Published in final edited form as:

Clin Trials. 2013 ; 10(3): . doi:10.1177/1740774513483934.

Adaptive adjustment of the randomization ratio using historical control data

Brian P. Hobbs^a, Bradley P. Carlin^b, and Daniel J. Sargent^c

^aDepartment of Biostatistics, M.D. Anderson Cancer Center, Houston, TX, 77030, USA

^bDivision of Biostatistics, University of Minnesota, Minneapolis, Minnesota, 55455, USA

^cDepartment of Health Sciences Research, Mayo Clinic, Rochester, MN, 55905, USA

Abstract

Background—Prospective trial design often occurs in the presence of “acceptable” [1] historical control data. Typically this data is only utilized for treatment comparison in a posteriori retrospective analysis to estimate population-averaged effects in a random-effects meta-analysis.

Purpose—We propose and investigate an adaptive trial design in the context of an actual randomized controlled colorectal cancer trial. This trial, originally reported by Goldberg et al. [2], succeeded a similar trial reported by Saltz et al. [3], and used a control therapy identical to that tested (and found beneficial) in the Saltz trial.

Methods—The proposed trial implements an adaptive randomization procedure for allocating patients aimed at balancing total information (concurrent and historical) among the study arms. This is accomplished by assigning more patients to receive the novel therapy in the absence of strong evidence for heterogeneity among the concurrent and historical controls. Allocation probabilities adapt as a function of the effective historical sample size (EHSS) characterizing relative informativeness defined in the context of a piecewise exponential model for evaluating time to disease progression. Commensurate priors [4] are utilized to assess historical and concurrent heterogeneity at interim analyses and to borrow strength from the historical data in the final analysis. The adaptive trial’s frequentist properties are simulated using the actual patient-level historical control data from the Saltz trial and the actual enrollment dates for patients enrolled into the Goldberg trial.

Results—Assessing concurrent and historical heterogeneity at interim analyses and balancing total information with the adaptive randomization procedure leads to trials that on average assign more new patients to the novel treatment when the historical controls are unbiased or slightly biased compared to the concurrent controls. Large magnitudes of bias lead to approximately equal allocation of patients among the treatment arms. Using the proposed commensurate prior model to borrow strength from the historical data, after balancing total information with the adaptive randomization procedure, provides admissible estimators of the novel treatment effect with desirable bias-variance trade-offs.

Limitations—Adaptive randomization methods in general are sensitive to population drift and more suitable for trials that initiate with gradual enrollment. Balancing information among study arms in time-to-event analyses is difficult in the presence of informative right-censoring.

Conclusions—The proposed design could prove important in trials that follow recent evaluations of a control therapy. Efficient use of the historical controls is especially important in

contexts where reliance on pre-existing information is unavoidable because the control therapy is exceptionally hazardous, expensive, or the disease is rare.

Keywords

adaptive designs; Bayesian analysis; historical controls

1 Introduction

Adaptive designs of clinical trials facilitate mid-trial modifications based on interim information from internal or external sources. Recently, methods have been proposed to facilitate adaptivity for prospective modification of many trial features including randomization [5, 6]. Friedman, Furberg, and DeMets [7] broadly refer to randomization methods that facilitate mid-trial adjustments to the allocation ratios as *adaptive*. Two types of adaptive randomization (AR) procedures are commonly implemented in clinical trials. *Baseline AR* designs are used to balance the study arms with respect to prognostic factors that are available at baseline [8, 9]. By contrast, *response-adaptive* or *outcome-adaptive* designs were developed for the purpose of assigning more patients to more effective or safer treatment regimens based on interim data from an ongoing trial [10–14]. While Korn and Freidlin [15] have questioned the usefulness of designs that use outcome-adaptive randomization for this purpose in the two-armed setting, the potential remains to benefit from adaptive designs that assign more new patients to learn about newer, less studied therapies in the presence of historical controls that satisfy Pocock's six "acceptability" criteria [1, p.177]. Such designs promise to enhance efficiency when implementing controlled clinical trials that follow recent evaluations of a control therapy by facilitating more precise estimates of the treatment effect. However, assumptions of exchangeability among the historical and concurrent control data may lead to poor frequentist operating characteristics and highly biased results.

In this article we propose an adaptive analog of an actual randomized controlled colorectal cancer trial originally reported by Goldberg et al. [2] that succeeded a similar trial reported by Saltz et al. [3]. The Goldberg trial used as its control arm an identical treatment to that found superior in the Saltz trial. The proposed adaptive design is based on a randomization method that adapts as a function of the relative informativeness of the historical data for evaluating the endpoint of interest (in this case time to disease progression). Actual patient-level data from the Saltz trial (historical) and enrollment dates from the Goldberg (current) are used to simulate the proposed adaptive design's frequentist properties. Historical data is formally incorporated into the Bayesian analysis of the Goldberg trial using commensurate priors [4, 16] in the context of a piecewise exponential model [17].

In the absence of strong evidence for heterogeneity among the historical and accumulating current controls, the proposed adaptive design assigns more new patients to learn more about novel therapies, thus attempting to impose balance between concurrent and historical information. After an initial stage of equal allocation, the allocation probability is adjusted as a function of the effective historical sample size (EHSS). EHSS is defined as a function of a measure of the degree to which estimates of model parameters are "shrunk" toward their historical counterparts.

Designs such as this promise to be important in contexts where reliance on pre-existing information is unavoidable, as occurs when the control therapy is exceptionally hazardous or expensive, or when the disease is rare. There are many sub-areas of oncology that could benefit from applications of these methodologies, including rare subtypes of sarcomas, rare progressions of renal cell carcinoma such as advanced sarcomatoid, and pediatric brain

tumors such as choroid plexus carcinoma, medulloblastoma, and pontine glioma, to mention a few.

This article proceeds as follows. Section 2 describes the colorectal cancer clinical trials that motivated the research. Section 3 introduces the concept of effective historical sample size in the context of Bayesian analysis. Section 4 introduces the probability models used to evaluate time to disease progression (TTP) in the proposed, adaptive trial. In Section 5 we formulate our proposed alternative, adaptive design. Section 6 discusses the adaptive design's frequentist properties. Finally, Section 7 concludes and suggests avenues for further development.

2 Colon cancer trials

The proposed adaptive design is motivated by two successive randomized controlled colorectal cancer clinical trials originally reported by [2] and [3]. The initial trial [3] randomized 683 patients with previously untreated metastatic colorectal cancer between May 1996 and May 1998 to one of three regimens: Irinotecan alone; Irinotecan and bolus Fluorouracil plus Leucovorin (IFL); or a regimen of Fluorouracil and Leucovorin (5FU/LV) ("standard therapy"). IFL resulted in significantly longer TTP and overall survival than both Irinotecan alone and 5FU/LV, and thus became the standard of care therapy. The subsequent trial [2] compared two new (at the time) drug combinations in 795 patients with previously untreated metastatic colorectal cancer, randomized between May 1999 and April 2001. Patients in the first drug group received the current "standard therapy," the IFL regimen identical to that used in the historical study. The second group received Oxaliplatin and infused Fluorouracil plus Leucovorin (abbreviated FOLFOX), while the third group received Irinotecan and Oxaliplatin (abbreviated IROX); both of these latter two regimens were new as of the beginning of the second trial.

Historical control data from the Saltz trial appears to satisfy Pocock's acceptability criteria [1]. Less than one year elapsed in the time between the last patient enrolling into the Saltz trial and the first patient enrolling into the Goldberg trial. Both trials used identically defined IFL therapies, both trials used similar inclusion and exclusion criteria, and both used identical criteria to assess TTP. For example, inclusion required an Eastern Cooperative Oncology Group performance status (PS) of 2 or less. Exploratory data analysis suggests that the trials enrolled patients from comparable populations. The first, second, and third quartiles of the sum of the longest diameters (in cm) of up to 9 tumors at baseline are 5.0, 8.5, 12.8 in the Saltz trial and 4.7, 7.9, 12.7 in the Goldberg trial; for age they are 54, 62, 69 in the Saltz trial and 53, 61, 69 in the Goldberg trial.

Our proposed adaptive trial compares TTP between the FOLFOX and IFL regimens. The historical controls consist of patients randomized to IFL in the Saltz study; the current data consists of patients randomized to IFL or FOLFOX in the Goldberg trial. For simplicity, we omit data from the Irinotecan alone and 5FU/LV arms in the Saltz study, and the IROX arm in the Goldberg study. We consider only patients that had measurable tumors at baseline, bringing the total sample size to 643: 224 historical and 419 current observations. Among the current patients, there are 208 controls (IFL) and 211 patients treated with the new regimen (FOLFOX).

3 Effective historical sample size

Balancing information among treatment arms as data accrues in an adaptive, controlled trial, requires interim assessment of the relative informativeness of the historical data. Before formulating our adaptive trial design we briefly introduce the concept of effective historical sample size (EHSS) in the context of Bayesian analysis, a concept that will be used in

Section 5 to adjust allocation probabilities among study arms for new patients enrolled thereafter, in the context of a hierarchical model. The concept of EHSS is related to the work of Morita et al. [18], who considered effective sample sizes of parametric *prior distributions* for *non-hierarchical* models.

Let $\mathbf{y}_0$ denote the vector of outcomes for n_0 patients assigned to the control therapy in the historical trial. Similarly, let $\mathbf{y}$ denote the vector of outcomes for n patients in the current trial. Let θ_0 and θ denote analogous model parameters and $L(\theta|\mathbf{y})$ and $L_0(\theta_0|\mathbf{y}_0)$ denote the likelihood functions corresponding to the historical and current data, respectively, where θ is the parameter of interest. Let $p^*(\theta)$ denote a suitable non-informative prior distribution for θ . If the historical information is ignored, inference proceeds with respect to the posterior distribution of $\theta|\mathbf{y}$:

$$q^*(\theta|\mathbf{y}) \propto p^*(\theta)L(\theta|\mathbf{y}). \quad (1)$$

We refer to (1) as the “reference” model.

Now consider borrowing strength from the historical data. We may model a conditional relationship between θ and θ_0 by assuming a conditional prior distribution, $p(\theta|\theta_0, \eta)$, that is dependent upon a hyperparameter, η , controlling the amount of cross-study borrowing. Let $p_0^*(\theta_0)$ denote a non-informative prior distribution for θ_0 . Inference proceeds with respect to the posterior distribution of $\theta|\mathbf{y}_0, \mathbf{y}, \eta$ induced by the Bayesian model:

$$q(\theta|\mathbf{y}_0, \mathbf{y}, \eta) \propto L(\theta|\mathbf{y}) \int_{\theta_0} p(\theta|\theta_0, \eta) p_0^*(\theta_0) L_0(\theta_0|\mathbf{y}_0) d\theta_0. \quad (2)$$

Model (2) facilitates borrowing of strength via joint modeling of the historical and current data. Thus, we refer to (2) as the “joint” model.

Let $\mathcal{P}^*(\mathbf{y})$ denote the posterior precision of $\theta|\mathbf{y}$ corresponding to inference under the reference model, $\mathcal{P}^*(\mathbf{y}) = [E_{\theta|\mathbf{y}}\{\theta - E_{\theta|\mathbf{y}}(\theta)\}^2]^{-1}$. Let $\mathcal{P}(\mathbf{y}_0, \mathbf{y}, \eta)$ denote the posterior precision of $\theta|\mathbf{y}_0, \mathbf{y}, \eta$ corresponding to inference under the joint model, $\mathcal{P}(\mathbf{y}_0, \mathbf{y}, \eta) = [E_{\theta|\mathbf{y}_0, \mathbf{y}, \eta}\{\theta - E_{\theta|\mathbf{y}_0, \mathbf{y}, \eta}(\theta)\}^2]^{-1}$. If the relationship among sample size and precision is reasonably linear under the reference model, then the rate of precision per patient may be approximated by $\mathcal{P}^*(\mathbf{y})/n$. Therefore, the effective sample size of the joint model’s posterior is approximately $n \frac{\mathcal{P}(\mathbf{y}_0, \mathbf{y}, \eta)}{\mathcal{P}^*(\mathbf{y})}$, suggesting that a sensible functional relationship between joint posterior precision and EHSS follows as,

$$EHSS \approx n \left\{ \frac{\mathcal{P}(\mathbf{y}_0, \mathbf{y}, \eta)}{\mathcal{P}^*(\mathbf{y})} - 1 \right\}. \quad (3)$$

In this formulation, EHSS is approximately the effective sample size of the joint posterior minus the current sample size n . If the joint model leads to little gain in precision, then the EHSS is small. In contrast, relatively large gains in precision will produce large values of EHSS. Of note, the above formulation considers the case where θ is a scalar. Mapping relative gains in precision among multiple parameters into a single EHSS presents additional complexity. Our example adaptive colon cancer trial presents one approach for time-to-event analysis using a piecewise exponential likelihood with multiple baseline hazards.

4 Probability models

Pocock [1] proposed Bayesian models for borrowing of strength from historical controls for time-to-event data in the context of parametric exponential models. He suggested that inference should proceed with fixed “magnitudes” of the historical bias, noting that one may wish to repeat the analysis with several alternative values expressing varying degrees of trust in the historical controls. For the case of one historical study, the general commensurate prior approach [4] extends the approach of Pocock and proposes a set of prior distributions for estimating the magnitude of historical bias from the observed data in the context of fully hierarchical and empirical Bayesian analysis for parametric inference with exponential families.

4.1 Piecewise exponential likelihood

Let $\mathbf{Y}$ and $\mathbf{c}$ denote vectors containing time-to-event variables and right-censoring indicators for n patients, respectively. We use a piecewise exponential model that assumes constant baseline hazards within finite partitions of the time axis. Such a flexible model accommodates numerous shapes of the baseline hazard over the partition intervals. Following the notation of Ibrahim, Chen, and Sinha [17, p.48], we partition the time axis into $J + 1$ intervals, $(0, s_1], (s_1, s_2], \dots, (s_{J-1}, s_J], (s_J, \infty)$ where $0 < s_1 < s_2 < \dots < s_J < \infty$, and s_J denotes the end of patient follow-up, the maximum possible time that a patient could be followed during the trial. For y in the j th interval, $s_{j-1} < y \leq s_j$, the model assumes a constant baseline hazard $\lambda_j > 0$, for $j = 1, \dots, J$. Suppose $\mathbf{x}_i$ is a vector of p patient-specific baseline covariates, where $i = 1, \dots, n$ indexes patient, and $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)$ is the corresponding vector of regression coefficients. Let $\boldsymbol{\lambda}$ denote the vector of baseline hazards, $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_J)$. Within the j th time partition interval, the hazard for the i th patient is assumed to be $h(\lambda_j, \mathbf{x}_i, \boldsymbol{\beta}) = \lambda_j \exp(\mathbf{x}_i \boldsymbol{\beta})$. Let $\boldsymbol{\theta}$ denote the vector of model parameters $\boldsymbol{\theta} = (\boldsymbol{\lambda}, \boldsymbol{\beta})$. The i th patient's full contribution to the likelihood follows as,

$$\begin{aligned}
 L_Y(\boldsymbol{\theta} | Y_i) &= \prod_{j=1}^{J+1} \{ \lambda_j \exp(\mathbf{x}_i \boldsymbol{\beta}) \}^{I(s_{j-1} < Y_i \leq s_j) (1 - c_i)} \\
 &\quad \times \exp \left[-I(s_{j-1} < Y_i \leq s_j) \left\{ \lambda_j (Y_i - s_{j-1}) + \sum_{l=1}^{j-1} \lambda_l (s_l - s_{l-1}) \right\} \exp(\mathbf{x}_i \boldsymbol{\beta}) \right].
 \end{aligned} \tag{4}$$

The likelihood in (4) assumes that the ratio of hazards for two individuals is constant over time, and thus represents a proportional hazards model.

4.2 Commensurate prior model

In this subsection we formulate a commensurate prior model [4] for the scenario presented by the motivating successive colon cancer trials, where historical data are available *only* for the control group. The general commensurate prior approach facilitates estimation of the extent to which analogous parameters from distinct data sources have similar (“commensurate”) posteriors through the specification of a particular hierarchical model. Commensurate prior distributions center model parameters defined in the context of the likelihood for the current data at their historical counterparts. Hobbs, Sargent, and Carlin [4] use sparsity-inducing spike-and-slab hyperpriors for the dispersion parameters to favor the concurrent information in the presence of heterogeneity.

Let $\mathbf{Y}_0$ and $\mathbf{c}_0$ denote vectors of length n_0 containing time-to-event variables and right-censoring indicators for patients assigned to the current control therapy in the historical study. Let $\mathbf{Y}$ and $\mathbf{c}$ denote data from n patients in the current trial. Furthermore, let d_i denote an indicator of the novel treatment for the i th patient in the current trial, $i = 1, \dots, n$. In this context, the hazard for a historical patient in the j th time partition interval is identical to the j th historical baseline hazard, $h(\lambda_{0,j}) = \lambda_{0,j}$. For the i th patient in the current trial, the j th baseline hazard, λ_j , is modified by novel treatment status, d_i : $h(\lambda_j, d_i, \xi) = \lambda_j \exp(d_i \xi)$, where ξ is the log acceleration factor corresponding to the novel treatment. Both historical and current baseline hazard parameters are defined with respect to a single partition of the time axis characterized by boundary points $0 < s_1 < s_2 < \dots < s_J < \infty$.

To borrow strength from the historical controls, the commensurate prior approach can be applied by assuming a normal prior for the j th log baseline hazard, $\log(\lambda_j)$, centered at its historical counterpart, $\log(\lambda_{0,j})$, with precision τ_j : $\log(\lambda_j) \sim N(\log(\lambda_{0,j}), 1/\tau_j)$. Given no additional prior information about the historical baseline hazards or novel treatment effect, inference may proceed with non-informative Gaussian prior distributions for the $\log(\lambda_{0,j})$ s and ξ . Let $\mathcal{D}$ and $\mathcal{D}_0$ denote the current and historical data: $\mathcal{D} = (\mathbf{Y}, \mathbf{d}, \mathbf{c})$ and $\mathcal{D}_0 = (\mathbf{Y}_0, \mathbf{c}_0)$, respectively. Let $\boldsymbol{\theta}$ denote the parameter vector $\boldsymbol{\theta} = \{\xi, \log(\lambda_1), \dots, \log(\lambda_J)\}$, let $\boldsymbol{\theta}_0$ denote the parameter vector $\boldsymbol{\theta}_0 = \{\log(\lambda_{0,1}), \dots, \log(\lambda_{0,J})\}$, and let $\boldsymbol{\tau}$ denote vector $\boldsymbol{\tau} = (\tau_1, \dots, \tau_J)$. Following from (4), the joint posterior distribution of $\boldsymbol{\theta} \boldsymbol{\tau}$ is proportional to

$$q(\boldsymbol{\theta} | \boldsymbol{\tau}, \mathcal{D}, \mathcal{D}_0) \propto N(\xi | 0, 100) \prod_{i=1}^n L_Y(\boldsymbol{\theta} | Y_i) \int_{\boldsymbol{\theta}_0} \prod_{j=1}^J N(\log(\lambda_j) | \log(\lambda_{0,j}), 1/\tau_j) \times N(\log(\lambda_{0,j}) | 0, 100) \prod_k^{n_0} L_{Y_0}(\boldsymbol{\theta}_0 | Y_{0,k}) d\boldsymbol{\theta}_0, \quad (5)$$

where $L_{Y_0}(\boldsymbol{\theta}_0 | \cdot)$ denotes the likelihood function corresponding to the piecewise exponential model for the historical data.

Following Hobbs et al. [4], we assume “spike and slab” prior distributions [19] for the τ_j s. This sparsity-inducing prior distribution is locally uniform between two limits, $0 < s_l < s_u$ except for a bit of probability mass concentrated at a point $\mathcal{K} > s_u$:

$$Pr(\tau_j < u) = p_0 \{(u - s_l) / (s_u - s_l)\}, s_l \leq u \leq s_u \text{ and } Pr(\tau_j > s_u) = 1 - p_0. \quad (6)$$

The additional hyperparameter, p_0 , denotes the prior probability of the existence within the slab, i.e., $s_l \leq \tau_j \leq s_u$. Given s_l and $\mathcal{K}$, smaller values of s_u impose more borrowing of strength from the historical information, requiring more evidence for heterogeneity to overcome the prior probability mass assigned to the spike, $1 - p_0$. The aforementioned authors [4] show that this prior, when properly calibrated, leads to desirable bias-variance trade-offs for estimating concurrent effects in the context of commensurate prior models for exponential families. The results of our simulation study discussed in Section 6 illustrate the advantages of the spike and slab commensurate prior model in this context. Markov chain Monte Carlo (MCMC) methods can be used to sample from the joint posterior (5).

5 Adaptive trial design

In this section we formulate our proposed alternative design of the Goldberg trial following completion of the Saltz trial, which uses permuted-block randomization in both non-adaptive and adaptive phases. During the initial, non-adaptive phase, new patients are allocated equally between the control (IFL) and novel (FOLFOX) treatments until a targeted number of events selected to ensure that sufficient information has accrued to assess heterogeneity among the concurrent and historical controls. Thereafter, the treatment

allocation probability is adjusted as a function of the EHSS and the numbers of patients assigned to each regimen. We use the piecewise exponential commensurate prior model (5) to jointly model the concurrent and historical data at each interim analysis. Given large EHSS, the adaptive design will randomize more new patients to the newer, less studied therapy.

5.1 Historical data

We use likelihood inference based on (4) to select the time axis partition for the full Bayesian analysis of the historical control data alone. The Akaike information criterion optimal partition contains two intervals with boundary point at $s = 243$ days. Thus, the piecewise exponential model contains two parameters characterizing baseline hazards within intervals $[0, 243)$ and $[243, \max(Y_0)]$, respectively. Figure 1 contains the associated TTP curves. The dashed line represents the Kaplan-Meier curve, while dotted lines correspond to associated 95% log-transformed [20, see e.g. p.105] pointwise confidence intervals. Results for the piecewise exponential analysis are plotted with a solid line. Posterior summaries corresponding to the Bayesian analysis are provided in Table 1. The posterior mean of median survival is approximately 205 days.

5.2 New trial

The Goldberg trial actually enrolled a total of $n = 419$ patients with measurable tumors assigned to either IFL (control) or FOLFOX (novel treatment) regimens. Figure 2 plots actual enrollment over time. Each enrollment is characterized by a dot representing the trial's cumulative sample size (y-axis) in calendar time (x-axis). The first patient was enrolled on May 20, 1999, the last patient was enrolled on April 25, 2001. We use the Goldberg et al. [2] article's submission date as the time of final analysis, which occurred approximately 28 months after close of enrollment.

Our proposed trial consists of two phases. During the first, non-adaptive phase, patients are randomized 1:1 to the treatment regimens until O_1 events are observed. In order to ensure periodic balance in the numbers of patients assigned to each regimen, permuted-block randomization [21] is used with blocks of size 2 patients. An initial interim analysis to assess historical and concurrent control heterogeneity and compute EHSS will occur at that time. An adaptive, permuted-block randomization procedure that is a function of EHSS will be used to assign new patients to treatment arms thereafter in blocks of size B patients. EHSS will be re-assessed after every B th enrollment. The final analysis, using the joint model, occurs 51 months after the trial's initiation. O_1 needs to be large enough so that sufficient information to assess control heterogeneity has accrued, yet small enough to justify the adaptive treatment allocation. In practice, O_1 and B should be selected in the context of a comprehensive assessment of the trial's operating characteristics in the presence of the historical data and expected patient enrollment. Our simulation study, discussed in Section 6, considers frequentist properties of the adaptive design under 3 different values of O_1 .

5.2.1 Data structure—Let $T = 1552$ denote the time of the trial's final analysis in days. Let t denote "trial time," $0 < t < T$, and let e_i , $0 < e_i < 706$, represent the time at which the i th patient enrolls. Let $Y_i(t)$ denote the random variable containing the value of the i th patient's TTP process at time t . Note that $Y_i(t)$ is recorded from e_i , so that $0 < Y_i(t) < T - e_i$. Let $\mathbf{Y}(t) = \{Y_1(t), \dots, Y_{n(t)}(t)\}$ denote the vector of values of the TTP processes for the $n(t)$ patients enrolled prior to time t , $n(t) = \sum_{i=1}^n I(e_i < t)$, where I is the indicator function. Let $c_i(t)$ denote an indicator of the i th patient's right-censoring status at time t , and let $\mathbf{c}(t) = \{c_1(t), \dots, c_{n(t)}(t)\}$ denote the vector of values of censoring processes for the $n(t)$ patients enrolled

prior to time t . Finally, let d_i denote an treatment indicator for the i th patient, with 0 indicating IFL and 1 FOLFOX, and let $\mathbf{d}(t) = \{d_1, \dots, d_{n(t)}\}$. Denote concurrent data at time t by $\mathcal{D}(t) = \{\mathbf{Y}(t), \mathbf{c}(t), \mathbf{d}(t)\}$, and historical data by $\mathcal{D}_0 = (\mathbf{Y}_0, \mathbf{c}_0)$.

5.2.2 Posterior inference—The piecewise exponential models used for analysis at time t contain two baseline hazard parameters, λ_1 and λ_2 , corresponding to intervals $[0, m(t))$ and $[m(t), T]$, where $m(t)$ denotes the median TTP among the concurrent events observed by time t . This balances the available concurrent information between the two partition intervals at each analysis. Following the notation of Section 4, let ξ denote the FOLFOX effect, and denote the current and historical parameter vectors by $\boldsymbol{\theta}$ and $\boldsymbol{\theta}_0$, where $\boldsymbol{\theta} = \{\xi, \log(\lambda_1), \log(\lambda_2)\}$, and $\boldsymbol{\theta}_0 = \{\log(\lambda_{0,1}), \log(\lambda_{0,2})\}$, respectively.

The reference model (1) in this context consists of the Bayesian piecewise exponential model that estimates $\boldsymbol{\theta}$ in the absence of the historical controls. It assumes non-informative Gaussian prior distributions (with small precision $w = 0.01$) for the FOLFOX effect, ξ , and the log baseline hazards, $\log(\lambda_1)$ and $\log(\lambda_2)$. At trial time t , the posterior distribution of $\boldsymbol{\theta} | \mathcal{D}(t)$ under the reference model follows from (4) as

$$q^*(\boldsymbol{\theta} | \mathcal{D}(t)) \propto N(\xi | 0, 1/w) \prod_{j=1}^2 N(\log(\lambda_j) | 0, 1/w) \prod_{i=1}^{n(t)} L_Y(\boldsymbol{\theta} | Y_i(t)). \quad (7)$$

The joint model (2) consists of a commensurate prior model (5) that borrows strength from the historical control data for estimating $\boldsymbol{\theta}$. Following from (4), (5), and (6), the posterior distribution of $\boldsymbol{\theta} | \mathcal{D}(t), \mathcal{D}_0$ under the joint model at trial time t is

$$\begin{aligned} q(\boldsymbol{\theta} | \mathcal{D}(t), \mathcal{D}_0) &\propto N(\xi | 0, 1/w) \prod_{i=1}^{n(t)} L_Y(\boldsymbol{\theta} | Y_i(t)) \\ &\times \int_{\boldsymbol{\tau}} \int_{\boldsymbol{\theta}_0} \prod_{j=1}^2 N(\log(\lambda_j) | \log(\lambda_{0,j}), 1/\tau_j) N(\log(\lambda_{0,j}) | 0, 1/w) p_{\boldsymbol{\tau}}(\boldsymbol{\tau}_j) \prod_k^{n_0} L_{Y_0}(\boldsymbol{\theta}_0 | Y_{0,k}) d\boldsymbol{\theta}_0 d\boldsymbol{\tau}, \end{aligned} \quad (8)$$

where $p_{\boldsymbol{\tau}}(u)$ denotes the spike-and-slab prior distribution in (6) evaluated at u .

5.2.3 Effective number of historical controls and adaptive randomization—We compute EHSS at interim analyses, with use thereafter to guide the randomization procedure. In the Appendix we investigate the linear association between number of events and posterior precision for $\log(\lambda_1) | \mathcal{D}(t)$ in the context of the reference model (7) in the presence of the historical data. EHSS in this context maps the joint model's (8) relative gain in precision for estimating the set of current model parameters, $\boldsymbol{\lambda}$, that assume commensurate prior distributions (defined conditionally upon analogous historical parameters), to a number of *additional* events effectively observed for control. The first interim assessment of EHSS will occur at the trial time t_1 at which at least O_1 events have been observed,

$t_1 = \arg \min_t \left[\sum_{i=1}^{n(t)} \{1 - c_i(t)\} = O_1 \right]$. EHSS will then be re-assessed after every B additional patients enroll. Thus, the h th re-assessment of EHSS will occur at trial time, t_h , where $t_h = \arg \min_t \{n(t) = n(t_{h-1}) + B\}$, for $h = 2, 3, \dots$. Let O_h denote the number of events

that have occurred by trial time t_h , $O_h = \sum_{i=1}^{n(t_h)} \{1 - c_i(t_h)\}$. Following Section 3, let $\mathcal{D}_j^* \{ \mathcal{D}(t_h) \}$ denote the posterior precision of $\log(\lambda_j) | \mathcal{D}(t_h)$ at the h th interim analysis under the reference model (7). Similarly, let $\mathcal{P}_j \{ \mathcal{D}(t_h), \mathcal{D}_0 \}$ denote the posterior precision of $\log(\lambda_j) | \mathcal{D}(t_h), \mathcal{D}_0$ at the h th interim analysis under the joint model (8). Recall that we have defined

the time axis partition for analysis at time t_h such that approximately $O_h/2$ of the observed events occur within each time axis interval. The effective historical sample size at the h th interim analysis follows from (3) as the sum of parameter-specific EHSSs (each based upon $O_h/2$ events) for the current baseline hazards,

$$EHSS_h = \max \left\{ \sum_{j=1}^2 \left(\frac{O_h}{2} \left[\frac{\mathcal{P}_j\{\mathcal{D}(t_h), \mathcal{D}_0\}}{\mathcal{P}_j^*\{\mathcal{D}(t_h)\}} - 1 \right] \right), 0 \right\}. \quad (9)$$

We introduced EHSS (3) in the context of a Bayesian analysis with fixed hyperparameter η . Because we now average over the uncertainty for estimating hyperparameter τ_j , sufficient heterogeneity may result in less precision for the joint posterior inference,

$\mathcal{P}_j\{\mathcal{D}(t_h), \mathcal{D}_0\} < \mathcal{P}_j^*\{\mathcal{D}(t_h)\}$. Thus, we define $EHSS_h$ to be non-negative valued.

After the h th interim analysis, we seek to obtain balance between the numbers of patients assigned to each regimen at the end of the trial relative to $EHSS_h$ in the context of a permuted-block randomization procedure with blocks of size B . Let $n^{FOX}(t)$ and $n^{IFL}(t)$ denote the numbers of patients assigned to FOLFOX and IFL at trial time t , respectively. Let $R(t)$ denote the number of remaining patients at trial time t , $R(t) = n - n^{IFL}(t) - n^{FOX}(t)$. Following the h th interim analysis, approximately $\pi_h B$ of the next B patients to enroll into the trial will be randomly assigned to FOLFOX and $(1 - \pi_h)B$ assigned to IFL, where π_h is defined to satisfy the following equation

$$n^{FOX}(t_h) + \pi_h R(t_h) = EHSS_h + n^{IFL}(t_h) + (1 - \pi_h)R(t_h). \quad (10)$$

Therefore, π_h is

$$\pi_h = \min \left\{ \frac{1}{2} \left(\frac{EHSS_h + n^{IFL}(t_h) - n^{FOX}(t_h)}{R(t_h)} + 1 \right), 1 \right\}. \quad (11)$$

The design can accommodate alternative default allocation ratios (besides 1 : 1) through modification of (10) to reflect the desired sample size imbalance at the trial's completion.

6 Simulations

We use simulation to evaluate the frequentist properties of the proposed alternative, adaptive design of the Goldberg trial. The simulation uses the actual patient-level data from $n_0 = 224$ historical controls assigned to IFL in the Saltz trial. Patient accrual in the simulated trials follows the actual enrollment dates (shown in Figure 2) for $n = 419$ patients randomized to FOLFOX or IFL during the Goldberg trial. Following Subsection 5.2, each simulated trial proceeds as follows. Patients are randomized 1:1 to IFL and FOLFOX using permuted-block randomization with blocks of size 2 until O_1 total events are observed at time t_1 . Thereafter, the probability of assigning a new patient to FOLFOX (11) adjusts in blocks of size $B = 30$ as a function of EHSS and the numbers of patients already assigned to each treatment arm. The joint commensurate prior model (8) used to compute EHSS at each interim analysis assumes spike and slab priors (6) for $\tau_j, j = 1, 2$, with the following hyperparameters: $s_i = 0.1$, $s_u = 2$, $\kappa = 5000$, and $p_0 = 0.5$. A final analysis occurs at time T , 51 months following initial enrollment using the joint commensurate prior model with $s_u = 5$ and $p_0 = 0.9$. The approach provides for more borrowing of strength from the historical data for computing

EHSS, while facilitating slightly less borrowing when estimating the FOLFOX treatment effect, ξ , at the final analysis.

Frequentist properties are considered under varying magnitudes of assumed historical bias. Let $\theta = \{\xi, \log(\lambda_1), \log(\lambda_2)\}$ denote a set of fixed parameters characterizing a true state of the current model, where the baseline hazards, λ_1 and λ_2 , are defined under $[0, 243)$ and $[243, T]$ (the AIC-optimal time axis partition for inference on the historical data alone). TTP for current patients are generated from θ by shifting the log baseline hazards, $\log(\lambda)$, from their historical estimates in Table 1 by $\Delta 100\%$, where $\lambda(\Delta) = (\exp\{-5.689(1 - \Delta)\}, \exp\{-4.858(1 - \Delta)\})$. The historical data is assumed to be unbiased for the case where $\Delta = 0$. Without loss of generality, we fix $\xi = 0$.

Figure 3 plots results for mean EHSS and mean proportion of patients (out of $n = 419$) assigned to FOLFOX as functions of trial time. Each column corresponds to a different value of $\Delta 100\%$, indicating relative magnitudes of assumed historical bias. Each graph in the bottom row of plots shows results for designs that initiate adaptive randomization after $O_1 = 60, 90$, and 120 observed events. Results are shown in the top row of plots for $O_1 = 60$. The plots show that increasing (decreasing) the baseline hazards leads to designs for which the initial computation of EHSS (determined by O_1), and initiation of the AR procedure thereafter, occurs earlier (later), on average. The top row of graphs show us that the actual historical data is most commensurate, on average, with the concurrent data when $\Delta 100\% = 0\%$ (no change) and 5% (5% increase). The joint model borrows more strength from the historical data for these two scenarios, which leads to larger initial EHSS, and increasing trends over time as more events facilitate larger relative gains in posterior precision. This induces monotonically increasing trends over time in the mean proportion of patients assigned to FOLFOX. Note that because increasing the baseline line hazards yields fewer right-censored observations, EHSS is ultimately slightly larger for $\Delta 100\% = 5\%$ when compared to $= 0\%$. In addition, AR initiates earlier for $\Delta 100\% = 5\%$. Thus, slightly more patients, on average, are assigned to FOLFOX when $\Delta 100\% = 5\%$. In contrast, the graphs show smaller initial EHSS with decreasing trends over time for the scenarios where $\Delta 100\% = -10\%, -5\%$, and 10% , leading to less allocation to FOLFOX, on average. For these three scenarios, we see that the trends over time in mean proportion of patients assigned to FOLFOX for the three choices of O_1 ultimately converge. This suggests that our adaptive design is robust among reasonable choices in the initial timing of the interim assessments of EHSS.

Because our simulation study uses the actual historical data (as opposed to generating data from a piecewise exponential model with parameters fixed at their estimates in Table 1), the results lack symmetry around the case where the historical data is assumed to be unbiased, $\Delta = 0$. The actual historical information appears to be most commensurate with accumulating concurrent information when current data is generated from the piecewise exponential model with $0 < \Delta < 0.05$.

In addition, we evaluated the adaptive design's frequentist properties for point estimation of the FOLFOX effect, ξ , using the posterior mean at the final analysis, $E_{\xi|\mathcal{D}(T), \mathcal{D}_0}(\xi)$. Specifically, we simulated expected bias and risk under squared error loss with respect to the conditional distribution of the data, $\mathcal{D}(T)$, given Δ and the actual historical data, $\mathcal{D}_0$:

$$E_{\mathcal{D}(T)|\Delta, \mathcal{D}_0} \{E_{\xi|\mathcal{D}(T), \mathcal{D}_0}(\xi) - \tilde{\xi}\}, \quad (12)$$

and $E_{\mathcal{D}(T)|\Delta, \mathcal{D}_0} \left[\{E_{\xi|\mathcal{D}(T), \mathcal{D}_0}(\xi) - \tilde{\xi}\}^2 \right].$

Results for the adaptive designs are compared to a fixed 1:1 allocation design with analysis under the associated reference (“no borrowing”) model and a “pooled” model, which simply pools the historical and current controls and randomizes new patients to FOLFOX with probability 0.77, so that the final analysis has approximately 321 total patients (historical and current) assigned to each regimen.

Figure 4 presents expected bias and risk under SEL for estimating ξ as functions of $\Delta 100\%$. Each column corresponds to a different value of O_1 . A total of 1000 replicated trials were simulated for each of 7 scenarios corresponding to values of $\Delta 100\%$ ranging from -15% to 15% . Each approach leads to an admissible estimator. Under no borrowing, $\mathcal{D}_0$ is ignored. Thus, the posterior mean of $\xi | \mathcal{D}(T)$ is unbiased. Increasing the baseline hazards results in fewer right-censored patients, which decreases risk. In contrast, the pooled design offers maximal variance reduction, and thus is associated with the largest reductions in risk at $\Delta 100\% = 0\%$. However, this approach precludes learning about intra-control heterogeneity, which leads to prohibitively biased estimators with sharply, uniformly increasing risk that exceeds no borrowing for $|\Delta 100\%| > 2\%$. The adaptive design offers a compromise. For $|\Delta 100\%| < 3.5\%$, it leads to estimators that dominate no borrowing for risk. Bias is attenuated for large values of $|\Delta 100\%|$, leading to near replication of no borrowing risk for $|\Delta 100\%| > 10\%$. No borrowing dominates for intermediate intra-control effects, $(-10\% < \Delta 100\% < -3.5\%) \cup (3.5\% < \Delta 100\% < 10\%)$. However, adaptive borrowing leads to considerable reductions in risk and bias when compared to naive pooling over this portion of the parameter space. The hierarchical model’s spike-and-slab hyperparameters can be adjusted to modulate risk tolerance in the presence of intermediate intra-control effects. In addition, the simulations suggest the estimator is relatively robust to the choice of O_1 .

7 Discussion

In this article we proposed an adaptive trial design that implements a randomization procedure for allocating patients aimed at balancing total information (concurrent and historical) among the study arms by assigning more patients to receive the novel therapy in the absence of evidence for heterogeneity among the concurrent and historical controls, and yields an estimator of the novel treatment effect that obtains desirable frequentist properties. The proposed design promises to enhance efficiency in the prospective evaluation of new therapies involving controlled clinical trials that follow recent evaluations of the control therapy in the presence of little intra-control heterogeneity without risking highly biased estimation in the presence of substantial intra-control heterogeneity.

The proposed adaptive trial was considered in a “non-sequential” setting involving a single treatment comparison at the end of the trial. That is, the design, while sequential in assessing evidence for heterogeneity among the historical and concurrent controls, did not incorporate decision rules for early stopping for efficacy or futility. Extending the proposed adaptive trial to allow for earlier stopping is an area for potential future development. Indeed, other areas for future development are many, since they include any setting where we wish to adjust trial enrollment adaptively based on an updated estimate of how much strength may be sensibly borrowed from external data sources. For instance, when historical information is also available on the new treatment, we may have two commensurability calculations, each having their own impact on the randomization ratio (11). Alternatively, in a one-arm device safety trial we may use commensurate priors to decide when to stop accrual based on the predictive probability of a favorable end result given the current estimate of the total effective sample size. While our approach is exclusively Bayesian, Freedman and Spiegelhalter [22] have considered clinical trials that use Bayesian methods for design and frequentist methods for final analysis.

Acknowledgments

The work of the first author was supported in part by the University of Texas M.D. Anderson's Cancer Center Support Grant NIH P30 CA016672. The work of the second author was supported in part by National Cancer Institute grant 1-R01-CA157458-01A1. The work of the third author was supported in part by NCI grant CA25224.

References

1. Pocock SJ. The combination of randomized and historical controls in clinical trials. *Journal of Chronic Diseases*. 1976; 29:175–188. [PubMed: 770493]
2. Goldberg RM, Sargent DJ, Morton RF, Fuchs CS, Ramanathan RK, Williamson SK, et al. A randomized controlled trial of fluorouracil plus leucovorin, irinotecan, and oxaliplatin combinations in patients with previously untreated metastatic colorectal cancer. *Journal of Clinical Oncology*. 2004; 22:23–30. [PubMed: 14665611]
3. Saltz LB, Cox JV, Blanke C, et al. Irinotecan plus fluorouracil and leucovorin for metastatic colorectal cancer. *The New England Journal of Medicine*. 2000; 343:905–914. [PubMed: 11006366]
4. Hobbs BP, Sargent DJ, Carlin BP. Commensurate priors for incorporating historical information in clinical trials using general and generalized linear models. *Bayesian Analysis*. 2012; 7:639–674.
5. Bretz F, Koenig F, Brannath W, Glimm E, Posch M. Adaptive designs for confirmatory clinical trials. *Statistics in Medicine*. 2009; 28:1181–1217. [PubMed: 19206095]
6. Berry, SM.; Carlin, BP.; Lee, JJ.; Müller, P. *Bayesian Adaptive Methods for Clinical Trials*. Boca Raton, FL: Chapman and Hall/CRC Press; 2010.
7. Friedman, L.; Furberg, C.; DeMets, DL. *Fundamentals of Clinical Trials*. 3. New York: Springer-Verlag; 1998.
8. Pocock SJ, Simon R. Sequential treatment assignment with balancing for prognostic factors in the controlled clinical trial. *Biometrics*. 1975; 31:103–115. [PubMed: 1100130]
9. Anderson, G.; LeBlanc, M.; Liu, PY.; Crowley, J. On use of covariates in randomization and analysis of clinical trials. In: Crowley, J.; Hoering, A., editors. *Handbook of statistics in clinical oncology*. 3. Boca Raton, FL: Chapman and Hall/CRC Press; 2011.
10. Berry DA, Eick S. Adaptive assignment versus balanced randomization in clinical trials: a decision analysis. *Statistics in Medicine*. 1995; 14:231–246. [PubMed: 7724909]
11. Thall PF, Wathen JK. Practical Bayesian adaptive randomisation in clinical trials. *European Journal of Cancer*. 2007; 43:859–866. [PubMed: 17306975]
12. Zhou X, Liu S, Kim ES, Herbst RS, Lee JJ. Bayesian adaptive design for targeted therapy development in lung cancer – a step toward personalized medicine. *Clinical Trials*. 2008; 5:181–193. [PubMed: 18559407]
13. Huang X, Ning J, Li Y, Estey E, Issa JP, Berry DA. Using short-term response information to facilitate adaptive randomization for survival clinical trials. *Statistics in Medicine*. 2009; 28:1680–1689. [PubMed: 19326367]
14. Lee JJ, Gu X, Liu S. Bayesian adaptive randomization designs for targeted agent development. *Clinical Trials*. 2010; 7:584–597. [PubMed: 20571130]
15. Korn EL, Freidlin B. Outcome-adaptive randomization: is it useful? *Journal of Clinical Oncology*. 2011; 29:771–776. [PubMed: 21172882]
16. Hobbs BP, Carlin BP, Mandekar SJ, Sargent DJ. Hierarchical commensurate and power prior models for adaptive incorporation of historical information in clinical trials. *Biometrics*. 2011; 67:1047–1056. [PubMed: 21361892]
17. Ibrahim, JG.; Chen, MH.; Sinha, D. *Bayesian Survival Analysis*. New York: Springer-Verlag; 2001.
18. Morita S, Thall PF, Müller P. Determining the effective sample size of a parametric prior. *Biometrics*. 2008; 64:595–602. [PubMed: 17764481]
19. Mitchell TJ, Beauchamp JJ. Bayesian variable selection in linear regression. *Journal of the American Statistical Association*. 1988; 83:1023–1032.

20. Klein, JP.; Moeschberger, ML. Survival Analysis: Techniques for Censored and Truncated Data. 2. New York: Springer-Verlag; 2003.
21. Matts JP, Lachin JM. Properties of permuted-block randomization in clinical trials. Controlled Clinical Trials. 1988; 9:327–344. [PubMed: 3203524]
22. Freedman LS, Spiegelhalter DJ. The assessment of subjective opinion and its use in relation to stopping rules for clinical trials. The Statistician. 1983; 32:153–160.

Appendix

Here we investigate the association between number of events and posterior precision for $\log(\lambda_1) | \mathcal{D}(t)$ in the context of the reference model (7). Specifically, we simulated 7,000 datasets from a piecewise exponential probability distribution. For each simulated dataset, $\mathcal{D}$, the sample size, n , and right-censoring probability, π^{cens} , were randomly generated from uniform distributions covering broad ranges of the sample spaces of relevance: $n \sim \text{Uniform}(50, 600)$ and $\pi^{cens} \sim \text{Uniform}(0.1, 0.8)$. Time-to-event observations were generated for each of the $\tilde{n}$ patients from θ containing baseline hazards fixed at their corresponding historical estimates in Table 1, $\log(\lambda_1) = -5.689$ and $\log(\lambda_2) = -4.858$, and treatment effect $\xi \sim 0$. For each dataset we recorded the number of events observed in the initial partition interval, and fit the reference model (7) to obtain the associated posterior precision of $\log(\lambda_1) | \mathcal{D}$. The results in Figure 5 reveal a strong linear association (Pearson correlation of 0.996) among posterior precision and number of observed events for the considered portion of the parameter space.

Figure 1.

Time to disease progression for patients assigned to IFL in the Saltz trial: Kaplan-Meier curve (dashed), piecewise exponential curve (solid), 95% log-transformed pointwise confidence intervals (dotted). Right-censored observations are marked by +.

Figure 2. Enrollment over time for patients randomized to the FOLFOX or IFL regimens in the Goldberg trial. The y-axis represents cumulative sample size. The x-axis represents calendar time.

Figure 3.

Simulated mean EHSS (top row) and mean proportion of patients assigned to FOLFOX (bottom row) as functions of trial time corresponding to designs that initiate adaptive randomization after 60 (solid), 90 (dashed), and 120 (dotted) observed events. Results are shown for five magnitudes of assumed historical bias, indicating the amounts of relative change in the concurrent log baseline hazards, $\log(\lambda_1)$ and $\log(\lambda_2)$, from their historical estimates in Table 1.

Figure 4. Expected bias (top row) and risk under squared error loss (bottom row) for estimating the FOLFOX effect, ξ , as functions of percent change in the log baseline hazards from their historical estimates in Table 1. Results are shown for three values of the number of observed events required before initiating the adaptive randomization procedure.

Figure 5. Scatterplot of simulation results. Each dot depicts the number of events (y-axis) as a function of posterior precision of $\log(\lambda_1)$ | $\mathcal{D}(t)$ (x-axis) under the reference model (7).

Table 1

Posterior means and standard deviations derived from the piecewise exponential analysis of time to disease progression for patients randomized to IFL in the Saltz trial, $n_0 = 224$.

	mean	sd
log baseline hazard 1	-5.689	0.094
log baseline hazard 2	-4.858	0.133