

Clinical Reading Questions #4

HST 190: Introduction to Biostatistics

These discussion questions are based on Janes et al. (2024).

1. What were the major goals of this analysis? Vaccines are usually evaluated by whether they prevent disease, but here the outcome is viral load among people who *did* become ill. Why do the authors argue that viral load is worth studying? What is it as a proxy for?
2. The four parent trials enrolled 136,382 participants in total, but the analysis cohort contains 2,216. Describe the steps that take restrict the authors' attention from the first number to the second. Since only participants diagnosed with COVID-19 contribute to this analysis, are the vaccine and placebo groups being compared still protected by randomization?
3. How is the primary outcome defined and measured, and why is it reported on the $\log_{10}$ scale rather than in raw copies/mL?
4. The authors fit a separate linear regression model within each trial rather than pooling all four together. What is their reasoning for taking such an approach? List the covariates the regression models adjust for, and explain what adjustment is meant to accomplish here?
5. For the Novavax trial the *primary* analysis is an unadjusted linear regression, with the covariate-adjusted analysis relegated to a sensitivity analysis; the other three trials do the reverse. Why did the authors make this choice for Novavax in particular?
6. Consider Figure 1: Describe what each element of the plot indicates (e.g., violin, points, triangles)? Compare it to Figure 2: What does Figure 2 add that Figure 1 cannot show, and what does Figure 1 show that Figure 2 fails to capture?
7. In Figure 2, the estimated reduction for NVX-CoV2373 ($2.78 \log_{10}$) is the largest of the four, yet its confidence interval is by far the widest, while Ad26.COV2.S has the smallest estimate and the narrowest interval. Using the sample sizes reported in the abstract, explain why this might be the case. Which estimate would you consider the most reliable?
8. Viral sequence data were missing for roughly 21% of participants, and the authors note that missingness was more common among those with low viral load and among vaccine recipients. Is this pattern a problem? How might the reported association between a $\log_{10}$ reduction in viral load and a 21%–29% decrease in transmission inform public health?

References

Janes, Holly, Leigh H Fisher, Jia Jin Kee, et al. 2024. "Association Between SARS-CoV-2 Viral Load and COVID-19 Vaccination in 4 Phase 3 Trials." *The Journal of Infectious Diseases* 230 (6): 1384–89. <https://doi.org/10.1093/infdis/jiae400>.