Vanderbilt University Biostatistics Comprehensive Examination

PhD Applied Exam Series 2

May 24–27, 2022

Instructions: Please adhere to the following guidelines:

- This exam is scheduled to be administered on Tuesday, May 24 at 9:00am, and will be due on Friday, May 27 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. Robert Greevy 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.
- There are three problems. Note that not all questions and sub-questions are weighted equally. 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 responses 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.
- This exam is open-everything, but remains 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 and Dr. Ben French.

Scientific Background

Influenza virus

Influenza is a contagious seasonal respiratory virus that can cause mild to severe illness. There are three classes of influenza strains known to infect humans (A, B, and C). Influenza type A viruses are divided into sub-types based on hemagglutinin (H) and neuraminidase (N) proteins located on the viral surface. Influenza type B viruses are instead divided into sub-types based on lineage (Victoria and Yamagata). Infection from influenza type C is generally associated with more mild symptoms than types A and B; further, influenza type C is not known to cause human epidemics. For this reason, research focused on mitigating the burden of influenza-associated illness focuses primarily on types A and B.

Protection from severe illness

Yearly vaccination for influenza is recognized as the mainstay for protection from severe influenza-associated illness, and operates by eliciting an antibody response. However, vaccination against one strain does not necessarily protect against infection from another strain. For this reason, vaccines typically include antigen from multiple sub-types. Historically, inactivated influenza vaccines have been *trivalent*, including antigen from two influenza A strains and one influenza B strain based on expert predictions on which strains and sub-types will most likely circulate the following season.

One concern is that influenza vaccination may not elicit sufficiently protective antibody responses among individuals with weaker immune systems. In such sub-populations, there are several sensible modifications to the standard practice of a single yearly dose of trivalent influenza vaccine that may be worth studying, including:

- Inclusion of antigen from both influenza B sub-types instead of just one (i.e., turn the trivalent vaccine into a *quadrivalent* vaccine).
- Formation of a higher vaccine dose (i.e., introduce more antigen at a single time in an attempt to elicit a stronger immune response).
- A two-dose regimen in the same influenza season (similar to the previous idea, but with antigen being introduced more gradually).

Evaluation of antibody response

To evaluate a vaccine's protective effects from severe infection requires a very large sample size. Particularly in the earlier phases of clinical research, vaccine strategies can be evaluated and compared in part by determining the degree to which they produce antibodies post-vaccination. Naturally, it takes time for the immune system to mount an antibody response; we typically assume that it takes no more than four weeks following influenza vaccination for the body to produce its maximal antibody response to that dose. The hemagglutination inhibition (HAI) assay measures antibody levels by leveraging the natural tendency of blood to arrange irregularly (agglutinate) when exposed to a virus sample unless there are antibodies circulating to *inhibit* this process. Specifically, serum antibody levels can be quantified by sequentially diluting the serum sample and exposing it to a fixed amount of virus; an individual's HAI titer is defined as the relative concentration of the final dilution at which agglutination does not occur (higher titers are therefore indicative of a greater concentration of antibodies). If one vaccine tends to produce better antibody responses over another, we consider it to be immunogenically superior. In a traditional assay, the initial dilution of the serum sample for HAI titer evaluation occurs at a ratio of 1:10, and each sample is sequentially diluted by a factor of two for a maximum of eleven subsequent dilutions (the idea being that the vast majority of samples will not require this many dilutions in order to achieve agglutination). On the other hand, it is also possible that an individual's antibody levels are so low that hemagglutination is observed at the initial dilution of 1:10. In such instances, it is standard procedure to impute titers as half the initial dilution (i.e., as a titer of 5).

Study Background

Study design

Older age is often associated with weaker vaccine response. An immunogenicity study was conducted at a single institution over a single influenza season to gain insights into strategies that should be compared to standard vaccine practices in future effectiveness studies more broadly. In that spirit, note that alterations to the standard-dose trivalent regimen described on the previous page could be mixed and matched. One complicating factor is that only *some* combinations had been authorized for evaluation at the time of this study. For instance, a high-dose trivalent vaccine had been authorized, but a high-dose quadrivalent vaccine had *not*. Therefore, not all $2^3 = 8$ combinations of strategies were eligible for study (nor would a $2 \times 2 \times 2$ factorial design be practical anyway). A double-blind, phase II randomized control trial of $N = 168$ adults over 65 years old was conducted to compare the short-term immunogenicity and reactogenicity (tendency to produce vaccine-related symptoms) associated with each of the following *three* vaccine regimens:

- Group 1: A two-dose series of standard-dose quadrivalent inactivated influenza vaccine (SD-QIV).
- Group 2: A two-dose series of high-dose trivalent inactivated influenza vaccine (HD-TIV).
- Group 3: A single high-dose trivalent inactivated influenza vaccine.

To be clear regarding vaccine regimens and corresponding timelines, patients were evaluated and measured according to the following schedule:

- Visit 1 (baseline): Collection of demographics and serum collection for evaluation of pre-vaccine HAI titers, followed by the first vaccine in assigned series (Group 1: SD-QIV; Group 2: HD-TIV; Group 3: HD-TIV).
- Visit 2 (four weeks following baseline): Serum collection for evaluation of HAI titers, survey regarding injection-site and systemic reactions to first dose, followed by second vaccine in assigned series (Group 1: SD-QIV; Group 2: HD-TIV; Group 3: Placebo).
- Visit 3 (four weeks following Visit 2): Serum collection for evaluation of HAI titers, along with survey regarding injection-site and systemic reactions to second dose.

Note: SD-QIV contained 15µg of antigen from each of the following influenza strains: A/H1N1, A/H3N2, B/Victoria, and B/Yamagata; HD-TIV contained 60µg of antigen from A/H1N1, A/H3N2, and B/Victoria (but notably, not B/Yamagata). A more detailed codebook can be found at the end of this document.

Study questions

The following questions are intentionally open-ended and focus on clinical questions. Your objective is to thoroughly and carefully answer the clinical questions, taking statistical considerations into account. Key aspects to weigh and address in your analytic plan include multiple time points, multiple study groups, and multiple antigens. Again, there is no one single correct answer.

1. Is high dose inactivated influenza vaccine immunogenicially superior to standard dose inactivated influenza vaccine?
2. Are two-doses of inactivated influenza vaccine immunogenicially superior to a single dose of inactivated influenza vaccine?
3. Do subjects experiencing greater degrees of reactogenicity (i.e., injection-site or systemic reactions) tend to achieve greater degrees of antibody response?

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 your analyses are. Describe or address any statistical considerations you would expect to be considered in the peer-review process.

Advice for the analysis report
Much of this advice also applies to your professional practice as a biostatistician.

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

(2) 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.

- 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.

(3) Unlike a research paper, an analysis report should have these four sections laid out for each question being answered. It can also have a more extensive section on data preparation/cleaning. 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.

(4) If you do multiple analyses for a question, be clear which is the main analysis and which are exploratory/ confirmatory analyses.

(5) You want your analysis report to be readable by both clinicians and statisticians.

(6) You want to summarize your findings in plain English.

(7) You want to include all the detail, 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 do 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.

Evaluation

Your exam submission will be evaluated on the following three criteria equally:

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

Codebook

The data set contains the variables listed in the table below. The data have been supplied to you in wide format. Note that you may or may not find it helpful to reshape the data set to the long format depending upon your analytic choices.

`id`	unique study participant identifier
`group`	vaccine series assignment (1=Group 1; 2=Group 2; 3=Group 3)
`age`	age at study enrollment (years)
`gender`	self-identified gender at study enrollment (0=female; 1=male)
`h1n1.1`	visit 1 HAI titer to A/H1N1
`h1n1.2`	visit 2 HAI titer to A/H1N1
`h1n1.3`	visit 3 HAI titer to A/H1N1
`h3n2.1`	visit 1 HAI titer to A/H3N2
`h3n2.2`	visit 2 HAI titer to A/H3N2
`h3n2.3`	visit 3 HAI titer to A/H3N2
`vic.1`	visit 1 HAI titer to B/Victoria
`vic.2`	visit 2 HAI titer to B/Victoria
`vic.3`	visit 3 HAI titer to B/Victoria
`yam.1`	visit 1 HAI titer to B/Yamagata
`yam.2`	visit 2 HAI titer to B/Yamagata
`yam.3`	visit 3 HAI titer to B/Yamagata
`ir.1`	Injection-site reactions* within one week of vaccine 1 (0=no; 1=mild; 2=moderate; 3=severe)
`ir.2`	Injection-site reactions* within one week of vaccine 2 (0=no; 1=mild; 2=moderate; 3=severe)
`sr.1`	Systemic reactions† within one week of vaccine 1 (0=no; 1=mild; 2=moderate; 3=severe)
`sr.2`	Systemic reactions† within one week of vaccine 2 (0=no; 1=mild; 2=moderate; 3=severe)

* pain, redness, tenderness, swelling

† fever, headache, fatigue, nausea, vomiting, myalgia