

Overview

- Biostatisticians in the drug and medical device industries are increasingly faced with data that are:
 - **highly multivariate**, with many important predictors and response variables
 - **temporally correlated** (longitudinal, survival studies)
 - **costly and difficult to obtain**, but often with **historical data** on previous but similar drugs or devices

Overview

- Biostatisticians in the drug and medical device industries are increasingly faced with data that are:
 - **highly multivariate**, with many important predictors and response variables
 - **temporally correlated** (longitudinal, survival studies)
 - **costly and difficult to obtain**, but often with **historical data** on previous but similar drugs or devices
- Recently, the FDA Center for Devices has encouraged **hierarchical Bayesian** statistical approaches –
 - Methods are not terribly novel: **Bayes (1763)!**
 - **But** their practical application has only become feasible in the last decade or so due to advances in computing via **Markov chain Monte Carlo** (MCMC) methods and related **WinBUGS** software

Role of Bayes in drug/device settings

- **Safety/efficacy studies:** Historical data and/or information from published literature can be used to reduce sample size, reducing time and expense. Unlimited looks at accumulating data are also permitted (due to different framework for testing).

Role of Bayes in drug/device settings

- **Safety/efficacy studies:** Historical data and/or information from published literature can be used to reduce sample size, reducing time and expense. Unlimited looks at accumulating data are also permitted (due to different framework for testing).
- **Equivalence studies:** Bayes allows one to make direct statements about the probability that one drug is equivalent to another, rather than merely “failing to reject” the hypothesis of no difference.

Role of Bayes in drug/device settings

- **Safety/efficacy studies:** Historical data and/or information from published literature can be used to reduce sample size, reducing time and expense. Unlimited looks at accumulating data are also permitted (due to different framework for testing).
- **Equivalence studies:** Bayes allows one to make direct statements about the probability that one drug is equivalent to another, rather than merely “failing to reject” the hypothesis of no difference.
- **Meta-analysis:** Bayes facilitates combining disparate but similar studies of a common drug or device.

Role of Bayes in drug/device settings

- **Safety/efficacy studies:** Historical data and/or information from published literature can be used to reduce sample size, reducing time and expense. Unlimited looks at accumulating data are also permitted (due to different framework for testing).
- **Equivalence studies:** Bayes allows one to make direct statements about the probability that one drug is equivalent to another, rather than merely “failing to reject” the hypothesis of no difference.
- **Meta-analysis:** Bayes facilitates combining disparate but similar studies of a common drug or device.
- **Hierarchical models:** Realistic models can be fit to complicated, multilevel data (e.g., multiple observations per patient, or multiple patients per clinical site), accounting for all sources of uncertainty.

Bayesian design of experiments

- In traditional sample size formulae, one often plugs in a “best guess” or “smallest clinically significant difference” for $\theta \Rightarrow$ “Everyone is a Bayesian at the design stage.”

Bayesian design of experiments

- In traditional sample size formulae, one often plugs in a “best guess” or “smallest clinically significant difference” for $\theta \Rightarrow$ “Everyone is a Bayesian at the design stage.”
- In practice, frequentist and Bayesian outlooks arise:

Bayesian design of experiments

- In traditional sample size formulae, one often plugs in a “best guess” or “smallest clinically significant difference” for $\theta \Rightarrow$ “Everyone is a Bayesian at the design stage.”
- In practice, frequentist and Bayesian outlooks arise:
 - Applicants may have a more Bayesian outlook:
 - to take advantage of historical data or expert opinion (and possibly stop the trial sooner), or
 - to “peek” at the accumulating data without affecting their ability to analyze it later

Bayesian design of experiments

- In traditional sample size formulae, one often plugs in a “best guess” or “smallest clinically significant difference” for $\theta \Rightarrow$ “Everyone is a Bayesian at the design stage.”
- In practice, frequentist and Bayesian outlooks arise:
 - Applicants may have a more Bayesian outlook:
 - to take advantage of historical data or expert opinion (and possibly stop the trial sooner), or
 - to “peek” at the accumulating data without affecting their ability to analyze it later
 - Regulatory agencies may appreciate this, but also retain many elements of frequentist thinking:
 - to ensure that in the long run they will only rarely approve a useless or harmful product

Bayesian design of experiments

- In traditional sample size formulae, one often plugs in a “best guess” or “smallest clinically significant difference” for $\theta \Rightarrow$ “Everyone is a Bayesian at the design stage.”
- In practice, frequentist and Bayesian outlooks arise:
 - Applicants may have a more Bayesian outlook:
 - to take advantage of historical data or expert opinion (and possibly stop the trial sooner), or
 - to “peek” at the accumulating data without affecting their ability to analyze it later
 - Regulatory agencies may appreciate this, but also retain many elements of frequentist thinking:
 - to ensure that in the long run they will only rarely approve a useless or harmful product

Applicants must thus design their trials accordingly!

Bayesian clinical trial design

- Again, the problem is essentially one of **sample size determination** using a Bayesian approach

Bayesian clinical trial design

- Again, the problem is essentially one of **sample size determination** using a Bayesian approach
- **Example 6.5: Safety Study B**, in which we must show freedom from severe drug-related adverse events (AEs) at 3 months will have a 95% lower confidence bound at least 85%.

Bayesian clinical trial design

- Again, the problem is essentially one of **sample size determination** using a Bayesian approach
- **Example 6.5: Safety Study B**, in which we must show freedom from severe drug-related adverse events (AEs) at 3 months will have a 95% lower confidence bound at least 85%.
- **Problem:** Using traditional statistical methods, we obtain an estimated sample size of over 100 – *too large!*

Bayesian clinical trial design

- Again, the problem is essentially one of **sample size determination** using a Bayesian approach
- **Example 6.5: Safety Study B**, in which we must show freedom from severe drug-related adverse events (AEs) at 3 months will have a 95% lower confidence bound at least 85%.
- **Problem:** Using traditional statistical methods, we obtain an estimated sample size of over 100 – *too large!*
- **But:** We have access to the following (1-month) data from Safety Study A:

	No AE	AE	total
count	110	7	117
(%)	(94)	(6)	

Bayesian clinical trial design

- Since we expect similar results in two studies, use Study A data for the prior \Rightarrow reduced sample size!

Bayesian clinical trial design

- Since we expect similar results in two studies, use Study A data for the prior \Rightarrow reduced sample size!
- **Model:** Suppose N patients in Study B, and for each,

$$\theta = \text{Pr}(\text{patient does not experience the AE})$$

Let $X = \#$ Study B patients with no AE (“successes”).
Assuming **independent** patients,

$$X|\theta \sim \text{Binomial}(N, \theta)$$

Bayesian clinical trial design

- Since we expect similar results in two studies, use Study A data for the prior \Rightarrow reduced sample size!
- **Model:** Suppose N patients in Study B, and for each,

$$\theta = \Pr(\text{patient does not experience the AE})$$

Let $X = \#$ Study B patients with no AE (“successes”).
Assuming **independent** patients,

$$X|\theta \sim \text{Binomial}(N, \theta)$$

- If the **prior** is $\theta \sim \text{Beta}(a, b)$, then the **posterior** is

$$\theta|X \sim \text{Beta}(X + a, N - X + b)$$

Note that a $\text{Beta}(a = 110, b = 7)$ prior (the **target** prior) delivers **equal weighting** of Studies A and B.

Bayesian clinical trial design

- Other prior possibilities:

Bayesian clinical trial design

- Other prior possibilities:
 - **Downweight** the prior sample size to $117w$, $0 \leq w \leq 1$ via a $Beta(110w, 7w)$ prior. This has the same overall success rate, but **decreases our confidence**: each “old” patient is only worth the fraction w of a new patient.

Bayesian clinical trial design

- Other prior possibilities:
 - **Downweight** the prior sample size to $117w$, $0 \leq w \leq 1$ via a $Beta(110w, 7w)$ prior. This has the same overall success rate, but **decreases our confidence**: each “old” patient is only worth the fraction w of a new patient.
 - **Shift** the target to a $Beta(110 + s, 7 - s)$ distribution for $0 \leq s < 7$. This has the same prior sample size (117), but shifts to more **optimistic** ($s > 0$) or **pessimistic** ($s < 0$) levels by increasing or decreasing the number of successes in the “old” dataset.

Bayesian clinical trial design

- To find optimal design:

Bayesian clinical trial design

- To find optimal design:
 - Draw θ_j from the prior, followed by X_j from the binomial likelihood, $j = 1, \dots, Nrep$.

Bayesian clinical trial design

- To find optimal design:
 - Draw θ_j from the prior, followed by X_j from the binomial likelihood, $j = 1, \dots, Nrep$.
 - Check to see if the 2.5% point of the simulated posterior is in fact greater than C .

Bayesian clinical trial design

- To find optimal design:
 - Draw θ_j from the prior, followed by X_j from the binomial likelihood, $j = 1, \dots, Nrep$.
 - Check to see if the 2.5% point of the simulated posterior is in fact greater than C .
 - The observed proportion of times this happens is the “Bayesian power”!

Bayesian clinical trial design

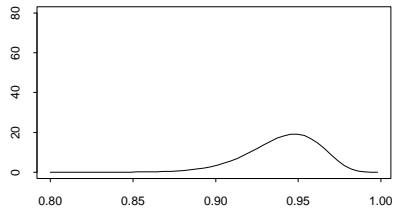
- To find optimal design:
 - Draw θ_j from the prior, followed by X_j from the binomial likelihood, $j = 1, \dots, Nrep$.
 - Check to see if the 2.5% point of the simulated posterior is in fact greater than C .
 - The observed proportion of times this happens is the “Bayesian power”!
- Repeat this over a grid of C values, several possible sample sizes N , and several priors (indexed by w or s). This then produces the “Bayesian sample size table”!

Bayesian clinical trial design

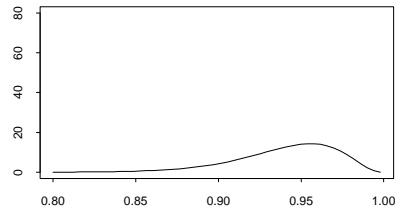
- To find optimal design:
 - Draw θ_j from the prior, followed by X_j from the binomial likelihood, $j = 1, \dots, Nrep$.
 - Check to see if the 2.5% point of the simulated posterior is in fact greater than C .
 - The observed proportion of times this happens is the “Bayesian power”!
- Repeat this over a grid of C values, several possible sample sizes N , and several priors (indexed by w or s). This then produces the “Bayesian sample size table”!
- The figure and table on the next two slides show the actual posteriors themselves and the Bayesian sample size table, respectively...

Results: Bayesian Design of Study B

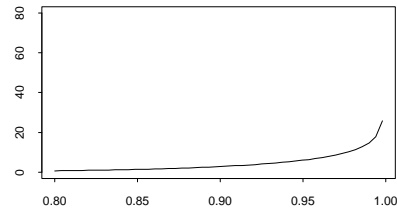
Priors and simulated posteriors, Chronicle B study, beta-binomial design
using weighted Beta(a,b) priors (target->downweighted)



Beta(110w,7w) prior with w = 1

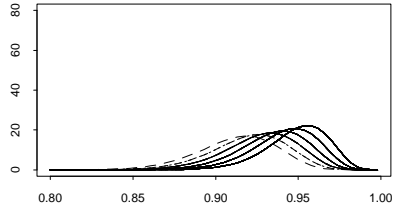


Beta(110w,7w) prior with w = 0.5

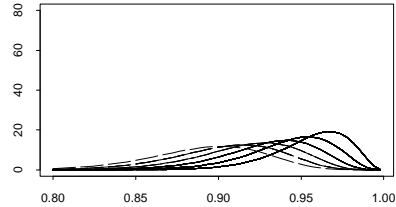


Beta(110w,7w) prior with w = 0.1

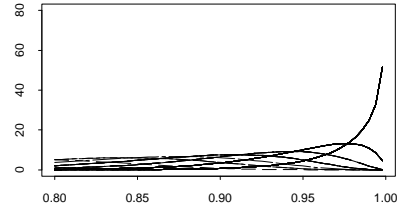
the three priors



Nrep= 100, N= 20, w = 1

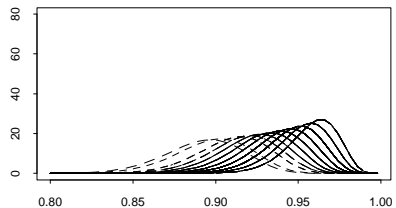


Nrep= 100, N= 20, w = 0.5

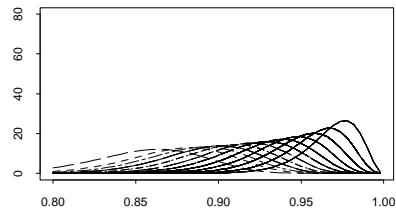


Nrep= 100, N= 20, w = 0.1

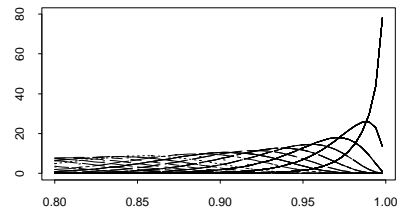
$N = 20$ posteriors



Nrep= 100, N= 50, w = 1



Nrep= 100, N= 50, w = 0.5



Nrep= 100, N= 50, w = 0.1

$N = 50$ posteriors

● S code to create this plot is available in
www.biostat.umn.edu/~brad/w.S

Results: Bayesian Design of Study B

lower limit	target		50% weight		10% weight	
	$N = 20$	$N = 50$	$N = 20$	$N = 50$	$N = 20$	$N = 50$
0.85	1.00	0.98	0.92	0.91	0.55	0.78
0.86	0.99	0.98	0.92	0.88	0.55	0.78
0.87	0.97	0.91	0.78	0.79	0.55	0.73
0.88	0.87	0.87	0.78	0.63	0.55	0.73
0.89	0.71	0.80	0.37	0.48	0.55	0.73
0.90	0.44	0.52	0.37	0.48	0.55	0.58

Posterior predictive probabilities that the 95% lower confidence bound will exceed the given limit, for

- two sample sizes: $N = 20$ and $N = 50$, and

Results: Bayesian Design of Study B

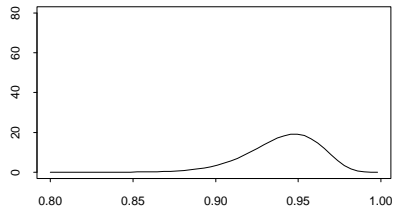
lower limit	target		50% weight		10% weight	
	$N = 20$	$N = 50$	$N = 20$	$N = 50$	$N = 20$	$N = 50$
0.85	1.00	0.98	0.92	0.91	0.55	0.78
0.86	0.99	0.98	0.92	0.88	0.55	0.78
0.87	0.97	0.91	0.78	0.79	0.55	0.73
0.88	0.87	0.87	0.78	0.63	0.55	0.73
0.89	0.71	0.80	0.37	0.48	0.55	0.73
0.90	0.44	0.52	0.37	0.48	0.55	0.58

Posterior predictive probabilities that the 95% lower confidence bound will exceed the given limit, for

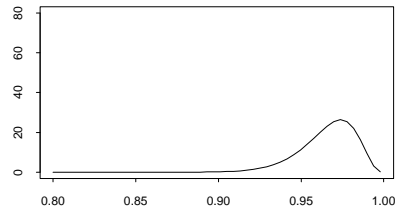
- two sample sizes: $N = 20$ and $N = 50$, and
- three $Beta(110w, 7w)$ priors: $w = 1$ (target), $w = 0.5$ (50% downweighted), and $w = 0.1$ (90% downweighted).

Results: Bayesian Design of Study B

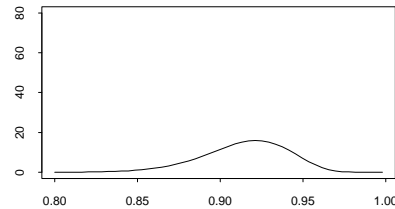
Priors and simulated posteriors, Chronicle B study, beta-binomial design using shifted Beta(a,b) priors ('target-optimistic-pessimistic')



Beta(110+s,7-s) prior with s = 0

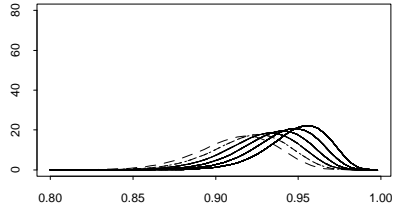


Beta(110+s,7-s) prior with s = 3

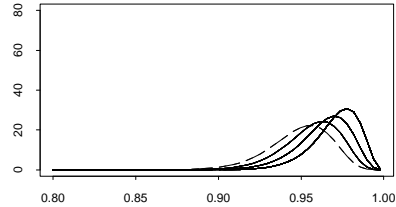


Beta(110+s,7-s) prior with s = -3

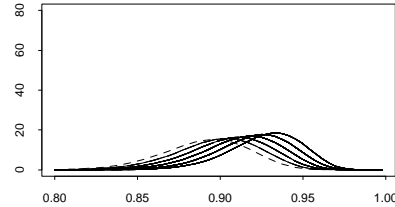
the three priors



Nrep= 100 . N= 20 . s = 0

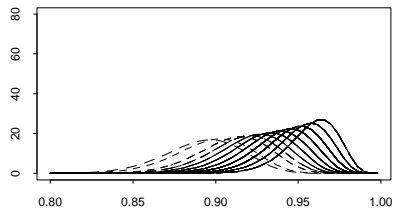


Nrep= 100 . N= 20 . s = 3

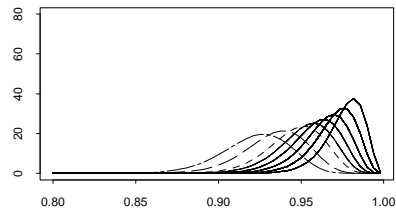


Nrep= 100 . N= 20 . s = -3

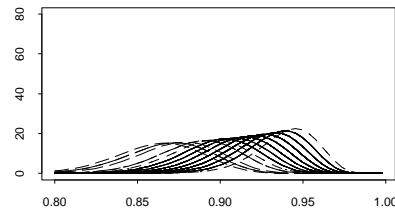
$N = 20$ posteriors



Nrep= 100 . N= 50 . s = 0



Nrep= 100 . N= 50 . s = 3



Nrep= 100 . N= 50 . s = -3

$N = 50$ posteriors

● S code to create this plot is available in

www.biostat.umn.edu/~brad/tlc.S

Results: Bayesian Design of Study B

lower limit	target		optimistic		pessimistic	
	$N = 20$	$N = 50$	$N = 20$	$N = 50$	$N = 20$	$N = 50$
0.85	1.00	0.98	1.00	1.00	0.94	0.82
0.86	0.99	0.98	1.00	1.00	0.54	0.75
0.87	0.97	0.91	1.00	1.00	0.22	0.45
0.88	0.87	0.87	1.00	0.99	0.00	0.25
0.89	0.71	0.80	1.00	0.99	0.00	0.12
0.90	0.44	0.52	1.00	0.98	0.00	0.00

Posterior predictive probabilities that the 95% lower confidence bound will exceed the given limit, for

- two sample sizes: $N = 20$ and $N = 50$, and

Results: Bayesian Design of Study B

lower limit	target		optimistic		pessimistic	
	$N = 20$	$N = 50$	$N = 20$	$N = 50$	$N = 20$	$N = 50$
0.85	1.00	0.98	1.00	1.00	0.94	0.82
0.86	0.99	0.98	1.00	1.00	0.54	0.75
0.87	0.97	0.91	1.00	1.00	0.22	0.45
0.88	0.87	0.87	1.00	0.99	0.00	0.25
0.89	0.71	0.80	1.00	0.99	0.00	0.12
0.90	0.44	0.52	1.00	0.98	0.00	0.00

Posterior predictive probabilities that the 95% lower confidence bound will exceed the given limit, for

- two sample sizes: $N = 20$ and $N = 50$, and
- three $Beta(110 + s, 7 - s)$ priors: $s = 0$ (target), $s = 3$ (optimistic), and $s = -3$ (pessimistic).

Addendum: Is our prior “ethical”?

- Based only on the $Beta(110w, 7w)$ prior with $w = 0.75$, the 95% lower confidence bound is $\theta_L = 0.882$. Also,

$$E_X[P(\theta_L > 0.88|X)] = 0.80 \text{ with } N = 50 .$$

But notice the prior **already** satisfies this condition; there is no need to collect more data \Rightarrow prior is “**unethical**”!

Addendum: Is our prior “ethical”?

- Based only on the $Beta(110w, 7w)$ prior with $w = 0.75$, the 95% lower confidence bound is $\theta_L = 0.882$. Also,

$$E_X[P(\theta_L > 0.88|X)] = 0.80 \quad \text{with } N = 50 .$$

But notice the prior **already** satisfies this condition; there is no need to collect more data \Rightarrow prior is “**unethical**”!

- **BUT:** It is permissible (and sensible) to use a **different** prior in the design and analysis stages! With $w = .60$,

$$\theta_L = 0.874 < 0.88 \quad (\text{prior is now ethical})$$

Now generate future (θ_j, X_j) pairs using the **unweighted** $Beta(110, 7)$ prior (i.e., future data will look **exactly** like the old data) \Rightarrow revised Bayesian power statement

$$E_X[P(\theta_L > 0.88|X)] = 0.81 \quad \text{with } N = 50 .$$

Software for Bayesian Clinical Trials

- BART and i BART, based on the [range of equivalence](#) ideas of Spiegelhalter et al. (see more below...)

Software for Bayesian Clinical Trials

- BART and iBART, based on the **range of equivalence** ideas of Spiegelhalter et al. (see more below...)
- Quite an assortment from **MD Anderson Cancer Center**:
 - CRMSimulator: a simplified, primarily pedagogical continual reassessment method program
 - EffTox: dose-finding based on efficacy and toxicity outcomes (Thall and Cook, 2004)
 - Multic Lean: a “lean” (just two outcomes, efficacy and toxicity) implementation of multiple comparison safety monitoring (Thall, Simon and Estey, 1995)
 - Predictive Probabilities: of superiority, futility, etc. for trials with binary or time-to-event outcomes – useful for **design** and **interim analysis**

Software for Bayesian Clinical Trials

- BART and iBART, based on the **range of equivalence** ideas of Spiegelhalter et al. (see more below...)
- Quite an assortment from **MD Anderson Cancer Center**:
 - CRMSimulator: a simplified, primarily pedagogical continual reassessment method program
 - EffTox: dose-finding based on efficacy and toxicity outcomes (Thall and Cook, 2004)
 - Multic Lean: a “lean” (just two outcomes, efficacy and toxicity) implementation of multiple comparison safety monitoring (Thall, Simon and Estey, 1995)
 - Predictive Probabilities: of superiority, futility, etc. for trials with binary or time-to-event outcomes – useful for **design** and **interim analysis**
- All have well-developed (often Windows) user interfaces, but none appear to be MCMC-driven.

Other relevant published work

- O'Hagan and Stevens (2001, *Medical Decision Making*): Bayesian design to demonstrate cost-effectiveness of a treatment using simple models (no MCMC) and **enormous** assumptions re: utility functions and likelihood/prior covariance matrices!

Other relevant published work

- O'Hagan and Stevens (2001, *Medical Decision Making*): Bayesian design to demonstrate cost-effectiveness of a treatment using simple models (no MCMC) and **enormous** assumptions re: utility functions and likelihood/prior covariance matrices!
- Wang and Gelfand (2002, *Statistical Science*): MCMC-based sample size determination for model performance or for distinguishing between two models, but where utility functions (as in e.g. Müller and Parmigiani, 1995) are not considered.

Other relevant published work

- O'Hagan and Stevens (2001, *Medical Decision Making*): Bayesian design to demonstrate cost-effectiveness of a treatment using simple models (no MCMC) and **enormous** assumptions re: utility functions and likelihood/prior covariance matrices!
- Wang and Gelfand (2002, *Statistical Science*): MCMC-based sample size determination for model performance or for distinguishing between two models, but where utility functions (as in e.g. Müller and Parmigiani, 1995) are not considered.
- and **several** other papers/short courses in the last decade or so, much of it by **MD Anderson staff!**

Other relevant published work

- O'Hagan and Stevens (2001, *Medical Decision Making*): Bayesian design to demonstrate cost-effectiveness of a treatment using simple models (no MCMC) and **enormous** assumptions re: utility functions and likelihood/prior covariance matrices!
- Wang and Gelfand (2002, *Statistical Science*): MCMC-based sample size determination for model performance or for distinguishing between two models, but where utility functions (as in e.g. Müller and Parmigiani, 1995) are not considered.
- and **several** other papers/short courses in the last decade or so, much of it by **MD Anderson staff!**
- Ideas are there, but practitioners need

Other relevant published work

- O'Hagan and Stevens (2001, *Medical Decision Making*): Bayesian design to demonstrate cost-effectiveness of a treatment using simple models (no MCMC) and **enormous** assumptions re: utility functions and likelihood/prior covariance matrices!
- Wang and Gelfand (2002, *Statistical Science*): MCMC-based sample size determination for model performance or for distinguishing between two models, but where utility functions (as in e.g. Müller and Parmigiani, 1995) are not considered.
- and **several** other papers/short courses in the last decade or so, much of it by **MD Anderson staff!**
- Ideas are there, but practitioners need
 - easy-to-follow “rulebook,” laying out the key issues

Other relevant published work

- O'Hagan and Stevens (2001, *Medical Decision Making*): Bayesian design to demonstrate cost-effectiveness of a treatment using simple models (no MCMC) and **enormous** assumptions re: utility functions and likelihood/prior covariance matrices!
- Wang and Gelfand (2002, *Statistical Science*): MCMC-based sample size determination for model performance or for distinguishing between two models, but where utility functions (as in e.g. Müller and Parmigiani, 1995) are not considered.
- and **several** other papers/short courses in the last decade or so, much of it by **MD Anderson staff!**
- Ideas are there, but practitioners need
 - easy-to-follow “rulebook,” laying out the key issues
 - corresponding suite of easy-to-use software!

Advanced models: Adding MCMC

- Previous methods and R software are fine when the posterior is available in closed form, as in previous beta/binomial and Dirichlet/multinomial models.

Advanced models: Adding MCMC

- Previous methods and R software are fine when the posterior is available in closed form, as in previous beta/binomial and Dirichlet/multinomial models.
- **But** many models in clinical trial design and analysis will require **Markov chain Monte Carlo (MCMC)** computational methods to **sample from** the posterior.

Advanced models: Adding MCMC

- Previous methods and R software are fine when the posterior is available in closed form, as in previous beta/binomial and Dirichlet/multinomial models.
- **But** many models in clinical trial design and analysis will require **Markov chain Monte Carlo (MCMC)** computational methods to **sample from** the posterior.
- Most popular software package for this: **WinBUGS**
 - Uses **R**-like syntax to specify models
 - Examples manual includes survival models (Weibull and Cox), longitudinal models, bioequivalence, meta-analysis, and others of biostatistical interest
 - freely available from <http://www.mrc-bsu.cam.ac.uk/bugs/welcome.shtml>

Example: Weibull survival model

- Let t_i be the time until death for subject i , with corresponding treatment indicator x_i ($= 0$ or 1 for control and treatment, respectively). Suppose

$$t_i \sim \text{Weibull}(r, \mu_i), \quad \text{where } \mu_i = e^{-(\beta_0 + \beta_1 x_i)} .$$

Example: Weibull survival model

- Let t_i be the time until death for subject i , with corresponding treatment indicator x_i ($= 0$ or 1 for control and treatment, respectively). Suppose

$$t_i \sim \text{Weibull}(r, \mu_i), \quad \text{where } \mu_i = e^{-(\beta_0 + \beta_1 x_i)} .$$

- Then the baseline hazard function is $\lambda_0(t_i) = r t_i^{r-1}$, and the median survival time for subject i is

$$m_i = [(\log 2) e^{\beta_0 + \beta_1 x_i}]^{1/r} .$$

Example: Weibull survival model

- Let t_i be the time until death for subject i , with corresponding treatment indicator x_i ($= 0$ or 1 for control and treatment, respectively). Suppose

$$t_i \sim \text{Weibull}(r, \mu_i), \quad \text{where } \mu_i = e^{-(\beta_0 + \beta_1 x_i)} .$$

- Then the baseline hazard function is $\lambda_0(t_i) = r t_i^{r-1}$, and the median survival time for subject i is

$$m_i = [(\log 2) e^{\beta_0 + \beta_1 x_i}]^{1/r} .$$

- The value of β_1 corresponding to a **15% increase in median survival in the treatment group** satisfies

$$e^{\beta_1/r} = 1.15 \iff \beta_1 = r \log(1.15) .$$

Range of equivalence

- The range of β_1 values within which we are indifferent as to use of treatment or control

Range of equivalence

- The range of β_1 values within which we are indifferent as to use of treatment or control
- lower limit β_I , the **clinical inferiority** boundary
 - We typically take $\beta_I = 0$, since we would never prefer a harmful treatment

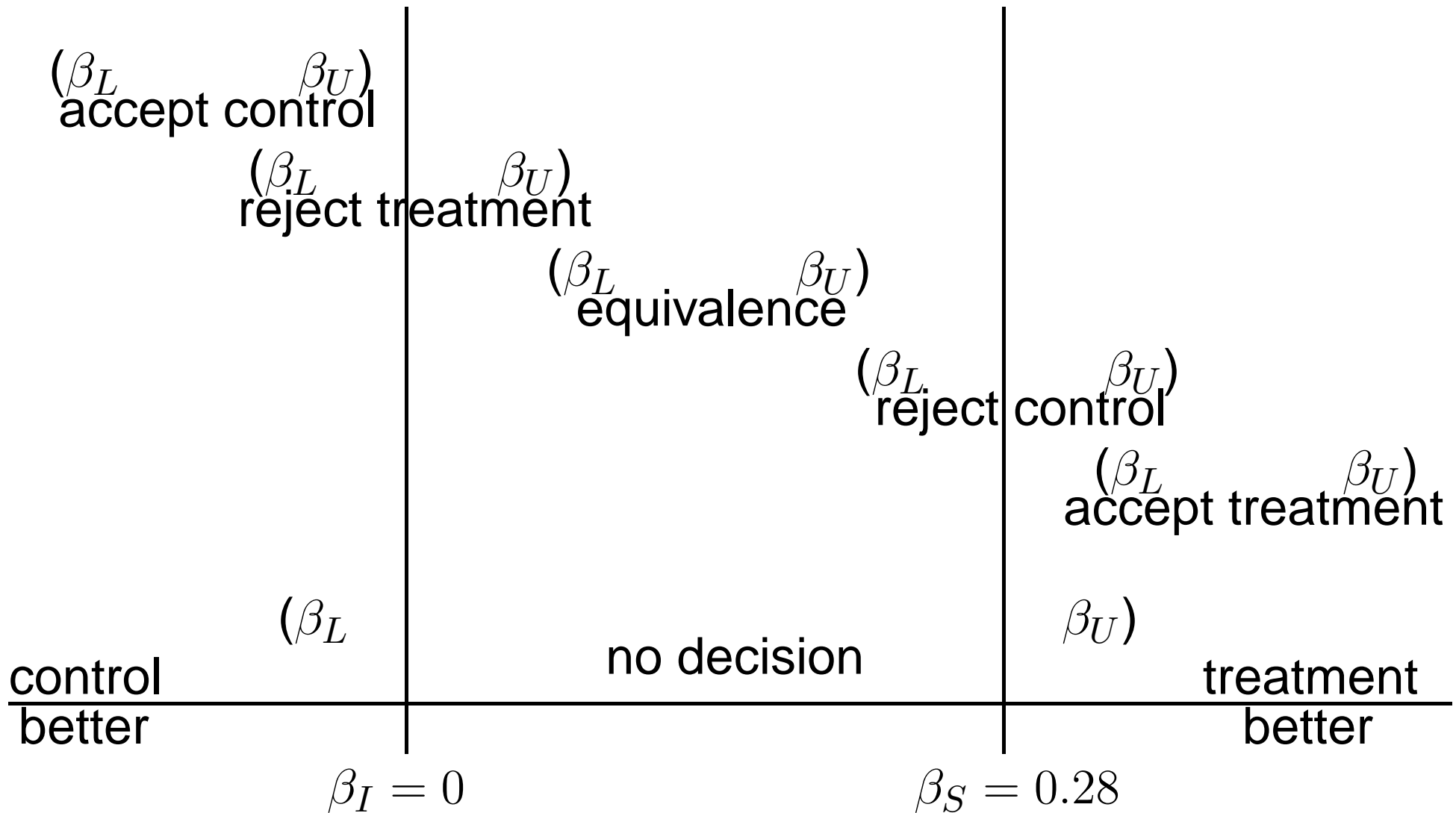
Range of equivalence

- The range of β_1 values within which we are indifferent as to use of treatment or control
- lower limit β_I , the clinical inferiority boundary
 - We typically take $\beta_I = 0$, since we would never prefer a harmful treatment
- upper limit β_S , the clinical superiority boundary
 - We typically take $\beta_S > 0$, since we may require “clinically significant” improvement under the treatment (due to cost, toxicity, etc.)
 - **Example:** If $r = 2$, then $\beta_S = 2 \log(1.15) \approx 0.28$ corresponds to 15% improvement in median survival

Range of equivalence

- The range of β_1 values within which we are indifferent as to use of treatment or control
- lower limit β_I , the clinical inferiority boundary
 - We typically take $\beta_I = 0$, since we would never prefer a harmful treatment
- upper limit β_S , the clinical superiority boundary
 - We typically take $\beta_S > 0$, since we may require “clinically significant” improvement under the treatment (due to cost, toxicity, etc.)
 - **Example:** If $r = 2$, then $\beta_S = 2 \log(1.15) \approx 0.28$ corresponds to 15% improvement in median survival
- The outcome of the trial can then be based on the location of the 95% posterior confidence interval for β_1 , say (β_L, β_U) , relative to the indifference zone!....

The six possible outcomes and decisions



- Note both "acceptance" and "rejection" are possible!

Community of priors

Spiegelhalter et al. (1994) recommend considering several priors, in order to represent the broadest possible audience:

- **Skeptical** Prior

Community of priors

Spiegelhalter et al. (1994) recommend considering several priors, in order to represent the broadest possible audience:

- **Skeptical** Prior

- One that believes the treatment is likely no better than control (as might be believed by a regulatory agency)

Community of priors

Spiegelhalter et al. (1994) recommend considering several priors, in order to represent the broadest possible audience:

- **Skeptical Prior**

- One that believes the treatment is likely no better than control (as might be believed by a regulatory agency)
- Perhaps obtained by taking mean 0 and setting $P(\beta_1 > \beta_S) = \epsilon$; in our setting $\epsilon = .05$ delivers a $N(0, 0.03)$.

Community of priors

Spiegelhalter et al. (1994) recommend considering several priors, in order to represent the broadest possible audience:

- **Skeptical** Prior

- One that believes the treatment is likely no better than control (as might be believed by a regulatory agency)
- Perhaps obtained by taking mean 0 and setting $P(\beta_1 > \beta_S) = \epsilon$; in our setting $\epsilon = .05$ delivers a $N(0, 0.03)$.

- **Enthusiastic (or Clinical)** Prior

Community of priors

Spiegelhalter et al. (1994) recommend considering several priors, in order to represent the broadest possible audience:

- **Skeptical** Prior

- One that believes the treatment is likely no better than control (as might be believed by a regulatory agency)
- Perhaps obtained by taking mean 0 and setting $P(\beta_1 > \beta_S) = \epsilon$; in our setting $\epsilon = .05$ delivers a $N(0, 0.03)$.

- **Enthusiastic (or Clinical)** Prior

- One that believes the treatment will succeed (typical of the clinicians running the trial)

Community of priors

Spiegelhalter et al. (1994) recommend considering several priors, in order to represent the broadest possible audience:

● Skeptical Prior

- One that believes the treatment is likely no better than control (as might be believed by a regulatory agency)
- Perhaps obtained by taking mean 0 and setting $P(\beta_1 > \beta_S) = \epsilon$; in our setting $\epsilon = .05$ delivers a $N(0, 0.03)$.

● Enthusiastic (or Clinical) Prior

- One that believes the treatment will succeed (typical of the clinicians running the trial)
- Perhaps obtained by taking mean β_S and the same variance as the skeptical prior; in our setting this delivers a $N(0.28, 0.03)$.

Community of priors (cont'd)

- Reference (or Noninformative) Prior

Community of priors (cont'd)

- Reference (or Noninformative) Prior
 - One that expresses no particular opinion about the treatment's merit

Community of priors (cont'd)

- Reference (or Noninformative) Prior
 - One that expresses no particular opinion about the treatment's merit
 - Often a improper uniform ("flat") prior is permissible

Community of priors (cont'd)

- Reference (or Noninformative) Prior
 - One that expresses no particular opinion about the treatment's merit
 - Often a **improper uniform ("flat")** prior is permissible
-
- Note it may be sensible to match the prior to the decision one hopes to reach; the prior should represent **"an adversary who will need to be disillusioned by the data to stop further experimentation"**. Thus:

Community of priors (cont'd)

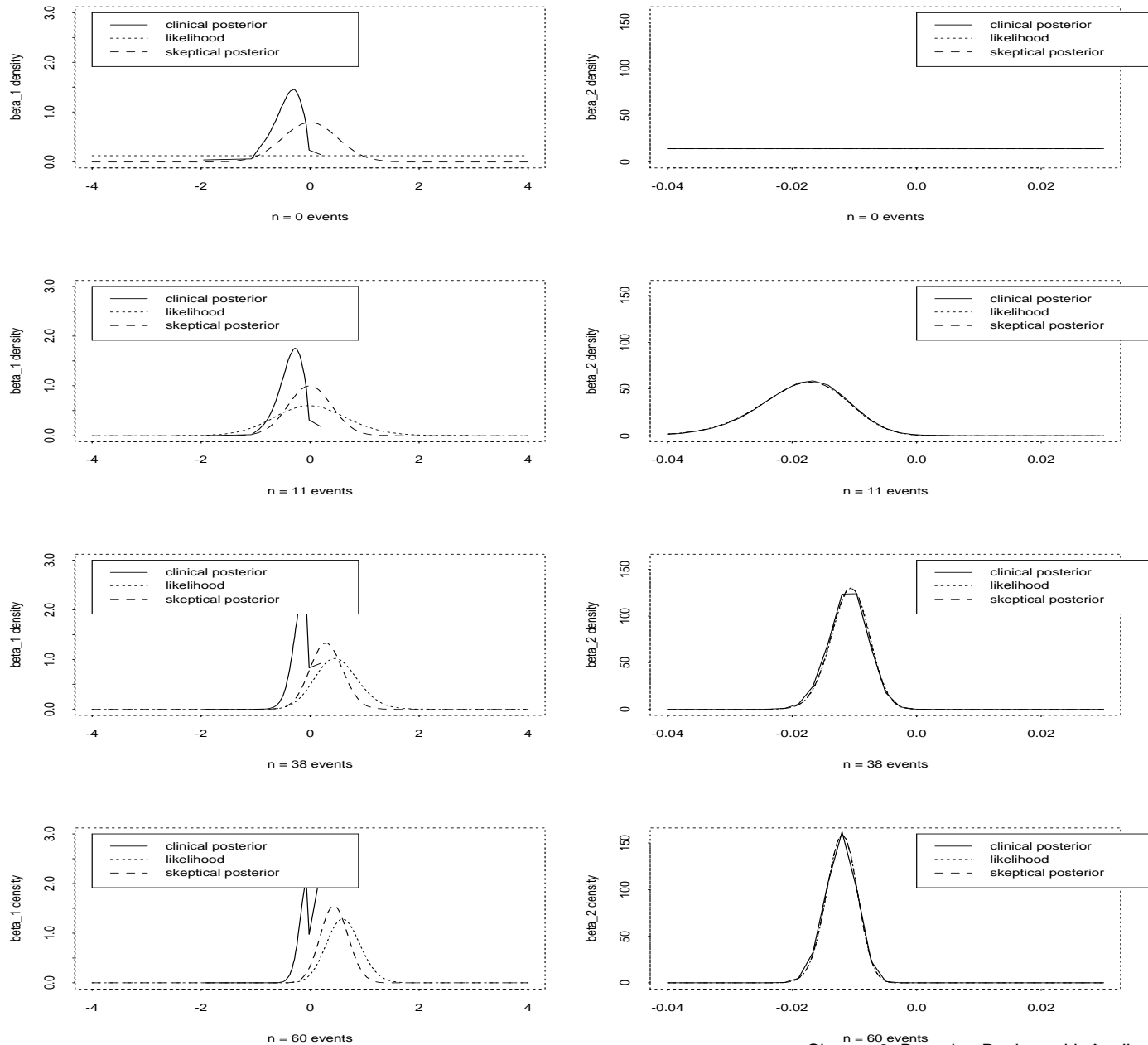
- Reference (or Noninformative) Prior
 - One that expresses no particular opinion about the treatment's merit
 - Often a **improper uniform** ("flat") prior is permissible
-
- Note it may be sensible to match the prior to the decision one hopes to reach; the prior should represent **"an adversary who will need to be disillusioned by the data to stop further experimentation"**. Thus:
 - To conclude a treatment difference, use the **skeptical** prior

Community of priors (cont'd)

- Reference (or Noninformative) Prior
 - One that expresses no particular opinion about the treatment's merit
 - Often a **improper uniform** ("flat") prior is permissible
-
- Note it may be sensible to match the prior to the decision one hopes to reach; the prior should represent "an adversary who will need to be disillusioned by the data to stop further experimentation". Thus:
 - To conclude a treatment difference, use the **skeptical** prior
 - To conclude no difference, use the **enthusiastic** prior

Monitoring plots: full posteriors

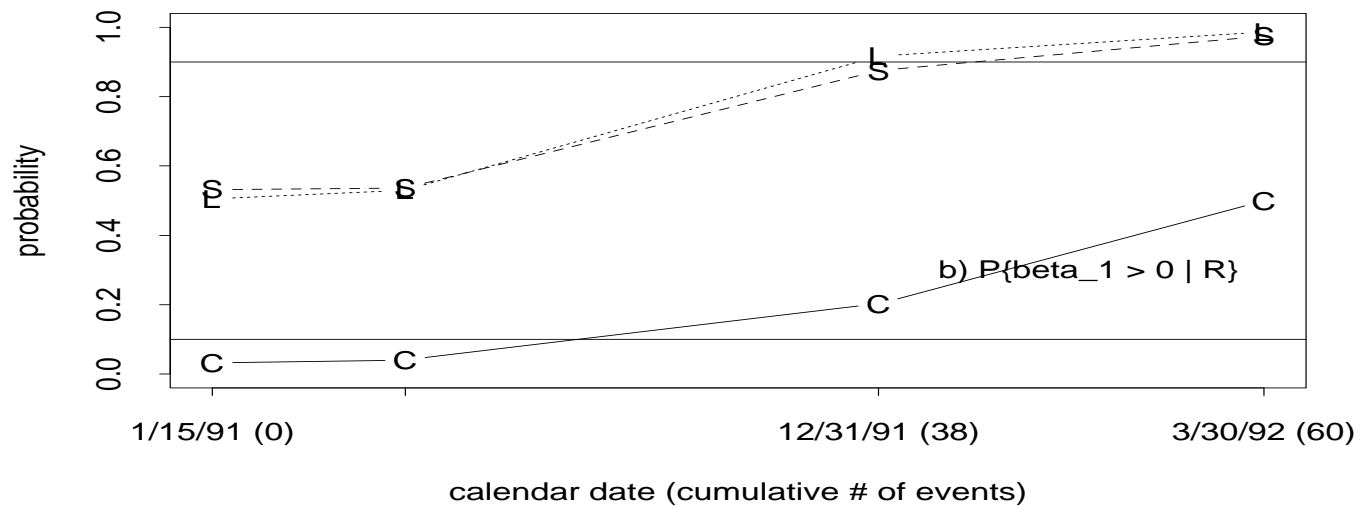
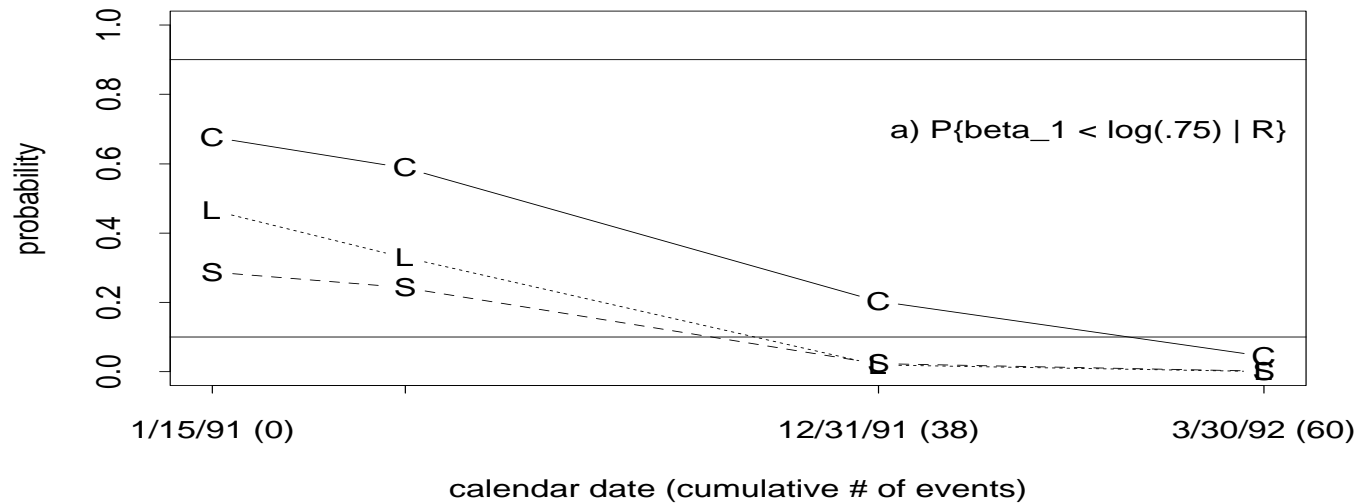
Posterior distributions; Covariate = Baseline CD4 Count
Monitoring dates = (1/15/91, 7/31/91, 12/31/91, 3/30/92)



Monitoring plots: tail areas

Posterior monitoring plot for beta_1; Covariate = Baseline CD4 Count

(C = clinical posterior, L = likelihood, S = skeptical posterior)



MCMC-based Bayesian design in BRugs

Within R, simulating the power or other operating characteristics in this setting works the same as before:

- Sample “true” β values from an assumed “true prior” (skeptical, enthusiastic, or in between)

MCMC-based Bayesian design in BRugs

Within R, simulating the power or other operating characteristics in this setting works the same as before:

- Sample “true” β values from an assumed “true prior” (skeptical, enthusiastic, or in between)
- Given these, sample fake survival times t_i (say, N from each study group) from the Weibull

MCMC-based Bayesian design in BRugs

Within R, simulating the power or other operating characteristics in this setting works the same as before:

- Sample “true” β values from an assumed “true prior” (skeptical, enthusiastic, or in between)
- Given these, sample fake survival times t_i (say, N from each study group) from the Weibull
- We may also wish to sample fake **censoring** times c_i from a particular distribution (e.g., a normal truncated below 0); for all i such that $t_i > c_i$, replace t_i by “NA”

MCMC-based Bayesian design in BRugs

Within R, simulating the power or other operating characteristics in this setting works the same as before:

- Sample “true” β values from an assumed “true prior” (skeptical, enthusiastic, or in between)
- Given these, sample fake survival times t_i (say, N from each study group) from the Weibull
- We may also wish to sample fake **censoring** times c_i from a particular distribution (e.g., a normal truncated below 0); for all i such that $t_i > c_i$, replace t_i by “NA”
- Compute (β_L, β_U) using BRugs commands within R

MCMC-based Bayesian design in BRugs

Within R, simulating the power or other operating characteristics in this setting works the same as before:

- Sample “true” β values from an assumed “true prior” (skeptical, enthusiastic, or in between)
- Given these, sample fake survival times t_i (say, N from each study group) from the Weibull
- We may also wish to sample fake **censoring** times c_i from a particular distribution (e.g., a normal truncated below 0); for all i such that $t_i > c_i$, replace t_i by “NA”
- Compute (β_L, β_U) **using BRugs commands within R**
- Determine the simulated trial’s outcome based on location of (β_L, β_U) relative to the indifference zone

MCMC-based Bayesian design in BRugs

Within R, simulating the power or other operating characteristics in this setting works the same as before:

- Sample “true” β values from an assumed “true prior” (skeptical, enthusiastic, or in between)
- Given these, sample fake survival times t_i (say, N from each study group) from the Weibull
- We may also wish to sample fake **censoring** times c_i from a particular distribution (e.g., a normal truncated below 0); for all i such that $t_i > c_i$, replace t_i by “NA”
- Compute (β_L, β_U) **using BRugs commands within R**
- Determine the simulated trial’s outcome based on location of (β_L, β_U) relative to the indifference zone
- Repeat this process $Nrep$ times; report empirical frequencies of the six possible outcomes

Bayesian power calculation in BRugs

We will likely wish to repeat the **entire** process for several sample sizes N and several priors.

- A **Bayesian power** calculation here might arise from using the **enthusiastic** prior as the “truth”

Bayesian power calculation in BRugs

We will likely wish to repeat the **entire** process for several sample sizes N and several priors.

- A **Bayesian power** calculation here might arise from using the **enthusiastic** prior as the “truth”
- For $Nrep = 1000$ (and using 100 burn-in and 1000 production MCMC iterations in each BRugs call), we obtained the following probabilities of rejecting the control when the enthusiastic prior is true:

N	Skeptical	Reference	Enthusiastic
25	.014	.207	.475
50	.087	.352	.615
75	.191	.378	.652
100	.288	.472	.682

Power increases with N and/or prior enthusiasm!

Type I error rate calculation

- A Bayesian version of this calculation would arise similar to the method of the previous slide, but now assuming the **skeptical** prior is true

Type I error rate calculation

- A Bayesian version of this calculation would arise similar to the method of the previous slide, but now assuming the **skeptical** prior is true
- A true frequentist Type I error calculation is also possible: simply **fix** $\beta_1 = 0$, and generate **only** the t_i and c_i for each of the $Nrep$ iterations.

Type I error rate calculation

- A Bayesian version of this calculation would arise similar to the method of the previous slide, but now assuming the **skeptical** prior is true
- A true frequentist Type I error calculation is also possible: simply **fix** $\beta_1 = 0$, and generate **only** the t_i and c_i for each of the $Nrep$ iterations.
- Note that while Bayesians are free to look at their data at any time without affecting the inference, multiple looks **will alter the frequentist Type I error behavior of the procedure**. If this is of interest, the algorithm must be modified to explicitly include these multiple looks, checking for early stopping after each look.

Type I error rate calculation

- A Bayesian version of this calculation would arise similar to the method of the previous slide, but now assuming the **skeptical** prior is true
- A true frequentist Type I error calculation is also possible: simply **fix** $\beta_1 = 0$, and generate **only** the t_i and c_i for each of the $Nrep$ iterations.
- Note that while Bayesians are free to look at their data at any time without affecting the inference, multiple looks **will alter the frequentist Type I error behavior of the procedure**. If this is of interest, the algorithm must be modified to explicitly include these multiple looks, checking for early stopping after each look.
- Early stopping for futility based on **predictive distributions** (“Bayesian stochastic curtailment”) may also be of interest – see Berry and Berry (2004)!

Example: Weibull model in BRugs

- **BRugs** is a suite of R routines written by WinBUGS head programmer Andrew Thomas for calling OpenBUGS from R; see mathstat.helsinki.fi/openbugs/

Example: Weibull model in BRugs

- **BRugs** is a suite of R routines written by WinBUGS head programmer Andrew Thomas for calling OpenBUGS from R; see mathstat.helsinki.fi/openbugs/
- Given web access, BRugs may be installed directly within R: **Packages – Install package(s) – CRAN mirror – BRugs – easy!**

Example: Weibull model in BRugs

- **BRugs** is a suite of R routines written by WinBUGS head programmer Andrew Thomas for calling OpenBUGS from R; see mathstat.helsinki.fi/openbugs/
- Given web access, BRugs may be installed directly within R: **Packages – Install package(s) – CRAN mirror – BRugs – easy!**
- The R program www.biostat.umn.edu/~brad/software/BRugs/Power.BRugs sources in the BRugs commands, generates the fake β and data values, calls OpenBUGS to compute (β_L, β_U)

Example: Weibull model in BRugs

- **BRugs** is a suite of R routines written by WinBUGS head programmer Andrew Thomas for calling OpenBUGS from R; see mathstat.helsinki.fi/openbugs/
- Given web access, BRugs may be installed directly within R: **Packages – Install package(s) – CRAN mirror – BRugs – easy!**
- The R program www.biostat.umn.edu/~brad/software/BRugs/Power.BRugs sources in the BRugs commands, generates the fake β and data values, calls OpenBUGS to compute (β_L, β_U)
- The OpenBUGS program www.biostat.umn.edu/~brad/software/BRugs/refmodel.txt passes the Weibull model and flat prior to OpenBUGS; **data** and **inits** files are also passed after creation in R

Power.BRugs (R code)

```
for (i in 1:nrep) {      # outer data simulation loop

# [Sample true parameters;
#  sample fake data (survival and censoring times) given parameters]

mydata <- pairlist(t = T, t.cens = T.cens, x = X, n = 2*N)
dput(mydata, "C:/joe/PowerCalcs/powerdata.txt")

modelCheck("C:/joe/PowerCalcs/refmodel.txt")
modelData("C:/joe/PowerCalcs/powerdata.txt")
modelCompile()
modelInits("C:/joe/PowerCalcs/powerinits.txt")
modelGenInits()

modelUpdate(100)
samplesSet("beta1")
dicSet()
modelUpdate(1000)
samplesAutoC("beta1", chain=1)
dicStats()
samplesDensity("beta1")
```

Power.BRugs (R code, cont'd)

```
LL <- samplesStats("beta1")$val2.5pc
UL <- samplesStats("beta1")$val97.5pc

if (UL < 0) acccontrol <- acccontrol + 1
  else
if (LL > hypbeta1) acctrtr <- acctrtr + 1
  else
if ((UL < hypbeta1) & (LL > 0)) equiv <- equiv + 1
  else
if ((UL < hypbeta1) & (LL < 0)) rejtrtr <- rejtrtr + 1
  else
if ((UL > hypbeta1) & (LL > 0)) rejcontrol <- rejcontrol + 1
  else
nodec <- nodec + 1

# Bind the summary statistics of the current iteration to beta1stat:
beta1stat <- rbind(beta1stat, stats(beta1))

} # end of outer data simulation loop

# [Write simulated power summaries to the screen]
```

refmodel.txt (BUGS code)

```
model {  
  
  for (i in 1:n) {  
    t[i] ~ dweib(2, mu[i]) I(t.cens[i], )  
    mu[i] <- exp(-beta0 - beta1*x[i])  
  }  
  
  beta0 ~ dnorm( 7.53, 25)  
  beta1 ~ dnorm(0, .0001) #non-informative  
  
}
```

Results

- Assuming:
 - Weibull shape $r = 2$, and $N = 50$ in each group
 - median survival of 36 days with 50% improvement in the treatment group
 - a $N(80, 20)$ censoring distribution
 - the enthusiastic prior as the “truth”

We obtain the following output from $Nrep = 100$ reps:

Results

- Assuming:
 - Weibull shape $r = 2$, and $N = 50$ in each group
 - median survival of 36 days with 50% improvement in the treatment group
 - a $N(80, 20)$ censoring distribution
 - the enthusiastic prior as the “truth”

We obtain the following output from $Nrep = 100$ reps:

- Here are simulated outcome frequencies for $N = 50$
 - accept control: 0
 - reject treatment: 0.07
 - equivalence: 0
 - reject control: 0.87
 - accept treatment: 0.06
 - no decision: 0
- End of BRugs power simulation