Optimizing Interim Analysis Timing for Bayesian Adaptive Commensurate Designs

Xiao Wu[1*], Yi Xu[2], Bradley P. Carlin[3]

[1]Department of Biostatistics, Harvard TH Chan School of Public Health, Boston, MA, [2]Xenon Pharmaceuticals Inc., Burnaby, BC, Canada, [3] Counterpoint Statistical Consulting, Edina, MN

*wuxiao@g.harvard.edu

Background

- In developing products for rare diseases, statistical challenges arise due to the limited number of patients available for participation in clinical trials.
- A specific example motivating this research was the consideration of a pediatric trial design testing the effect of a new drug for a rare genetic disease.
- Bayesian adaptive clinical trial designs offer the possibility of increased efficiency via their incorporation of evidence from historical data, and flexibility in the specification of interim looks.

Objectives

To develop Bayesian adaptive **commensurate** designs that borrows adaptively from **historical information** and also uses a particular payoff function to optimize the timing of the study's **interim analysis**. Importantly, we also propose calibration procedures to maintain acceptable long-run frequentist properties (Type I error and power) for the designs.

- Historical information: In the case of pediatric trials, it was reasonable from a clinical perspective to utilize information from adult populations.
- Commensurability: The primary endpoint used to measure the treatment effect and that the magnitude of change in the primary endpoint would expect to be similar between adults and children in both the placebo and treatment arms.
- Interim analysis: The interim analysis is desirable since the fixed sample size design may require a sample size that is not realistic.

Notations

- Let n_k be the sample size for the pediatric study in group k, where $k = 1$ is the placebo group and $k = 2$ is the treatment group.
- Let n_{0k} be the sample size for the two historical adult groups respectively. Since we will typically have much more adult data than pediatric, we set $n_1 = n_2 < n_{01} = n_{02}$.
- Let $n \equiv n_1 + n_2$ be the planned pediatric sample size if the trial runs to completion.
- The analysis prior uses a *commensurate prior* framework [1], and assumes

$$\theta_k|\theta_{0k} \sim N(\theta_{0k}, 1/\tau_k), \text{ and } \theta_{0k} \sim N(0, \sigma_0^2), \ k = 1, 2.$$

- Suppose we assume that Y_{kj} and Y_{0kj}, the observed percent reductions for each pediatric and adult patient, are also normally distributed, that is,

$$Y_{kj} \overset{ind}{\sim} N(\theta_k, 1/\omega) \text{ and } Y_{0kj} \overset{ind}{\sim} N(\theta_{0k}, 1/\omega_0)$$

- Early winner: If at the interim look, the probability that the novel treatment arm ($k = 2$) is better exceeds some prespecified probability p_U, i.e., if $P(\theta_2 > \theta_1|Data) > p_U$.
- Final winner: If, after all patients have been randomized and reported results, the probability that the treatment arm is the best exceeds some prespecified probability p_0, i.e., if $P(\theta_2 > \theta_1|Data) > p_0$.
- Early futility: If at the interim look, the probability that the novel treatment arm ($k = 2$) is better than some prespecified minimally tolerable response rate θ_{min} falls below some prespecified probability p_L, i.e., if $P(\theta_2 > \theta_{min}|Data) < p_L$.
- Effective historical sample sizes (EHSS)

$$EHSS_k = min\left(max\left[n_{0k}\left(\frac{Prec(\theta_k|\mathbf{D}', \mathbf{D}_0)}{Prec(\theta_k|\mathbf{D}')} - 1\right), 0\right], n_{0k}\right),$$

where $Prec(\theta_k|\mathbf{D}')$ and $Prec(\theta_k|\mathbf{D}', \mathbf{D}_0)$ are the posterior precision of the pediatric response in each group using both the pediatric data alone and the full model (commensurate prior with adult historical data), where $\mathbf{D}'$ and $\mathbf{D}_0$ denote the interim pediatric and full adult data, respectively.

Proposed Approach

1. Fix $\theta_1 = \theta_2 = 0$, so that the null hypothesis is true (no difference in pediatric spleen volume reduction between treatment and placebo). Generate Monte Carlo pediatric observations Y_{kj}, $j = 1, \ldots, n'_k$, $k = 1, 2$, and combine with the actual adult observations Y_{0kj}, $j = 1, \ldots, n_{0k}$, $k = 1, 2$.
2. Perform the interim analysis at the data, compute the effective historical sample sizes ($EHSS_k$) for $k = 1, 2$. Monitor the $EHSS$ to ensure it is not unacceptably large.
3. Use the early winner and futility rules above to see if the trial can stop now; if so, write this down and skip Step 4.
4. Generate the remaining pediatric observations Y_{kj}, $j = n'_k + 1, \ldots, n_k$, $k = 1, 2$, and then use the "final winner" rule above to see if the trial can now choose a definite winner.
5. Repeat Steps 2–4 N_{rep} times, and estimate the Type I error of our design. Grid search on the choices of p_L, p_U, and p_0 in the stopping rules for the study designs with the desired test size (e.g., 5.0%).
6. Keep $\theta_1 = 0$ but change $\theta_2 = 20$ (or any known value meet the target efficacy), so that now the alternative hypothesis is true (clinically significant improvement in pediatric spleen volume reduction on treatment as compared to placebo). Repeat Steps 2–5 above, estimating the power of our design, and check if it is above the desired level (e.g., 80.0%).
7. Rather than fix θ_1 and θ_2 as in Steps 1 and 6, repeatedly *sample* them from a particular *design prior*, for example

$$\theta_1 \sim N(\theta_{1,des}, \sigma^2_{1,des}) \text{ and } \theta_2 \sim N(\theta_{2,des}, \sigma^2_{2,des}) \quad (1)$$

for the children where $\sigma_{1,des}$ and $\sigma_{2,des}$ are known, and set $\theta_{1,des} = 0$ and $\theta_{2,des} = \Delta$. In all cases, we repeat Steps 2–6 above again, estimating the marginal probabilities of early stopping $\hat{P}$ under our design prior, including early futility and early winner under the design prior. The numerator (benefit) of the payoff function can be defined for optimizing beneficial goals, i.e. to estimate the marginal probabilities of making correct decisions at IA, define

$$\hat{P}_1 = \frac{\#\text{ of early futility stops}}{N_{rep}} \text{ under } H_0 \text{ and } \hat{P}_2 = \frac{\#\text{ of treatment early winners}}{N_{rep}} \text{ under } H_a \ . \quad (2)$$

Use these quantities to compute the trial payoff as

$$\text{Payoff} = \frac{w\hat{P}_1 + (1-w)\hat{P}_2}{\hat{P}n' + (1-\hat{P})n}, \quad (3)$$

where $w \in (0, 1)$ is a preselected weight that trades off the two types of decisions in (2). The denominator (cost) can be explained as the expected sample size of the study design.

8. Repeat *all* the steps above (Steps 1–7) and choose the n' value that maximizes the Payoff as computed in equation (3). This n' is optimal under design prior (1), and the resulting design has correctly calibrated and acceptable Type I error and power.

Simulations

- To illustrate the method, we simulated the historical study data hypothetically from a normal distribution with the endpoint being % reduction in spleen volume. The historical study sample size are assumed to be $n_{01} + n_{02} = 50$. The simulated historical data have mean difference $\Delta_0 = 25$ and corresponding standard deviation SD = 22. We consider a current pediatric study with planned sample size of $n_1 + n_2 = 40$.

Simulation Results

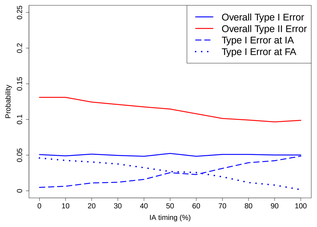

Figure: IA timing (%) vs Type I and Type II error. We calibrate Type I error (one-sided at the size of 5.0%) by grid searching the suitable p_U value and fix $p_0 = 0.975$ and $p_L = 0.25$.

Table: EHSS borrowed from historical study at interim look (Placebo/Treated)

IA time	$\Delta = 0$	$\Delta = 15$	$\Delta = 25$	$\Delta = 35$
30%	19.86 / 10.11	19.94 / 17.75	20.09 / 20.17	20.01 / 17.62
40%	19.61 / 7.55	19.74 / 16.64	19.72 / 19.88	19.63 / 16.26
50%	19.25 / 5.73	19.35 / 15.30	19.42 / 19.49	19.31 / 15.11
60%	18.59 / 4.63	18.81 / 14.15	18.81 / 18.81	18.81 / 13.83
70%	18.18 / 3.91	18.29 / 12.85	18.22 / 18.32	18.14 / 12.65
80%	17.42 / 3.47	17.44 / 11.74	17.67 / 17.65	17.48 / 11.63
90%	16.73 / 3.08	16.68 / 10.66	16.96 / 16.94	16.76 / 10.79

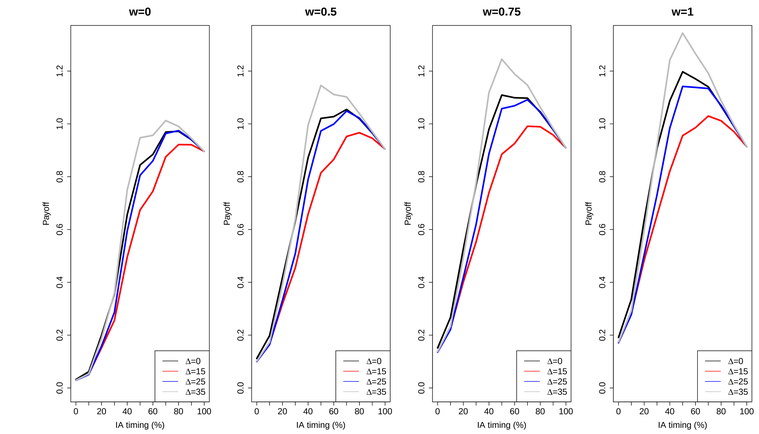

Figure: IA timing (%) vs Payoff. Each panel represents the values of payoff function with respect to different weights $w = 0, 0.5, 0.75, 1$. IA timing under different design prior with effect size: 1) $\Delta = 0$, no efficacy, 2) $\Delta = 15$, at minimal efficacy, 3) $\Delta = 25$, high efficacy, and 4) $\Delta = 35$, surprised high efficacy.

Table: Optimal IA timing under different choices of design priors and choices of weights in the payoff

effect size	$w = 0$	$w = 0.5$	$w = 0.75$	$w = 1$
0	32 (80%)	28 (70%)	20 (50%)	20 (50%)
15	32 (80%)	32 (80%)	28 (70%)	28 (70%)
25	32 (80%)	28 (70%)	28 (70%)	20 (50%)
35	28 (70%)	20 (50%)	20 (50%)	20 (50%)

Conclusions

1. We develop a Bayesian commensurate prior formulation to design a clinical trial with an optimally placed single interim look.
2. Our findings suggest optimal IA times tended to be different from the optimal time that minimized the expected sample size alone.
3. The **optimIA** `R` package provided an implementation of our approach is available at `https://github.com/wxwx1993/Bayesian_IA_Timing`.

References

[1] Brian P Hobbs, Bradley P Carlin, Sumithra J Mandrekar, and Daniel J Sargent. Hierarchical commensurate and power prior models for adaptive incorporation of historical information in clinical trials. *Biometrics*, 67(3):1047–1056, 2011.

HARVARD T.H. CHAN SCHOOL OF PUBLIC HEALTH

SANOFI GENZYME