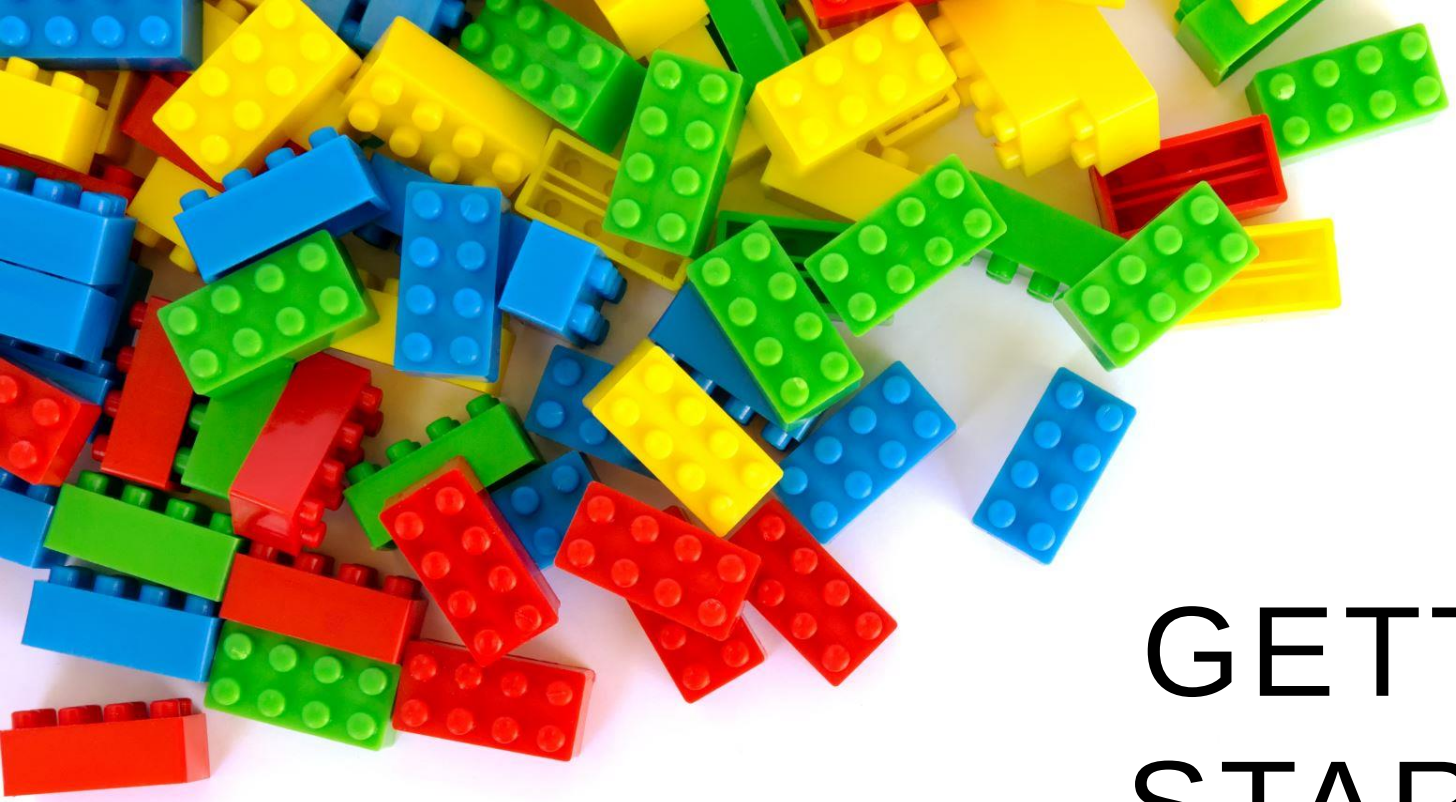
AUGUST 26, 2024

DATA COLLECTION PRACTICES

Allyson Calderon

OVERVIEW

- 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 and need InfoEd access:

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

For REDCap access:

- https://redcap.ucdenver.edu/surveys/?s=5ZLmzX

Getting Started:

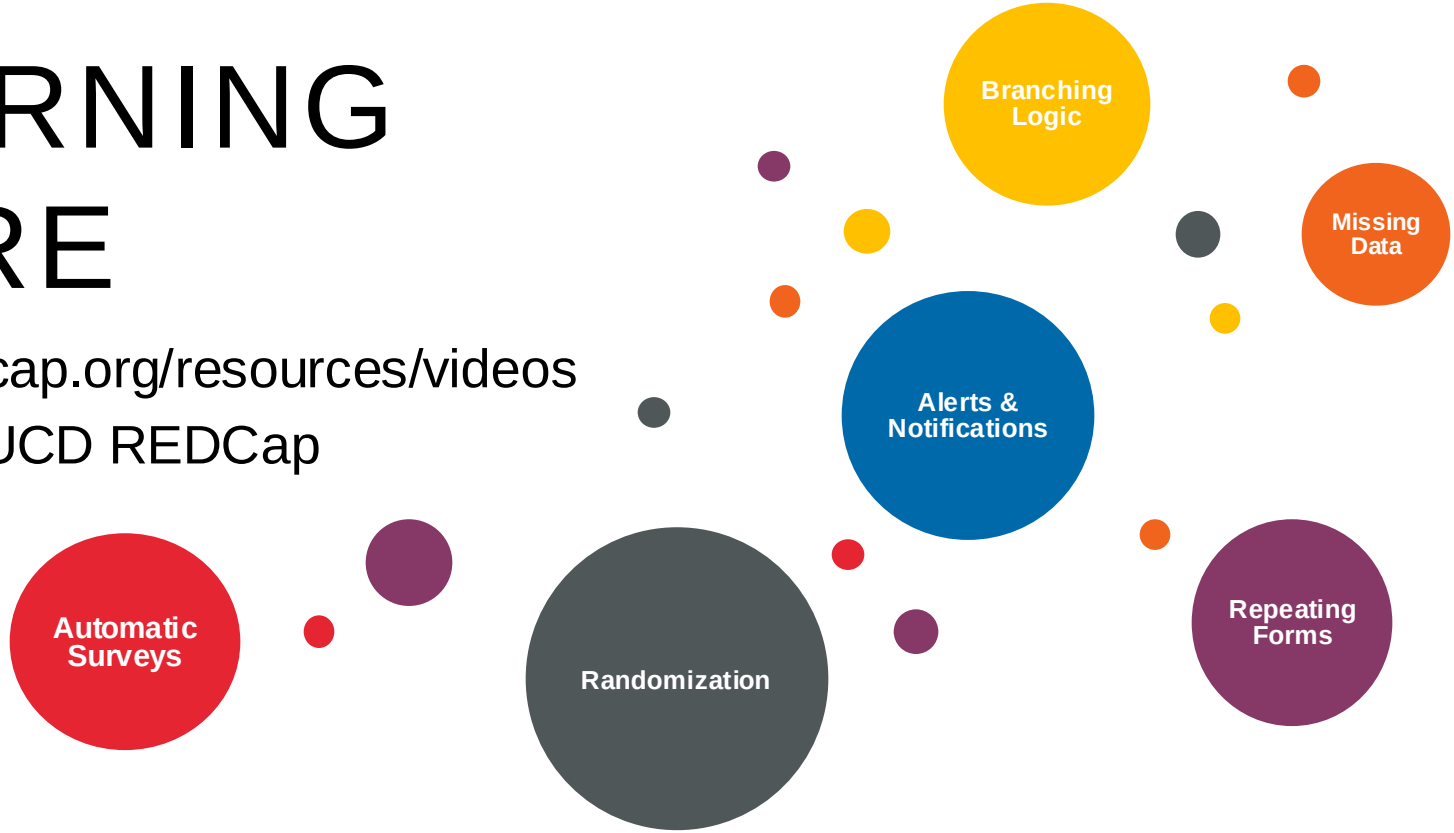
1. Complete the full online REDCap tutorial
2. Online tutorial takes approx. 2 hours
3. View REDCap Resources for additional project design/setup help

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

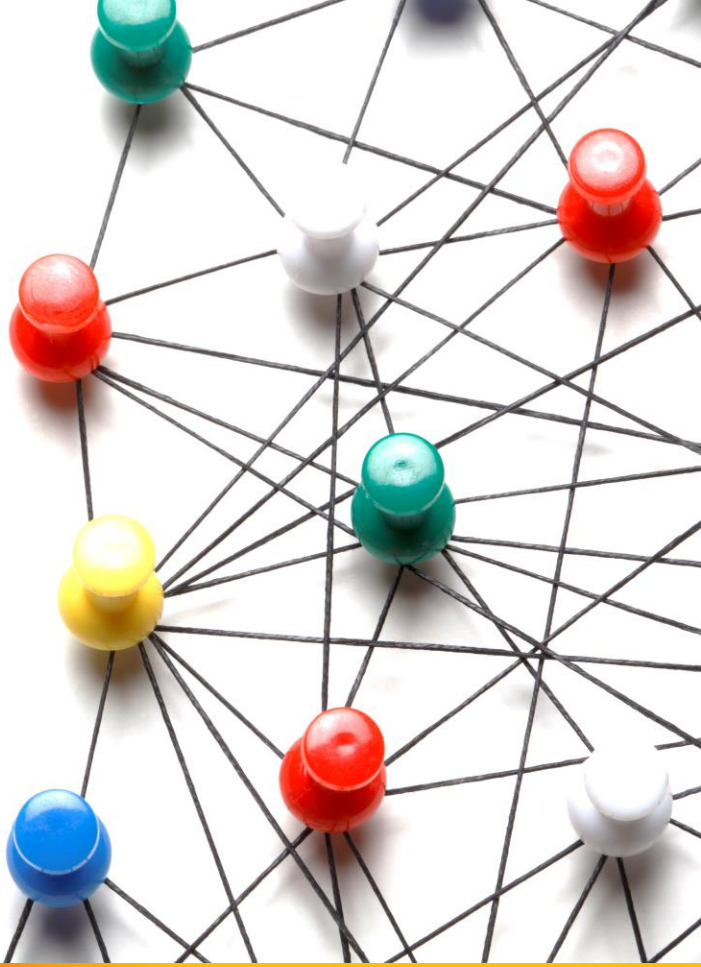
LEARNING MORE

Projectredcap.org/resources/videos

YouTube: UCD REDCap

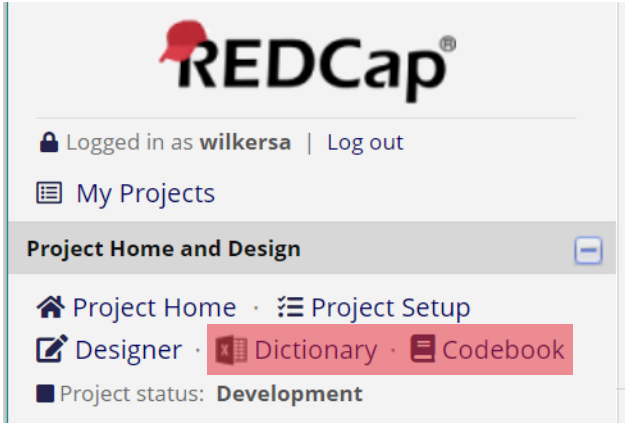


DATA DICTIONARY



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

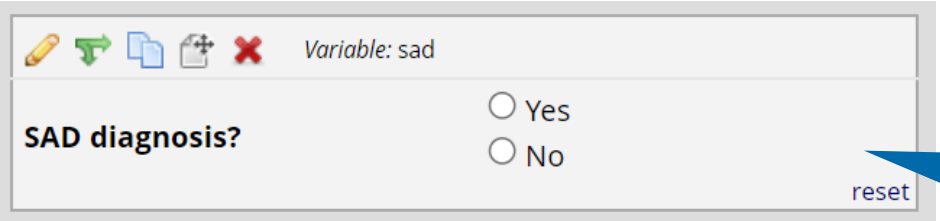


	#	Variable / Field Name	Field Label *Field Note*	Field Attributes (Field Type, Validation, Choices, Calculations, etc.)
Instrument: **Demographics** (demographics)				Collapse
	1	[record_id]	Record ID	text
	2	[mrn]	Medical Record Number	text (number), Identifier Custom alignment: LH
	3	[dob]	Date of 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 1 Male 2 Female 3 Other

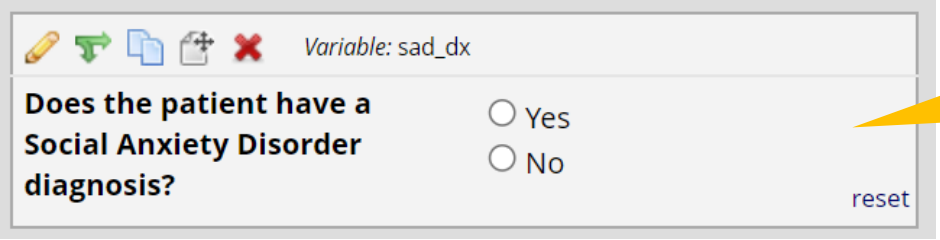
THE DATA DICTIONARY IS A BRIDGE

FIELDS & VARIABLES

FIELD NAMES



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

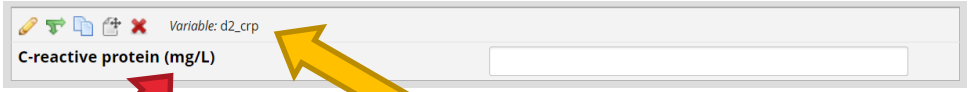


The field label is more descriptive!

NAMING VARIABLES

- Keep it short, but descriptive
- No spaces, but can use _
- Put units and/or a longer description in the field **label** not the variable name

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.



Variable is short, but describes the data to be entered

Units are included in field label

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

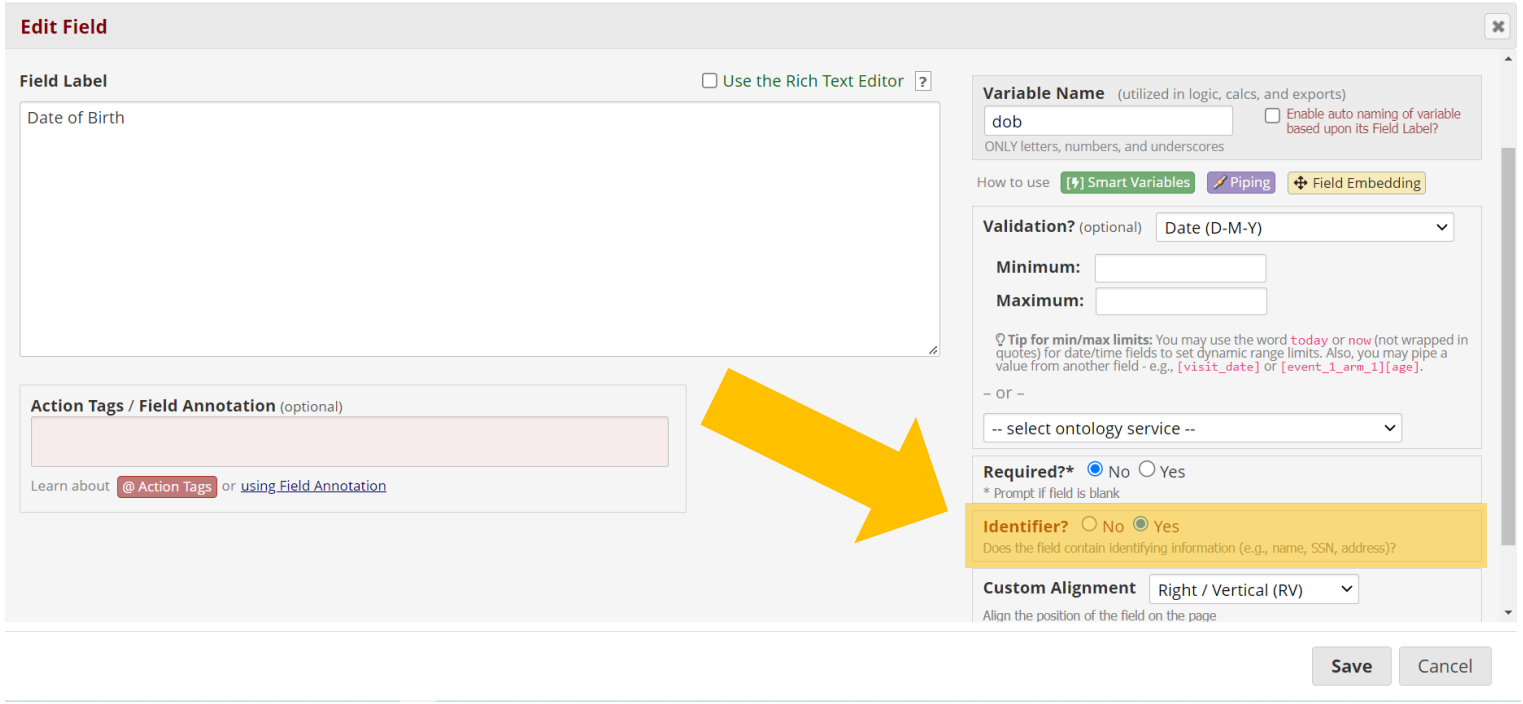
- Private Insurance
- State (Medicaid, Medicare)
- No Insurance/Cash/Uninsured
- VA/Military

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!

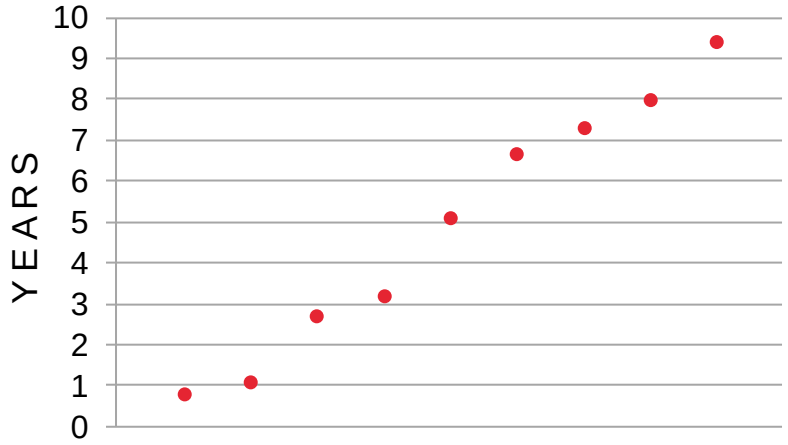
INDICATE IDENTIFIERS



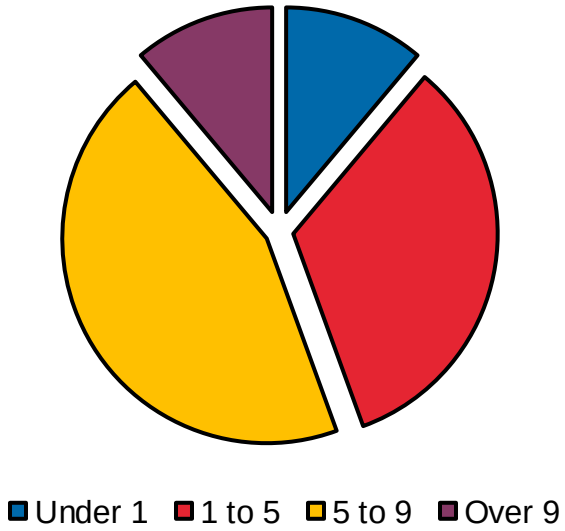
BUILDING A DATABASE

DATA CAPTURE

CONTINUOUS



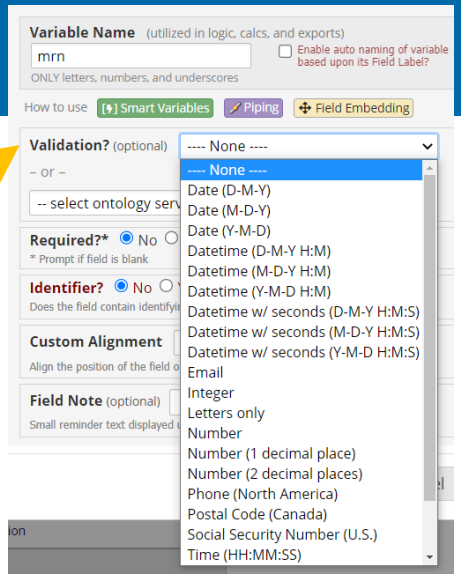
CATEGORICAL



CONSISTENCY

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

Use validation!

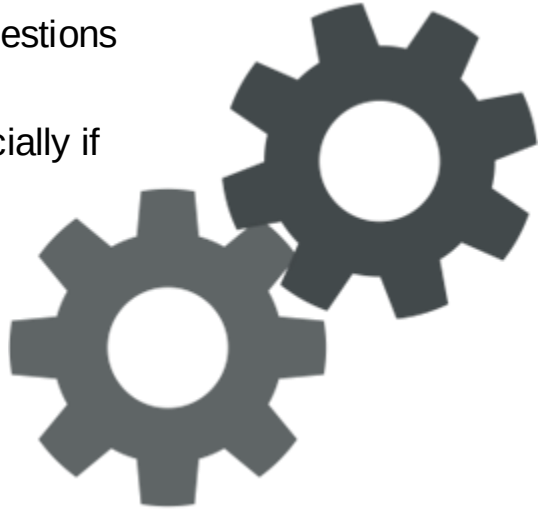


	record_id	mrn	dob	visitdate	age	gender	race	race_2___	race_2___	race_2___	race_2___	race_2___	race_2___	d2_crp	demographics_complete
2	1	111111	8/8/2010	8/8/2022	12	1	1	1	0	0	0	0	0	3	2
3	2	111112	8/8/2008	8/8/2022	13	2	2	0	1	0	0	0	0	2.7	2
4															
5															

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

Repeating Instrument

Example

Not Repeating

Record ID	Record ID	Enrollment Data	Demographics	Clinical Information	Week 2 Survey	Week 2 Photo	Week 4 Survey	Week 4 Photo	Week 6 Survey	Week 6 Photo	Week 8 Survey	Week 8 Photo
1												
2												

Repeating

Record ID	Demographics	Pre-op Patient Info	Clinical Details	Post-Op Day	Hospital Discharge	Follow-up	Complications
1 (Subject ID 01-001)							
2 (Subject ID 01-002)							

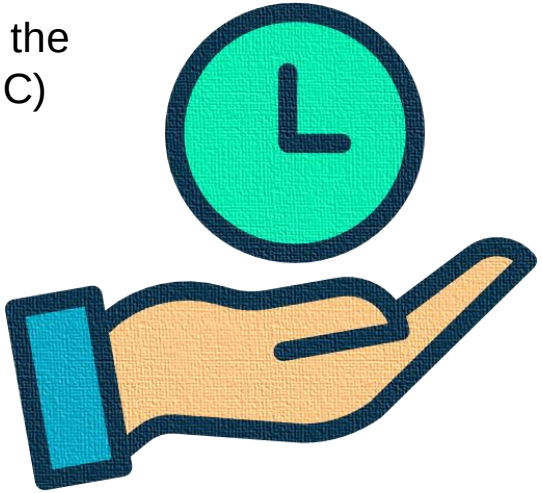
TIPS & TRICKS

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

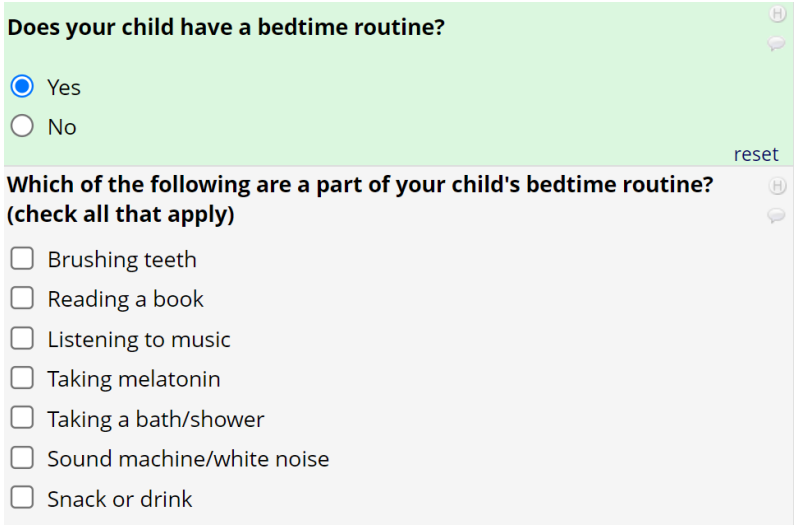
- Minimize free text fields
  - Sensitive to spaces, punctuation, spelling variations, typos, etc.
  - Use drop-downs, radio buttons, & check boxes when possible

[allergy]	Allergy	text
[allergy_2]	Allergy	radio 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!



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!

THANK YOU

QUESTIONS?