

Vanderbilt University Biostatistics Comprehensive Examination

PhD Applied Exam Series 2

May 30–June 2, 2023

Instructions: Please adhere to the following guidelines:

- This exam is scheduled to be administered on Tuesday, May 30 at 9:00am, and will be due on Friday, June 2 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 and Dr. Svetlana Eden by the deadline. This level of 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 our inboxes.
 - 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 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: Testing for and treatment of influenza in hospitalized children

Influenza is a contagious seasonal respiratory virus and an important cause of acute respiratory illnesses (ARI) in children. The Infectious Diseases Society of America states that antiviral treatment with oseltamivir is recommended for all hospitalized individuals with either confirmed or suspected influenza, particularly if such treatment can be initiated within 48 hours of symptom onset. To be clear, these treatment guidelines are unambiguous about the fact that a positive influenza test is not required for treatment with oseltamivir to be indicated if influenza infection is suspected. Nevertheless, adherence to these guidelines—put forth over a decade ago at the time of this writing—has been lacking. Moreover, influenza testing for individuals hospitalized for ARI is not always consistent even when their symptoms suggest likely influenza infection (e.g., fever and/or cough during peak influenza season). It is of scientific interest to characterize sources of variation in the testing for and treatment of influenza.

Disclaimer: This data set and the description thereof are motivated by/adapted from a real study, but both have undergone substantial modification for pedagogical purposes and to protect patient information.

Study Background: BRIC

The Burden of Respiratory Illness Consortium (BRIC) consists of investigators from five medical centers across the United States (situated in Miami, Boston, Los Angeles, Minneapolis, and Austin). The investigative team conducts a wide variety of studies related to the incidence, burden, and cost associated with hospitalization for respiratory illnesses in children and adults.

The particular data set that you have been supplied originated from a study conducted by BRIC and comprises 2,086 children <11 years of age who were hospitalized across each of the five study sites for acute respiratory illnesses between September, 2018 and August, 2019. To be eligible for inclusion in the study, patients were required to present with *at least one* of fever or cough at the time of admission. Information was obtained by retrospective medical chart reviews: demographic information abstracted included age, sex, race/ethnicity, and insurance status; medical history included receipt of influenza vaccine and various underlying medical conditions understood to be associated with high risk of severe influenza illness; factors pertaining to clinical presentation (fever, cough, days of symptoms prior to admission) were also obtained. A record was noted of whether each patient was admitted during peak influenza season (November, 2018 through February, 2019). Records were further kept of whether each patient was ever tested for influenza while hospitalized, and whether each was ever treated for influenza with oseltamivir while hospitalized. There is no well-defined temporal ordering of these two outcome variables (some oseltamivir-treated patients were tested for influenza prior to treatment initiation, and some were treated with oseltamivir prior to being tested).

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. How are patient characteristics associated with influenza testing among children hospitalized for acute respiratory illness?
2. How are patient characteristics associated with antiviral treatment for influenza with oseltamivir among children hospitalized for acute respiratory illness?

Note: Missing data and between-site variation in influenza testing and treatment should not be ignored.

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 a large extent, but these data are not so clean that they are ready to be fed into off-the-shelf software without further processing 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 where 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. One 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); 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	Unique study participant identifier
studysite	Study site (0=Miami; 1=Boston; 2=Los Angeles; 3=Minneapolis; 4=Austin)
agemo	Age at time of admission (months)
male	Sex (0=Non-male; 1=Male)
raceeth	(0=Non-Hispanic White; 1=Non-Hispanic Black; 2=Hispanic; 3=Other; 4=Missing)
insurance	(0=Public; 1=Private; 2=Both; 3=Self-pay; 4=Missing)
vaccinated	Ever received influenza vaccine (0=No; 1=Yes; 2=Unsure; 999=Refused to answer)
peak	Admitted during peak influenza season (0=No; 1=Yes)
ldx	Underlying lung disease (0=No; 1=Yes)
nmdx	Underlying neuromuscular disease (0=No; 1=Yes)
chd	Underlying congenital heart disease (0=No; 1=Yes)
immuno	Immunocompromised (0=No; 1=Yes)
fever	Fever at admission (0=No; 1=Yes; 999=Missing)
cough	Cough at admission (0=No; 1=Yes; 999=Missing)
dayssymp	Days symptomatic prior to admission (numeric; 999=Missing)
tested	Tested for influenza while hospitalized (0=No; 1=Yes)
oseltamivir	Treated for influenza with oseltamivir while hospitalized (0=No; 1=Yes)

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.