Improving Power in Group Sequential, Randomized Trials by Adjusting for Prognostic Baseline Variables and Short-Term Outcomes

Tianchen Qian[1], Elizabeth Colantuoni[2], Aaron Fisher[3], Michael Rosenblum[2]

[1]Department of Statistics, Harvard University. [2]Department of Biostatistics, Johns Hopkins University. [3]Department of Biostatistics, Harvard University

Summary

Research Question: In adaptive enrichment trial designs, how do prognostic baseline variables and short-term outcomes, accrual rate, and delayed outcomes impact trial power, sample size and duration?

- We derive formulas for the precision gain (measured by the asymptotic relative efficiency against the unadjusted estimator) from adjusting for baseline variables and short-term outcomes using semiparametric estimators in randomized trials.
- Compared to the unadjusted estimator, adjusting for prognostic baseline variables and short-term outcomes **increase power** and **reduce sample size and duration** of the trial. The adjustment is most valuable when the variables are highly correlated with final outcome, when the delay to observe final outcome is long, when accrual rate is fast, or when there is small or no treatment effect heterogeneity.
- We use a targeted maximum likelihood estimator (TMLE) to adjust for pre- and post-treatment covariates, and combine it with a new multiple testing procedure. This method is **guaranteed to strongly control the familywise type I error rate**, asymptotically.

Our **motivating clinical application is a trial of a new treatment for preventing Alzheimer's disease progression**.

- We use a data set of 286 patients from the Alzheimer's Disease Neuroimaging Initiative (ADNI). The primary outcome is the Clinical Dementia Rating Sum of Boxes (CDRsb) measured 2 years from enrollment.
- In simulation studies, compared to the standard unadjusted estimator, by using the adjusted estimator that leverages prognostic covariates, we **simultaneously increased the power by 10%, reduced sample size by 3%, and reduced duration by 0.6 years**.

Adaptive Enrichment Design

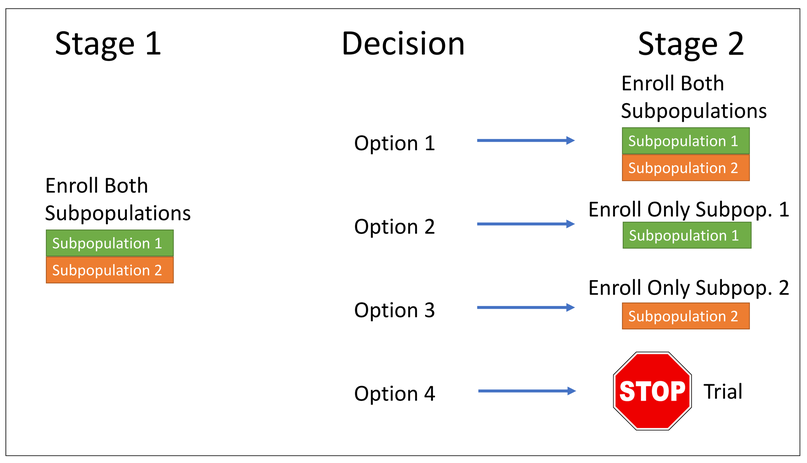
- Adaptive enrichment designs involve preplanned rules for modifying enrollment criteria based on accrued data in an ongoing trial.
- We consider two prespecified subpopulations (defined by ApoE4 genotype) that partition the overall population.
- At each interim analysis, for each subpopulation a decision is made of whether to continue accrual or stop.
- Randomization is 1:1 to treatment or control throughout the trial.
- Strong familywise Type I error control guaranteed by [3].

Null Hypotheses Denote by Δ_1, Δ_2, and Δ_0 the average treatment effect in subpopulation 1, subpopulation 2, and the combined population, respectively. We test the following null hypotheses:

$$H_{01}: \Delta_1 \leq 0; \qquad H_{02}: \Delta_2 \leq 0; \qquad H_{00}: \Delta_0 \leq 0.$$

We simulate the trials under a) both H_{01} and H_{02} are true; b) only H_{02} is true; c) only H_{01} is true; d) neither H_{01} nor H_{02} is true.

Figure: Example of a two-stage adaptive enrichment design.



(Image source: Michael Rosenblum's short-sourse on Adaptive Enrichment Designs. Johns Hopkins University, August 30, 2017.)

Table: Example of a five-stage trial in our simulation, assuming no early stopping due to efficacy. Cumulative Sample Size (Cum. S. S.) is formatted as final outcome observed (+ pipeline*)

Interim Analysis (k)	1	2	3	4	5
Cum.S.S. Subpop. 1	184 (+368)	368 (+368)	552 (+368)	768 (+216)	984 (+0)
Cum.S.S. Subpop. 2	314 (+628)	628 (+314)	942 (+0)	942 (+0)	942 (+0)
Cum.S.S. Comb. Pop.	498 (+996)	996 (+682)	1494 (+368)	1710 (+216)	1926 (+0)
Futility Boundary ($l_{1,k}$)	0.25	0.24	0.23	0.23	-
Futility Boundary ($l_{2,k}$)	0.41	0.39	∞	-	-

Efficacy boundaries are calculated using the covariances of the test-statistics for each simulated trial [1].

* Number of pipeline is calculated assuming $d_L = 1$ year.

Acknowledgements

This research was supported by the FDA BAA (HHSF223201400113C), the US National Institute of Neurological Disorders and Stroke (5R01 NS046309-07 and 5U01 NS062851-04), and the Patient-Centered Outcomes Research Institute (ME-1306-03198). This paper's contents are solely the responsibility of the author and do not represent the views of these agencies.

Notation and Theoretical Result

For a subject, define the following:

- $\boldsymbol{S}$: subpopulation (no ApoE4 ϵ4 alleles / at least one allele)
- $\boldsymbol{W}$: baseline variables (baseline CDR, age, $A\beta_{42}$, ADA scale, etc.)
- $\boldsymbol{A}$: binary treatment indicator
- $\boldsymbol{L}$: short-term outcome, delay d_L from enrollment (CDRsb at 1 year);
- $\boldsymbol{Y}$: primary outcome, delay d_Y from enrollment (CDRsb at 2 year).

At an interim analysis, denote by:

- $\boldsymbol{p_l}$: proportion of enrollees with L observed;
- $\boldsymbol{p_y}$: proportion of enrollees with Y observed.

- Define $E_0(\cdot) = E(\cdot \mid A = 0)$, $\mathrm{Var}_0(\cdot) = \mathrm{Var}(\cdot \mid A = 0)$; similarly define $E_1(\cdot)$ and $\mathrm{Var}_1(\cdot)$.
- $\boldsymbol{\gamma}$ is the treatment effect heterogeneity:

$$\gamma = \frac{\mathrm{Var}\{E_1(Y \mid W) - E_0(Y \mid W)\}}{\Sigma_{a\in\{0,1\}}\mathrm{Var}_a\{Y\}}.$$

- $\boldsymbol{R^2_W}$ is the prognostic value in W:

$$R^2_W = \frac{\Sigma_{a\in\{0,1\}}\mathrm{Var}_a\{E_a(Y \mid W)\}}{\Sigma_{a\in\{0,1\}}\mathrm{Var}_a(Y)}.$$

- $\boldsymbol{R^2_{L|W}}$ is the prognostic value in L after adjusting for W:

$$R^2_{L|W} = \frac{\Sigma_{a\in\{0,1\}}\mathrm{Var}_a\{E_a(Y \mid L, W) - E_a(Y \mid W)\}}{\Sigma_{a\in\{0,1\}}\mathrm{Var}_a(Y)}.$$

In the ADNI data, $R^2_W \approx 0.20$ and $R^2_L \approx 0.48$.

Result Qian et al. [2]: Assume randomization ($A \perp\!\!\!\perp W$) and independent censoring on L, Y. For estimating the average treatment effect of A on Y, the asymptotic relative efficiency between an efficient RAL estimator and the unadjusted estimator is

$$\frac{\mathrm{AVar(unadjusted)}}{\mathrm{AVar(efficient)}} = \{1 + (p_y/2)\gamma - [1 - (p_y/p_l)]R^2_{L|W} - R^2_W\}^{-1}.$$

Adjusted Estimator: TMLE

We use the Targeted Maximum Likelihood Estimator (TMLE) [4] to adjust for prognostic baseline variables and short-term outcomes. The advantages of using TMLE in a randomized trial are as follows.

- Guaranteed to strongly control the familywise Type I error rate, using the testing procedure based on corresponding Wald statistics. (Assuming outcome data missing completely at random, or missing at random with correctly modeled missingness probability.)
- Improve power, reduce sample size, and reduce trial duration compared to the unadjusted estimator.
- R^2_W and $R^2_{L|W}$ can be estimated empirically to predict the precision gain from covariate adjustment.
- Available in R package ltmle.

Simulation Setup

Our goal is to evaluate the performance of an adaptive enrichment design with a delayed response when we vary the prognostic values in baseline variables and short-term outcome, accrual rates, delay time, and estimator used. The performance is evaluated based on Type I error, power, expected sample size and average duration of the trial, and is based on two estimators: the unadjusted estimator (the difference between the sample means of the primary outcome between the two study arms), and an adjusted estimator (TMLE).

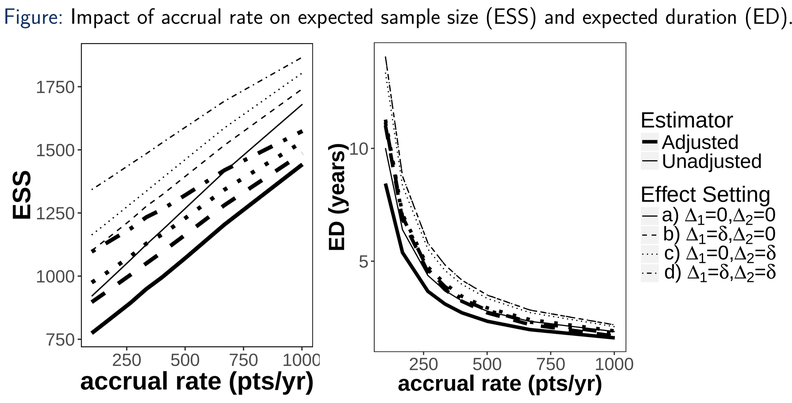
We vary the following in our simulation studies: the prognostic value of baseline variables W and short-term outcome L represented by the R-squared formulas; the delay time d_L of the short-term outcome; the delay time d_Y of the final outcome; and the accrual rate.

Table: Summary of simulation setup. Default value of parameter: $R^2_W = 0.20$, $R^2_L = 0.48$, $d_L = 1$ years, $d_Y = 2$ year, accrual rate 167 patients/year. Ranges of values $x - y$ indicate the design characteristic(s) varied in the corresponding simulation study.

Simulation study	R^2_W	R^2_L	d_L (years)	d_Y (years)	accrual rate (patients/year)
1	$0 - 0.6$	0	default	default	default
2	0	$0 - 0.6$	default	default	default
3	default	0	default	$0 - 4$	default
4	default	default	$0 - d_Y$	$0.1, 1, 2, 3, 4$	default
5	default	default	default	default	$50 - 500$

Simulation Results

Figure: Impact of accrual rate on expected sample size (ESS) and expected duration (ED).



Simulation Results (cont.)

Figure: Impact of prognostic value on expected sample size (ESS), expected duration (ED). Performance of the unadjusted estimator does not depend on R^2_W or R^2_L, and is marked next to the vertical axis.

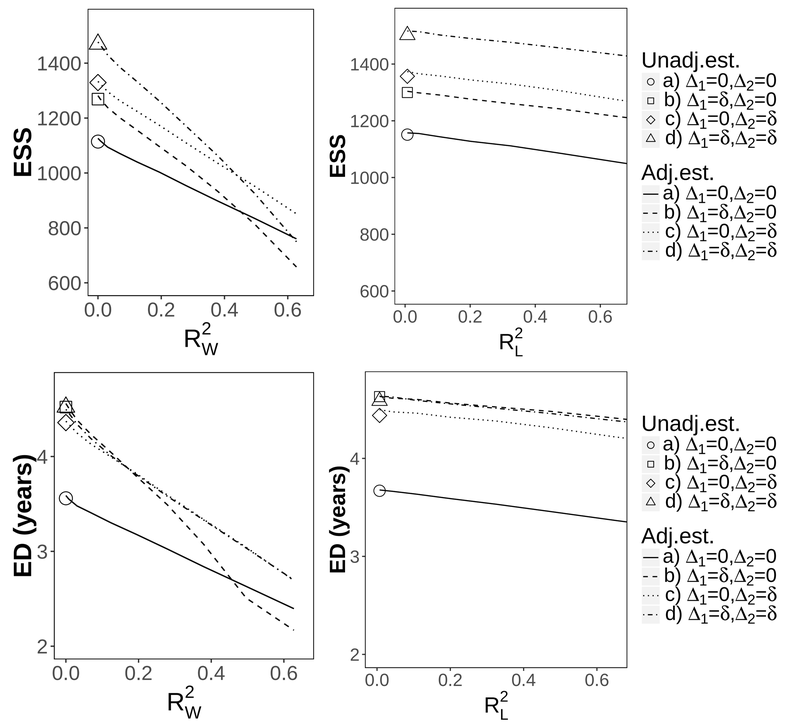
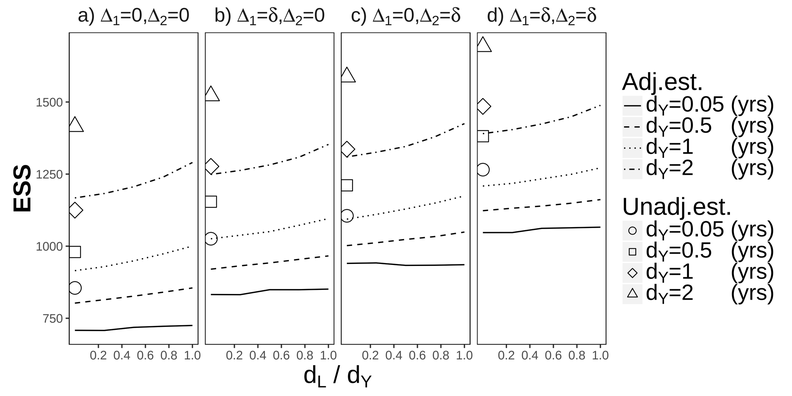


Figure: Impact of delay time to final outcome (d_Y) and to short-term outcome (d_L) on expected sample size (ESS). Horizontal axis is the ratio of the delays to short-term outcome and to final outcome. Performance of the unadjusted estimator does not depend on d_L or d_Y, and is marked next to the vertical axis. Figure for expected duration (ED) has similar trends and is omitted here.



Future Research

- **Extension to sequential decision time points.** The theory and simulation results presented here are for traditional randomized trials with a single intervention. When there are multiple time points of intervention (e.g., mHealth setting), how to properly adjust for time-varying covariates and how much precision gain such adjustment brings is ongoing research.
- **Extension to individual treatment effect.** The current setup focuses on estimating the average treatment effect. When there are multiple time points of intervention (e.g., mHealth setting), it is possible to consider estimation of individual treatment effect. We are working on random effects models with the presence of time-varying covariates in order for regression coefficients to still have valid causal interpretation.

References

[1] Qian, T., E. Colantuoni, A. Fisher, M. Rosenblum, A. D. N. Initiative, et al. (2017). Sensitivity of adaptive enrichment trial designs to accrual rates, time to outcome measurement, and prognostic variables. *Contemporary Clinical Trials Communications 8*, 39–48.

[2] Qian, T., M. Rosenblum, and H. Qiu (2016). Improving power in group sequential, randomized trials by adjusting for prognostic baseline variables and short-term outcomes. *Johns Hopkins University, Dept. of Biostatistics Working Papers, Working Paper 285*.

[3] Rosenblum, M., T. Qian, Y. Du, H. Qiu, and A. Fisher (2016). Multiple testing procedures for adaptive enrichment designs: combining group sequential and reallocation approaches. *Biostatistics 17*(4), 650–662.

[4] van der Laan, M. J. and S. Gruber (2012). Targeted minimum loss based estimation of causal effects of multiple time point interventions. *The international journal of biostatistics 8*(1).

Contact Information

- Email: qiantianchen@fas.harvard.edu
- Phone: +1 (609) 216-4600
- Link to this poster: `https://goo.gl/ZqWV99`