Supplementary Online Materials for
"A Practical Bayesian Adaptive Design
Incorporating Data from Historical Controls"

Matthew A. Psioda, Mat Soukup, and Joseph G. Ibrahim

Web Appendix A: Data Simulation for Piecewise Constant Proportional Hazards Model

In this appendix we provide an explicit recipe for how to generate the observed datasets for the interim and final analysis opportunities. Let n be the total sample size, ρ be the randomization probability for the treated group, π_{cov} be the distribution for the covariates, π_{strat} be the distribution for the stratification factor, π_{enr} be the distribution for enrollment times, and π_{csr} be the distribution for censorship times. Before constructing the interim dataset $\mathbf{D}_1$ and final dataset $\mathbf{D}_2$, one must simulate the *complete* dataset $\mathbf{D}$ containing event times and censorship times for all subjects. The complete dataset contains all data, observed and unobserved, and should not be confused with the final dataset which only contains observed data. We simulate the complete data for subject i as follows:

1. Simulate the enrollment time $\tilde{r}_i$ from π_{enr}.
2. Simulate the treatment indicator z_i from Bernoulli (ρ).
3. Simulate the covariates $\mathbf{x}_i$ from π_{cov}.
4. Simulate the stratum s_i from π_{strat}.
5. Compute $\phi_i = \exp\left(z_i\gamma + \mathbf{x}_i^T\boldsymbol{\beta}\right)$ and $\theta_{i,k} = \lambda_{s_i,k}\phi_i$ for $k = 1, ..., K_{s_i}$.
6. Simulate the event time $\tilde{t}_i$ as follows:
   (i) Set $k = 1$.
   (ii) Simulate $\tilde{t}_i \sim \text{Exponential}\,(\theta_{i,k}) + t_{s_i,k-1}$.
   (iii) If $\tilde{t}_i > t_{s_i,k}$ then increment k by one and return to step (ii), otherwise terminate.
7. Simulate the censorship time $\tilde{c}_i$ from π_{csr}.
8. Calculate the complete data observation time $\tilde{y}_i = \min\left(\tilde{t}_i, \tilde{c}_i\right)$ and event indicator $\tilde{\nu}_i = I\left(\tilde{t}_i \leq \tilde{c}_i\right)$.
9. Calculate the complete data elapsed time $\tilde{e}_i = \tilde{r}_i + \tilde{y}_i$.

Next, we describe how to determine $\mathbf{D}_j$ from the complete dataset $\mathbf{D}$. To obtain $\mathbf{D}_j$, we proceed as follows:

1. Determine the time T_j of event number ν_j in the complete dataset $\mathbf{D}$.
2. Remove any subject with $\tilde{r}_i \geq T_j$.
3. For all remaining subjects,
   (i) if $\tilde{e}_i > T_j$, set $y_i = T_j - \tilde{r}_i$ and $\nu_i = 0$.
   (ii) if $\tilde{e}_i \leq T_j$, set $y_i = \tilde{y}_i$ and $\nu_i = \tilde{\nu}_i$.

Web Appendix B: Justification for Baseline Hazard Perturbation

For the example design application using the proportional hazards model that is presented in the paper, we computed the type I error rate and power over a discrete set of possible new trial parameter values. The set of parameter values was obtained by fixing the hazard ratio regression parameters at their respective historical data posterior means and *uniformly* perturbing the historical data posterior means for all the baseline hazard parameters (ranging from a 45% hazard reduction to a 45% hazard increase). By choosing w_0 so that the type I error rate and power are bounded over this discrete subspace of the parameter space, one would hope the type I error rate and power are bounded over the entire parameter space. In this appendix, we present our rationale and simulation results that illustrate that this property holds for the proportional hazards model with piecewise constant baseline hazard and discuss the generalized linear regression model case.

Rationale for focusing on a discrete subspace of the overall parameter space

For this discussion we focus on ensuring bounded control of the type I error rate. The process of ensuring bounded control of power is completely analogous. Let Ω represent the subspace of the null parameter space that satisfies $\gamma = \gamma_0$ and let $\Omega_\Delta = \{\boldsymbol{\theta} : \mathrm{E}\left[W \mid \mathbf{D}_0, a_0, \boldsymbol{\theta}\right] = \Delta, \boldsymbol{\theta} \in \Omega\}$. Thus Ω_Δ is the set of all $\boldsymbol{\theta}$ resulting in the same expected value for W where the expectation is taken with respect to $\mathbf{D}_1$ (the data at the time of the interim analysis opportunity). It is straightforward to see that $\{\Omega_\Delta : \Delta \in [0, \infty)\}$ is a partition of Ω. Let $\boldsymbol{\theta}_\Delta^+$ be the parameter value that corresponds to the maximum type I error rate among the set $\boldsymbol{\theta} \in \Omega_\Delta$. If the type I error rate is acceptable at $\boldsymbol{\theta}_\Delta^+$, then it will be acceptable for all $\boldsymbol{\theta} \in \Omega_\Delta$ (i.e., all $\boldsymbol{\theta}$ that produce the same expected degree of prior-data conflict) by nature of the adaptive design. Thus, if the design ensures that the type I error rate is acceptable for all $\boldsymbol{\theta}_\Delta^+$, this suggests that the type I error rate is acceptably bounded over the entire parameter space Ω. Assuming one could identify $\boldsymbol{\theta}_{\Delta_j}^+$ for some set $\{\Delta_j : j = 1, ..., J\}$ where Δ_1 is small, $\Delta_j < \Delta_{j+1}$, and Δ_J is larger than any reasonable value of w_0, one could ensure approximate bounded control of the type I error rate over the entire parameter space by ensuring bounded control over the set $\{\boldsymbol{\theta}_{\Delta_j}^+ : j = 1, ..., J\}$. The key challenge is to identify that set.

A simulation study justifying uniform perturbation of the baseline hazard

We devised a simulation study to illustrate that for any Δ, $\boldsymbol{\theta}_\Delta^+$ can be identified by perturbing the intercept for the regression model. In the context of the proportional hazards model with piecewise constant baseline hazard, this corresponds to a uniform perturbation of all components of the baseline hazard. Since data from the SAVOR trial are not publicly available, for this appendix we simulated control data from a hypothetical historical trial. These data and the SAS programs used for this simulation are available as noted in the paper.

We simulated subject-level data from a two-stratum proportional hazards model with exponential baseline hazard $\lambda_1 = 0.020$ in the lower risk stratum and $\lambda_2 = 0.035$ in the higher risk stratum. In terms of the likelihood for the data, this model is equivalent to a proportional hazards model with a single stratum and two baseline hazard components. We considered a hazard ratio regression model with two parameters, denoted as β_1 and β_2, corresponding to one semi-continuous covariate (values uniformly distributed between -0.6 and 0.6 with 0.2 step size) and one binary covariate (values 0 or 1 equally likely), respectively. We set $\beta_1 = \beta_2 = 0.4$. We simulated a historical trial with $n_0 = 6000$ control subjects and administratively censored the data such that $\nu_0 = 605$ events were observed. Posterior means for the baseline hazard parameters λ_1 and λ_2 were 0.018 and 0.034, respectively. Historical posterior means for the hazard ratio regression parameters β_1 and β_2 were 0.385 and 0.451, respectively. The sufficient statistics for the simulated historical dataset are given in Table 1.

Next, using this historical dataset we identified acceptable design parameters for a new trial (a hypothetical CVOT) having the same inferential goals as the example application in the paper (i.e., to rule out a hazard ratio of 1.3 for treatment versus control). For the new trial, we simulated a total enrollment of $n_1 = 5000$ subjects with a linearly increasing enrollment rate over three years. We took $\nu_1 = 306$ and $\nu_2 = 612$ as the target number of events at which the interim and final analysis opportunities would occur, respectively. This choice for ν_1 corresponds to a 50% reduction in the required number of events compared to a non-adaptive

Table 1: Sufficient Statistics for Simulated Historical Dataset

—— Covariate ——		—— Baseline Hazard ——			
Continuous	Binary	Stratum # 1		Stratum # 2	
-0.6	0	4	373.98	17	480.74
-0.4	0	13	792.09	25	826.12
-0.2	0	12	841.95	30	780.38
0.0	0	12	851.20	31	737.08
0.2	0	13	778.46	27	730.99
0.4	0	13	783.50	22	814.55
0.6	0	10	408.15	18	350.95
-0.6	1	7	413.71	13	386.24
-0.4	1	16	711.04	38	835.56
-0.2	1	23	810.80	29	640.34
0.0	1	33	883.36	36	678.12
0.2	1	19	798.65	34	670.04
0.4	1	26	768.33	41	662.64
0.6	1	18	370.69	25	338.99

trial ($\nu_1/\nu_2 = 0.50$). We identified $a_0 = 0.52$ by solving the equation $\nu_2 - \nu_1 = a_0 \cdot \nu_0$ which results in an effective number of events at the interim analysis equal to ν_2. We set the tolerability bounds for the type I error rate and power to be $(\delta_e, \delta_p) = (0.025, 0.05)$. For our simulations to determine w_0, we set the new trial hazard ratio regression parameters equal to their posterior means. For the baseline hazard parameters, we considered perturbations of the posterior means ranging from a 45% decrease to a 45% increase with a step size of 1%, uniformly perturbing both baseline hazard parameters by the same amount at a given time. The value $w_0 = 0.45$ ($e^{w_0} = 1.57$) was determined to be optimal.

Having identified the optimal decision rule for early stoppage in the trial using the discrete subspace of the overall parameter space, we then evaluated the type I error rate ($\gamma = \log 1.3$) and power ($\gamma = 0$) over the entire nuisance parameter space using a grid of values for $\boldsymbol{\psi}$ constructed by considering all possible perturbations of the historical posterior means with the most extreme perturbation to any one parameter being a 35% increase or decrease. These perturbations include perturbations to single parameters as well as collections of parameters (e.g., decrease λ_1 by 5% and decrease β_1 by 10%). Perturbed parameter values that produced a very large value of $a_0 \cdot (\ell(\hat{\boldsymbol{\psi}}|\mathbf{D}_0) - \ell(\boldsymbol{\psi}|\mathbf{D}_0))$ were discarded since they would necessarily produce a large expected value for the statistic W.

Figure 1 presents the estimated type I error rate for every value of $\boldsymbol{\theta}$ that we considered based on a design that always stops at the interim analysis opportunity as well as the optimal design that discards the prior information when $W > 0.45$. Vertical cross sections in this plot correspond to sets Ω_Δ and for any such cross section one can see that all points fall within or below the confidence bands for the type I error rate curve associated with uniform baseline hazard perturbation. Figure 2 contains analogous information for power. Note that all points fall within or above the confidence bands for the power curve associated with uniform perturbation of the baseline hazard. The confidence bands are included to give the reader some idea of the margin of error associated with 100,000 simulation studies per point estimate.

A note on generalized linear models

Uniformly perturbing the baseline hazards essentially amounts to perturbing the overall intercept for the regression model in the sense that one could parametrize the baseline hazard in terms of α_0 and α_1 such that $\lambda_1 = \exp(\alpha_0)$ and $\lambda_2 = \exp(\alpha_0 + \alpha_1)$. We are essentially perturbing α_0. Thus, in the context of a generalized linear regression model, the appropriate perturbation would be to simply perturb the model intercept while keeping all other regression parameters fixed at their historical posterior means. We also performed simulation studies based on a linear regression model (data not shown), further confirming the

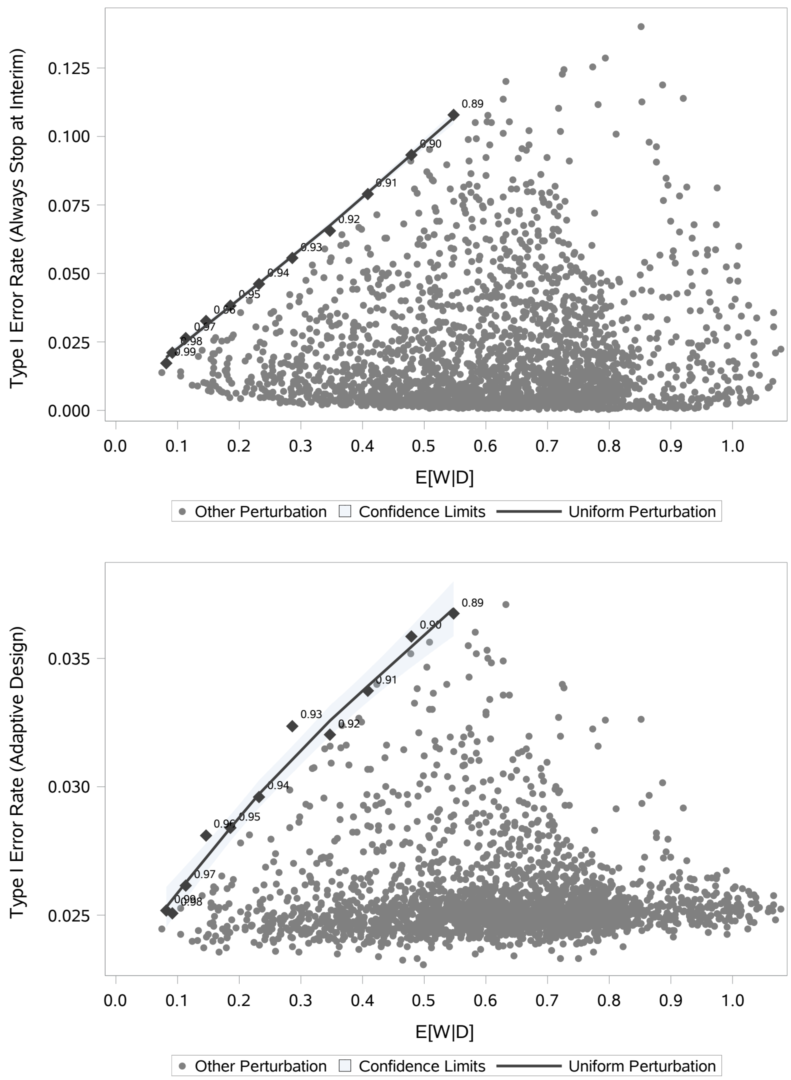

Figure 1: Plot of type I error rate by $\mathrm{E}\left[W \mid \mathbf{D}_0, a_0, \boldsymbol{\theta}\right]$. Each point corresponds to a different value of $\boldsymbol{\theta}$. Vertical cross sections correspond to different sets Ω_{Δ}. Each point estimate is based on 100,000 simulation studies. The type I error rate curve for uniform baseline hazard perturbations was smoothed using LOESS methods. The shaded region around the LOESS curve corresponds to pointwise 95% confidence bands. The degree of perturbation in the baseline hazard is annotated alongside each point estimate for uniform baseline hazard perturbations.

appropriateness of perturbing the overall model intercept to obtain the discrete subspace of the overall parameter space to be used for design simulations.

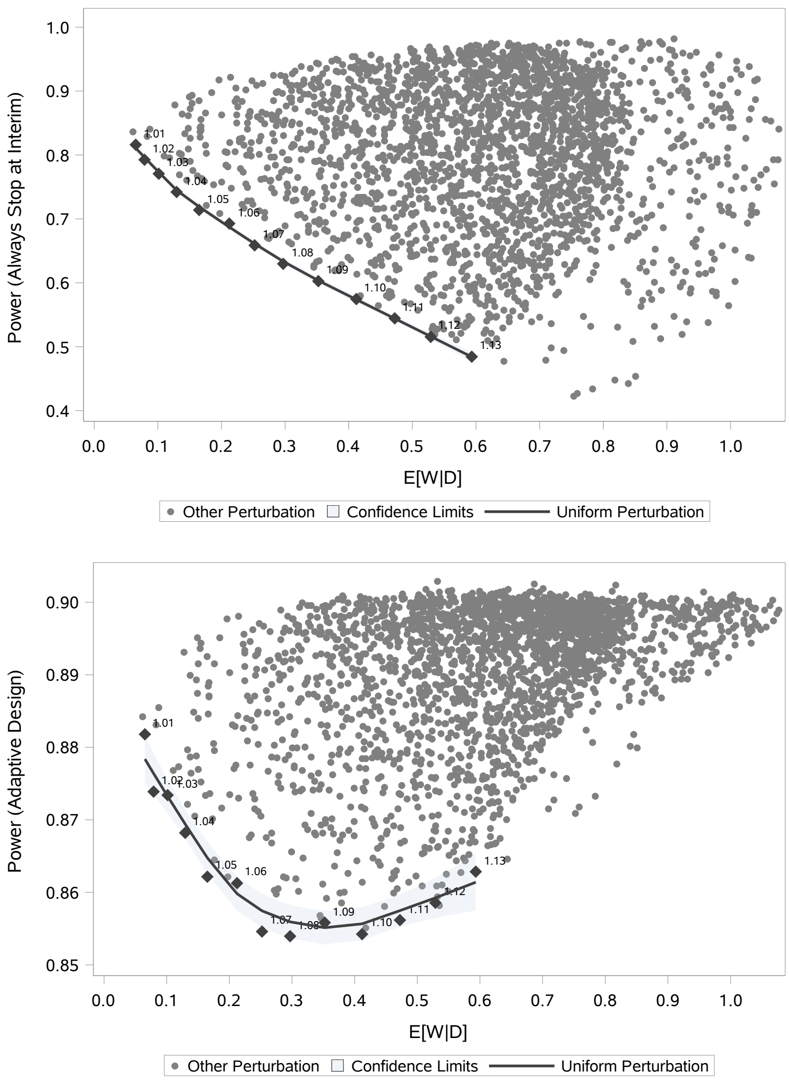

Figure 2: Plot of power by $\mathrm{E}\left[W \mid \mathbf{D}_0, a_0, \boldsymbol{\theta}\right]$. Each point corresponds to a different value of $\boldsymbol{\theta}$. Vertical cross sections correspond to different sets Ω_Δ. Each point estimate is based on 100,000 simulation studies. The power curve for uniform baseline hazard perturbations was smoothed using LOESS methods. The shaded region around the LOESS curve corresponds to pointwise 95% confidence bands. The degree of perturbation in the baseline hazard is annotated alongside each point estimate for uniform baseline hazard perturbations.