Bayesian Clinical Trial Methodologies

Brian P. Hobbs

*bphobbs@mdanderson.org

Department of Biostatistics, M.D. Anderson Cancer Center

February 7, 2012

Outline

1. Thesis Research
   - prospective incorporation of historical data from a small number of historical studies
2. Platform-based designs for effective screening multiple agents in Phase II

Bayes Theorem

- The general Bayesian approach involves combining "prior knowledge" about the distributions of the unknown model parameters with observed data to provide direct estimation of "evidence" for the parameter of interest
- Prior distributions *may* summarize our preexisting understanding or beliefs regarding unknown model parameters $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_K)$
- Inference is conducted on the posterior distribution of $\boldsymbol{\theta}$ given the observed data $\mathbf{y} = (y_1, \ldots, y_N)'$, via the Bayes theorem

$$p(\boldsymbol{\theta}|\mathbf{y}) = \frac{p(\boldsymbol{\theta}, \mathbf{y})}{p(\mathbf{y})} = \frac{f(\mathbf{y}|\boldsymbol{\theta})p(\boldsymbol{\theta})}{\int f(\mathbf{y}|\boldsymbol{\theta})p(\boldsymbol{\theta})d\boldsymbol{\theta}} .$$

- Hierarchical models specify priors distributions in stages conditional on a set of hyperparameters: $p(\boldsymbol{\theta}|\boldsymbol{\eta})$
- In contrast, frequentist hypothesis-tests based on P-values offer indirect evidence for the parameters of interest that is based on conditional probabilities of the *observed data* given a fixed values of the parameters

Frequentist Operating Characteristics

- Frequentist operating characteristics refer to statistical properties of Bayesian procedures
- Assess "posterior performance" under fixed values of the model parameters characterizing true states of nature, $\boldsymbol{\theta}^{tr}$
  - $\lambda^{tr} < -\delta$ implies failure,
  - $-\delta \leq \lambda^{tr} \leq \delta$ implies equivalence
  - $\lambda^{tr} > \delta$ implies efficacy

Decision rules:

$$\begin{aligned} p(\lambda < -\delta|\mathbf{y}) > 0.95, &\quad \textit{failure} \\ p(\lambda \in [-\delta, \delta]|\mathbf{y}) > 0.90, &\quad \textit{equivalence} \\ p(\lambda > \delta|\mathbf{y}) > 0.95, &\quad \textit{efficacy} \\ \textit{otherwise}, &\quad \textit{inconclusive} \end{aligned}$$

- $\phi(|\boldsymbol{\theta}^{tr})$ denote the prob. of a decision rule given $\boldsymbol{\theta}^{tr}$, i.e.

$$\phi(\lambda < -\delta|\boldsymbol{\theta}^{tr}) = \int I\left\{p(\lambda < -\delta|\mathbf{y}) > 0.95\right\} f(\mathbf{y}|\boldsymbol{\theta}^{tr}) d\mathbf{y}$$

Thesis Research

Commensurate Prior Methodology

Brian P. Hobbs, Bradely P. Carlin, Daniel J. Sargent

- Proposed a framework of Bayesian hierarchical models for incorporating historical data into the analysis of a prospective trial from a *small number* of historical studies.

Papers:

- Hobbs, B.P., Carlin, B.P., Mandrekar, S., and Sargent, D.J. (2011). Hierarchical commensurate and power prior models for adaptive incorporation of historical information in clinical trials, to appear *Biometrics*
  - 2010 John Van Ryzin Award (ENAR)
- Hobbs, B.P., Sargent, D.J., and Carlin, B.P. (2011). Commensurate priors for incorporating historical information in clinical trials using general and generalized linear models. *Bayesian Analysis* (in process)

Pocock's "Acceptability" Criteria for Historical Controls

Ingelfinger (1973): "...ethical, as well as scientific, considerations require that medicine depend on the most reliable and the best controlled data available –the kind of data that is sought by *randomized* clinical study."

Pocock (1976): "...presence of *acceptable* historical data cannot be ignored in the full comparative evaluation of a new treatment."

1. Precisely defined, identical therapies
2. Historical group part of a recent clinical study which contained identical requirements for patient eligibility
3. Uniform treatment evaluation
4. Distributions of important patient characteristics comparable
5. Same organization with largely the same clinical investigators
6. No other indications leading one to expect differing results,
   - i.e. enthusiasm of investigators (patient accrual rates)

- The model *must* account for unknown bias in the historical controls

Example: Randomized Controlled Colorectal Cancer Trials

- Two successive randomized controlled colorectal cancer trials:

Saltz et al. (2000) trial randomized $N_0 = 683$: May 1996 and May 1998

1. Irinotecan alone (arm A)
2. Irinotecan and bolus Fluorouracil plus Leucovorin (arm B; IFL) *significantly longer progression free survival*
3. Fluorouracil and Leucovorin (arm C; 5FU/LV) *standard therapy*

Goldberg et al. (2004) trial randomized $N = 795$: May 1999 and April 2001

1. Irinotecan and bolus Fluorouracil plus Leucovorin (IFL) *regulatory standard in March 2000*
2. Oxaliplatin and infused Fluorouracil plus Leucovorin (FOLFOX) *new regimen*
3. Irinotecan and Oxaliplatin (IROX) *new regimen*

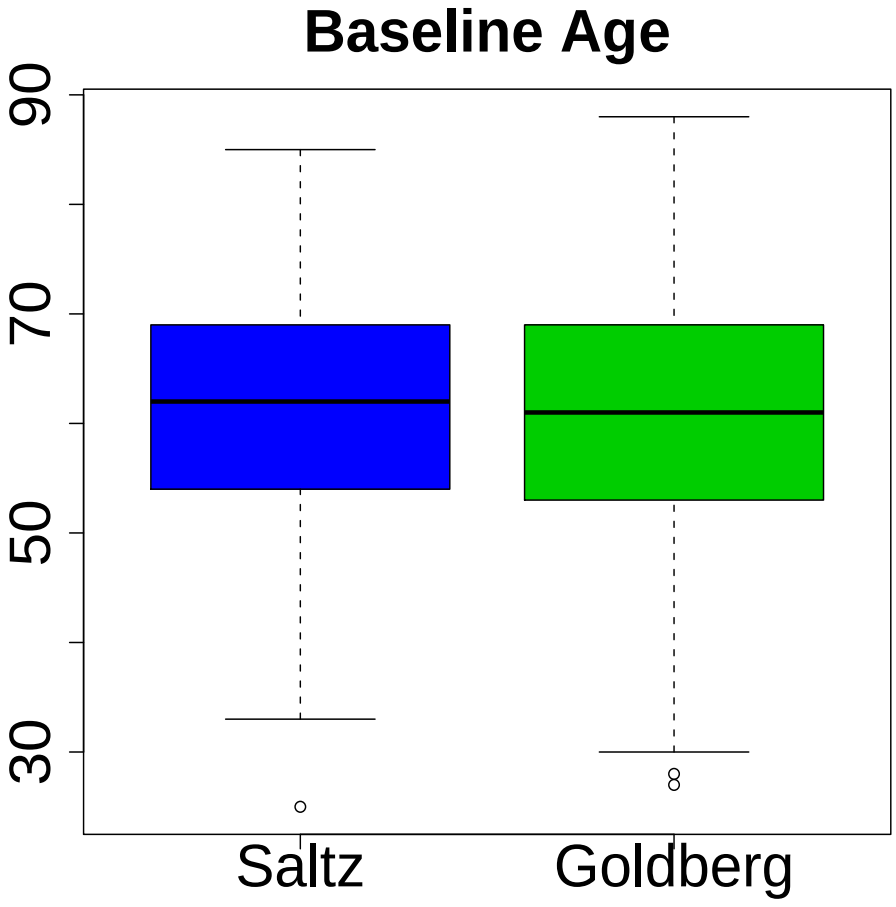
Randomized Controlled Colorectal Cancer Trials (cont'd)

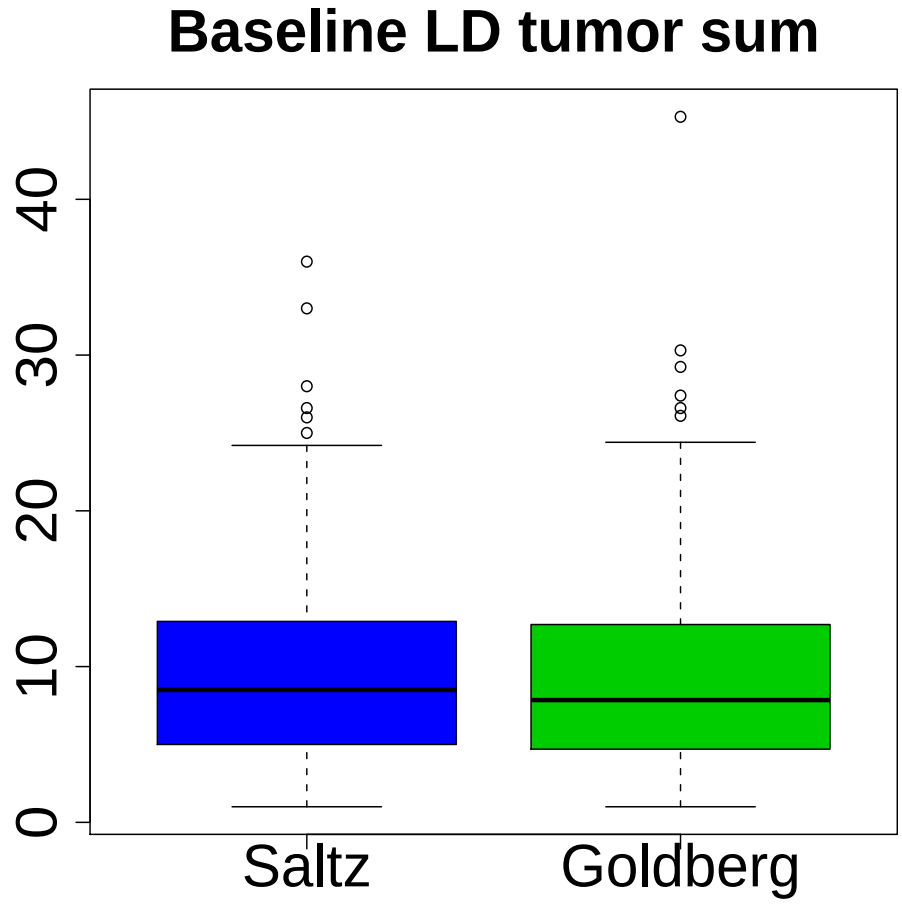
Incorporate IFL from Saltz trial prospectively, into Goldberg analysis

Consider acceptability criteria:

- Identical therapies: (doses in mg/m2) IFL was irinotecan 125 and bolus FU 500 plus LV 20 weekly for 4 weeks every 6 weeks
- Inclusion Criteria:
  - histologically proven unresectable colorectal adenocarcinoma
  - Eastern Cooperative Oncology Group performance status ≤ 2
  - adequate organ function
- Exclusion Criteria:
  - Prior therapy for metastatic disease
  - Adjuvant fluorouracil in previous 12 months
- Identically measured baseline and time-varying prognostic factors
  - Tumor measurements at regular 6-week cycles

Randomized Controlled Colorectal Cancer Trials (cont'd)





Boxplots of age (left) and ld tumor sum (right) at baseline for the Saltz et al. and Goldberg et al. trials.

Randomized Controlled Colorectal Cancer Trials (cont'd)

- Uniformly defined disease progression
  - 25% or greater increase in measurable tumor or the appearance of new lesions
- 1 year from end of Saltz to start of Goldberg
- 5 years between initiation of Saltz to completion of Goldberg

Conventional Methodology: Random Effects Meta-analysis

- Synthesis of information about *population average effects* from multiple sources
- Not intended for small number of sources (< 4), (Gelman 2006)
- yields *no meaningful gain in precision* in this context
- Consequently, historical data incorporated by assuming Homogeneity a priori risking *subjective analysis with highly biased results!*

Motivation

Prospective Trial's Objective:

- to evaluate the treatment comparison in the *current analysis*, incorporate historical evidence prospective *if it emerges as commensurate*.

Define *commensurate* "lack of strong evidence for heterogeneity"

Formulate prior distribution:

- Not necessarily to reflect our prior beliefs *before the trial* ...
- To deliver *desirable frequentist operating characteristics*
- Facilitate alternative, more desirable bias-variance trade-offs than those offered by pre-existing methodologies for incorporating historical data from a small number of historical studies.

Background

Consider general case without covariates:

- compare novel treatment to a previously studied control

$$\text{current trial:} \quad y_i = \mu + d_i\lambda + \epsilon_i, \quad \epsilon_i \sim N(0, \sigma^2)$$

- d_i is an indicator of novel treatment

$$\text{historical trials:} \quad \mathbf{y}_{0,h} \overset{iid}{\sim} N(\mu_{0,h}, \sigma^2_{0,h}),$$

- $h = 1, ..., H$ historical studies
- λ is of primary interest for treatment evaluation
- historical data is incorporated into the analysis of the current trial for purpose of facilitating more precise estimate of λ

Conventional Random-effects Meta-analytic Approach

Random-effects meta-analysis[a] assumes exchangeability:

$$\mu_{0,1}, ..., \mu_{0,H}, \mu \sim N(\xi, \eta^2)$$

- between-study *heterogeneity* and within-study variability
- ξ and η^2 characterize the population mean and between-study variance
- *shrinkage parameter*

$$B = \sigma^2/(\sigma^2 + \eta^2),$$

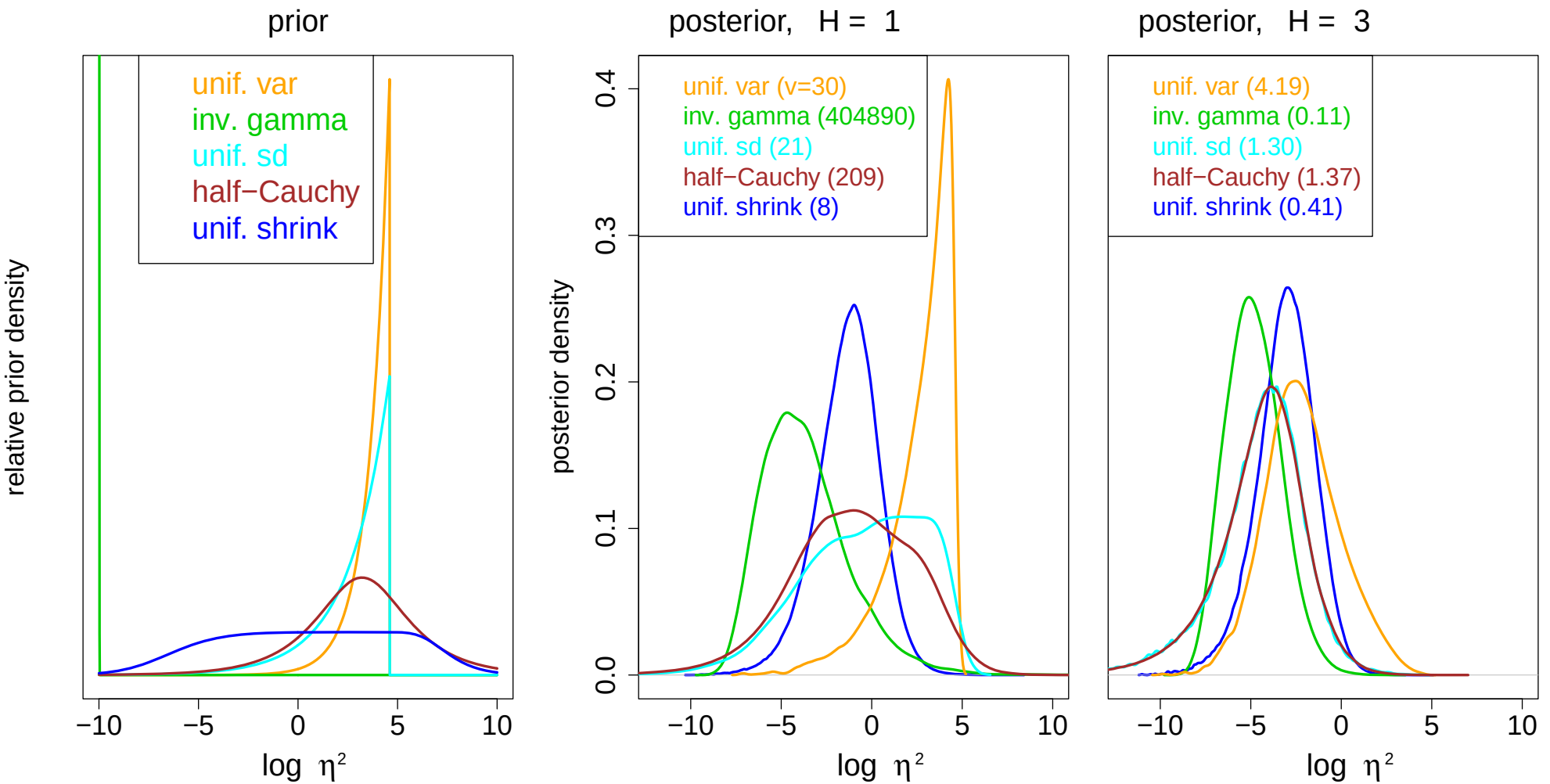
  weight placed on the prior mean ξ for the posterior mean μ
- denote *unknown biases* by $\Delta_h = \mu - \mu_{0,h}$,
- parameter vector $\boldsymbol{\theta} = (\lambda, \boldsymbol{\Delta}, \sigma^2, \sigma^2_{0,1}, ..., \sigma^2_{0,H})$, and $\mathbf{Y} = (\mathbf{y}, \mathbf{y}_{0,1}, ..., \mathbf{y}_{0,H})$ denote the collection of response data

a = Spiegelhalter et al., 2004

Common "noninformative" and "weakly-informative" priors for η^2

prior	form
uniform variance[a]	$p(\eta^2) = U(0, a),\ a = 100$
inverse gamma[a]	$p(\eta^2) = \Gamma^{-1}(\epsilon, \epsilon),\ \epsilon = 0.001$
uniform standard deviation[a]	$p(\eta) = U(0, \sqrt{a})$
half-Cauchy[b]	$p(\eta) \propto (\eta^2 + b)^{-1},\ b = 25$
uniform shrinkage[c]	$p(\eta^2) \propto \sigma^2/\{(\sigma^2 + \eta^2)^2\},\ \sigma^2_{0,h} = \sigma^2,$

a = Spiegelhalter et al., 2004; b = Gelman, 2006; c = Daniels, 1999



Prior & posteriors for $\log(\eta^2)$ under full homogeneity: $\boldsymbol{\Delta}^{tr} = \mathbf{0}$, for $n = 180$, $n_0 = 60$, and $\sigma^2 = \sigma^2_{0,h} = 1$. Parens = stand dev on the scale of η^2

$$\propto \int_{\mathbf{Y}} \left\{ \int_{\boldsymbol{\theta}} \eta^2 p(\eta^2, \boldsymbol{\theta}) L(\mathbf{Y}|\boldsymbol{\theta}) d\boldsymbol{\theta} \right\} L(\mathbf{Y}|\boldsymbol{\theta}^{tr}) d\mathbf{Y}$$

Commensurate Prior Model: One Historical Study

- *commensurate prior* for μ (HCMS, 2011)

$$\mu \sim N(\mu|\, \mu_0, 1/\tau)$$

- μ is a "non-systematically biased" representation of μ_0
- unknown bias: $\Delta = \mu - \mu_0$
- *initial prior*, $p(\mu_0)$, characterizes info. before observing hist. data
- one-to-one relationship between τ and η^2: $\tau = 1/(2\eta^2)$
- joint posterior: $q(\boldsymbol{\theta}|\tau, \mathbf{y}, \mathbf{y}_0) \propto$

$$N(\mu|\, \mu_0, 1/\tau)p(\mu_0)p(\sigma, \sigma_0)\prod_{j=1}^{n_0} N(y_{0j}|\, \mu_0, \sigma_0^2)\prod_{i=1}^{n} N(y_i|\, \mu + d_i\lambda, \sigma^2)$$

- Pocock (1976) repeated analysis under several fixed values of $1/\tau$
- HCMS (2011) consider fully Bayesian approaches as well as alternative fully Bayesian power prior formulations

Commensurate Prior Model: Multiple Historical Studies

- assume homogeneity among the hist. studies (or fixed degree of heterogeneity): $\mu_{0,1}, ..., \mu_{0,H} = \mu_0$
- commens. prior for μ cond. on the hist. pop. mean
- constraining the H historical means, $\mu_{0,h}$, to be equal to each other but perhaps *not* to μ inserts an asymmetry into the model that is not present in the usual exchangeability model

Relationship between τ and η^2 is more complex

- denote $\omega_{0,h} = \sigma^2_{0,h}/n_{0,h}$, and let $\omega = \sigma^2/(n - \sum d_i)$
- τ^{-1} characterizes the meta-analytic between-study variability, plus the diff. between the summed variability among the sample means, and the pop. mean when heterogeneity is estimated η^2 versus when full homogeneity is assumed,

$$\tau^{-1} = \eta^2 + \left\{ 1/(\omega + \eta^2) + \sum_{h=1}^{H} 1/(\omega_{0,h} + \eta^2) \right\}^{-1} - \left(1/\omega + \sum_{h=1}^{H} 1/\omega_{0,h} \right)^{-1}$$

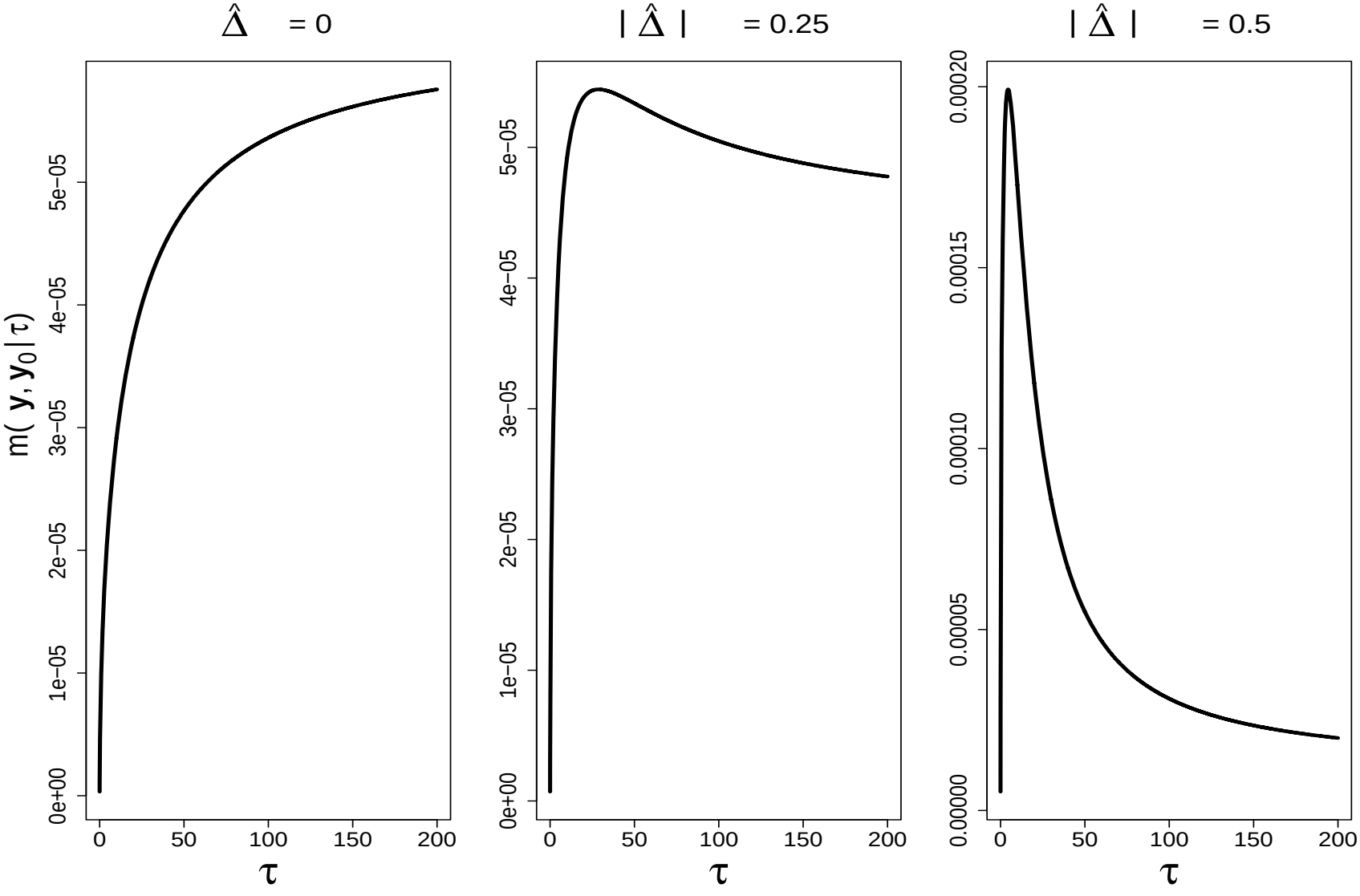
Estimation of τ

I. Parametric empirical Bayesian (EB)

- EB inference for $\boldsymbol{\theta}$ proceeds by replacing the scalar hyperparameter τ with its marginal maximum likelihood estimate (MMLE)

$$m(\mathbf{y}, \mathbf{y}_0|\tau) \propto N\left\{\hat{\Delta} \,\middle|\, 0, \frac{\sigma^2}{n - n_d} + v_0 + \frac{1}{\tau}\right\}$$

- known sampling variances
- $n_d = \sum_{i=1}^{n} d_i, \qquad \bar{y}_d = \sum_{i=1}^{n} d_i y_i / n_d, \qquad \hat{\mu} = \frac{n\bar{y} - n_d \bar{y}_d}{n - n_d}$
- $v_0 = (\sum_{h=1}^{H} 1/\omega_{0,h})^{-1}, \qquad \hat{\mu}_0 = v_0 \left(\sum_{h=1}^{H} \bar{y}_{0,h}/\omega_{0,h}\right)$
- $\hat{\Delta} = \hat{\mu} - \hat{\mu}_0$



Maginal density of $\mathbf{y}, \mathbf{y}_0 | \tau$ as a function of τ for three values of $|\hat{\Delta}|$ $\sigma^2 = 1$, $v_0 = 1/n_0$, $n_0 = 60$, $n = 180$, $n_d = 90$.

Estimation of τ (cont'd)

I. Parametric empirical Bayesian (EB) (cont'd)

- fix $\nu^* = 1/\tau^*$ at the value that MMLE restricted to a pre-specified interval, $0 < l_\nu < u_\nu$, capturing the "effective range of borrowing of strength"

$$\begin{aligned}\nu^* &= \arg \max_{\nu \in [l_\nu, u_\nu]} \{m(\mathbf{y}, \mathbf{y}_0 \mid 1/\nu)\} \\ &= \max \left[\min \left\{ \hat{\Delta}^2 - \frac{\sigma^2}{n - n_d} - v_0,\ u_\nu \right\},\ l_\nu \right]\end{aligned}$$

- bounding precludes full homogeneity when evidence for heterogeneity is not strong
- select limits via formal evaluation of the induced frequentist operating characteristics and bias-variance trade-offs in context
- paper uses $\nu^* \in [2(0.05^2), 2(10^2)]$ equiv. to $\eta \in [0.05, 10]$ for one hist. study

Estimation of τ (cont'd)

II. Fully Bayesian approaches

- EB procedure yields *approximate full homogeneity* when evidence for heterogeneity is not strong
- EB inference typically "underestimates" variability in $\boldsymbol{\theta}$, since posterior uncertainty in ν^* is unacknowledged in the analysis
- fully Bayesian approaches take *full account of uncertainty* in the parameter estimates
- we consider two families of priors
  1. conditionally conjugate gamma distribution
  2. a variant of the "spike and slab" distribution introduced for variable selection (Mitchell and Beauchamp, 1988)

Estimation of τ (cont'd)

II. Fully Bayesian approaches (cont'd)

1. Gamma prior: $p(\tau) = \Gamma(c\tilde{\tau}, c)$
   - conjugate full conditional posterior:

$$q(\tau \mid \boldsymbol{\theta}, \mathbf{y}, \mathbf{y}_0) \propto \Gamma\left(c\tilde{\tau} + 1/2,\ \Delta^2/2 + c\right).$$

   - $\tilde{\tau} > 0$, is prior guess at τ
   - $c > 0$, represents "degree of confidence", with a smaller value corresponding to weaker prior belief

Estimation of τ (cont'd)

II. Fully Bayesian approaches (cont'd)

2. "Spike and Slab" prior

- locally uniform between $0 \leq \mathcal{S}_l < \mathcal{S}_u$
- except for a bit of prob. mass concentrated at point, $\mathcal{K} > \mathcal{S}_u$

$$
\begin{aligned}
Pr(\tau < \mathcal{S}_l) &= 0, \\
Pr(\tau < u) &= p_0 \left\{ (u - \mathcal{S}_l) / (\mathcal{S}_u - \mathcal{S}_l) \right\}, \ \mathcal{S}_l \leq u \leq \mathcal{S}_u \\
\text{and } Pr(\tau > \mathcal{S}_u) &= Pr(\tau = \mathcal{K}) = 1 - p_0
\end{aligned}
$$

- where p_0 denotes the prior prob. that $\mathcal{S}_l \leq \tau \leq \mathcal{S}_u$

In presence of lack of evidence for heterogeneity,

- marginalized likelihood prefers a large value for τ,
- but is virtually flat over a vast portion of the parameter space, providing little information to distinguish among values.
- one carefully selected large value of τ (a "spike"), may characterize commensurability

Point Estimation of λ

- current trial's objective is to compare a *novel treatment* to the previously studied control therapy
- posterior inference on the novel treatment effect parameter, λ, is of primary interest
- evaluate the proposed models for estimating λ under squared error loss (SEL)
- ignoring historical data yields marg. posterior:

$$q(\lambda|\mathbf{y}) \propto N\left[(\bar{y}_d - \bar{y})/(1 - n_d/n)\ ,\ \sigma^2/\{n_d(1 - n_d/n)\}\right]$$

- otherwise let $\mathcal{D} = (\mathbf{y}, \mathbf{y}_0)$, it follows as

$$q(\lambda, \tau|\mathbf{y}, \mathbf{y}_0) \propto \quad \propto \tau N\left(\lambda \,\middle|\, \hat{\lambda}_\tau, V_\tau\right) N\left\{\hat{\Delta} \,\middle|\, 0, \frac{\sigma^2}{n - n_d} + v_0 + \frac{1}{\tau}\right\} p(\tau)$$

Point Estimation of λ (cont'd)

- where

$$V_\tau = \left[\left\{ \left(\frac{n}{n_d} \right)^2 \left(\frac{\sigma^2}{n} + v_0 + \frac{1}{\tau} \right) \right\}^{-1} + \frac{n_d(1 - n_d/n)}{\sigma^2} \right]^{-1},$$

and

$$\hat{\lambda}_\tau = \left\{ \frac{\bar{y} - \hat{\mu}_0}{\left(\frac{n}{n_d} \right) \left(\frac{\sigma^2}{n} + v_0 + \frac{1}{\tau} \right)} + \frac{\bar{y}_d - \bar{y}}{\sigma^2/n_d} \right\} / V_\tau^{-1}$$

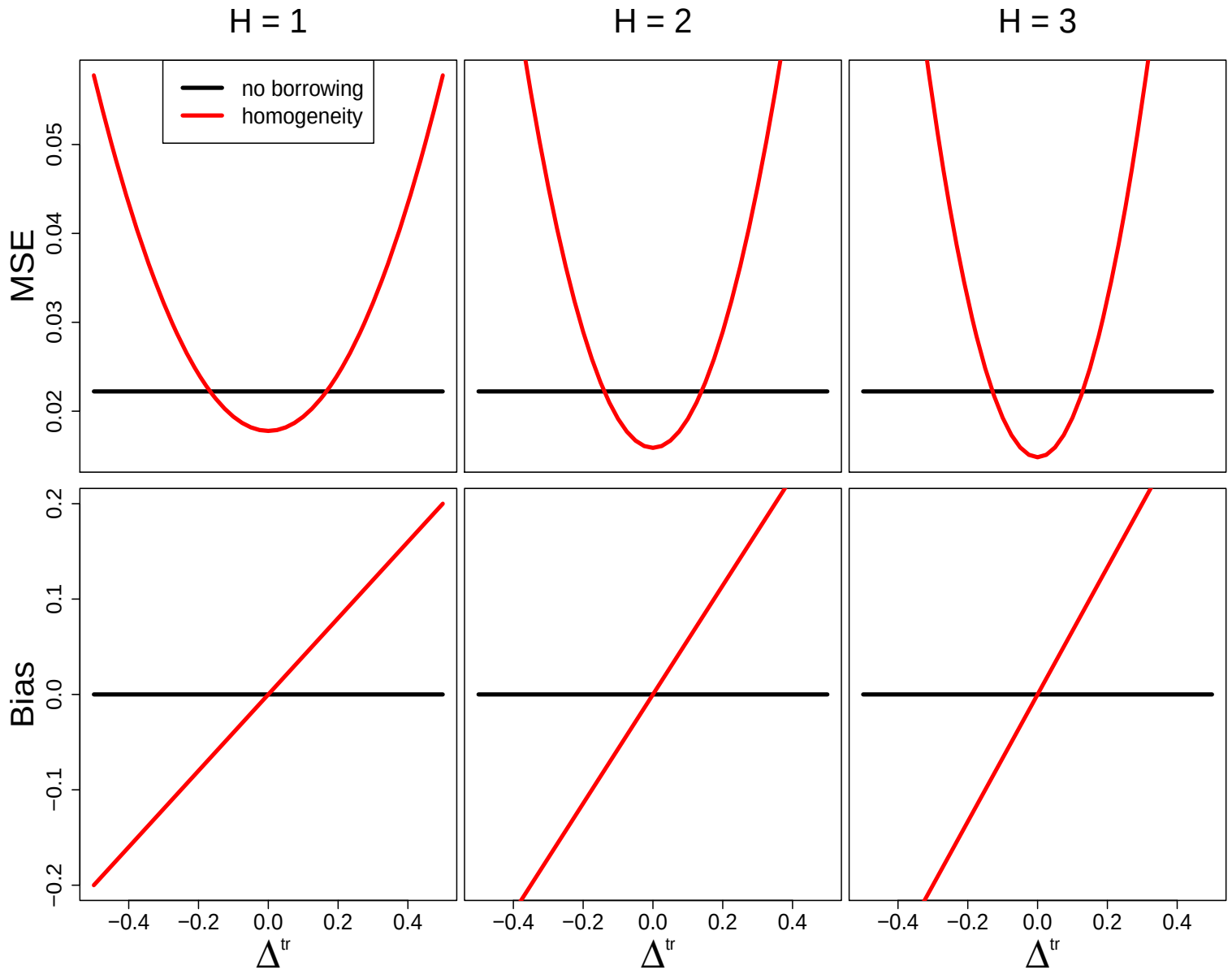
Point Estimation of λ (cont'd)

- let $\boldsymbol{\theta}^{tr} = (\lambda^{tr}, \Delta^{tr}, \sigma^2_{tr}, v_0^{tr})$, denote a set of fixed parameters
- *preposterior risk under squared error loss* (Carlin and Louis, 2000)

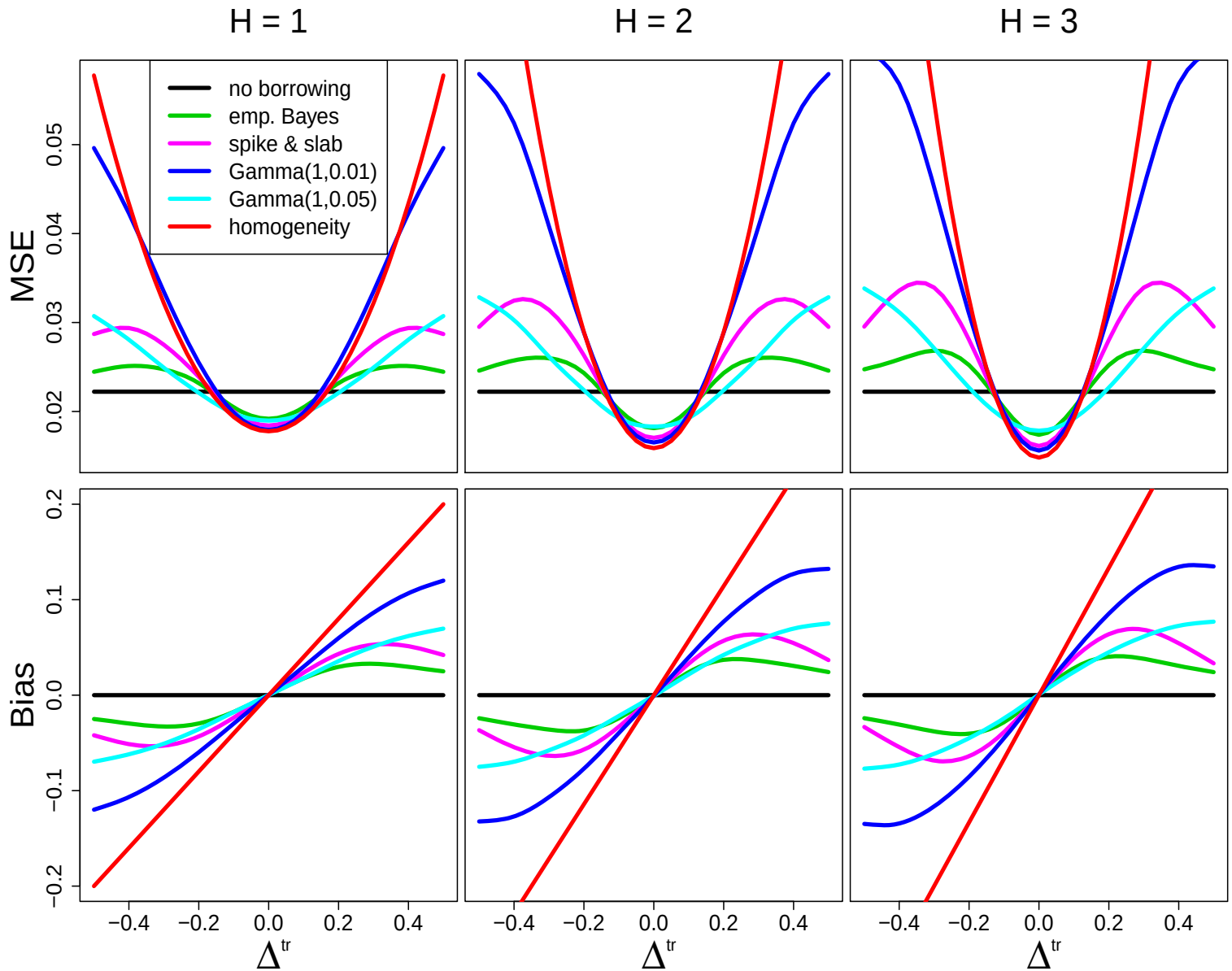
$$E_{\mathcal{D}|\boldsymbol{\theta}^{tr}}\left[\left\{E_{\lambda|\mathcal{D}}(\lambda) - \lambda^{tr}\right\}^2\right]$$

- *expected bias* is $E_{\mathcal{D}|\boldsymbol{\theta}^{tr}}\left[E_{\lambda|\mathcal{D}}(\lambda)\right] - \lambda^{tr}$

inference	$E_{\lambda\|\mathcal{D}}(\lambda)$
no borrowing	$\hat{\lambda}_0$
fully Bayesian	$\int \hat{\lambda}_\tau q(\tau\|\mathcal{D})d\tau$
emp. Bayes	$\hat{\lambda}_{1/\nu^*}$
full homogeneity	$\hat{\lambda}_\infty$



Preposterior risk under squared error loss (MSE) and expected bias as functions of Δ^{tr} for $H = 1, 2$, and 3 historical studies.



Preposterior risk under squared error loss (MSE) and expected bias as functions of Δ^{tr} for $H = 1, 2$, and 3 historical studies.

Comparison to Meta-analysis

Compare frequentist properties of our proposed commensurate prior models with results for the meta-analysis models for the case when sampling level variances are *unknown*

- results are shown for $H = 1, 2$, and 3 historical studies
- $\sigma^{tr} = \sigma^{tr}_{0,1} = ... = \sigma^{tr}_{0,H} = 1$
- $n = 180$, $n_d = 90$, and $n_{0,h} = 60$, $h = 1, .., H$
- under no borrowing the marginal posterior for $\lambda|\mathbf{y}$ follows as

$$q(\lambda|\mathbf{y}) \propto t\left\{n-2, (\bar{y}_d - \bar{y})/(1-n_d/n), \frac{s^2 + 1/(1/n - 1/n_d)(\bar{y} - \bar{y}_d)^2}{n_d(n-2)(1-n_d/n)}\right\}$$

	Percent change from no borrowing in preposterior risk under SEL								
	$\Delta^{tr} = 0$			$\Delta^{tr} = 0.25$			$\Delta^{tr} = 0.5$		
	$H = 1$	2	3	$H = 1$	2	3	$H = 1$	2	3
unif. var.	−1			0			0		
unif. shrink	−4			−2			0		
unif. sd	−4			−1			1		
half-Cauchy	−3			−1			1		
inv. gamma	−12			−1			8		
emp Bayes spike & slab $\Gamma(1, 0.01)$									
homog.	−19			25			152		

	Expected Bias		
unif. var.	0.00	0.00	0.00
unif. shrink	0.00	0.01	0.01
unif. sd	0.00	0.01	0.01
half-Cauchy	0.00	0.01	0.01
inv. gamma	0.00	0.03	0.03
emp Bayes spike & slab $\Gamma(1, 0.01)$			
homog.	0.00	0.10	0.19

	Percent change from no borrowing in preposterior risk under SEL								
	$\Delta^{tr} = 0$			$\Delta^{tr} = 0.25$			$\Delta^{tr} = 0.5$		
	$H = 1$	2	3	$H = 1$	2	3	$H = 1$	2	3
unif. var.	−1			0			0		
unif. shrink	−4			−2			0		
unif. sd	−4			−1			1		
half-Cauchy	−3			−1			1		
inv. gamma	−12			−1			8		
emp Bayes	−13			7			8		
spike & slab	−13			1			9		
$\Gamma(1, 0.01)$	−16			0			25		
homog.	−19			25			152		

	Expected Bias		
unif. var.	0.00	0.00	0.00
unif. shrink	0.00	0.01	0.01
unif. sd	0.00	0.01	0.01
half-Cauchy	0.00	0.01	0.01
inv. gamma	0.00	0.03	0.03
emp Bayes	0.00	0.03	0.02
spike & slab	0.00	0.03	0.02
$\Gamma(1, 0.01)$	0.00	0.05	0.07
homog.	0.00	0.10	0.19

	Percent change from no borrowing in preposterior risk under SEL								
	$\Delta^{tr} = 0$			$\Delta^{tr} = 0.25$			$\Delta^{tr} = 0.5$		
	$H = 1$	2	3	$H = 1$	2	3	$H = 1$	2	3
unif. var.	−1	−3	−11	0			0		
unif. shrink	−4	−9	−14	−2			0		
unif. sd	−4	−14	−21	−1			1		
half-Cauchy	−3	−14	−21	−1			1		
inv. gamma	−12	−20	−25	−1			8		
emp Bayes	−13	−17	−22	7			8		
spike & slab	−13	−17	−22	1			9		
$\Gamma(1, 0.01)$	−16	−22	−24	0			25		
homog.	−19	−28	−32	25			152		

	Expected Bias								
unif. var.	0.00	0.00	0.00	0.00			0.00		
unif. shrink	0.00	0.00	0.00	0.01			0.01		
unif. sd	0.00	0.00	0.00	0.01			0.01		
half-Cauchy	0.00	0.00	0.00	0.01			0.01		
inv. gamma	0.00	0.00	0.00	0.03			0.03		
emp Bayes	0.00	0.00	0.00	0.03			0.02		
spike & slab	0.00	0.00	0.00	0.03			0.02		
$\Gamma(1, 0.01)$	0.00	0.00	0.00	0.05			0.07		
homog.	0.00	0.00	0.00	0.10			0.19		

	Percent change from no borrowing in preposterior risk under SEL								
	$\Delta^{tr} = 0$			$\Delta^{tr} = 0.25$			$\Delta^{tr} = 0.5$		
	$H = 1$	2	3	$H = 1$	2	3	$H = 1$	2	3
unif. var.	−1	−3	−11	0	−1	−1	0	1	4
unif. shrink	−4	−9	−14	−2	−2	−1	0	3	7
unif. sd	−4	−14	−21	−1	1	3	1	8	14
half-Cauchy	−3	−14	−21	−1	0	5	1	7	16
inv. gamma	−12	−20	−25	−1	4	9	8	20	31
emp Bayes	−13	−17	−22	7	11	20	8	9	16
spike & slab	−13	−17	−22	1	5	11	9	10	11
$\Gamma(1, 0.01)$	−16	−22	−24	0	5	8	25	35	38
homog.	−19	−28	−32	25	61	86	152	337	475

	Expected Bias								
unif. var.	0.00	0.00	0.00	0.00	0.00	0.02	0.00	0.01	0.02
unif. shrink	0.00	0.00	0.00	0.01	0.02	0.03	0.01	0.02	0.03
unif. sd	0.00	0.00	0.00	0.01	0.03	0.04	0.01	0.03	0.04
half-Cauchy	0.00	0.00	0.00	0.01	0.03	0.05	0.01	0.02	0.04
inv. gamma	0.00	0.00	0.00	0.03	0.05	0.07	0.03	0.05	0.07
emp Bayes	0.00	0.00	0.00	0.03	0.04	0.05	0.02	0.02	0.03
spike & slab	0.00	0.00	0.00	0.03	0.03	0.04	0.02	0.02	0.02
$\Gamma(1, 0.01)$	0.00	0.00	0.00	0.05	0.06	0.07	0.07	0.08	0.08
homog.	0.00	0.00	0.00	0.10	0.14	0.17	0.19	0.28	0.34

Linear Regression Models

In the context of two successive clinical trials that identically measure $p-1$ covariates:

$$y_0 \sim N_{n_0}(X_0\beta_0, \sigma_0^2) \text{ and } y \sim N_n(X\beta + d\lambda, \sigma^2)$$

- X_0 and X denote $n_0 \times p$ and $n \times p$ design matrices
- λ is the (scalar) novel treatment effect
- formulate linear model by replacing τ with a vector $\boldsymbol{\tau} = (\tau_1, ..., \tau_p)$ containing a commensurability parameter, τ_g, for each associated pair of parameters in β_g and β_{0g},

$$p(\beta_g|\beta_{0g}) \propto N\left(\beta_g|\beta_{0g}, \tau_g^{-1}\right),$$

for $g = 1, ..., p$

- β_gs are assumed a-priori independent

Linear Regression Models (cont'd)

- let $\hat{\boldsymbol{\beta}}_\lambda = (X^T X)^{-1} X^T (\mathbf{y} - \mathbf{d}\lambda)$ and $\hat{\boldsymbol{\beta}}_0 = (X_0^T X_0)^{-1} X_0^T \mathbf{y}_0$,
- precision matrix that results from averaging over the hist. likelihood:

$$V_{\boldsymbol{\tau}} = diag\{\boldsymbol{\tau}\} - diag\{\boldsymbol{\tau}\} \left(X_0^T X_0 / \sigma_0^2 + diag\{\boldsymbol{\tau}\} \right)^{-1} diag\{\boldsymbol{\tau}\}$$

Flat initial prior yields cond. posterior $\boldsymbol{\beta} | \mathbf{y}, \mathbf{y}_0, \lambda, \sigma^2, \sigma_0^2, \boldsymbol{\tau}$,

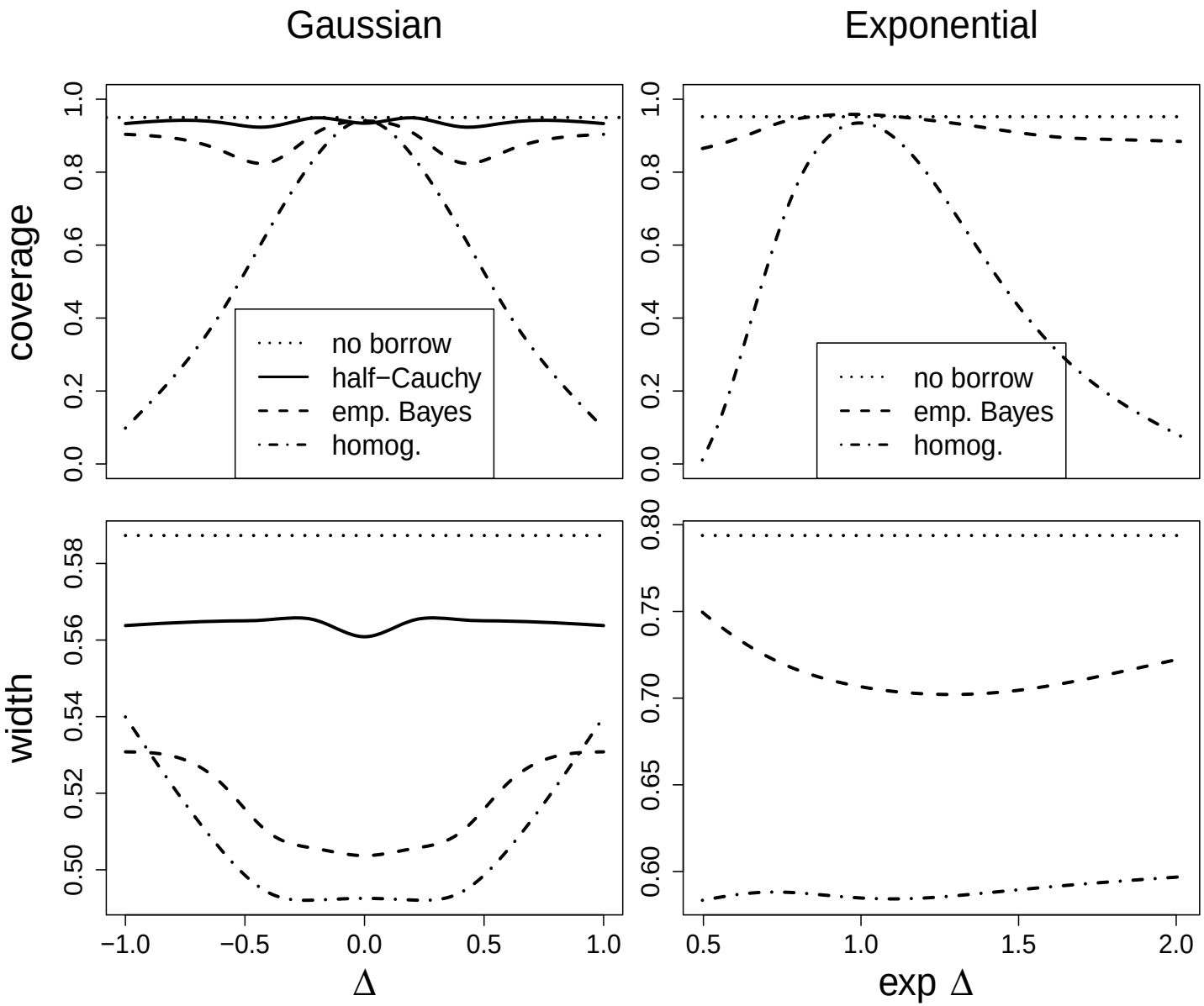
$$\propto N_p \left(\boldsymbol{\beta} \Big| \left(V_{\boldsymbol{\tau}} + \frac{X^T X}{\sigma^2} \right)^{-1} \left(V_{\boldsymbol{\tau}} \hat{\boldsymbol{\beta}}_0 + \frac{X^T X}{\sigma^2} \hat{\boldsymbol{\beta}}_\lambda \right), \left(V_{\boldsymbol{\tau}} + \frac{X^T X}{\sigma^2} \right)^{-1} \right),$$

- $\tau_g \to \infty$, for all $g = 1, ..., p$, $V_{\boldsymbol{\tau}} \to X_0^T X_0 / \sigma_0^2$, "fully borrowing"
- $\boldsymbol{\tau} \to \mathbf{0}$, the marginal posterior for $\boldsymbol{\beta}$ converges "no borrowing"
- HSC (2012) considers General and Generalized Linear Mixed Models

Simulation Study

Simulate freq. and Bayesian OC for case of one historical study:

- unknown bias, $\Delta = \mu - \mu_0$
- inference on novel treatment effect, λ
- Gaussian data:
  - compare to no borrowing, half-Cauchy meta-analytic, and homogeneity models
  - $n_0 = 90$ hist. patients and equal allocation of $n = 180$
  - fixed true parameters $\mu^{tr} = 0$, and $\sigma^{tr} = \sigma_0^{tr} = 1$
- Exponential data:
  - compare to no borrowing and homogeneity models
  - $\mu^{tr} = 2$, $n_0 = 200$, and equal allocation of $n = 100$



Coverage and width of 95% HPD intervals for λ by Δ^{tr} or $\exp(\Delta^{tr})$ for Gaussian and exponential data for one historical study.

Simulation Study (cont'd)

True states of nature: (Freedman et al., 1984)

- $\lambda^{tr} < -\delta$ implies failure,
- $-\delta \leq \lambda^{tr} \leq \delta$ implies equivalence
- $\lambda^{tr} < \delta$ implies efficacy

Decision rules:

$$\begin{aligned} q(\lambda < -\delta | \mathbf{y}, \mathbf{y}_0) > 0.95, &\quad \textit{failure} \\ q(\lambda \in [-\delta, \delta] | \mathbf{y}, \mathbf{y}_0) > 0.90, &\quad \textit{equivalence} \\ q(\lambda > \delta | \mathbf{y}, \mathbf{y}_0) > 0.95, &\quad \textit{efficacy} \\ \textit{otherwise}, &\quad \textit{inconclusive} \end{aligned}$$

- $\phi(|\boldsymbol{\theta}^{tr})$ denote the prob. of a decision rule given $\boldsymbol{\theta}^{tr}$, i.e.

$$\phi(\lambda < -\delta | \boldsymbol{\theta}^{tr}) = \int I\left\{q(\lambda < -\delta | \mathbf{Y}) > 0.95\right\} L(\mathbf{Y} | \boldsymbol{\theta}^{tr}) d\mathbf{Y}.$$

Simulation Study (cont'd)

Compare prob. allocated to the correct and incorrect decision spaces for fixed true values $\delta = 0.33$, $\sigma^{tr} = \sigma_0^{tr} = 1$, and $\mu_0^{tr} = 0$

$$
\begin{aligned}
M(c) = & \int_{-\infty}^{-\delta} \int_{-\infty}^{\infty} \phi(\lambda < -\delta | \Delta^{tr}, \lambda^{tr}) p(\Delta^{tr}) p(\lambda^{tr}) d\Delta^{tr} d\lambda^{tr} \\
& + \int_{-\delta}^{\delta} \int_{-\infty}^{\infty} \phi(\lambda \in [-\delta, \delta] | \Delta^{tr}, \lambda^{tr}) p(\Delta^{tr}) p(\lambda^{tr}) d\Delta^{tr} d\lambda^{tr} \\
& + \int_{\delta}^{\infty} \int_{-\infty}^{\infty} \phi(\lambda > \delta | \Delta^{tr}, \lambda^{tr}) p(\Delta^{tr}) p(\lambda^{tr}) d\Delta^{tr} d\lambda^{tr} \\
& - \int_{-\infty}^{-\delta} \int_{-\infty}^{\infty} \phi(\lambda > -\delta | \Delta^{tr}, \lambda^{tr}) p(\Delta^{tr}) p(\lambda^{tr}) d\Delta^{tr} d\lambda^{tr} \\
& - \int_{-\delta}^{\delta} \int_{-\infty}^{\infty} \phi(\lambda < -\delta \cup \lambda > \delta | \Delta^{tr}, \lambda^{tr}) p(\Delta^{tr}) p(\lambda^{tr}) d\Delta^{tr} d\lambda^{tr} \\
& - \int_{\delta}^{\infty} \int_{-\infty}^{\infty} \phi(\lambda < \delta | \Delta^{tr}, \lambda^{tr}) p(\Delta^{tr}) p(\lambda^{tr}) d\Delta^{tr} d\lambda^{tr} \\
& - c \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \phi(\mathit{inconclusive} | \Delta^{tr}, \lambda^{tr}) p(\Delta^{tr}) p(\lambda^{tr}) d\Delta^{tr} d\lambda^{tr}
\end{aligned}
$$

- metric $M(c) \in [-1, 1]$, a generalized version of expected $0 - 1$ loss

Results $M(1/2) \in [-1, 1]$ for $p(\Delta^{tr}) = N(0, s^2)$, and $p(\lambda^{tr}) = N(0, 1/4)$

	M	$+Eff$	$+Fail$	$+Eq$	$-Eff$	$-Fail$	$-Eq$	$-Inc/2$
				s=0				
no borrow	0.19	0.13	0.13	0.20	0.00	0.00	0.01	0.26
half-Cauchy	0.20	0.13	0.14	0.20	0.00	0.00	0.01	0.26
emp. Bayes	**0.29**	0.14	0.15	0.24	0.00	0.00	0.01	0.23
homog.	0.32	0.15	0.15	0.25	0.00	0.00	0.01	0.22
				s=1				
no borrow	0.19	0.13	0.13	0.20	0.00	0.00	0.01	0.26
half-Cauchy	0.21	0.13	0.14	0.20	0.00	0.00	0.01	0.25
emp. Bayes	**0.24**	0.14	0.15	0.21	0.00	0.00	0.02	0.24
homog.	0.14	0.14	0.15	0.12	0.00	0.00	0.07	0.21
				s=30				
no borrow	0.19	0.13	0.13	0.20	0.00	0.00	0.01	0.26
half-Cauchy	0.21	0.13	0.14	0.20	0.00	0.00	0.01	0.25
emp. Bayes	**0.23**	0.14	0.14	0.21	0.00	0.00	0.02	0.24
homog.	0.07	0.14	0.14	0.08	0.00	0.00	0.07	0.21

Example: Randomized Controlled Colorectal Cancer Trials

- Two successive randomized controlled colorectal cancer trials:

Saltz et al. (2000) trial randomized $N_0 = 683$: May 1996 and May 1998

1. Irinotecan alone (arm A)
2. Irinotecan and bolus Fluorouracil plus Leucovorin (arm B; IFL) *significantly longer progression free survival*
3. Fluorouracil and Leucovorin (arm C; 5FU/LV) *standard therapy*

Goldberg et al. (2004) trial randomized $N = 795$: May 1999 and April 2001

1. Irinotecan and bolus Fluorouracil plus Leucovorin (IFL) *regulatory standard in March 2000*
2. Oxaliplatin and infused Fluorouracil plus Leucovorin (FOLFOX) *new regimen*
3. Irinotecan and Oxaliplatin (IROX) *new regimen*

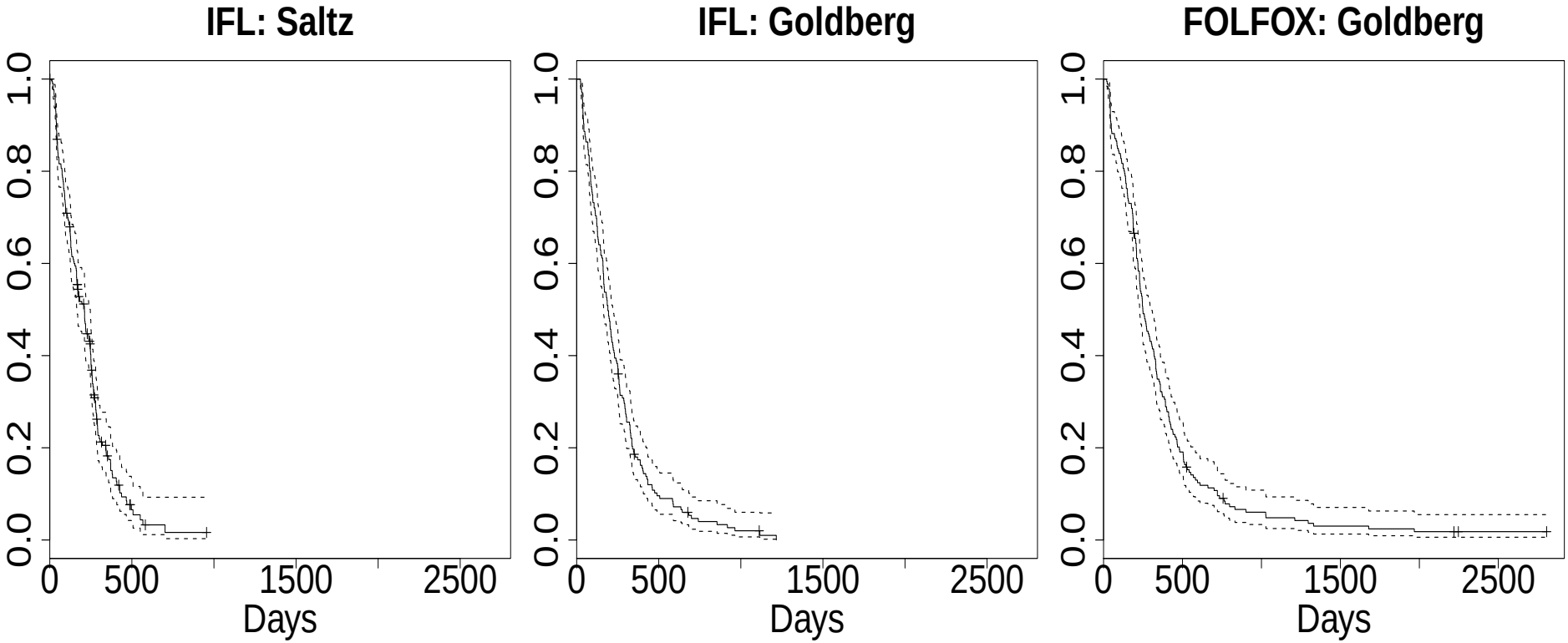
Incorporate IFL from Saltz trial prospectively, into Goldberg analysis

Randomized Controlled Colorectal Cancer Trials (cont'd)

Compare disease progression rates for FOLFOX and IFL using Weibull reg.

- historical data: arm B (IFL) from the Saltz trial, $n_0 = 224$
- current data: IFL, $n = 176$, and FOLFOX in the Goldberg trial, $n = 186$
- covariate: sum of the longest diameter (ld) in cm of 1 to 9 tumors at baseline (ld sum)

Randomized Controlled Colorectal Cancer Trials (cont'd)



Separate Kaplan-Meier survival curves corresponding subjects on IFL in the Saltz et al. trial (left), IFL in the Goldberg et al. trial (center), and FOLFOX in the Goldberg et al. trial.

Suggests survival experience for subjects on IFL was similar in both the Saltz et al. and Goldberg et al. trials, and FOLFOX is associated with prolonged time-to-progression.

Randomized Controlled Colorectal Cancer Trials (cont'd)

Scale-Shape commensurate prior for Weibull reg. (Time-to-Event)

- triples $(t_{0j}, \delta_{0j}, X_{0j})$ for $j = 1, ..., n_0$ and (t_i, δ_i, X_i) for $i = 1, ..., n$
- $t_{0j}, t_i > 0$ are obs. failure times; δ_{0j}, δ_i are noncensoring indicators
- log-linear models, (Kalbfleisch and Prentice, 2002)

$$y_0 = log(t_0) = X_0\beta_0 + \sigma_0 e_0, \quad \text{where} \quad e_0 = (y_0 - X_0\beta_0)/\sigma_0$$

$$y = log(t) = X\beta + d\lambda + \sigma e, \quad \text{where} \quad e = (y - X\beta - d\lambda)/\sigma$$

- Weibull regression assumes e_0 and e follow the *extreme value* dist.

$$f(u) = exp\left[u - exp\left(u\right)\right]$$

- we assume commensurate priors for *both* the regression coefficients, $\boldsymbol{\beta}$, and log transformation of the shape parameters, σ

Randomized Controlled Colorectal Cancer Trials (cont'd)

Weibull model fits to colorectal cancer data $n_0 = 224$, $n = 362$

	Separate analyses				Pooled analysis	
	Historical		Current			
	est	sd	est	sd	est	sd
Intercept	5.503	0.058	5.555	0.067	5.533	0.045
BL ldsum	−0.043	0.051	−0.115	0.045	−0.092	0.034
FOLFOX	–	–	0.417	0.092	0.453	0.077
$\log(\sigma)$	−0.291	0.060	−0.153	0.039	−0.186	0.033

Randomized Controlled Colorectal Cancer Trials (cont'd)

Commensurate prior Weibull model fits to colorectal cancer data $n_0 = 224$, $n = 362$

	EB		spike & slab		Gamma$(1, 0.01)$	
	est	sd	est	sd	est	sd
Intercept	5.541	0.054	5.547	0.058	5.546	0.058
BL ldsum	-0.100	0.040	-0.103	0.042	-0.105	0.042
FOLFOX	0.435	0.085	0.431	0.086	0.432	0.085
$\log(\sigma)$	-0.152	0.038	-0.158	0.038	-0.158	0.038
τ_1	200	–	153.2	84.4	124.8	107.3
τ_2	200	–	153.8	83.9	123.8	106.7
τ_3	40.0	–	126.1	96.1	102.5	93.7

Discussion

Before applying this methodology in practice one must consider carefully:

- bias-variance trade-off in the context of other important factors such as,
  - disparities in the sample sizes among the historical and current studies
  - width of the equivalence region
  - priors on Δ^{tr} and λ^{tr}
  - the design (ie, randomized versus single-arm) of the historical study
- differences in patient populations between the historical and new study and other known/unknown confounding factors that can be potential sources of bias when borrowing from the historical data.

Platform clinical trial designs for efficient drug development strategies

Brian P. Hobbs and J. Jack Lee

Goal: Develop randomized trial designs that process multiple agents simultaneously for effectively screening drugs in phase IIB, identifying biomarkers and predictive markers for guided targeted therapy, and treating more patients with more effective treatments during the trial

Methodology: Bayesian group sequential design incorporating decision rules to add treatment arms as well as drop poorly (Fail) or well (Graduate) performing treatment arms. Simulation is used to compare the design's operating characteristics under Equal Randomization (ER) and three adaptive randomization methods

Background

Motivation: Berry, D. A., (2004) *Statistical Science*, p.184

"The greatest room for innovation and for improving drug development is effectively dealing with the enormous numbers of molecules that are available as potential drugs..... building the foundation for a phase II trial for evaluating drugs that is more a process than a trial."

Lee, J. J. and Feng, L. (2005) *J. Clinical Oncology*

- Drug development programs that screen drugs one-at-a-time are inefficient:
  - multiple protocols involves operational "white space"
  - excessive number of patients are assigned standard therapy

Platform Design:

- A "process" involving a single protocol for continuous monitoring of outcomes for the purpose of efficient screening of multiple treatment regimens for efficacy for further evaluation in phase III and identification of biomarkers for targeted agent therapy
- involves "seamless" modifications to the study arms as poorly and well performing arms are replaced by new regimens

Model

Compare a standard regimen to p experimental regimens

Model: Logistic Regression

- Y = scalar indicator of successful treatment
- $\boldsymbol{\beta} = (\beta_1, ..., \beta_p)$ exp. regimens trt effects, α = intercept

$$Y \sim Bernoulli(\pi_j), \ logit(\pi_j) = \alpha + \beta_j$$

where $j = 0, ..., p$ and $\beta_0 = 0$

- Denote the number of patients assigned to the jth regimen by n_j

Priors:

- $\pi_j \sim Beta(0.5, 0.5)$, $j = 0, ..., p$

Posterior: Denote the posterior by $q(\alpha, \boldsymbol{\beta}|Y)$

Model (cont'd)

- $t = \textit{study time}$, $0 < t < T$, and let $p_t = \#$ of exp. regimens at t
- $n_{j,t} = \#$ of patients assigned to the jth regimen at t
- $\mathbf{Y}_{j,t} = 1 \times n_{j,t}$ response vector for jth regimen at t
- $\mathbf{Y}_t = (\mathbf{Y}_{0,t}, ..., \mathbf{Y}_{p_t,t})$ and $\mathbf{n}_t = (n_{0,t}, ..., n_{p_t,t})$

Decision Criteria: Thall, P. F. and Simon, R. (1994) *Biometrics*

- Interim adaptive decision will be based on posterior probability

$$\lambda_j(\delta, \mathbf{Y}_t) = Pr(\pi_j > \pi_0 + \delta | \mathbf{Y}_t) = \int_{-\infty}^{logit(1-\delta)} \int_{\beta^*}^{\infty} q(\alpha, \beta | \mathbf{Y}_t) d\alpha d\beta,$$

where $\beta^* = \log\left\{\frac{(\pi_0+\delta)(1-\pi_0)}{(1-\pi_0-\delta)\pi_0}\right\}$

Design Parameters:

- $N^{init} =$ min. $\#$ required to evaluate jth regimen
- $N^{fail}, N^{grad} =$ min. $\#$ required to Fail or Graduate jth regimen
- $N^{max} =$ max. $\#$ treated with the jth regimen

Model (cont'd)

- $p_L, p_U \in [0,1]$ = posterior probability decision cutoffs

Decision Rules: At time t

- If $\lambda_j(\delta, \mathbf{Y}_t) \geq p_U$ and $n_{j,t} \geq N^{grad}$, Graduate jth regimen
- If $\lambda_j(\delta, \mathbf{Y}_t) \leq p_L$ and $n_{j,t} \geq N^{fail}$, Fail jth regimen
- If $p_L \leq \lambda_j(\delta, \mathbf{Y}_t) \leq p_U$ and
  - $n_{j,t} < N^{max}$, treat another patient with jth regimen
  - $n_{j,t} \geq N^{max}$, declare jth regimen Inconclusive
  - Inconclusive $\neq$ Equivalence

Lookahead Rules:

- If $N^{init} \leq n_{j,t} \leq N^{fail}$ and $\lambda_j(\delta, \mathbf{Y}'_t) \leq p_L$, where $\mathbf{Y}'_{j,t} = (\mathbf{Y}_{j,t}, \mathbf{1}_{N^{fail} - n_{j,t}})$ Fail the jth regimen
- If $p_L \leq \lambda_j(\delta, \mathbf{Y}^*_{t,0}), \lambda_j(\delta, \mathbf{Y}^*_{t,1}) \leq p_U$, where $\mathbf{Y}^*_{j,t,0} = (\mathbf{Y}_{j,t}, \mathbf{0}_{N^{max} - n_{j,t}})$ and $\mathbf{Y}^*_{j,t,1} = (\mathbf{Y}_{j,t}, \mathbf{1}_{N^{max} - n_{j,t}})$, declare jth regimen Inconclusive

Simulation Study

Run platform design for a period of 5 years

Design Parameters:

- Assume enrollment rate of 10 patients per month (≈ 600 total)
- Fix $\pi_0 = 0.2$, randomly generate $\pi_j \sim Unif(0.095, 0.605)$ for each exp. regimen, $j > 0$
- $\delta = 0.1$, $p_L = 0.2$, and $p_U = 0.9$
- Require $n_{0,t} \geq 10$ before evaluating decision criteria
- $N^{init} = 2$, required to evaluate decision criteria for exp. regimens
- $N^{max} = 78$, provides 90% power for two-sided test of $\pi_0 = 0.2$ vs $\pi_j = 0.4$ given Type I error $= 0.1$
- $N^{fail} = 14$, commonly used for 1st stage of two-stage design
- $N^{grad} = 20$, allow promising regimens to remain in platform longer

Simulation Study (cont'd)

Cycle in new regimens: After decision at time t if

$$\left\{ \frac{T-t}{120/365} - \sum_{j=1}^{p_t} \max(N^{fail} - n_{j,t}, 1) \right\} \geq N^{fail},$$

- Unresolved exp. regimens at time T are considered Inconclusive

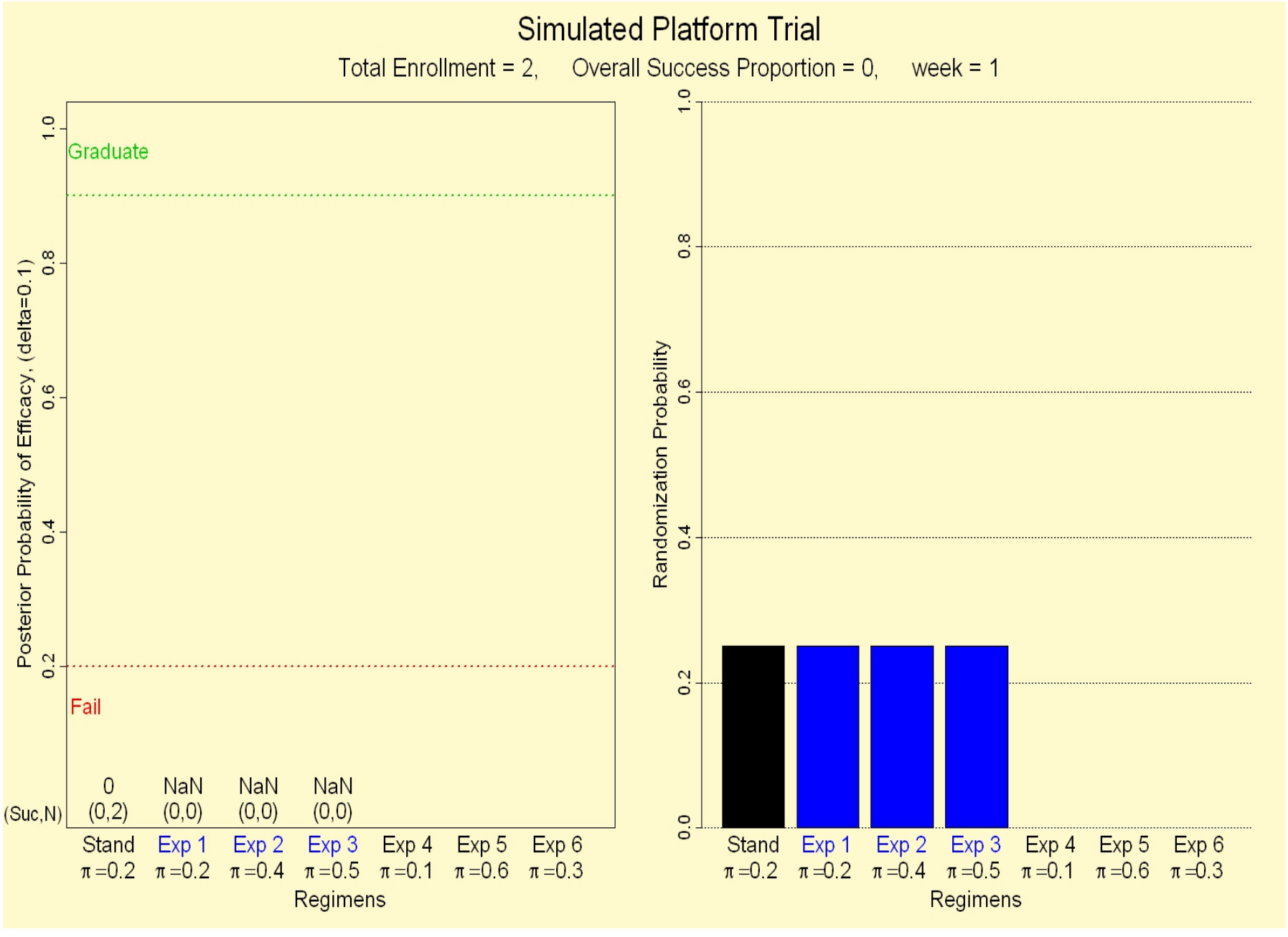
Objective: Identify "optimal"

1. Randomization method (Equal Rz + 3 adaptive methods)
2. Number of experimental regimens evaluated (simulated 1 to 7)

Operating Characteristics:

- Frequentist properties: probability of decision rules
- Total number of successful responses and arms evaluated
- Time to decision

Simulated Platform Design: Equal Randomization (ER)



Bayesian Adaptive Randomization (BAR)

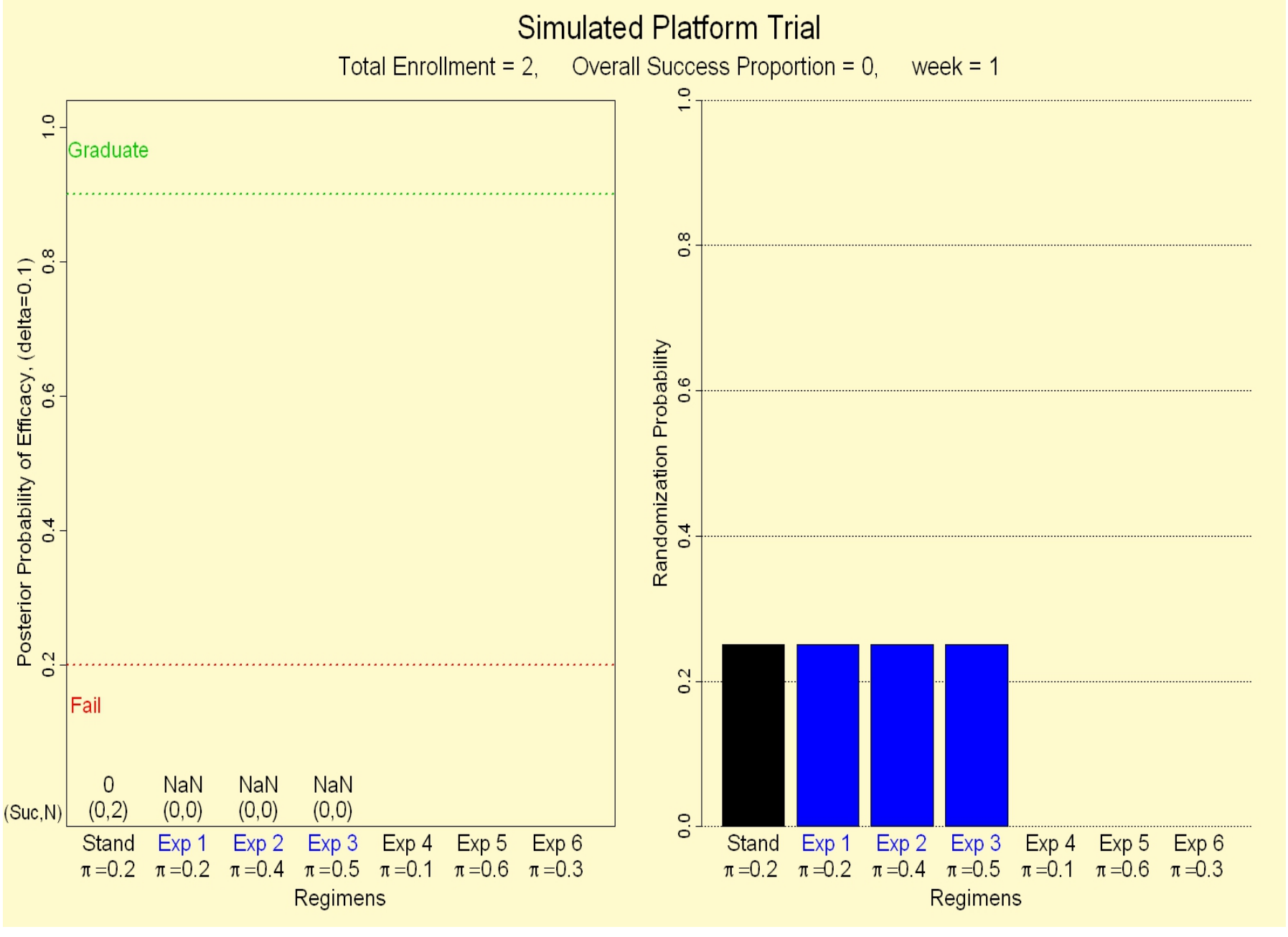
Thall, P. F. and Wathen, J. K. (2007) *EJC*; Lee, J. J., Gu, X., and Liu, S. (2010) *Clinical Trials* (incorporate biomarkers)

- Consider randomizing a new patient at time t to standard regimen or p_t experimental regimens
- Let $\mathbf{Y}_{t^-}$ and $\mathbf{n}_{t^-}$, denote the data for all active regimens prior to t
- $c =$ scalar tuning parameter
- Randomize new patient to the jth regimen with probability,

$$\phi_j^{BAR}(\mathbf{Y}_{t^-}, \mathbf{n}_{t^-}) = \frac{Pr\left(\bigcap_{k\neq j}^{p_t} \beta_k < \beta_j | \mathbf{Y}_{t^-}, \mathbf{n}_{t^-}\right)^c}{\sum_{m=0}^{p_t}\left\{Pr\left(\bigcap_{k\neq m}^{p_t} \beta_k < \beta_m | \mathbf{Y}_{t^-}, \mathbf{n}_{t^-}\right)^c\right\}}$$

- Following suggestion of Thall and Wathen (2007) we fix $c = 1/2$ in the simulation study

Simulated Platform Design: BAR



Randomization Methods

Summary of example simulation results

	(S_0, n_0)	Total Enrollment	Overall Success	Weeks
ER	$(20, 145)$	351	0.26	161
BAR	$(4, 17)$	135	0.39	60
PAR	$(15, 68)$	233	0.34	101
WBAR	$(5, 24)$	160	0.32	69

Decisions for experimental arms

	Exp. 4 $\pi = 0.1$	Exp. 1 $\pi = 0.2$	Exp. 6 $\pi = 0.3$	Exp. 2 $\pi = 0.4$	Exp. 3 $\pi = 0.5$	Exp. 5 $\pi = 0.6$
ER	Fail	Fail	Inconcl	Grad	Grad	Grad
BAR	Fail	Fail	Grad	Grad	Grad	Grad
PAR	Fail	Fail	Inconcl	Grad	Grad	Grad
WBAR	Fail	Fail	Fail	Grad	Grad	Grad