Biostatistics Comprehensive Exam Part III

Identifying heterogeneous treatment effects for online single-session interventions for adolescent depression

Dr. Clement Ma

Due: August 2, 2024 at 5:00 pm

Background

The COVID-19 pandemic significantly impacted the mental well-being of youth around the world. Youth experiencing depression require accessible care. The COVID-19 pandemic severely limited the ability of youth to access adequate treatment for depression. Schleider et al (2022) conducted a 3-arm randomized controlled trial to examine the efficacy of three online single-session interventions (SSIs) in reducing depression symptoms. A total of N=2,452 youth aged 13-16 from the U.S. were randomized to one of three interventions: (1) **Project ABC**, a behavioural activation SSI; (2) **Project Personality**, an SSI teaching that traits are malleable; and (3) **Placebo control.** The primary outcome was the short-form Children's Depression Inventory 2 (CDI-SF) measured at baseline and 3-months post intervention. Results showed that both Project ABC and Project Personality significantly reduced 3-month depression symptoms vs. control, adjusting for baseline depression.

However, treatment effects may vary between youth. Heterogeneous treatment effects (HTE) is defined as "nonrandom variation in the magnitude or direction of a treatment effect across levels of a covariate (that is, a patient attribute or set of attributes) against a clinical outcome" (Kent et al, 2018). For example, female youth may benefit more from the intervention than male youth. It is unknown whether certain subgroups of youth will benefit more or less from each of the two interventions.

Objectives

The overall research goal is to develop **two separate models** to identify subgroups of participants with heterogeneous effects for each SSI. Your specific objectives are:

1. To identify subgroups of participants with heterogeneous treatment effects for the **Project ABC SSI vs. control** using baseline demographic and clinical predictors.
2. To identify subgroups of participants with heterogeneous treatment effects for the **Project Personality SSI vs. control** using baseline demographic and clinical predictors.

Data and Data Dictionary

For this project, we will focus only on the CDI-SF outcome measured at baseline and 3-months post-intervention. The attached CSV dataset contains a subset of **N=1,501 youth with complete data** for the outcomes and baseline characteristics. This dataset differs from the dataset used for the primary analysis in the paper which analyzed all N=2,452 youth using multiple imputation.

A brief data dictionary is provided below.

Variable	Definition	Notes
b_response_id	Unique participant identifier	
condition	Treatment arm allocation	
b_cdi_mean	Baseline CDI-SF score	
f1_cdi_mean	3-month follow-up CDI-SF score	Primary outcome
b_dem_sex	Biological sex	
b_dem_orientation	Sexual orientation	
b_screener_age	Age in years	
b_dem_race_*	Indicator variables for race / ethnicity	Variables are not mutually exclusive. Youth can select more than one race
b_covid_family_*	Indicator variables for specific challenges during the COVID-19 pandemic	Variables are not mutually exclusive. Youth can select more than one challenge
b_covid_cope_1_*	Indicator variables for coping strategies to improve mental health during the COVID-19 pandemic	Variables are not mutually exclusive. Youth can select more than one coping strategy
b_dem_gender_*	Indicator variables for gender identity	Variables are not mutually exclusive. Youth can select more than one gender identity

Methods

The original primary analysis used linear regression to compare the 3-month CDI-SF score across the three arms, controlling for baseline CDI-SF:

lm(f1_cdi_mean ~ condition + b_cdi_mean)

Your strategy for identifying heterogeneous treatment effects must mirror the original analysis approach. For example, do not use a repeated measures type analysis as this would deviate from the primary analysis in Schleider et al. Your models should also adjust for the baseline CDI-SF score.

You will need to build separate models, one for each intervention versus control. The identified subgroups of participants may differ for each intervention.

Some baseline characteristics have many categories and/or are not mutually exclusive. You will need to decide on an appropriate analytical strategy to handle these data.

Select an appropriate analytic strategy and apply it. Do not attempt to try multiple analytic strategies to compare methods. Your analytic strategy must use an existing R package (provide citation in methods). Do not attempt to create your own custom analytic strategy. Your goal is to identify and apply an existing, published analytic strategy with available R software.

Depending on your selected analytical strategy, you may or may not discover heterogeneous treatment effects. Report your model findings regardless of whether you were able to identify subgroups or not.

Your exam will be evaluated based on your analytical approach and presentation of results, and not whether you identified the same subgroups (if any) as the examiner (Dr. Ma) or other students.

Report Format

Your report should be submitted as a PDF by email to Dr. Clement Ma (clement.ma@camh.ca) by the deadline and copy Dr. Olli Saarela (olli.saarela@utoronto.ca). The maximum word count is 3,000 words (provide word count), excluding the title page, tables, appendices, and references.

Overall, the report should be written in the style for a clinical journal. Your target audience is a clinician-scientist. Use the Schleider et al paper as an example of the writing style. Avoid excessive statistical jargon. Explain statistical concepts in words rather than detailed equations.

The maximum combined number of tables and figures is 7. Your report should include the following sections:

1. Introduction
    a. Provide a brief background of the research problem
    b. State the study objectives
2. Methods
    a. Briefly summarize the study population and data. The actual participant numbers should be described in the Results section.
    b. Describe the statistical approach. This should be in sufficient detail that another statistician could perform the analysis. Provide a brief summary of your selected analytical strategy for identifying heterogeneous treatment effects. If available, include appropriate references to the statistical method and/or R software.
    c. Describe your approach to validate your model findings using the provided dataset
3. Results
    a. Describe your findings. Provide some interpretation of the model results.
    b. Describe your validation of the proposed models
    c. Include relevant tables and figures to support and display your findings
4. Discussion
    a. Summarize your key findings
    b. Discuss the strengths and limitations of your analysis
5. References
    a. A formal literature review is not expected. Include references for the statistical method and R software (if available).
6. Appendices
    a. Your report should be self-contained. Assume I will not read this section.
7. R code
    a. Submit a separate R script or R Markdown file with your analysis code. Provide brief comments in your code to guide the examiner (Dr. Ma) through your program.

References

1. Schleider JL, Mullarkey MC, Fox KR, Dobias ML, Shroff A, Hart EA, Roulston CA. A randomized trial of online single-session interventions for adolescent depression during COVID-19. Nat Hum

Behav. 2022 Feb;6(2):258-268. doi: 10.1038/s41562-021-01235-0. Epub 2021 Dec 9. PMID: 34887544; PMCID: PMC8881339.

2. Kent DM, Paulus JK, van Klaveren D, D'Agostino R, Goodman S, Hayward R, Ioannidis JPA, Patrick-Lake B, Morton S, Pencina M, Raman G, Ross JS, Selker HP, Varadhan R, Vickers A, Wong JB, Steyerberg EW. The Predictive Approaches to Treatment effect Heterogeneity (PATH) Statement. Ann Intern Med. 2020 Jan 7;172(1):35-45. doi: 10.7326/M18-3667. Epub 2019 Nov 12. PMID: 31711134; PMCID: PMC7531587.