Sample size analysis for the NIH HEAL proposal

Zheyang Wu

February 8, 2023

Contents

1 Overview

This document clarify the methods and assumptions and provides results in studying sample sizes for the NIH HEAL proposal.

Overall considerations:

- The design is a one-arm clinical trial – all individuals will be treated by mindfulness therapy.
- The sample size is planned as: about 50 patient subjects to be recruited in the first stage (UG3), followed by 300 subjects in the second stage (UH3). Longitudinal measurements for each individuals will be collected over 4 time points (baseline, 4-week, 8-week, 6-months). This study aims to address what statistical power and prediction capacity could be reached accordingly.

- Sample sizes are estimated based on two different criteria: 1) Statistical power; 2) Predictive accuracy.

- Quantitative response: PEG, a pain relief score between 0 – 10; mean = 5.85 and SD = 2.43 (by Morone data at the baseline).

- Binary response: Consider 30%, 50%, and 70% pain relief/improvement after mindfulness treatment ((Morone et al., 2016) and Morone's proposal).

- The proposal involves detecting useful predictors; the number of candidate factors is at least 50.

- We can consider a dropping/attrition rate (e.g., 20% in Morone's proposal), so the final sample size = (calculated sample size)/(1 - attrition rate).

The effect sizes of the predictors are unknown. We assume a series of values.

An effect size can be described by any one of these:

1) The slope/coefficient and standard deviation of the predictor (for the binary predictor of 1 and 0, the slope gives the change of responses, for which a large magnitude indicates the scientific significance of the predictor). For description purpose, we can state the amount of changes (for quantitative response) or the odds ratio (for binary response) per unit increase of the predictor (the unit can be original data unit or the SD unit).

2) A scaled effect size: coefficient of determinant / Cohen's f^2 / correlation coefficient, etc.

2 Sample size by statistical power

Analyze sample size by statistical power. Consider longitudinal analysis with each individual measured over 4 time points (0, 1, 2, and 6 months). Apply the generalized linear mixed models (GLMM) and R package SIMR for power analysis (Green and MacLeod, 2016).

Fixed effects: intercept, predictor, time, the interaction of predictor by time (imitating Morone et al. (2016) and Morone proposal without group clustering of subjects). Consider the time is a continuous variable.

Random effects: intra-subject longitudinal measurements (assume correlation=0.1).

Studies: 1) Type of responses: continuous or binary; 2) Type of predictors: continuous (assuming N(0,1)) or binary (assuming balanced distribution over subjects).

Results

3 Sample size by predictive model-related criteria

A "minimum" sample size can be calculated based on some generic model-fitting and prediction-related criteria (Riley et al., 2019a,b).

3.1 Quantitative outcomes

Apply multiple regression model. The minimum sample size is calculated to satisfy all four recommended predictive model-related criteria:

1. Small overfitting is defined by an expected shrinkage of predictor effects by 10% or less.
2. Small absolute difference of 0.05 in the model's apparent and adjusted R-squared value.
3. Precise estimation of the residual standard deviation with a multiplicative margin of error (MMOE) less than 1.1.
4. Precise estimation of the average outcome value within 95% confidence interval.

See Table ? for the sample sizes over R^2 and the number of predictors.

3.2 Binary outcomes

Apply multivariable logistic model. The minimum sample size is calculated to satisfy all three recommended predictive model-related criteria:

1. Small overfitting defined by an expected shrinkage of predictor effects by 15% or less.
2. Small absolute difference of 10% in the model's apparent and adjusted Nagelkerke's R^2.
3. Precise estimation (within +/- 10%) of the average outcome risk in the cohort of the study.

See Table ? for the sample sizes over AUC and the number of predictors.

4 Sample size by predictive accuracy

It is difficult to analytically calculate sample size based on 1) direct prediction accuracy and 2) complex predictive model (to my best knowledge, such calculations are not available in literature yet). Therefore, we utilize simulations to estimate the sample size under these challenging scenarios.

Here, we utilize the linear mixed model (LMM) and the the generalized linear mixed model (GLMM) to account for data heterogeneity (e.g., the clustering effects among diverse ethnic groups). The predictions were carried out based on a 5-fold cross-validation procedure.

We can study how the prediction accuracy depends on the sample size and the percentage of all possible predictors included into the prediction model.

- "Basic" predictors: 1) gender, 2) age, 3) education level (high/low), 4) long duration of pain (yes/no), 5) baseline pain score, 6) presence of certain concomitant diseases (yes/no). Cherkin et al. (2016); Morone et al. (2016).
- Extra biopsychosocial predictors: number = 50.
- Assuming all predictors, if we can find, can explain most the data variations. We can study how the prediction accuracy depends on the sample size and the percentage of predictors included into the prediction model.

Prediction accuracy measures:

- Quantitative response: the correlation coefficient between the predicted and the observed outcomes.
- Binary response: AUC.

References

Cherkin, D. C., Sherman, K. J., Balderson, B. H., Cook, A. J., Anderson, M. L., Hawkes, R. J., Hansen, K. E. and Turner, J. A. (2016). Effect of mindfulness-based stress reduction vs cognitive behavioral therapy or usual care on back pain and functional limitations in adults with chronic low back pain: a randomized clinical trial. *Jama*, **315** 1240–1249.

Green, P. and MacLeod, C. J. (2016). Simr: an r package for power analysis of generalized linear mixed models by simulation. *Methods in Ecology and Evolution.*

Morone, N. E., Greco, C. M., Moore, C. G., Rollman, B. L., Lane, B., Morrow, L. A., Glynn, N. W. and Weiner, D. K. (2016). A mind-body program for older adults with chronic low back pain: a randomized clinical trial. *JAMA internal medicine*, **176** 329–337.

Riley, R. D., Snell, K. I., Ensor, J., Burke, D. L., Harrell Jr, F. E., Moons, K. G. and Collins, G. S. (2019a). Minimum sample size for developing a multivariable prediction model: Part i–continuous outcomes. *Statistics in medicine*, **38** 1262–1275.

Riley, R. D., Snell, K. I., Ensor, J., Burke, D. L., Harrell Jr, F. E., Moons, K. G. and Collins, G. S. (2019b). Minimum sample size for developing a multivariable prediction model: Part ii-binary and time-to-event outcomes. *Statistics in medicine*, **38** 1276–1296.