Clinical Reading Questions #9

HST 190: Introduction to Biostatistics

These discussion questions are based on Torkildsen et al. (2026).

1. What were the major goals of the OVERLORD-MS trial? The authors note that comparative evidence between rituximab and ocrelizumab had until now come from observational studies. What specific shortcoming of that evidence does a randomized head-to-head trial address, and why does the cost difference between the two drugs make the question worth asking?

2. This is a *noninferiority* trial, not a superiority trial. What is the hypothesis being tested, and what is the pre-specified noninferiority margin? What rule do the authors use to declare noninferiority? Explain why a conventional test of "no difference" would not have answered the scientific question.

3. The margin was chosen to "preserve a substantial proportion of the established effect" of anti-CD20 therapy. Why must a noninferiority margin be justified in advance rather than chosen after seeing the data? What might go wrong, scientifically, if the margin is set too wide?

4. The primary endpoint is absence of new or enlarging lesions on T2-weighted MRI between month 6 and month 24. Why did the investigators use month 6, rather than baseline, as the reference point? MRI lesion activity is a *surrogate* for the outcome patients care about—what do the authors say in the Discussion to justify its use, and is anything lost by using it?

5. The account of randomization and blinding notes an implementation error that reversed the allocation sequence at some sites, after which site-specific lists with a 9:1 ratio were generated to restore the intended overall 3:2 allocation. What could go wrong when an allocation sequence is implemented incorrectly, and why does blinding of participants, clinicians, and the MRI reading center matter independently of randomization?

6. The primary analysis used the modified intention-to-treat (mITT) population, which yielded a risk difference of -2.6 percentage points (95% CI: [-9.4, 4.3]); meanwhile, the per-protocol (PP) "sensitivity" analysis yielded 2.3 percentage points (95% CI: [-6.0, 10.6]). Explain the difference between the populations considered by these distinct estimation strategies. When considering a *superiority* trial, intention-to-treat (ITT) is often seen as the conservative choice—explain why that argument does not carry over to a noninferiority trial.

7. Seven participants (3%) had missing data on the primary endpoint, handled by multiple imputation under a missing-at-random (MAR) assumption; meanwhile, missing SDMT scores were imputed by a "best-case" approach. What is the difference in intent between these two strategies for handling missingness.

8. The authors state that secondary endpoints were not multiplicity-adjusted and "should not be interpreted as causal"—given that infections occurred in 82% of the rituximab group versus 69% of the ocrelizumab group, what may a clinician conclude about infection risk from this trial? What further evidence would be necessary to draw a stronger conclusion?

References

Torkildsen, Øivind, Hilde Kjelgaard Brustad, Einar August Høgestøl, et al. 2026. "Rituximab Versus Ocrelizumab in Newly Diagnosed Relapsing Multiple Sclerosis." *New England Journal of Medicine* 395 (1): 44–53. https://doi.org/10.1056/NEJMoa2600993.