### General Guidelines

These guidelines are meant to generally help grade the homework and avoid it being too much a matter of opinion. If in doubt of confusion about anything please ask, and in general I favor leniency on grading of the homework. I am not concerned about grammar or elegant language, if you can fundamentally understand what is written, that is fine.

### 

### Replicability I

Write a 300 word reflection on inclusion of biological variables in research, such as sex, weight, underlying conditions. Explain how these may pertain to your area of research, limitations that may make it more challenging to include these features in papers, benefits of presenting the information to others. When reading literature have you always found this information is clear, and how might that impact your understanding of the research and the reproducibility of the paper as it relates to your field? Be specific.

**Rubric:**

Note:

| 4 | 3 | 2 | 1 |
| --- | --- | --- | --- |
| Student addresses all areas of question:   1. How key biological variables relate to their research 2. Limitations that mitigate inclusion in research 3. Benefits of presenting the information for others 4. Discussion of inclusion in literature 5. The impact of understanding key biological variables in a study   Student responses display critical thinking and tie each part of the question to reproducibility. Student responses are engaged with the topic and specific to their experience/research domain. | Student addresses most areas of the question, but not all five, and demonstrates some degree of engagement with inclusion of key biological variables in their area of research. | Student addresses three of fewer areas of the question and presents a limited or perfunctory engagement with relating the question to their field and to reproducibility. Answer is unclear. | Student answers one or fewer areas of question and does not engage thoughtfully with how this topic relates to their research domain. Answer is unclear and does not demonstrate engagement with the question. |

### Replicability II

Review a publishing guideline of your choosing, for example, Cell’s Star Methods,<https://www.cell.com/pb-assets/journals/research/cell/methods/Methods_Guide_general-1678470557763.pdf>, Nature’s Checklist <https://www.nature.com/documents/Reporting_checklist_new.doc> the Arrive guidlelines

<https://arriveguidelines.org/arrive-guidelines/inclusion-and-exclusion-criteria/3c> or another of your choosing and write a 300 word reflection on a) how the guideline might help someone doing your specific research improve reproducibility and b) drawbacks or barriers to following the guideline for someone in your field. Be specific about how the guidelines relate to your area of work.

**Rubric:**

Note: As above, there is not a necessary right or wrong answer, we want to see that they are engaged in the question and that their reasoning makes sense.

| 4 | 3 | 2 | 1 |
| --- | --- | --- | --- |
| Student provides **both** benefits and drawbacks of a checklists for reproducibility. Student connects the answer to their area of research  Following Benefits/Drawbacks are suggestions, but if student writes something else that makes sense, that is good too.  Benefits might include: May highlight areas that researchers would miss or otherwise overlook; Can help organize reporting in methods; can help others quickly identify key parts of information  Drawbacks might include: May be done perfunctorily. May structure according to a standard but not actually provide required information; Information required may not be essential to a specific project; Information may be too long to include in a paper methods section’s word counts | Student provides some information on benefits and drawbacks, but information on how this relates to their field may be missing or unclear. | Student only addresses a benefit or a drawback. Information on how it relates to their field is unclear. | Student does not engage question and answer does not make sense. |

### RDM Assignment

**Part 1: Based on the Schiermeir article** ([**https://www.nature.com/articles/d41586-018-03071-1**](https://www.nature.com/articles/d41586-018-03071-1)) and the class slides:

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

Note: As above, there is not a right or wrong answer, we want to see that they are engaged in the question and that their reasoning makes sense.

**Part 2:**

Based on your experiences in Task 1 (Alien Case Study), improve the organization and documentation of the data so that future data requests of your lab can be quickly and easily met.

PROVIDE:

* File and folder naming conventions for all data types
* How you would approach the organization of files and folders
* Create a data inventory of all the data and its documentation

NOTE: YOU DO NOT NEED TO RENAME ALL FILES

| 4 | 3 | 2 | 1 |
| --- | --- | --- | --- |
| Student clearly describes three distinct research data management practices, explains how they would help with their own work, and why they are important for reproducibility more broadly.  Student provides tangible ways to improve the Alien dataset, including a file and folder naming convention in ways that are consistent with what was taught in class (e.g., avoids acronyms, special characters, spaces, etc), (2) and provides an approach to organizing files (e.g. hierarchy or file structure). | Student provides clear explanation but misses an element, e.g. only describes two practices, or describes importance for work but not the broader importance  Student provides ways to improve the Alien dataset, but misses an element (e.g., only provides a file naming convention, but not an approach to organization) | Student demonstrates some understanding but either only provides one example or fails to clearly explain the importance of these practices.  Student demonstrates some understanding, but does not provide all components. | Student response demonstrates a lack of understanding of what constitutes a research data management practice, or why would would engage in such practices.  Student response demonstrates a lack of understanding of data organization. |

### Search I

You are interested in the role of BRCA1 in cancer among smokers. Build a thorough search to avoid missing relevant abstracts in PubMed. This should include the use of synonyms, boolean operators like AND, OR and NOT; parentheses to group like terms. Provide the string you searched in PubMed and the number of results. Your search and results must work (ie not merely be an AI hallucination) to get credit.

Also answer the following:

What are places you could look for useful terms?

If you were getting too many results, what are some ways you could narrow down the search?

Answers to stand alone question:

Useful Terms: professional knowledge, thesaurus, MeSH Terms, in the abstract of a useful paper.

Narrow down: Exclude terms that are too broad in definition; add criteria such as population information (age, specimen type), condition being examined, etc.

Search Rubric:

| 4 | 3 | 2 | 1 |
| --- | --- | --- | --- |
| Search includes multiple synonyms to increase the likelihood of abstracts. These synonyms are grouped in a parenthesis with OR between them. Sets of parentheses can include “AND” between them but this is optional. Search terms are grouped by concept (like smoking terms vs. cancer terms vs brca1 terms).  Ex.:  (brca1 OR IRIS OR PSCP OR BRCAI OR BRCC1 OR FANCS OR PNCA4 OR RNF53 OR BROVCA1 OR PPP1R53)  (cancer OR neoplasms)  (tobacco OR nicotine OR vaping OR smoking)  Note in above search: synonyms for BRCA1 would be considered excellent and not totally necessary | Search has some synonyms but some terms seem irrelevant. Search generally looks correct but very high (e.g. over 500) or very low (e.g. under 20) number of results returned. | Search attempts to use synonyms but they are not grouped in parentheses, which causes errors.  Alternately, Groupings are not according to different concepts, for example the search may be:  BRCA1 OR (neoplasms OR cancer) OR (smoking OR nicotine OR vaping OR cigarettes) (mutations OR gene replication)  Here the problem is mutations/replication issues should likely be grouped with cancer if included.  Finally, use of OR is incorrect between synonyms (for example they do not write “OR”). | Search does not include terms for BRCA1, cancer and smoking. |

### Appraisal

**I**dentify an article of interest to you and provide a link in your submission. Using what we learned in class, briefly summarize two strength and two weakness with regards to reproducibility in bias, and refer to checklists relevant to your article’s domain.

Suggested articles:

Defining the proteomic landscape of cultured macrophages and their polarization continuum<https://onlinelibrary-wiley-com.ezproxy.med.nyu.edu/doi/10.1111/imcb.12687>

[A Patient-Derived Glioblastoma Organoid Model and Biobank Recapitulates Inter- and Intra-tumoral Heterogeneity - PMC (nyu.edu)](https://www-ncbi-nlm-nih-gov.ezproxy.med.nyu.edu/pmc/articles/PMC7556703/)

Next

Write a reflection on why you believe researchers may choose to include or exclude the information that this paper provided. How could researchers set themselves up for success in reporting relevant information in their papers and what sorts of barriers make it a challenge?

Rubric

| 4 | 3 | 2 | 1 |
| --- | --- | --- | --- |
| Student identifies two strengths and two weaknesses for reproducibility in a paper and describes them clearly. Student explains clearly why researchers may make similar decisions assessing both reproducibility and pragmatic limitations. Student engages with question of how barriers may be overcome | Student names less than two strengths and less than two weaknesses for reproducibility. Student does not engage with why they believe researchers may make these decisions and ways barriers may be overcome are muddled or unclear. | Student names one strength and one weakness but does not engage with why researchers may include or exclude this information. | Student’s response does not address the question, just provides circular language without addressing issues of reproducibility in paper. |

### Computational Reproducibility

Set up a free GitHub account and create your first repository. Go to Google Colab and create a simple analysis file. In that analysis file, print today’s date3nd write a brief reflection (no more than 300 words) on how using repositories, tracking changes and publishing like GitHub can help with computational reproducibility. Post this file to your GitHub repository and submit the link to the repository with your homework in it.

| 4 | 3 | 2 | 1 |
| --- | --- | --- | --- |
| A four means the student successfully: 1) uploads the file to github  The provide a clear engaged explanation of how repositories, track changes and posting code can help with reproducibility  They perform the basic arithmetic successfully | Student uploads successfully and includes code but is unclear on how repositories, posting code and track changes impact reproducibility. | Student provides an unclear explanation and does not address all of the areas of computational reproducibility in their description | Student’s response does not address the prompt and includes only vague circular language. |

### Authentication Module

| 4 | 3 | 2 | 1 |
| --- | --- | --- | --- |
| Student describes specific key variables (chemical, antibody, DNA, or cell line)  Student describes what will be done to insure quality (e.g. monitoring for issues, checking with online resources)  Student describes source of key variable | Student describes what variable they are working with but is unclear as to what will be done with regards to one of: quality monitoring or source. | Student describes key variable and attempts to describe authentication but is unclear as to actual monitoring or source. | Student only mentions a variable, but does not discuss an authentication plan. |

### 

### 

### Final Assessment

1. Compare and contrast computational reproducibility and experimental replicability.
2. Choose one issue underlying replicability and explain what it is, and how it impacts replicability.
3. Describe one aspect of research data management and how it is important for reproducibility/replicability?
4. Perform a comprehensive literature search in PubMed that combines synonyms. Provide the exact search you put into the system and the number of results (note: if you get very few hits (e.g. under 5) or very many hits (e.g. over 1000) there is probably an error.
5. Read the methods section of an article and identify one strength and one weakness related to the reproducibility or replicability of the research, and explain why they impact reproducibility/replicability.
6. Describe one best practice of computational reproducibility and how the tool learned in class could be used to achieve it.

| Question number | 4 | 3 | 2 | 1 |
| --- | --- | --- | --- | --- |
| 1 | 2 points - for correct definition of Comp repro  2 points - correct definition of experimental replicability |  |  |  |
| 2 | 2 points for defining a concept  2 points for explaining how it impacts replicability |  |  |  |
| 3 | 2 points for defining a concept  2 points for explaining how it impacts replicability |  |  |  |
| 4 | Boolean logic is sound (AND OR and use of parentheses is correct), and sufficient keywords related to the topic are searched [correct Boolean is 3 points, correct synonyms is 1 point] |  |  |  |
| 5 | 2 points for explaining a strength and impact  2 points for a weakness and impact |  |  |  |
| 6 | 2 points for naming a comp repro issue.  2 points for explaining how R meets it. |  |  |  |