Rigor and Reproducibility Spring 2020

Instructors:

Alisa Surkis, PhD, MLS
Fred LaPolla, MLS
Sheenah Mische, PhD

Location:

G101, Lapidus Library, MSB Ground Floor, 550 First Ave

Time:

Wednesdays, 9-10:30, 2/5/20-4/1/20

Class Goals

1. Students will understand and be able to address key areas that affect replicability in their work, including study design, analysis, and reporting
2. Assess rigor by being able to find and appraise existing research
3. Understand the how, what, and why of documentation and how it relates to scientific rigor and reproducibility
4. Recognize which key resources need to be authenticated and the role the NYULH cores play in ensuring reproducibility
5. Be able to explain how using code for data analysis adheres to the principles of computational reproducibility

Learning Objectives

Students will be able to...

1. Outline and describe issues of replicability at each phase of the research process.
2. Appraise an article and identify strengths and weaknesses with regards to rigor and reproducibility.
3. Perform a comprehensive literature search and summarize the relevance of effective searching for rigorous research.
4. Outline the elements of best practices in research data management
5. Outline and explain the elements of computational reproducibility and explain their value.

Calendar and Class Topics

Date	Time	Class Topic*
2/5/2020	9-10:30am	Replicability, Part 1
2/12/2020	9-10:30am	Replicability, Part 2
2/19/2020	9-10:30am	Search Mechanics
2/26/2020	9-10:30am	Research Data Management
3/4/2020	9-10:30am	Search Appraisal and Mechanics Continuation
3/11/2020	9-10:30am	Computational Reproducibility, Part 1
3/18/2020	9-10:30am	Authentication of Key Resources
3/25/2020	9-10:30am	Computational Reproducibility, Part 2
4/1/2020	9-10:30am	Final Assessment

***Order of classes subject to change (check current syllabus in Brightspace)**

Classes

1. Replicability, Part 1

Class Objectives:

- Understand the difference between replicability and reproducibility
- Be able to articulate the difficulties in assessing replicability
- Understand what is driving concerns about lack of reproducibility/replicability in science
- Understand how underpowered studies contribute to irreproducibility
- Understand the prevalence and impact of failure to include randomization and blinding in study design

Before class:

Read the following:

- Lithgow GJ, Driscoll M, Phillips P. A long journey to reproducible results. Nature. 2017 Aug 22;548(7668):387-388. https://www.nature.com/news/a-long-journey-to-reproducible-results-1.22478
- Yong E. A waste of 1,000 Research Papers. The Atlantic 2019 May 17. https://www.theatlantic.com/science/archive/2019/05/waste-1000-studies/589684/

Class Assessment:

Watch this video: https://www.youtube.com/watch?v=a4fUU85ABwc
Write a paragraph (~200 words) addressing the following:

- Have labs you've worked in generally employed randomization? Do you think randomization are generally employed in your area of study? If you have employed randomization in a study, did you consider all the factors discussed in the video?

2. Replicability, Part 2

Class Objectives:

- Understand the prevalence and impact of failure to consider sex as a biological variable
- Be able to recognize p-hacking and HARKing and how they affect replicability
- Recognize how publication bias may impact the "truth" of what can be learned from publications
- Understand the rationale for preregistration
- Be aware of guidelines and tools that facilitate reproducibility/replicability through transparent reporting
- Know the different components of the NIH Rigor & Reproducibility guidelines

Before class:

Read the following:

- Amrhein V, Greenland S, McShane B. Scientists rise up against statistical significance. Nature. 2019 Mar;567(7748):305-307. https://www.nature.com/articles/d41586-019-00857-9
- Ioannidis JPA. The Importance of Predefined Rules and Prespecified Statistical Analyses: Do Not Abandon Significance. JAMA. 2019 Apr 4. https://jamanetwork.com/journals/jama/article-abstract/2730486

Class Assessment:

Write a short essay (500 words maximum) on the use of $p < 0.05$ for statistical significance. Discuss whether you think this value should stay the same, change, or that we should abandon the use of statistical significance. Make an argument for why you think this will improve current issues around p-values.

3. Research Data Management

Class Objectives:

- Describe the key elements necessary for inclusion in a data management plan
- Apply research data management and documentation best practices to the following stages of the data lifecycle:

    - Data collection
    - Data organization
    - Data storage
    - Data preservation
    - Data sharing
- Articulate the importance of applying research data management best practices to their own research

Before class:

Read the following:

- Schiermeier Q. Data management made simple. Nature. 2018 Mar 15;555(7696):403-405. Available from: https://www.nature.com/articles/d41586-018-03071-1
- Everyone needs a data-management plan. Nature. 2018 Mar 15;555(7696):286. Available from: https://www.nature.com/articles/d41586-018-03065-z

Class Assessment:

Based on the Schiermeir article and the class slides:

- Write a short essay (500 word maximum) describing three separate measures you could implement to improve your research data management practices currently.
- Explain how and why these three measures are important within:
    - the context of your research,
    - within the biomedical research landscape more broadly.

4. Search Mechanics

Class Objectives:

- Students will be able to articulate how a comprehensive search fulfills NIH requirements on rigor
- Students will be able to construct a comprehensive search that employs:
    - Proper functioning Boolean logic
    - Appropriate syntax for PubMed and additional biomedical literature databases
- Students will be able to explain the benefits and drawbacks of using multiple databases.

Before class:

Read the following:

- Leenaars, M., Hooijmans, C. R., van Veggel, N., ter Riet, G., Leeflang, M., Hooft, L., … Ritskes-Hoitinga, M. (2012). A step-by-step guide to systematically identify all relevant animal studies. *Laboratory Animals*, *46*(1), 24–31. http://doi.org/10.1258/la.2011.011087

- Gordon, W. (2019) “6 Google Tricks That Will Turn You Into an Internet Detective” *New York Times.* https://www.nytimes.com/2019/08/21/smarter-living/6-google-tricks-that-will-turn-you-into-an-internet-detective.html?fallback=0&recId=1PmaNrnACnHIEZiBdNlo9XTArXb&locked=1&geoContinent=NA&geoRegion=PA&recAlloc=home-geo&geoCountry=US&blockId=home-living-vi&imp_id=320560633&action=click&module=Smarter%20Living&pgtype=Homepage

Class Activities:

- Overview of NIH Rigor and Reproducibility as it relates to searching.
- Explanation of Boolean Logic.
- Hands on work in PubMed including work on:
    - Creating synonyms
    - Finding entry terms (MeSH)
    - Correctly formatting Boolean operators
- Hands on work creating a search in Web of Science including work on:
    - Citation networks
- Hands on work in Google Scholar including work on:
    - Searching a specific journal

Class Assessment:

Students will be given two search scenarios, they must **choose one** and create a search that demonstrates the skills taught in class for a sensitive search. This should include:

- Proper use of Boolean Logic
- Use of entry terms if applicable
- Use of synonyms and different concepts

5. Search Appraisal and Mechanics continuation

Class Objectives:

- Students will be able to systematically appraise an article from the literature
- Students will be able to identify tools for critical appraisal of literature
- Students will be able to summarize bias as it relates to published literature

Before class:

Read the following:
Hooijmans CR, Rovers MM, de Vries RB, Leenaars M, Ritskes-Hoitinga M, Langendam MW. SYRCLE's risk of bias tool for animal studies. BMC medical research methodology. 2014;14:43. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4230647/

Krauth D, Woodruff TJ, Bero L. Instruments for assessing risk of bias and other methodological criteria of published animal studies: a systematic review. Environmental health perspectives. 2013;121(9):985-92. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3764080/

Viudez-Martínez, A, García-Gutiérrez, MS, Manzanares, J. Gender differences in the effects of cannabidiol on ethanol binge drinking in mice. *Addiction Biology*. 2019;e12765. https://doi-org.ezproxy.med.nyu.edu/10.1111/adb.12765

Chapter 5 of : Users Guide to Medical Literature https://mhebooklibrary-com.ezproxy.med.nyu.edu/doi/book/10.1036/9780071808729?contentTab=true, page 86-93

Class discusses literature from homework:

- Summarize research question
- Discuss author's affiliations, and any COI
- Discuss study design
- Discuss potential bias in the methods described
- Discuss what information is present or missing to make an article reproducible and rigorous

Class Assessment:

Students will be given a short article to read and evaluate using an assigned Critical Appraisal tool.

6. Computational Reproducibility and Intro to R and RStudio for Analysis I

Class Objectives:

- List best practices for computational reproducibility
- Describe the benefits of conducting analysis in a reproducible way
- Be able to create an R Project, read a dataset into R Studio and run a basic function on a part of it.

Before class:

Please download R https://www.r-project.org/ and R Studio to your computer or be prepared to use a classroom laptop.

Read the following:

- Wilson G, Bryan J, Cranston K, Kitzes J, Nederbragt L, Teal TK. Good enough practices in scientific computing. PLoS computational biology. 2017;13(6):e1005510. https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1005510

7. Authentication of Key Resources

Class Objectives:

- Articulate the difference between key resource and non-key resource
- Be aware of authentication plan resources for:

    Antibody validation | Cell line authentication | Purified proteins | Plasmids | Nucleic Acids | Specialty chemicals | Genetically Modified Animals or Cells

- Describe an authentication plan to identify and validate key resources essential for the experimental design
- Be aware of DART Cores as resource for multidisciplinary research expertise employing best practices for scientific rigor and responsible conduct of research

8. Computational Reproducibility and Intro to R and RStudio for Analysis II

Class Objectives:

- Explain the purpose of version control with regards to reproducibility
- Be able to connect an R Project to a GitHub repository
- Be able to create an R Markdown file and explain how and why using markdown contributes to reproducibility

Class Assessment:

Create an R Markdown file that uses the data from class and demonstrates text, code and output providing a simple analysis (such as descriptive statistics like mean or median of a variable, and a basic chart of a variable). The text should be an explanation of what you are doing in your analysis, the reason you are using the commands you are and an explanation of how using a Markdown file makes your analysis more reproducible.

9. Final Assessment

Evaluation

- Pass/Fail
- Attendance at all classes*
- Completion of all assignments by the due date (the start of the following class)*
- A grade of 80% or higher on the final assessment

*Students who either must miss class or are unable to complete an assignment on time must contact the instructors at hsl.dataservices@nyulangone.org for permission, which is at the professor's discretion.