AUGUST 26, 2024

DATA PREPARATION FOR YOUR BIOSTATISTICIAN

ROCS Biostats Team

Center for Children's Surgery
SCHOOL OF MEDICINE
UNIVERSITY OF COLORADO ANSCHUTZ MEDICAL CAMPUS

Research Outcomes in
Children's Surgery (ROCS)
UNIVERSITY OF COLORADO ANSCHUTZ MEDICAL CAMPUS

Children's Hospital Colorado
Here, it's different.™

1

OVERVIEW

1 ROCS Biostatistics Services

2 Rules of Collaboration

3 Excel Data Cleaning

4 Developing a Data Dictionary

5 Preparing to Discuss Your Analysis

6 Data Sharing Guidelines

Center for Children's Surgery
SCHOOL OF MEDICINE
UNIVERSITY OF COLORADO **ANSCHUTZ MEDICAL CAMPUS**

Research Outcomes in
Children's Surgery (ROCS)
UNIVERSITY OF COLORADO **ANSCHUTZ MEDICAL CAMPUS**

Children's Hospital Colorado
Here, it's different.™

2

Suhong Tong, MS

Research Senior Instructor, Dept of Pediatrics

Areas of Expertise:
- Large data
- Longitudinal data
- Time series
- Survey analysis with population complex design
- Structural equation modeling
- Quality improvement analysis
- Survival analysis

Kaci Pickett-Nairne, MS

Research Instructor, Dept of Pediatrics

Areas of Expertise:
- Survival analysis
- Dynamic prediction
- Statistical consulting

Emily Cooper, MS

Research Instructor, Dept of Pediatrics

Area of Expertise:
- Large Data
- Statistical consulting

Samantha Bothwell, MS

Research Instructor, Dept of Pediatrics

Area of Expertise:
- Time Series
- Longitudinal Data
- Clustering Analysis
- Spatial Statistics
- Statistical Consulting

Kaitlin Olson, MS

Research Instructor, Dept of Pediatrics

Area of Expertise:
- Missing data
- Statistical consulting

Center for Children's Surgery
SCHOOL OF MEDICINE
UNIVERSITY OF COLORADO ANSCHUTZ MEDICAL CAMPUS

Research Outcomes in Children's Surgery (ROCS)
UNIVERSITY OF COLORADO ANSCHUTZ MEDICAL CAMPUS

Children's Hospital Colorado
Here, it's different.™

3

ROCS BIOSTATISTICS SERVICES

The biostatistician can collaborate in all parts of protocol development and implementation, not just the statistics!

- Sample size calculation
- Analysis plan and protocol development
- Big data analysis (PHIS, TQIP)
- Retrospective data consulting
- Prospective data consulting
- Randomization schedules

- Database structure and review (REDCap)
- Clinical trial and general study design
- Abstract and poster development
- Manuscript and grant preparation

Center for Children's Surgery
SCHOOL OF MEDICINE
UNIVERSITY OF COLORADO ANSCHUTZ MEDICAL CAMPUS

Research Outcomes in
Children's Surgery (ROCS)
UNIVERSITY OF COLORADO ANSCHUTZ MEDICAL CAMPUS

Children's Hospital Colorado
Here, it's different.™

4

RULES OF COLLABORATION WITH ROCS BIOSTATISTICIANS

1. Involve your biostatistician early and communicate often!
2. Data should be stored in REDCap!
   - If your data is already in Excel, the data should be cleaned and include a data dictionary
   - Be considerate of the time we need to complete your analysis!
   - For simple analyses, we need at least 4 weeks from receiving **clean** data
3. Your biostatistician should be considered a co-investigator
   - Considered co-authors on manuscripts, usually 2nd author
4. Notify us of any co-authored submissions and acceptance, even for abstracts

Center for Children's Surgery
SCHOOL OF MEDICINE
UNIVERSITY OF COLORADO ANSCHUTZ MEDICAL CAMPUS

Research Outcomes in Children's Surgery (ROCS)
UNIVERSITY OF COLORADO ANSCHUTZ MEDICAL CAMPUS

Children's Hospital Colorado
Here, it's different.™

5

RULES OF COLLABORATION WITH ROCS BIOSTATISTICIANS

Hours of work doesn't equal weeks of work!

Request	Minimum Time Required
Power/Sample size calculation	10-20 hours
Protocol and analysis plan development	20-40 hours
Data analysis (depending on complexity)	40-120 hours +
Abstract assistance	10-40 hours

Center for Children's Surgery
SCHOOL OF MEDICINE
UNIVERSITY OF COLORADO ANSCHUTZ MEDICAL CAMPUS

Research Outcomes in Children's Surgery (ROCS)
UNIVERSITY OF COLORADO ANSCHUTZ MEDICAL CAMPUS

Children's Hospital Colorado
Here, it's different.™

6

EXCEL DATA CLEANING

Eight Tips for Creating Clean Data in Excel:

1. Create concise variable names
   - Ideally 4-20 characters long
   - Should not include special characters or spaces
2. Leave cells blank to indicate missing values
   - May also use a standard value like 9999 or NA
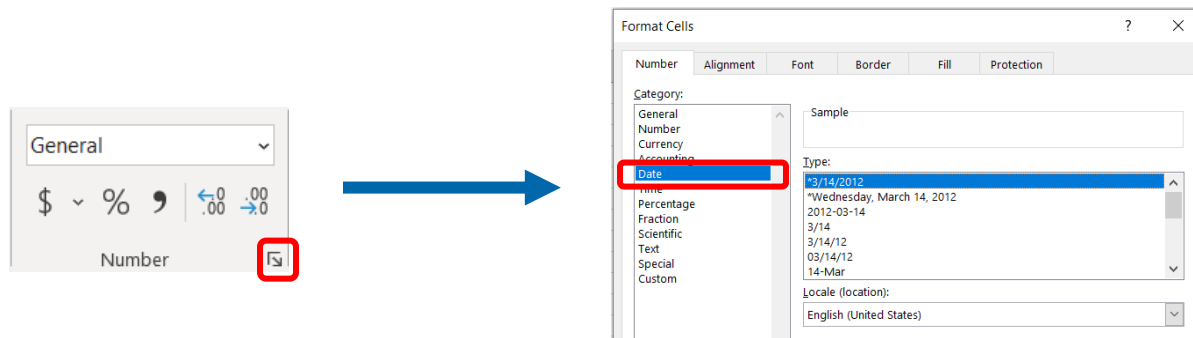3. For categorical variables, use shorthand notation to label categories
   - Use numbers (1, 2) or individual letters (M, F)
   - Be consistent in using uppercase **or** lowercase!

Center for Children's Surgery
SCHOOL OF MEDICINE
UNIVERSITY OF COLORADO **ANSCHUTZ MEDICAL CAMPUS**

Research Outcomes in
Children's Surgery (ROCS)
UNIVERSITY OF COLORADO **ANSCHUTZ MEDICAL CAMPUS**

Children's Hospital Colorado
Here, it's different.™

7

EXCEL DATA CLEANING

Eight Tips for Creating Clean Data in Excel:

4. Ensure dates use identical formatting, such as MM-DD-YYYY or MM/DD/YYYY

   - Use Excel date formatting by using Home > Number > Format Cells > Date

5. For select all that apply questions, each option should be a separate yes/no variable

Center for Children's Surgery
SCHOOL OF MEDICINE
UNIVERSITY OF COLORADO ANSCHUTZ MEDICAL CAMPUS

Research Outcomes in
Children's Surgery (ROCS)
UNIVERSITY OF COLORADO ANSCHUTZ MEDICAL CAMPUS

Children's Hospital Colorado
Here, it's different.™

EXCEL DATA CLEANING

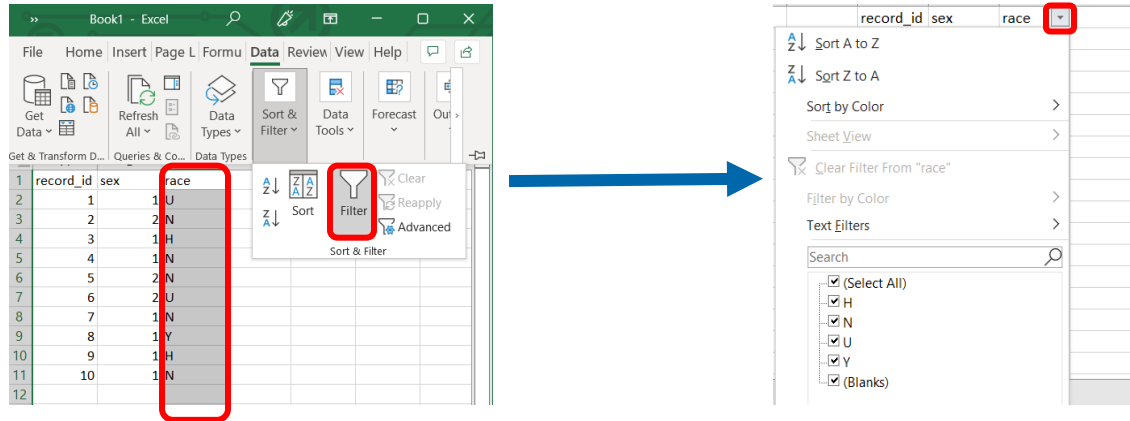*Eight Tips for Creating Clean Data in Excel:*

6. Free text should only be used for notes that won't be used in analysis!
7. Don't highlight, color-code, or hide rows/columns of your data!
   - Instead, add variables to indicate highlight or color-code
   - Unhide the rows and columns
8. Always keep an original copy of your data, even if it's messy!

Center for Children's Surgery
SCHOOL OF MEDICINE
UNIVERSITY OF COLORADO **ANSCHUTZ MEDICAL CAMPUS**

Research Outcomes in
Children's Surgery (ROCS)
UNIVERSITY OF COLORADO **ANSCHUTZ MEDICAL CAMPUS**

Children's Hospital Colorado
Here, it's different.™

EXCEL DATA CLEANING

How to identify data entry errors:

- For numeric variables, check the minimum and maximum values using =MIN() and =MAX()

- For categorical variables, select a column and use Data > Sort & Filter > Filter and click the down arrow

EXCEL DATA CLEANING

Data cleaning can take a lot of time!

It is usually an iterative process between you and your biostatistician. Ask your biostatistician for a meeting to discuss data cleaning expectations before you start!

Center for Children's Surgery
SCHOOL OF MEDICINE
UNIVERSITY OF COLORADO ANSCHUTZ MEDICAL CAMPUS

Research Outcomes in Children's Surgery (ROCS)
UNIVERSITY OF COLORADO ANSCHUTZ MEDICAL CAMPUS

Children's Hospital Colorado
Here, it's different.™

11

DEVELOPING A DATA DICTIONARY

Your data dictionary is usually a separate Excel document or sheet!

It should be up-to-date and include the same information as a REDCap codebook!

Variable Name	Variable Label	Variable Attributes
record_id	Unique record identifier	
mrn	Medical record number	
dob	Date of birth	MM-DD-YYYY
sex	Patient sex	1=Female, 2=Male
ethnicity	Ethnicity	H=Hispanic or Latino, N=Not Hispanic or Latino, U=Unknown
height	Height (cm)	
iss	Injury severity score	0-75

Center for Children's Surgery
SCHOOL OF MEDICINE
UNIVERSITY OF COLORADO ANSCHUTZ MEDICAL CAMPUS

Research Outcomes in
Children's Surgery (ROCS)
UNIVERSITY OF COLORADO ANSCHUTZ MEDICAL CAMPUS

Children's Hospital Colorado
Here, it's different.™

12

PREPARING TO DISCUSS YOUR ANALYSIS

Main types of data analysis we see in ROCS:

1. Descriptive statistics

2. Hypothesis testing

3. Correlations

4. Regression analyses

5. Survival analyses

6. Publication-ready figures

If you don't know what type of analysis you'd like, we're here to help!

Center for Children's Surgery
SCHOOL OF MEDICINE
UNIVERSITY OF COLORADO ANSCHUTZ MEDICAL CAMPUS

Research Outcomes in
Children's Surgery (ROCS)
UNIVERSITY OF COLORADO ANSCHUTZ MEDICAL CAMPUS

Children's Hospital Colorado
Here, it's different.™

13

PREPARING TO DISCUSS YOUR ANALYSIS

Things to discuss with your biostatistician:

- What is your primary aim?
    - When possible, provide your IRB protocol to your biostatistician!
- Is there similar published literature?
    - Consider looking at adult studies!
    - Helpful for identifying covariates/potential confounders and information on standard statistical tests
- What have previous researchers found?

Center for Children's Surgery
SCHOOL OF MEDICINE
UNIVERSITY OF COLORADO ANSCHUTZ MEDICAL CAMPUS

Research Outcomes in
Children's Surgery (ROCS)
UNIVERSITY OF COLORADO ANSCHUTZ MEDICAL CAMPUS

Children's Hospital Colorado
Here, it's different.™

14

PREPARING TO DISCUSS YOUR ANALYSIS

Things to discuss with your biostatistician:

- Will some patients have repeat events, admissions, or measurements?
- What is your primary outcome?
    - If your outcome is continuous, what is a clinically meaningful difference?
- What are your inclusion and exclusion criteria?

Center for Children's Surgery
SCHOOL OF MEDICINE
UNIVERSITY OF COLORADO ANSCHUTZ MEDICAL CAMPUS

Research Outcomes in
Children's Surgery (ROCS)
UNIVERSITY OF COLORADO ANSCHUTZ MEDICAL CAMPUS

Children's Hospital Colorado
Here, it's different.™

15

PREPARING TO DISCUSS YOUR ANALYSIS

Things to discuss with your biostatistician:

- Should the analysis be stratified? Should the analyses be performed separately within groups?
  - Are there known cofounders?
- Do you want to perform a sub-analysis? Do you want to repeat the analyses in a smaller group?
  - Is there a subset of your study population that may respond better to your intervention?

Center for Children's Surgery
SCHOOL OF MEDICINE
UNIVERSITY OF COLORADO **ANSCHUTZ MEDICAL CAMPUS**

Research Outcomes in
Children's Surgery (ROCS)
UNIVERSITY OF COLORADO **ANSCHUTZ MEDICAL CAMPUS**

Children's Hospital Colorado
Here, it's different.™

16

PREPARING TO DISCUSS YOUR ANALYSIS

Regardless of what you find in similar literature, understand that a specific analysis may not be feasible for your study!

Your biostatistician has the expertise to know when and how to appropriately implement analyses! We will consider:

• Sample size

• Study type

• Statistical assumptions

Center for Children's Surgery
SCHOOL OF MEDICINE
UNIVERSITY OF COLORADO ANSCHUTZ MEDICAL CAMPUS

Research Outcomes in Children's Surgery (ROCS)
UNIVERSITY OF COLORADO ANSCHUTZ MEDICAL CAMPUS

Children's Hospital Colorado
Here, it's different.™

17

DATA SHARING GUIDELINES

Data containing PHI should not be shared with ROCS biostatisticians via email!

Some alternative methods:

- OneDrive

- Microsoft Teams

- REDCap Send-It

Reach out to ARC for specific information on data sharing best practices!

THANK YOU!

Questions?

Please scan the QR code to give us feedback on Session 2: Conducting a Study, Data Collection Practices, & Data Prep for Your Biostatistician

Center for Children's Surgery
SCHOOL OF MEDICINE
UNIVERSITY OF COLORADO **ANSCHUTZ MEDICAL CAMPUS**

Research Outcomes in
Children's Surgery (ROCS)
UNIVERSITY OF COLORADO **ANSCHUTZ MEDICAL CAMPUS**

Children's Hospital Colorado
Here, it's different.™

19