QBS 131/Biol 073: Foundations of Epidemiology II
Study Design Exercises and Study Design Presentation

Purpose: The purpose of this quarter long project is to help you engage with the theory behind study design. The *Study Design Exercises* will help you to organize your ideas and to receive timely feedback on your progress throughout the quarter. Satisfactory completion and response to feedback provided by your instructor, TA, and peers will be factored into the assignments portion of your grade. The *final study design presentation* will be graded out of 120 points total and account for 15% of your final grade in the course.

A few important things to note:

- At the start of the quarter, read ***all*** of the study design assignments, including the final presentation assignment in order to ensure that you understand how the assignments will build upon one another. You may make changes to your study question over the course of the quarter in consultation with the instructor, but the process will be most efficient if you are thoughtful in your initial choice of study question and are able to follow it through to the end of the quarter.
- Students are encouraged to use office hours to solicit feedback or additional assistance from the instructor in designing their individual studies for the final project.
- Feedback provided on study design assignments should be incorporated into your final presentation. Points will be deducted from your presentation score if you are not responsive to comments on your exercises.
- You are being given a substantial amount of leeway in designing your study.
    - An exhaustive literature review is not required, though some exploration of the relevant literature is expected. Students are strongly encouraged to pick a topic that aligns with their own research interests in order to synergize substantive knowledge gains related to the course project with their rotations, thesis, dissertation, or other projects.
    - Students need not consider budget for the study design project. In reality, budget is a critical consideration when proposing new research. For purposes of the project, we will focus on exploring the epidemiologic theory underlying study design. Students will therefore not be penalized for designing an exorbitantly expensive study, though the exercise may prove more useful if you stick to a fiscally feasible study.
    - You have some flexibility to use your imagination in terms of the exposure, outcome, and overall context for the study that you design. For example, some projects in the past have focused on designing a study to investigate a vaccine that does not currently exist (e.g., vaccine for HIV) or placed the study in the future (e.g., a retrospective cohort study set 15 years in the future to assess the influence of e-cigarettes on long term cardiovascular health). Please clearly state your assumptions/inventions in your early exercises as you design your study, and remember to acknowledge these things when you present at the end of the quarter.

Study Design Assignment #1: Matching your scientific question to the best study design and preventing information bias

Provide a concise but complete response (At least 2, but not more than 4 pages) to the following questions.

1. Describe 3 scientific questions/hypotheses that you would be interested in investigating through an epidemiologic study. One or two sentences for each question is sufficient.
    a.
    b.
    c.
2. Choose one of the scientific questions that you have proposed above and answer the following questions. (Though you may change your study question later)
    a. What is your conceptual exposure? Is this exposure rare or common?
    b. What is your conceptual outcome? Is this outcome rare or common?
    c. Briefly describe how you might use each of the four major study designs in epidemiology (cohort, case-control, cross-sectional (or ecologic if you like), or randomized trial) to assess this question. For purposes of this exercise, I'd like you to stretch your ideas about study design, so do your best to come up with a way to use every one of the study designs to address your question of interest. Feel free to be a bit creative for this part of the question (you will assess feasibility and logistics in the next part of the question).

3. Focus in on what is pragmatic or logistically possible to answer the following questions about your scientific question and study design.
    a. Which study design from part 2c seems most feasible? Why does this design seem best for addressing your scientific question?
    b. What will you use as your operational exposure and outcome? Or what are some reasonable options for operational exposure and outcome? Note that this should match up to the study design you've identified as most feasible in 3a.
    c. What other data will you need to collect for you study (i.e., what are the important covariates for your study)?

4. Place your chosen question into the broader context of the existing literature.
    a. Identify 2-4 relevant papers from the primary literature to provide background and motivation for your proposed study. Provide the citations and a 1-2 sentence summary of the critical background information contained in each study.
    b. What knowledge gap does your proposed study address? (i.e., Will it add to our scientific knowledge by answering a completely new question? Will it help us understand a new mechanism to explain a previously observed association? Will it extend the research to a new population?)

5. Are there any relevant sources of information bias to consider for your study as designed (consider all potential types of information bias)? How might you prevent these or improve exposure/outcome/covariate data collection to minimize these concerns?

Study Design Assignment #2: Creating a strategy for the control of confounding

Provide a concise but complete response (1-2 pages) to the following questions.

1. Re-state your study question and chosen study design. (If you've decided to change your study question or design, please make a note of this in your write-up)

2. Based on the papers that you reviewed for Questions 3 & 4 in Exercise 1, list the important potential confounders of your exposure-outcome association. Will any of these be particularly challenging to measure?

3. How might you integrate prevention or control of confounding into your study design or analysis?

4. Based on your answers in part 3 and 4 of Study Design Assignment #1, create a preliminary DAG to define set {S} to describe which confounders you may wish to include in a multivariable model. You may use Daggity, R, Powerpoint, etc., to make your DAG. A clear and legibly hand drawn DAG is also acceptable. (Note: If you chose a randomized trial, please use the DAG to help you describe the confounding structure that will be accounted for via randomization).

    a. Describe/define each individual component in the DAG.
    b. Was it difficult to assess directionality of any of the arrows? What additional information would you like to have?

    (Note: The goal is not necessarily to produce a perfect and final DAG, but to help you gain an appreciation of the importance of this step and some of the challenges in DAG creation. You need not do a full literature review to support your DAG, but do use information from the papers identified in part 4 of assignment 1 to help you identify potential confounders and inform your DAG.)

Study Design Assignment #3: Preventing selection bias, identifying a source population, estimating sample size, and understanding participant burden

Provide a concise but complete response (2-3 pages) to the following questions.

1. Re-state your study question and chosen study design. (This is for the benefit of your peer reviewers.)

2. Precisely how might selection bias occur in the type of study you are designing (e.g., what causes selection bias in a case-control study)?

3. Based on your study design and chosen source population, describe any relevant considerations for preventing selection bias in your study (i.e., how might selection bias occur specifically in the study that you have designed?).

4. What would be the ideal source population or data source for this study? Why?

5. Identify and briefly describe 2-3 real data sources (e.g., NHANES, Danish National Registry, etc.) or populations (e.g., patients at DHMC, workers in a specific industry, etc.) that you might target for this study. Which of these seems most useful for your particular study?

6. Based on associations observed in prior studies, what is your sense of how large your study will need to be (100, 1000, 10,000, 100,000)? Perform a power or sample size calculation for you study. A helpful module on this topic is available at: http://sphweb.bumc.bu.edu/otlt/mph-modules/bs/bs704_power/BS704_Power_print.html

7. Review Lingler et al. 2015 and the related Perceived Research Burden Assessment tool for a description of direct (e.g.,) and indirect (e.g., inconvenience) burdens potentially experienced by research participants : https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4487419/.Consider both direct risks and indirect burdens, and describe what will be the participant burden for subjects in your study?

8. Are there any ethical concerns to consider in working with your population of interest? (Hint: if your study involves human subjects (or their data), then the answer is yes. If you have not completed training in the protection of human subjects, you may wish to briefly familiarize yourself with the NIH policies related to the protection of human subjects: https://grants.nih.gov/policy/humansubjects.htm.) Does your research use vulnerable populations or others requiring special protections (this specifically relates to the following NIH policy: https://grants.nih.gov/policy/humansubjects/policies-and-regulations/vulnerable-populations.htm)?

Study Design Presentation

1. All slides (regardless of your presentation date) must be submitted on Canvas by 2:00pm on March 1, 2022. Use the provided power point template to create your slides. While you are welcome to make your slides graphically pleasing, you are being graded on content, not style. Please name your file *LastNameFirstName_SD*. If you are preparing your presentation using Keynote on a Macbook, please submit a PDF version of your slides.
2. During your presentation, the instructor, TA, and two of your peers will use a simplified version of the final grading rubric for the presentation (on the next page) to provide feedback on your presentation. Peer reviews will be masked. A guest evaluator will also be invited to attend the presentations. Be aware that this individual will not have seen your study design exercises and will provide a fresh perspective on your project. Prof. Romano will consider all grader feedback and create a final evaluation of your presentation based on the rubric on the next page.
3. Time is typically very tight for presentations and questions and answers, so as a courtesy to your fellow students, please practice your presentation to ensure that you will stay within your allotted time. The exact amount of time will depend on the number of students enrolled in the course, and Prof. Romano will let you know around the midterm what the time allotment will be.

QBS 131/Biol 073: Foundations of Epidemiology II
Study Design Project Presentation (120 points total)

Presenter Name:_____

	SCORING CRITERIA	**SCORE**					
CONTENT	*(capture specific concerns or notable omissions in comments)*	Poor	Fair	Good	Very Good	Excellent	Points
Study motivation	Rationale for study/explanation of knowledge gap						5
Specific aims	Scientific question & study objectives are clear						5
Study overview	Operational exposure matches scientific question and captures the conceptual exposure (for RCT is placebo appropriate?)						5
	Operational outcome matches scientific question and captures the conceptual outcome						5
	Study design is clearly and completely defined						5
	Chosen study design matches study question						5
Population	Source population is clearly defined						5
	Subject selection is appropriate overall						5
	Cohort study has reasonable exposed:unexposed ratio, case-control has reasonable case:control, RCT has reasonable number of arms						5
	Sample size is scientifically sound						5
Confounding	Plan to collect information about covariates/baseline characteristics is clear						5
	Design measures to control for confounding are appropriate (for RCT – is randomization appropriate)						5
	Analytic measures to control for confounding are appropriate						5
	Assessment of potential sources of residual confounding is reasonable						5
Statistical methods	Approach matches scientific question and is rigorous to answer that question						5
Bias	Selection bias has been reasonably minimized and remaining limitations stated						5
	Information bias has been minimized and remaining limitations stated						5
Generalizability	Assessment of external validity is accurate						5
ORGANIZATION							
Format	Followed presentation template						5
Time	Stayed within allotted time (filled the time and did not run over)						5
Q&A	Responses to questions are clear, thoughtful, and reflect mastery of course concepts.						10
FEEDBACK	Degree to which feedback given over the quarter was integrated into final project						10
COMMENTS:					**Total**		**/120**