

# Biostatistics PhD Comprehensive Exam: Applied

June 7 - 10, 2021

---

**Instructions:** Please adhere to the following guidelines:

- The PhD Applied Comprehensive Exam will be administered on Monday, June 7 at 9:00am (central time); you have until Thursday, June 10 at 12:00pm (central time) to complete the exams and place your responses into your respective Box folder. You may (should) place draft solutions in your Box folder throughout the examination period; the latest version submitted prior to the deadline will be considered the final version. In addition, please also email your final version to Drs. Andrew Spieker ([andrew.spieker@vumc.org](mailto:andrew.spieker@vumc.org)) and Robert Greevy ([robert.greevy@vumc.org](mailto:robert.greevy@vumc.org)) prior to the deadline (dual submission helps ensure the exam is received).
  - There are four problems of varying difficulty. Note that not all problems are weighted equally. You are advised not to spend too much time on any one problem.
  - Answer each question clearly and to the best of your ability. Partial credit will be awarded for partially correct answers.
  - Please provide responses that are specific and thorough but also concise.
  - This is an open-book and open-notes examination, but it is an *individual effort*; do not discuss any part of this exam with anyone. Vanderbilt University's academic honor code applies.
  - This examination involves the files `SUM0.csv` (the data from the trial) and `SUM0.Dictionary.docx` (the data dictionary); both were provided to you at the time of exam distribution. It is expected that you will use statistical software in producing a response to this question. Wherever you do use statistical software, please attach your code in a clearly labeled appendix.
  - Please email any clarifying questions to:  
Dr. Robert Greevy ([robert.greevy@vumc.org](mailto:robert.greevy@vumc.org)), and  
Dr. Andrew Spieker ([andrew.spieker@vumc.org](mailto:andrew.spieker@vumc.org)).
-

Disclaimer: The following study is adapted from an actual diabetes intervention study. Data values have been simulated to protect patient information and for pedagogical needs.

**Background:** Management of Type II diabetes (DM) is multifaceted. Beyond medication therapy and compliance with it, other lifestyle factors including diet and exercise play an essential role. However, a person's lifestyle factors may be strongly influenced by those closest to them. Recent research is showing behavioral interventions for diabetes self-care are more effective when applied to a social unit consisting of the patient and someone close to them (e.g. a friend, spouse, sibling, etc.).

The Social Unit Management Orientation (SUMO) intervention is designed to improve DM self-care long term by incorporating the social unit. The intervention was compared to a control condition of educational instruction provided to the patient only in a randomized trial. A cohort of patients who were at high risk of renal function decline were randomized to SUMO or control and followed for five years. Serum creatinine measurements were recorded yearly and converted to estimated glomerular filtration rate (eGFR) using the CKD-EPI 2009 formula.

---

**Questions:** The following questions are intentionally open-ended. They focus on the clinical questions of interest. Your objective is to thoroughly and carefully answer the clinical questions, taking statistical considerations into account.

1. What is the effect of SUMO on the primary endpoint of eGFR at 5 years?
2. What is the effect of SUMO on eGFR over time?
3. Are there noteworthy effect modifiers for the effects of SUMO on eGFR?
4. The inclusion of race in the CKD-EPI formula is controversial. Many researchers at Vanderbilt are using the CKD-EPI formula without the adjustment factor for race. How does the inclusion/exclusion of race from the eGFR calculation impact these analyses?

Your task is to create an analysis report that states each question and summarizes 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 analyses are to the assumptions. Describe or address any statistical considerations you would expect to be considered in the peer-review process. Please see the following page for advice on formatting your report.

---

**Formatting advice for the analysis report:** The following advice is true for both the exam and for your professional practice.

- (1) 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.
- (2) 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.
- (3) If you do multiple analyses for a question, be clear which is the main analysis and which are exploratory/confirmatory analyses.
- (4) You want your analysis report to be readable by both clinicians and statisticians.
- (5) You want to summarize your findings in plain English.
- (6) You want to include all the detail, but not bury the main points between lots of details.
- (7) 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).