Better Research Interactions for Every Family

Statistical Analysis Plan
Project # 3469
Marshall Brown
6/1/2022

A feasibility study for an educational intervention to improve equity and respect in recruitment conversations with families who are eligible for a neonatal clinical trial. A pre-post study design will be used. Approximately 200 parents of eligible infants will be approached for consent, half using the current recruiting protocol ("pre") and half using the novel intervention ("post"). Conversations will be administered by 6 members, some of whom will administer more conversations than other members.

A subsample of these conversations will be recorded (n_record). This will be split equally between "pre" and "post" periods. Each recorded recruitment conversation will then be assessed by 3 experts (2 independent, and the member who administered the educational content). They will fill out a questionnaire (approximately 15 questions) answered with a Likert score. These questions fit into logical groupings to assess different aspects of equity and respect during the recruitment conversation.

The number of observations will be the number of conversations recorded multiplied by 3 (n_record*3). There are several sources of correlation between these observations (recruitment conversation, member administering the recruitment conversation, expert making the assessment) and so GEE will be used to estimate the intervention effect on each question and test the null hypothesis of no effect.

Analysis Plan

Descriptive statistics will be used to summarize eligible parents who engaged in conversations. Frequencies and percentages will be reported as appropriate for categorical variables. Means, medians, and standard errors will be calculated for continuous variables. Outcomes (5- point Likert scale) will be summarized in each intervention arm at the question level, as a total across all questions, and totals within each pre-defined question subgroup.

At the question level, ordinal logistic regression will be used to assess the effect of the intervention on the distribution of scores. Models will be fit using generalized estimating equations (GEE) to account for correlation among results assessed on the same conversation. Results will be reported as intervention odds ratios and 95% confidence intervals. To test the overall effect of the intervention on the score (and in each subgroup of questions) we take the total of the question scores to be continuous and use GEE to fit standard linear regression to test the mean effect of intervention.

Models above assume each conversation administer has the same mean outcome and that experts evaluate themselves the same as others. These two assumptions will be evaluated and GEE models adjusted as needed. Further, since we plan to run many statistical tests (15 questions individually, score totals, subgroups), the p-value threshold to determine significance will be adjusted using Holm's method to control the family-wise type I error rate.

List of Figures/Tables

Figure 1: Flow diagram showing sample size allocation in pre/post cohort, exclusion, and number of videos sampled for assessment.

Figure 2: Distributions of scores in pre/post cohorts for each question, question totals and subgroups.

Figure 3: Estimated odds ratios and confidence intervals and of the pre/post effect tested for each question.

Table 1: demographics of those approached vs videos sampled.

Table 2: Likert score outcomes (counts, percentages) in pre/post, intervention odds ratio, 95% CI's and p-values for each question.

Table 3. Outcome totals (Mean, SD, IQR) and subgroups, estimates of mean effects (95% CIs) .

Power calculations

To assess statistical power, we pair simulation with calculations based on cluster randomized trial designs with a continuous outcome. These equations assume a mixed effects model consistent with the GEE methods used above and depend on the assumed effect size, outcome standard deviation, and intra-cluster correlation. To achieve 80% power at a 0.05 level when the true mean difference is 5 and standard deviation of 5, 8 videos (with 3 observations) need to be recorded per trial arm (8*3*2 = 48 total observations) if the intra-recording correlation is low (rho = 0.1) while 15 recordings are needed per trial arm (15*3*2 = 90 total observations) if intra-recording correlation is high (rho = 0.8).

Input parameters were selected based on simulated ordinal 5-level data (with a uniform distribution across levels) generated 15 times (one for each question) to get totals across many realizations. This leads to an approximate distribution of total scores with mean 45 and standard deviation 5.

References

Heagerty, P.J. and Zeger, S.L. (1996) Marginal regression models for clustered ordinal measurements. *JASA*, **91** 1024–1036.

Matthews JNS. Introduction to Randomized Controlled Clinical Trials (2nd Ed.) Chapman & Hall: New York, 2006.

Donner A and Klar N. Design and Analysis of Cluster Randomization Trials in Health Research. Arnold: London, 2000.

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, **6**, 65–70. https://www.jstor.org/stable/4615733.