictionary in the “Codebook” tab








	#	Variable / Field Name	Field Label <i>Field Note</i>	Field Attributes (Field Type, Validation, Choices, Calculations, etc.)						
Instrument: Demographics (demographics)				^ Collapse						
	1	[record_id]	Record ID	text						
 	2	[mrn]	Medical Record Number	text (number), Identifier Custom alignment: LH						
 	3	[dob]	Date of Birth	text (date_dmy), Identifier						
 	4	[visitdate]	Date of Visit	text (date_dmy), Identifier Custom alignment: RH						
 	5	[age]	Age at Visit	calc Calculation: rounddown(datediff([dob], [visitdate], 'y')) Custom alignment: RH						
 	6	[gender]	Gender	radio <table border="1" data-bbox="1559 1096 1760 1305"> <tr> <td>1</td> <td>Male</td> </tr> <tr> <td>2</td> <td>Female</td> </tr> <tr> <td>3</td> <td>Other</td> </tr> </table>	1	Male	2	Female	3	Other
1	Male									
2	Female									
3	Other									

The Data Dictionary is a Bridge








What is SAD an abbreviation for?
Seasonal Affective Disorder or
Social Anxiety Disorder?

     Variable: sad

SAD diagnosis? Yes
 No

reset

Both the variable and the
field label are more
descriptive!

     Variable: sad_dx

**Does the patient have a
Social Anxiety Disorder
diagnosis?** Yes
 No

reset

Indicate Identifiers

Edit Field ✕

Field Label ☐ Use the Rich Text Editor ?

Date of Birth

Action Tags / Field Annotation (optional)

Learn about [@ Action Tags](#) or [using Field Annotation](#)

Variable Name (utilized in logic, calcs, and exports)

dob ☐ Enable auto naming of variable based upon its Field Label?

ONLY letters, numbers, and underscores

How to use [\[?\] Smart Variables](#) [Piping](#) [Field Embedding](#)

Validation? (optional) Date (D-M-Y) ▾

Minimum:

Maximum:

Tip for min/max limits: You may use the word `today` or `now` (not wrapped in quotes) for date/time fields to set dynamic range limits. Also, you may pipe a value from another field - e.g., `[visit_date]` or `[event_1_arm_1][age]`.

– OR –

-- select ontology service -- ▾

Required?* No Yes

* Prompt if field is blank

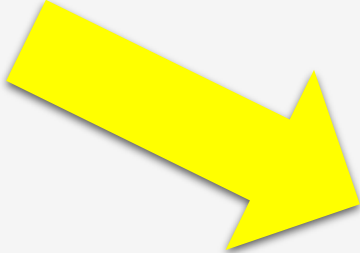
Identifier? No Yes

Does the field contain identifying information (e.g., name, SSN, address)?

Custom Alignment Right / Vertical (RV) ▾

Align the position of the field on the page

Save **Cancel**



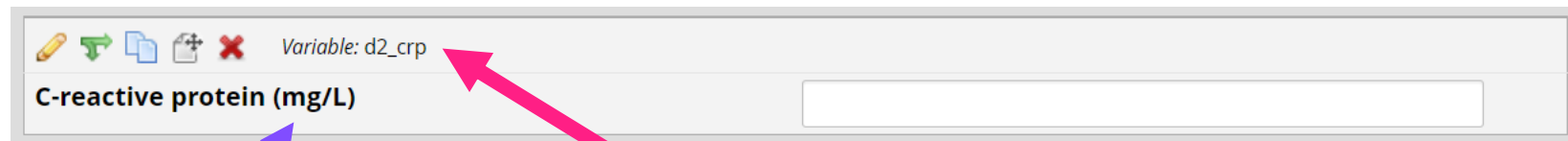
VARIABLES



Naming Variables

- Keep it short, but descriptive
- No spaces, but can use _
- Put units and/or a longer description in the field **label**

You want to make a field for C-reactive protein lab results. There is already a “crp” field in your Data Dictionary from Research Day 1.



The screenshot shows a data dictionary entry for a variable named 'd2_crp'. The variable name is displayed in the top right corner of the entry box. Below it, the field label 'C-reactive protein (mg/L)' is shown. To the right of the label is an empty input field. A blue arrow points from the text 'Units are included in field label' to the '(mg/L)' part of the label. A red arrow points from the text 'Variable is short, but describes the data to be entered' to the 'd2_crp' variable name.

Units are included in field label

Variable is short, but describes the data to be entered

Demographic Variables

Age (What is your age?)

Gender

(Which gender do you identify with?)

- Male
- Female
- Other _____
- Prefer not to say

Race

(Which race do you identify with?)

- American Indian or Alaska Native
- Asian
- Black or African American
- Native Hawaiian or Other Pacific Islander
- Latino or Hispanic
- White
- Two or more

Demographic Variables

Insurance Type

- Private Insurance
- State (Medicaid, Medicare)
- No Insurance/Cash/Uninsured
- VA/Military
- *Depending on population, you may need to include others*

Valuable Variables

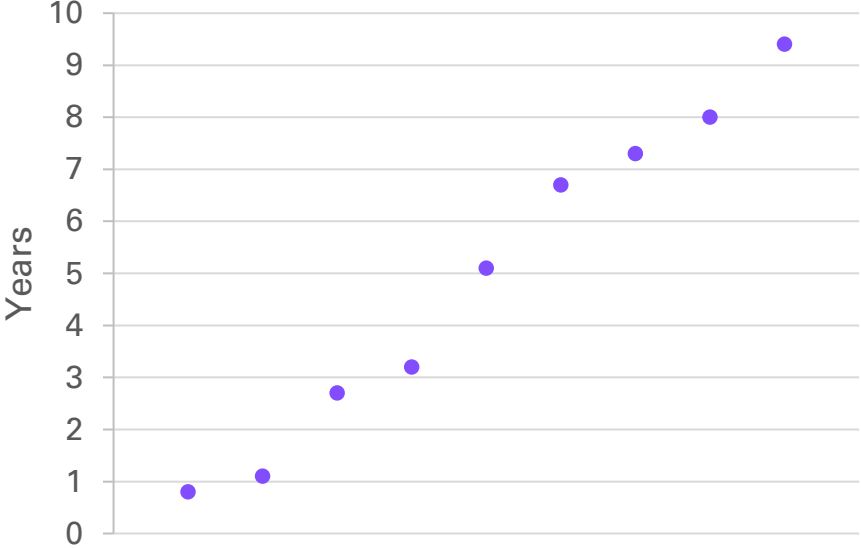
- What information do you need to address your aims?
- Does it add important information about your study population? (gender, age, race)
- Does it fall under the scope of information you are going to analyze?
- No data fishing!



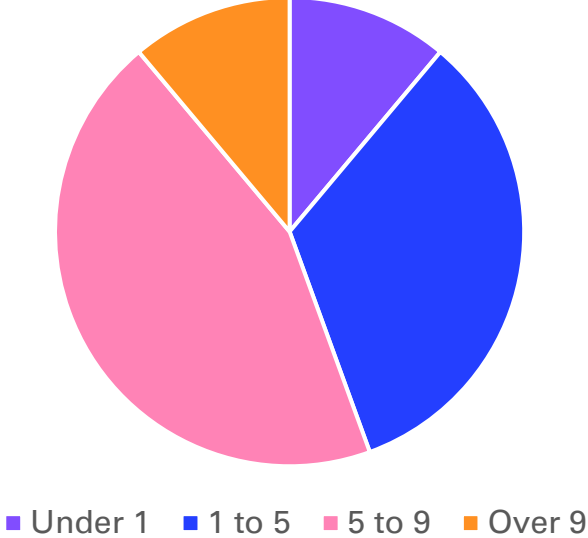
• + BUILDING A DATABASE • +

Data Capture

Continuous



Categorical



Consistency

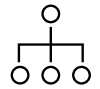
- Dates
 - D/M/Y
 - M/D/Y
- Units
- Coding Data
 - **Multiple Choice** vs. **Checkboxes**

Use validation!

The screenshot shows a configuration panel for a variable named 'mrn'. The 'Validation?' dropdown menu is open, displaying various validation types such as Date (D-M-Y), Datetime (D-M-Y H:M), and Integer. A pink arrow points to the 'Validation?' dropdown.

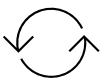
	record_id	mrn	dob	visitdate	age	gender	race	race_2	race_2	race_2	race_2	race_2	race_2	d2_crp	demographics_complete
1															
2	1	111111	8/8/2010	8/8/2022	12	1	1	1	0	0	0	0	0	3	2
3	2	111112	8/8/2008	8/8/2022	13	2	2	0	1	0	0	0	0	2.7	2
4															
5															

Be Efficient!



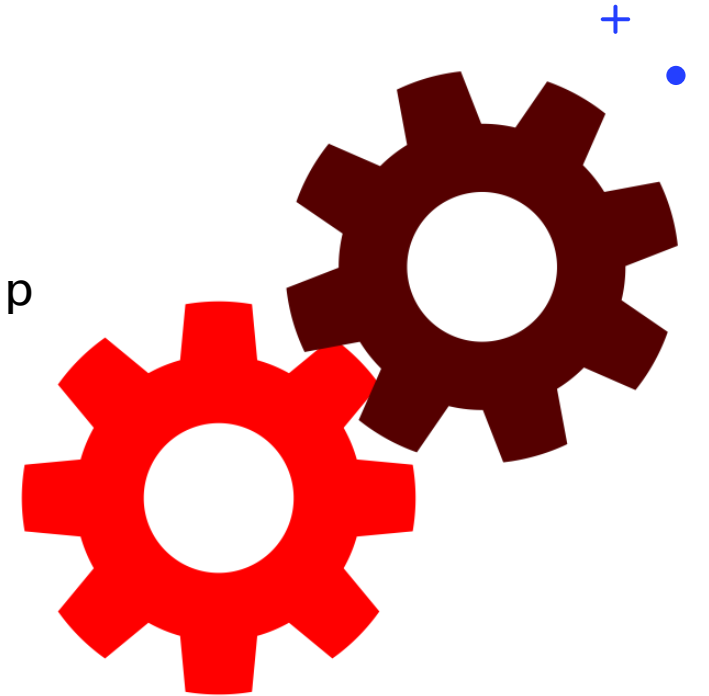
Branching Logic

- Use to “expand” answers and ask conditional follow-up questions
- Pros: Helps keep data & surveys clean
- Cons: Unintended errors in logic can happen easily, especially if there are a lot of branches



Repeating Instruments

- Use for repeat data that do not have defined time points
- Longitudinal studies



TIPS & TRICKS

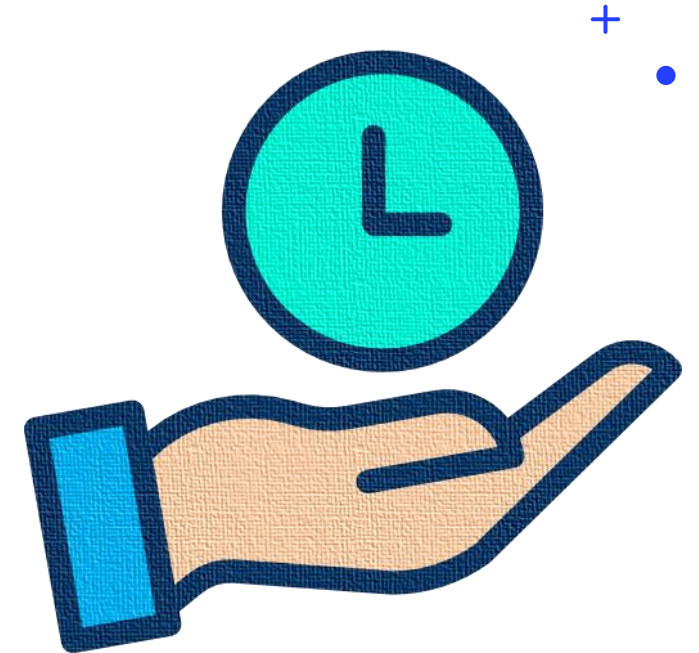


Time Savers



Shared Library

- Thousands of curated instruments approved by the REDCap Library Oversight Committee (REDLOC)
- Examples:
 - ICD10 Major Depression Inventory
 - PROMIS Emotional Distress – Anger - Short Form 8a
 - Pelvic Floor Distress Inventory Short Form 20 Scored
 - Toronto Empathy Questionnaire
- <https://projectredcap.org/resources/library/>



Avoiding Messy Data

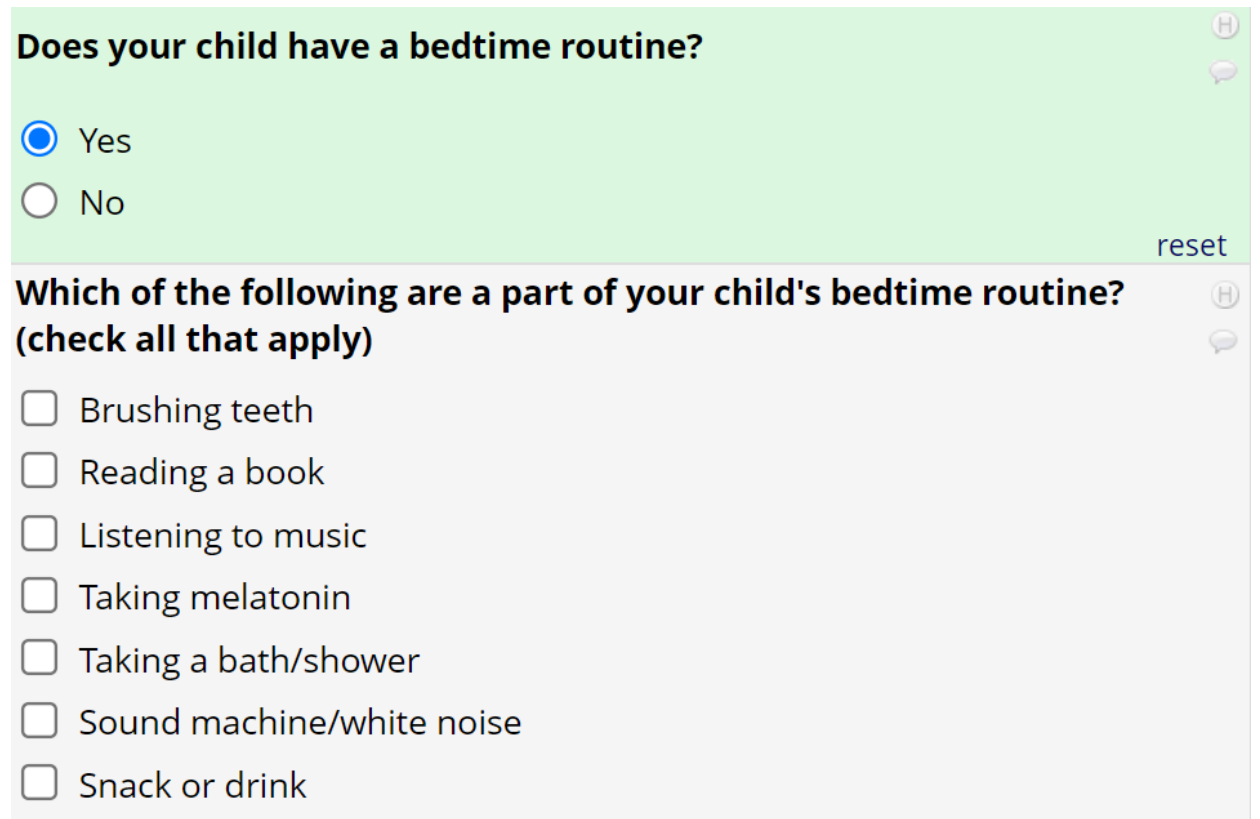
- Minimize free text fields
 - Sensitive to spaces, punctuation, spelling variations, typos, etc.
 - Use drop-downs, radio buttons, & check boxes when possible

[allergy]	Allergy	text						
[allergy_2]	Allergy	radio <table border="1"><tr><td>1</td><td>Sulfonamide</td></tr><tr><td>2</td><td>Penicillin</td></tr><tr><td>3</td><td>Tetracycline</td></tr></table>	1	Sulfonamide	2	Penicillin	3	Tetracycline
1	Sulfonamide							
2	Penicillin							
3	Tetracycline							

	A	B
1	allergy	allergy_2
2	sulfa drugs	1
3	tetracycline	3
4	sulfonamide	1
5	penicillin	2
6	tetra cycline	3
7	penecilin	2
8	sulpha	1

Avoiding Messy Data

Use a “screening” question before checkbox fields – add branching logic!



+

•

○

Does your child have a bedtime routine?

Yes

No

reset

**Which of the following are a part of your child's bedtime routine?
(check all that apply)**

Brushing teeth

Reading a book

Listening to music

Taking melatonin

Taking a bath/shower

Sound machine/white noise

Snack or drink

Data Collection: Surveys



Question Phrasing

- Survey vs. Data Entry
- Cultural sensitivity



Reading Level

- Aim for <6th grade
- Flyers, educational materials



Clear & Concise!

- Avoid double negatives & double-barreled questions
- Add a "N/A" option or screening question

Ask someone unfamiliar with your study to look at your materials!

DO:

- ✓ Be consistent, clear, and concise
- ✓ Be descriptive in field labels
- ✓ Consider reading level
- ✓ Consult your biostatistician!

DON'T:

- ✓ "Fish" for data
- ✓ Use free text fields if possible
- ✓ Use double-barreled questions
- ✓ Be afraid to get an outside perspective!

DATA COLLECTION PRACTICES

+



THANK YOU

QUESTIONS?