Vanderbilt University Biostatistics Comprehensive Examination

PhD Applied Exam Series 2

May 28–May 31, 2024

Instructions: Please adhere to the following guidelines:

- This exam is scheduled to be administered on Tuesday, May 28 at 9:00am, and will be due on Friday, May 31 at 5:00pm. This deadline is strict: late submissions will not be accepted.
- To turn in your exam, please use your assigned Box folder and e-mail your exam to Dr. Andrew Spieker by the deadline. This redundancy is designed to ensure that your exam is received by the deadline. If you would like to e-mail exam drafts along the way, that is perfectly acceptable—do not be concerned about spamming my inbox.
- You are advised to pace yourself and to not spend too much time on any one problem. Further, note that there is no one single correct answer to any question on this examination. The questions are open-ended.
- Answer each question clearly and to the best of your ability. Partial credit will be awarded for partially correct answers.
- Be as specific as possible in your responses.
- You may consult reference material (e.g., course notes, textbooks), though the work you turn in must be your own (this means no generative AI). This is an *individual effort*. Do not communicate about the exam with anyone. Vanderbilt University's academic honor code applies.
- Please direct clarifying questions by e-mail to Dr. Andrew Spieker.

Background: Hypertension

Hypertension is a major risk factor for a number of adverse health outcomes, and its prevalence is on the rise. Antihypertensive drug prescriptions can be effective in reducing blood pressure; however, adherence to medication as prescribed is critical in order to realize benefit. What's more, behaviors such as maintaining a healthy diet, maintaining a healthy weight range, being physically active, avoiding tobacco and alcohol products, and getting adequate sleep are also known to be important in achieving a healthy blood pressure range. Interventions to support positive behaviors among patients with hypertension are timely and important.

Disclaimer: This data set and the description thereof are motivated by/adapted from real-world studies, but with substantial modification for pedagogical purposes and to protect patient information.

Study Background: BP Buddy for support

"BP Buddy" is a comprehensive application that was developed to encourage positive self-care behaviors in patients with hypertension. It features a secure, comprehensive, interactive portal in which patients receive tailored feedback on one or more self-care targets (described in the background) and can log/track progress dynamically and visually. In addition to providing tailored encouragement, some features of BP Buddy include medication, diet, sleep, and fitness tracking—many of which exist as separate applications, but not in one centralized location. BP Buddy is compatible with most modern mobile devices and web browsers.

Adult patents were recruited from eight clinics across Tennessee, and were eligible if they were prescribed at least one antihypertensive drug, self-identified a need to improve on at least one of the behaviors described in the background, and had access to a mobile phone and/or web browser compatible with BP Buddy. After providing informed consent, patients were randomized on a 1:1 basis to receive either access to BP Buddy (the experimental intervention) or an active control featuring daily reminders (by text message and/or e-mail) for patients to take their medication as directed. The intervention assigned to a patient became available to them until their final study visit (a maximum of 365 days from randomization) or until they requested withdrawal from the intervention. Participants assigned to receive access to the BP Buddy portal were allowed to adaptively select as few or as many self-care behaviors to target as they wished, and the BP Buddy application was adaptable to include only the features selected by the patient (the caveat being that at least one feature must have been selected). Patients could select and/or deselect the components at will throughout the study. The goal of keeping the functionality adaptive in this way was to support patient autonomy and encourage continued engagement. However, information on how, when, and why patients turned functionality on and off was not logged. The three major study outcomes included systolic blood pressure (SBP), a validated hypertension behavior score (HTNB-10; higher scores are better) that aggregates information on all of the behaviors listed in the background, and self-reported medication adherence (ascertained at each visit by directly asking the patient "what percent of the time do you take your medication as directed?"). These measures were evaluated at two follow-up visits: one approximately three months post-baseline (Visit 2) and one approximately twelve months post-baseline (Visit 3—end of follow-up, after which access to BP Buddy and/or medication reminders ceases). For patients on the BP Buddy arm, the number of days with at least three minutes of activity on the BP Buddy portal was logged. The documentation for the data is featured on Page 4 and contains important information about study variables.

Study aims

The following questions are intentionally open-ended, with focus on broad scientific goals. Your objective is to thoroughly and carefully answer the questions, taking statistical considerations into account. Again, there is no one single correct answer.

1. What is the effect of BP Buddy relative to the active control on the three major study outcomes over time?
2. How are patient characteristics associated with BP Buddy portal activity?

Exam task and formatting instructions

Your task is to create an analysis report in which you address the scientific questions and summarize your findings. Clearly describe your methods in detail and state assumptions explicitly. Where possible, explore how well those assumptions are met and/or how sensitive the results of your analyses are to important assumptions you make. Describe or address any statistical considerations you would expect to be considered in the peer-review process. While there is no page/word limit, your report should be parsimonious.

Guidance for the analysis report (much of which also applies to your professional practice as a biostatistician)

(1) Pace yourself. Don't begin by running a bunch of models; instead, start by considering the study questions and background. Carefully weigh the advantages and disadvantages of different approaches. No one approach will be perfect, but considering various trade-offs before looking at the data will help you avoid common pitfalls and will leave you better equipped to articulate your reasons for choosing your approach.

(2) The data provided to you have been processed and cleaned to an extent, but these data are not so clean that they are ready to be fed into off-the-shelf software without further processing (e.g., handling missingness and/or reshaping) on your side. You will need to make judgment calls. You are expected to clearly articulate and defend your decisions on data processing.

(3) The study has limitations that may make it challenging to find the perfect solution to answer the scientific questions as stated. You should do the best you can, but your interpretation of the findings should be stated appropriately. In your discussion, you may want to describe some of these limitations (e.g., key variables omitted from the study) and how they impact your conclusions.

(4) Use clear section and subsection headers to delineate sections (e.g., introduction, methods, results, and discussion) so that so it is easy for the reader to find what they are looking for.

- I recommend beginning with an introduction section and—in the spirit of item (2)—a section on data preparation/cleaning. Unlike a research paper, though, an analysis report can include methodology/results/discussion sections laid out *separately* for each question being answered.
- For example, each clinical question deserves a main heading. Subheadings can include a summary of your findings, sections for methods, results, and discussion.
- You don't have to be strict in separating those the way you do in a journal article. For example, you could have a subsection on sensitivity analyses in which you describe the methods and results together in that subsection. Having them together often reads better as long as it's clear when you're reading methods and when you're reading results.

(5) Your goal is to answer the questions the way you would as a practicing statistician; it's not to show off all the methods you know. If you do multiple analyses for a question, be clear which is the main analysis and which are exploratory/confirmatory analyses.

(6) You want your analysis report to be readable by both clinicians and statisticians; you also want to summarize your findings in plain English.

(7) You want to be sufficiently detailed, but not bury the main points between lots of details.

(8) You'll want to make your code available, but you do not want it to clutter up your report. The preferred way to accomplish this is to make it so that you have to click a tab to reveal the code in a .html file report (RStudio notebooks/Quarto); another is to have it as a separate file with clear section headings as comments (knitr .pdf report). **Code should be annotated with comments that are designed to make it clear what the key pieces are doing. Do not assume the reader will be able to figure it out without guidance.**

Codebook

The data set contains the variables listed in the table below.

Variable name	Description
ID	Participant (HTNS-PPP: S indicates study site, PPP* indicates participant)
VISIT	Visit number (2 = three months, 3 = twelve months)
BPBUDDY	Randomization group** (0 = Active control; 1 = BP Buddy)
AGE	Age at randomization (years)
MALE	Gender (0 = Non-male, 1 = Male)
EDUC	Education at randomization (years)
INC	Income (0 = <$60,000, 1 = ≥$60,000)
DUR	Duration of hypertension at randomization (years)
SBP0	Systolic blood pressure at randomization (mm Hg)
HTNB0	Blood pressure behavior score at randomization (1-10)
MDADH0	Medication adherence at randomization (0-100)
SBP	Systolic blood pressure (mm Hg)
HTNB	Blood pressure behavior score (1-10)
MDADH	Medication adherence score (0-100)
STUDYTIME	Number of days between randomization and Visit 3 or withdrawal (whichever is first)
ACTIVITY	Number of days with at least three minutes of BP Buddy activity

* Observations marked with PPP = "000" are test observations to ensure functionality of data capture software (one for each study site); they are not observations from real participants. In fact, some of the test data entered do not take on plausible values.

** All participants retained their randomization status throughout the study duration.

Evaluation

Your exam submission will be evaluated on the following general areas:

- The statistical validity and thoughtfulness of your approach, along with accuracy of implementation.
- How well your responses address the scientific and clinical questions.
- The appropriateness/quality of your writing and presentation.