Phase III Overview

Phase III Quick Outline of Work

- **Present progress and respond to Phase II feedback**
  - Read and adjust project proposal based on reviewer feedback
  - Create high level overview presentation of project analysis plan and progress
  - Identify project adjustments and concerns for discussion
- **Transform and ready all data for analysis**
  - Conduct data transformations: handling missing data and outliers, aggregating and normalizing values, creating derived variables, augmenting with other sources or annotations, etc.
- **Proof of Concept check that steps are runnable**
  - Setup and run the environment, packages, datasets, and functions for your major analysis steps to demonstrate that you can finish your project by completing at least an single example run for each step (do not yet need to have all different comparison or combinations of parameters run)
- **Submit and review Phase III reports**

Phase III Submissions and Deadlines [**year long overview (https://canvas.illinois.edu/courses/57094/pages/assignments-overview)**]

Deliverable	Type	Due Date	Pts	Submit	Evaluators
Phase III: Initial Analysis Proof of Concept					
Phase III Progress Presentation / Discussion	18 min present & discuss	Jul 15 - Aug 6	5	[**Zoom** ⬏ (https://illinois.zoom.us/j/81984434295?pwd=WjF6YkpST0pZUkZTRVdSZjE4OVN1Zz09)]	CD

Phase III Report	around 5 pg team report	Fri, Aug 15	20	[link]	SP, ME, CD
Phase III Peer Evaluations	3 report evals	Wed, Aug 27	9	[link]	CD

Progress Presentation Sessions

- **Phase Objectives**
  - PP-1. Present your project importance and progress quickly to a generalized audience
  - PP-2. Prepare responses to written reviews and respond to feedback and questions from a live audience
  - PP-3. Identify project risks and weaknesses and solicit assistance from other scientists/researchers
  - PP-4. Participate in and provide valuable contributions to data-related scientific discussions science
- **Session Format.** Each one-hour Zoom session will focus on three projects with twenty minutes allocated to each project. Course staff and all students and mentors from the three teams will be encouraged to attend, share their video, and engage in the discussion throughout. With a team's twenty minutes, the following is expected to occur:
  - Team presentation [~ 12 minutes]
    - [~4 min] A quick overview of the project importance, goals, and progress through previous phases
    - [~3 min] Responses based on the most valuable feedback from the previous phases, either answers to important questions asked or significant alterations to project plan based on reviewer comments
    - [~4 min] Progress made toward the objectives for the current phase, along with any challenges encountered.
    - [~1 min] Possible discussion questions for the faculty audience or peers about project struggles or unknowns
  - Group Discussion [~ 7 mins]
    - Audience questions and feedback and discussion of team concerns
- The **5pts** of this session will be based on student participation in the session throughout the hour.

Data Transformations and Proof of Concept Analysis Runs

- **Phase Objectives**
  - P3-1. Complete any data pre-processing steps required before analysis, which could include dataset harmonization, feature generation, missing data imputation, outlier removal, etc.
  - P3-2. Prove the feasibility of the entire analysis plan by demonstrating the core analysis step with the proposed software and packages on a selected subset of the primary data.
  - P3-3. Identify and communicate the important metrics of the proof of concept analysis to convey to a scientific audience.
- **Overview**. The high level goal of Phase III is to demonstrate that you can run the analysis code you stated you would use in your proposal on the data you have selected to analyze. You don't need to complete all runs on all data, but we want to see evidence that you performed this "proof of concept" run of the core analysis. You can demonstrate this milestone by adding key code snippets as supplemental materials and reporting on direct observations from the outputs (model evaluation, quality, coefficients, significances, etc.).
- **Evaluation** of the report for its descriptions of the proof of concept analysis will be based on if it:
  - Provides descriptions of any data transformation and augmentation steps
  - Clearly justifies and describes specifics of methods used for primary analysis
  - Demonstrates that the core analysis steps have been run on a selected dataset
  - Contains detail (supplemental code snippet) about analysis implementation
  - Provides tables or figures that conveys initial analysis results

Initial Clinical Interpretation of Results

- **Phase Objectives**

  - P3-4. Connect the metrics and values from the data analysis to the clinical question being studied
- **Overview**. In Phase III, we ask you to show how the first pass results can be summarized and related to the clinical research question.
- **Evaluation** of the report for its clinical interpretation will be based on if it:
  - Contains initial summarization and interpretation of analysis results
  - Provides clear connection of potential clinical value of data analysis
  - Discusses modifications of proposed analysis work plan to strengthen feasibility and impact

Updated and Revised Phase III Report

- **Phase Objectives**

  - P3-5. Demonstrate initial progress on data analysis to convince reviewers that proposal can be completed with the available resources.
- **Phase III Report**: At the end of Phase III, all teams are expected to revise and update their previous submissions into a more complete report. By the end of Phase III, the report should contain at least the following sections:
  - **Project Title & Authors** (*Revised*)

- **Proposal Abstract** (*Revised*)
- **Introduction** -
  - Literature review of **clinical relevance of** issue (*Revised*)
  - **Background** about primary dataset(s) (*Revised*)
  - Literature review of **analysis methods and benchmarks** (*Revised*)
- **Methods**
  - Methods for downloading, extracting, and pre-processing primary dataset (*Revised*)
  - Description and **figure** of data selection inclusion or exclusion steps (*Revised*)
  - Methods for transforming and augmenting data (***New***)
  - Packages and functions (and possible code examples in supplementary text) for running analysis steps (***New***)
- **Results**
  - Initial description and **table** summary statistics for key sample populations (cohorts) and important variables related to the proposed data analysis (*Revised*)
  - Example results from running single instances of analysis steps (***New***)
- **Proposed Analysis Plan** (*Revised*)
  - Step-by-step description of multi-phase approach to answer clinical question with data analysis of selected data set
    - Indicating choice of analysis methods and evaluation criteria for each analysis step
  - **Figure** diagram to summarize that high-level analysis approach
  - Identification of major revisions to proposed analysis since the previous phase (***New***)
- **References** - cited throughout and listed at the end (*Revised*) (**IMPORTANT**: for each reference, please also provide short links that take readers to a corresponding webpage)
- The length of this report is expected to be equivalent to around 5 pages of 12 pt single-spaced text, not counting any figures, tables, or bibliography sections. As always, it is expected to conform to a medical journal or technical report style and be submitted as a pdf following the team number and shortened title naming convention, e.g. "Team 01 - Analysis of Breast Cancer Readmission.pdf". Additional material can be included as a single (or zipped directory) supplementary file.
- **Report Evaluation**: Submissions will be evaluated by faculty and peers on the following:
  - Clinical rationale of analysis plan is clear and understandable to a general audience
  - Updates previous phase content (abstract, literature review, dataset description, pro-processing methods, etc)
  - Organizes report with clear flow, integrating materials across phases
  - Figures and tables are numbered, cited, and interpretable by reader
  - Supplementary content is cleanly cited and organized
  - Follows professional journal style with correct tone, spelling, and grammar

Critical Evaluation of Peer Reports

- **Phase Objectives**

- E-1. Read, understand, and think critically about data analysis reports
  - E-2. Corroborate and assess the soundness of proposed and reported research in domains outside your expertise by finding and comparing to external literature sources
  - E-3: Provide meaningful and professional peer review feedback that resembles a medical journal review process
- **Phase Peer Evaluations**: We will assign every student to review three submitted reports each phase and provide valuable feedback to their peers. The purpose of this exercise is to give reviewers exposure to the efforts and outputs of other teams and exercise the ability to read and think critically about analyses in other domains presented to them and practice communicating their questions or suggestions. For the teams reviewed, this provides additional outside perspectives on the presentation and direction of their project that they have the chance to consider and respond to. We expect peer reviews to contain **Meaningful Feedback**, defined as
  - advice for fixing content errors (not grammatical errors) in the presentation, organizing the information in different ways to make it easier for the audience to follow, or suggestions for alternative methodology, research questions, or interpretation of findings which may constitute a future improvement to the work.
- Some resources for how to perform and write a meaningful review can be found at the paper, "**How to Review a Clinical Research Paper (https://www.ahajournals.org/doi/full/10.1161/STROKEAHA.118.021286)** " or the **JEE reviewer guidelines (https://cimed-dsp.github.io/files/JEE_reviewer_guidelines.pdf)** .