MEDICAL COLLEGE OF WISCONSIN

knowledge changing life

STUDY EVALUATION WITH FOCUS ON DIVERSITY AND INCLUSION METRICS

Leonard E. Egede, MD, MS
Professor of Medicine & Inaugural Milwaukee Community Chair in Health Equity Research at the Medical College of Wisconsin
Chief, Division of General Internal Medicine
Director, Center for Advancing Population Science
Medical College of Wisconsin, Milwaukee, Wisconsin, USA

NIH Workshop on Inclusive Participation in Clinical Research
March 30-31, 2023

Evaluating Study DEI Metrics

- How did you do with minority recruitment relative to projections?
  - Proportion recruited by subgroups relative to planned enrollment
- Race/ethnicity breakdown
- Sex/Gender breakdown
- Race by sex/gender breakdown
- Unexpected low accrual
- Loss to follow up by DEI groups

Process Evaluation - 1

- What worked well and what did not go so well with minority recruitment?
- Feedback from your study team members (coordinators, health educators, community partners)
- Feedback from study sites
- Feedback from participants
- Were incentives adequate?
- Were partners happy with flow, engagement, and incentives?
- Any obvious lapses that should be addressed for future studies?

Process Evaluation - 2

- Evaluate objective and subjective measures of minority recruitment strategies
- Evaluate data on success by strategies and sites
- Evaluate data on decline to participate by race/ethnicity and potential predictors of non-participation
- Conduct structured interviews for completers (why did you agree to participate and stay in the study?)
- Structured interviews for dropouts (why did you decide to drop out?)

Statistical Considerations 1 – Post hoc Power Analysis

- Power is typically calculated prospectively (sample size needed to achieve a stated effect size and significance level)
- Evidence suggest many studies are under-powered
- Prospective sample size calculation may have underestimated expected differences between groups
- Enrollment numbers may be less than planned
- Post hoc power analysis is revised power calculation based on the observed value of the effect size between groups
- Typically used when a statistically nonsignificant result is obtained.
  - Differentiates low power from a truly small effect
  - if post hoc power is high, then nonsignificance is due to a small effect size

Statistical Considerations 2 – Heterogeneity of Treatment Effect

- Randomized controlled trials (RCTs) and observational studies of comparative effectiveness usually report an average treatment effect (ATE)
- HTE is defined as nonrandom variability in the direction or magnitude of a treatment effect, measured using clinical outcomes
- There are two main goals of HTE analyses:
  - (1) to estimate treatment effects in clinically relevant subgroups (subgroup analysis)
  - (2) to predict whether an individual might benefit from a treatment (predictive learning)
- Consider evaluating HTE by race/ethnicity, age, sex/gender, SES

Statistical Considerations 3 – Heterogeneity of Treatment Effect

- Approaches:

1-Estimate benefit separately in subgroups of patients, based on the assumption that a subgroup is more homogeneous than the entire study population

2-Use a statistical model that estimates the relationship between multiple baseline characteristics and outcome. Such a model assigns a multivariable risk score to every patient

3-Construct a statistical model from the RCT data that formally incorporates interaction terms between treatment exposure and predetermined baseline factors

4-Newer machine learning methods that search across all possible combinations of potential predictor variables and interactions to predict variability in treatment response

Statistical Considerations 4 – Subgroup Analyses

- Most commonly used analytic approach for examining HTE
- Evaluates the treatment effect for several subgroups, one variable at a time, usually a baseline or pretreatment variable
- A test for interaction is conducted to evaluate if a subgroup variable has a statistically significant interaction with the treatment indicator.
- If the interaction is significant, then the treatment effect is estimated separately at each level of the subgroups (e.g., men and women).
- Interaction test generally has low power to detect differences in subgroup effects
  - sample size roughly four times as large is required for detecting a difference in subgroup effects of the same magnitude as ATE for a 50:50 subgroup split
  - a sample size approximately 16 times as large is required for detecting a difference that is half of ATE (at significance level 0.05).

Statistical Considerations 5 – Secondary Data Analysis

- Uses clinical trials data to answer other questions
- Useful for exploratory analysis and hypothesis generation
- Useful for graduate students and postdocs to gain experience
- Useful as pilot data for future grants
- Challenges
  - Sample size and power may not be adequate
  - Data is correlated, so need to use appropriate statistical tests
  - Need to account for randomization/group assignment
  - May need to deal with selection bias; internal vs. external validity
  - May need to account for multiplicity of testing
- Pooling data from multiple trials may increase sample size

Incorporating Lessons Learned into Future Studies

- What needs to change based on your experience?
- Where do you invest resources to bolster future recruitment efforts?
- How do you leverage your participants to build word of mouth for future studies?
- What is the true cost of the study?
- How do you use experience to modify budget and resource allocation for future studies?

References

- Russell V. Lenth. Post Hoc Power: Tables and Commentary. The University of Iowa Department of Statistics and Actuarial Science Technical Report No. 378. July 2007

- Developing a Protocol for Observational Comparative Effectiveness Research: A User's Guide. Velentgas P, Dreyer NA, Nourjah P, et al., editors. Rockville (MD): Agency for Healthcare Research and Quality (US); Jan. 2013

- Derek C. Angus, MD, MPH[1,2]; Chung-Chou H. Chang, PhD[2] Heterogeneity of Treatment Effect: Estimating How the Effects of Interventions Vary Across Individuals. *JAMA*. 2021;326(22):2312-2313. doi:10.1001/jama.2021.20552

QUESTIONS/COMMENTS?