**ABSTRACT (425 words max):**

**Background:**

Clinical trials often involve some form of interim monitoring to determine futility and/or efficacy before planned trial completion. While many commonly used parametric options for interim monitoring exist, nonparametric based interim monitoring methods are also needed to account for more complex trial designs and analyses. One recently proposed nonparametric interim monitoring method is the upstrap.

**Methods:**

Upstrapping involves repeatedly resampling existing interim data with replacement and determining the likelihood of trial success based on comparison to preselected p value and proportion threshold criteria. To evaluate upstrapping we conducted a simulation study considering different sample sizes (40, 160, 600, 2000), interim stopping points (0.25, 0.50, and 0.75), and power scenarios (5%, 50%, 80%, 95%). We first compared trial rejection rates across a selection of threshold combinations to validate the upstrapping method. We then considered several different calibration strategies for these thresholds, comparing them based on type I error rate, power, type II error rate, and level of significance. Finally, we applied upstrapping methods to simulated clinical trial data, directly comparing their performance (measured by expected sample size, overall rejection rate, and interim stopping rate) with more traditional group sequential design methods.

**Results:**

The method validation results showed that upstrapping, when threshold values are appropriately calibrated, produces reasonable results across a variety of simulations settings and is more likely to stop the trial early for futility in the null case scenario than in the alternative scenario. Although there are many potential approaches to calibration, using simulated data to select threshold parameters with minimal restrictions performed best. For futility only interim monitoring, upstrapping performed similarly well across performance metrics compared to group sequential designs. For efficacy only interim monitoring, upstrapping methods produced promising results with regards to expected sample size and were overall much more likely to stop early and/or reject the null hypothesis at trial conclusion. However, upstrapping for efficacy only monitoring did not perform as well as group sequential methods with respect to controlling type I error rate. For combined futility and efficacy monitoring, trends in expected sample size, early stopping rate, and rejection rate were less clear but upstrapping methods performed poorly in terms of controlling type I error rate.

**Conclusions:**

Upstrapping is a valuable tool for performing interim monitoring, particularly for futility monitoring. When properly calibrated, upstrapping is found to be comparable to group sequential methods in accurately predicting trial futility. Upstrapping is a promising new method for performing interim monitoring in the context of clinical trial designs that prompt nonparametric approaches to interim monitoring.