

# Phase I Overview

## Phase I Quick Work Outline

- **Build your team:**
  - Find your teammate
  - Complete the initial student and team [survey](https://canvas.illinois.edu/courses/68869/assignments/1593327) (required for each student)
  - Approach a faculty project mentor and have them complete mentor [agreement](https://forms.gle/YV86zGjgSsab8xDB7)
- **Explore project topics and datasets**
  - Explore possible datasets and project ideas ([mentors](https://canvas.illinois.edu/courses/68869/pages/data-science-project-mentors)), previous [years](https://canvas.illinois.edu/courses/68869/pages/previous-class-project-examples), other [collections](https://canvas.illinois.edu/courses/68869/pages/medical-data-sets)
  - Identify three datasets suitable for your project. These must be immediately available or with only 1-2 weeks delay.
  - Perform literature reviews related to dataset clinical topic
- **Write and Review Phase I Reports**
  - [\[Submit\]](https://canvas.illinois.edu/courses/68869/assignments/1593326) a one page report that summarizes findings about literature review and datasets
  - Perform peer [\[reviews\]](https://canvas.illinois.edu/courses/68869/assignments/1593325) on three Phase I Reports that will be assigned to you

## Timeline and Key Dates [\[year long overview](https://canvas.illinois.edu/courses/68869/pages/assignments-overview)

[\[https://canvas.illinois.edu/courses/68869/pages/assignments-overview\]](https://canvas.illinois.edu/courses/68869/pages/assignments-overview)

Table with timelines and key dates

| Deliverable   | Type                    | Due Date    | Pts | Submit   |
|---|-------------------------|-------------|-----|--|
| <b>Phase I: Team Formation, Possible Data Sets, Literature &amp; Methods Review</b> <a href="https://canvas.illinois.edu/courses/68869/pages/phase-i-overview">[link to Phase I</a><br><a href="https://canvas.illinois.edu/courses/68869/pages/phase-i-overview">]</a> |                         |             |     |  |
| <a href="https://canvas.illinois.edu/courses/68869/assignments/1593327">Student Survey w/ Team/Topic Identification Survey</a>  | individual survey       | Fri, Mar 28 | *   | <a href="https://forms.gle/Hh22EcF2erk8NgCd6">[link to student survey</a><br><a href="https://forms.gle/Hh22EcF2erk8NgCd6">]</a>           |
| <a href="https://canvas.illinois.edu/courses/68869/assignments/1593326">Phase I Report</a>  | around 1 pg team report | Fri, Apr 18 | 10  | <a href="https://forms.gle/qqJDJHP6u6nexT8J9">[link to Phase I report</a><br><a href="https://forms.gle/qqJDJHP6u6nexT8J9">]</a>           |
| <a href="https://canvas.illinois.edu/courses/68869/assignments/1593325">Phase I Peer Evaluations</a>  | 3 report evals          | Wed, Apr 30 | 6   | <a href="https://forms.gle/Sc9vuR7s5qiPJrnB9">[link to Phase I peer evaluations</a><br><a href="https://forms.gle/Sc9vuR7s5qiPJrnB9">]</a> |
| <a href="https://canvas.illinois.edu/courses/68869/assignments/1593324">Mentor Confirmation Form</a>  | agreement per team      | Fri, May 16 | *   | <a href="https://forms.gle/YV86zGjgSsab8xDB7">[link to mentor confirmation form</a><br><a href="https://forms.gle/YV86zGjgSsab8xDB7">]</a> |

Evaluator Key: **SP** = Student Peer; **ME** = Mentor; **CD** = Course Directors, **GP** = Guest Panel

# Teams, Topics, Self Assessment

- **Phase Objectives**
  - P1-1. Perform a self-assessment to identify experiences and skills you are missing related to data analysis
  - P1-2. Identify a peer partner with a similar clinical interest and complementary skills and experiences related to data analysis
  - P1-3. Identify a clinical issue that would benefit from a novel data analysis of related accessible datasets
  - P1-4. Connect with faculty mentor who brings expertise to analysis or clinical domain
- **Teams of 2:** For the DSP course, we want students to form teams of two that will work together throughout the year. This offers students the opportunity to find a partner with complementary skills and also exercise their ability to communicate and coordinate in a team setting. For students who wish to form groups of three, we request that you submit a paragraph to the course directors justifying why or how your project will benefit from having three medical student members. Example justifications could include:
  - the members share a specific specialty of interest
  - the members have a useful balance of skills that together will make the project stronger
  - the members have ideas for a more in-depth analysis that would require additional effort
- **Clinical Topic for Project:** Student teams are welcome to select a clinical topic of their choosing to conduct their data analysis project. The only requirements of the topic is that 1) the students must find a rich, related dataset that will be **fully accessible to them by Phase II** and 2) the analysis of this dataset will be able to be related to current understandings of clinical knowledge and practices. To gather ideas about possible topics and datasets, we encourage the students to look at
  - Faculty volunteer [mentor interests and datasets \(https://canvas.illinois.edu/courses/68869/pages/data-science-project-mentors\)](https://canvas.illinois.edu/courses/68869/pages/data-science-project-mentors)
  - [Previous class projects \(https://canvas.illinois.edu/courses/68869/pages/previous-class-project-examples\)](https://canvas.illinois.edu/courses/68869/pages/previous-class-project-examples) topics
  - Course [Template Projects \(https://canvas.illinois.edu/courses/68869/pages/templates-projects-overview\)](https://canvas.illinois.edu/courses/68869/pages/templates-projects-overview)
  - Collections of medical [datasets \(https://canvas.illinois.edu/courses/68869/pages/medical-data-sets\)](https://canvas.illinois.edu/courses/68869/pages/medical-data-sets)
- **Individual Survey Submission:** This first deliverable is due in two week and requires:
  - Name and netid of team partner, as well, as likely clinical topic
  - Self assessment of data analysis skills as well as team plan for division of labor
  - Updates on progress of finding a mentor
  - Likely availability timeslots for later phase progress presentations

## Literature Review

- **Phase Objectives**
  - P1-5. Summarize necessary background needed to understand clinical issue and cite evidence for the issue's importance
  - P1-6. Synthesize what is currently understood about clinical issue from medical literature
  - P1-7. Identify what types of data analysis methods are commonly used to explore the clinical issue
- **Literature Review Summary:** For phase I, you want to search, sift, and summarize literature related to your clinical topic in order to build an introduction to your project report. This written literature summary should accomplish the following tasks:
  - Provide literature citations for this evidence of the clinical issue significance
  - Provide literature citations to reviews and papers establishing current understanding
  - Provide literature citations of analysis examples or overviews
  - Clearly define the clinical topic to be addressed by data analysis
  - Provide necessary background, important vocabulary, and their definitions to a non-specialist audience
  - Quantify the importance/significance of the clinical issue in human, financial, time, or other costs and impacts on society or medical practice

- Provide an overview of the current state of understanding of the research related to the clinical issue. Identify areas specific to your project interests where the knowledge in literature has conflicting views, ambiguity, or regions that would benefit from further investigation
- What types of analysis methods have been applied to the datasets that you have selected for assessment? What are the general ideas and assumptions these methods rely on?

## Three Dataset Assessment

- **Phase Objectives**

- P1-8. Assess accessible datasets for their potential benefits and shortcomings for extracting insights related to clinical issue
- **Dataset Assessment:** In this phase we expect students to report on at least three possible datasets that would be potentially useful for the data analysis they intend to propose. These datasets must be available to the students by Phase II of the course, and must contain a sizable amount of data for analysis (at least hundreds of samples and dozens of relevant features). The three different datasets can each be for a separate, slightly different primary analysis, or can be alternatives or complementary/augmentary datasets that would support and enrich the intended primary analysis. For this Phase, students do not need to actually gain access and download the datasets, but should be able to answer the following questions about each one:
  - Which groups were responsible for collecting and/or releasing the data?
  - When, how, and from what populations was the data collected?
  - How large is the dataset? Number of samples/records available? Number and types of features available?
  - How reputable/robust is the collection of data considered by the community?
  - Why is the dataset especially well-suited for your team's prospective data analysis?
  - What are some drawbacks or shortcomings of using this data set for the prospective data analysis?

## Report Format

- **Phase Objectives**

- P1-9. Produce a concise, clear, and well documented written report of the literature and datasets findings related to the clinical issue
- **Phase 1 Report:** At the end of Phase I, all teams are expected to turn in a written report. This report will be iteratively refined over the phases of the course, eventually taking the form of a medical journal or technical report submission. For Phase I, the report should contain at least the following four sections:
  - Project Title & Authors
  - Introduction - containing background and literature findings
  - Datasets - describing the three possible datasets the project could use
  - References - cited throughout and listed at the end
- This length of this report is expected to be equivalent to around 1 page of 12 pt single-spaced text, not counting any figures, tables, or bibliography sections. As stated above, it is expected to generally conform to a medical journal or technical report style. A useful collection of science/medical journal templates are provided at [nextgenediting.com](https://www.nextgenediting.com/free-author-templates-science-medical-journals/) (https://www.nextgenediting.com/free-author-templates-science-medical-journals/). Additionally, for tips on style, consider reading this [guide for writing technical papers](https://homes.cs.washington.edu/~mernst/advice/write-technical-paper.html) (https://homes.cs.washington.edu/~mernst/advice/write-technical-paper.html) by Michael Ernst. Finally, we request you submit the report as a pdf with your team number and a shortened title, e.g. "Team 01 - Analysis of Breast Cancer Readmission.pdf". Report submissions will be evaluated by faculty and peers on the following:
  - Organization and clarity and flow of thought
  - Correct usage of grammar, punctuation, and spelling
  - Adhering to professional journal/report formatting and style

## Critical Evaluation

- **Phase Objectives**
  - P1-10. Read, understand, and think critically about data analysis reports
  - P1-11. Corroborate and assess the soundness of proposed and reported research in domains outside your expertise by finding and comparing to external literature sources
  - P1-12: Provide meaningful and professional peer review feedback that resembles a medical journal review process
- **Phase 1 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).