

Software for Bayesian Adaptive Drug and Device Trials

ENAR 2011 Tutorial

Miami, Florida, March 21, 2011

presented by

Bradley P. Carlin

University of Minnesota

ably assisted by

Laura Hatfield, Harrison Quick, and Lindsay Renfro

Textbook companions for this tutorial

- Recommended (“BCLM”): *Bayesian Adaptive Methods for Clinical Trials* (ISBN 978-1-4398-2548-8) by S.M. Berry, B.P. Carlin, J.J. Lee, and P. Müller, Boca Raton, FL: Chapman and Hall/CRC Press, August 2010.
- Other books of interest:
 - Your favorite math stat and linear models books
 - *Bayesian Methods for Data Analysis*, 3rd ed., by B.P. Carlin and T.A. Louis, Boca Raton, FL: Chapman and Hall/CRC Press, 2009.
 - *Bayesian Approaches to Clinical Trials and Health-Care Evaluation*, by D.J. Spiegelhalter, K.R. Abrams, and J.P. Myles: Chichester: John Wiley & Sons, 2004.
 - *Bayesian Computation with R*, by J.H. Albert, New York: Springer, 2007.

Ch 2: Basics of Bayesian Inference

- As usual, we start with a **likelihood** (or **model**) $f(\mathbf{y}|\boldsymbol{\theta})$ for the observed data $\mathbf{y} = (y_1, \dots, y_n)$ given the unknown parameters $\boldsymbol{\theta} = (\theta_1, \dots, \theta_K)$
- Add a **prior** distribution $\pi(\boldsymbol{\theta}|\boldsymbol{\lambda})$, where $\boldsymbol{\lambda}$ is a vector of **hyperparameters**.
- The **posterior** distribution for $\boldsymbol{\theta}$ is given by

$$\begin{aligned} p(\boldsymbol{\theta}|\mathbf{y}, \boldsymbol{\lambda}) &= \frac{p(\mathbf{y}, \boldsymbol{\theta}|\boldsymbol{\lambda})}{p(\mathbf{y}|\boldsymbol{\lambda})} = \frac{p(\mathbf{y}, \boldsymbol{\theta}|\boldsymbol{\lambda})}{\int p(\mathbf{y}, \boldsymbol{\theta}|\boldsymbol{\lambda}) d\boldsymbol{\theta}} \\ &= \frac{f(\mathbf{y}|\boldsymbol{\theta})\pi(\boldsymbol{\theta}|\boldsymbol{\lambda})}{\int f(\mathbf{y}|\boldsymbol{\theta})\pi(\boldsymbol{\theta}|\boldsymbol{\lambda}) d\boldsymbol{\theta}} = \frac{f(\mathbf{y}|\boldsymbol{\theta})\pi(\boldsymbol{\theta}|\boldsymbol{\lambda})}{m(\mathbf{y}|\boldsymbol{\lambda})}. \end{aligned}$$

We refer to this formula as *Bayes' Theorem*.

Basics of Bayesian Inference

- Since λ will usually not be known, a second stage (**hyperprior**) distribution $h(\lambda)$ will be required, so that

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

- Alternatively, we might replace λ in $p(\boldsymbol{\theta}|\mathbf{y}, \lambda)$ by an estimate $\hat{\lambda}$; this is called **empirical Bayes** analysis
- For prediction of a future value y_{n+1} , we would use the **predictive** distribution,

$$p(y_{n+1}|\mathbf{y}) = \int p(y_{n+1}|\boldsymbol{\theta})p(\boldsymbol{\theta}|\mathbf{y})d\boldsymbol{\theta} ,$$

which is nothing but the posterior of y_{n+1} .

Gibbs sampling

- **Gibbs Sampler:** Suppose the joint distribution of $\theta = (\theta_1, \dots, \theta_K)$ is uniquely determined by the **full conditional distributions**, $\{p_i(\theta_i | \theta_{j \neq i}), i = 1, \dots, K\}$.
- Given an arbitrary set of starting values $\{\theta_1^{(0)}, \dots, \theta_K^{(0)}\}$,

$$\text{Draw } \theta_1^{(1)} \sim p_1(\theta_1 | \theta_2^{(0)}, \dots, \theta_K^{(0)}),$$

$$\text{Draw } \theta_2^{(1)} \sim p_2(\theta_2 | \theta_1^{(1)}, \theta_3^{(0)}, \dots, \theta_K^{(0)}),$$

⋮

$$\text{Draw } \theta_K^{(1)} \sim p_K(\theta_K | \theta_1^{(1)}, \dots, \theta_{K-1}^{(1)}),$$

- Under mild conditions,

$$(\theta_1^{(t)}, \dots, \theta_K^{(t)}) \xrightarrow{d} (\theta_1, \dots, \theta_K) \sim p \text{ as } t \rightarrow \infty .$$

Gibbs sampling (cont'd)

- For t sufficiently large (say, bigger than t_0), $\{\boldsymbol{\theta}^{(t)}\}_{t=t_0+1}^T$ is a **(correlated)** sample from the true posterior.
- Can use a sample mean to estimate the posterior mean,

$$\hat{E}(\theta_i|\mathbf{y}) = \frac{1}{T - t_0} \sum_{t=t_0+1}^T \theta_i^{(t)} .$$

- The time from $t = 0$ to $t = t_0$ is commonly known as the *burn-in* period; one can safely **adapt** (change) an MCMC algorithm during this pre-convergence period, since these samples will be discarded anyway
- Most popular software package for this: **WinBUGS**
 - Uses **R**-like syntax to specify models
 - freely available from <http://www.mrc-bsu.cam.ac.uk/bugs/welcome.shtml>

Gibbs sampling (cont'd)

- In practice, we may actually run m *parallel* Gibbs sampling chains, instead of only 1, for some modest m (say, $m = 5$). Discarding the burn-in period, we obtain

$$\hat{E}(\theta_i | \mathbf{y}) = \frac{1}{m(T - t_0)} \sum_{j=1}^m \sum_{t=t_0+1}^T \theta_{i,j}^{(t)},$$

where now the j subscript indicates chain number.

- If the full conditional $p(\theta_i | \theta_{j \neq i}, \mathbf{y})$ is not available in closed form, it will typically still be available **up to proportionality constant**. So WinBUGS uses:
 - **adaptive rejection sampling** (log-concave densities)
 - **slice sampling** (bounded domains)
 - **Metropolis sampling** (all other cases)

Bayesian estimation

- **Point estimation:** Choose an appropriate measure of centrality: the posterior **mean**, **median**, or **mode**.
- **Interval estimation:** Consider q_L and q_U , the $\alpha/2$ - and $(1 - \alpha/2)$ -quantiles of $p(\theta|\mathbf{y})$:

$$\int_{-\infty}^{q_L} p(\theta|\mathbf{y})d\theta = \alpha/2 \quad \text{and} \quad \int_{q_U}^{\infty} p(\theta|\mathbf{y})d\theta = \alpha/2 .$$

Then clearly $P(q_L < \theta < q_U|\mathbf{y}) = 1 - \alpha$; our confidence that θ lies in (q_L, q_U) is $100 \times (1 - \alpha)\%$. Thus this interval is a $100 \times (1 - \alpha)\%$ **credible set** (“Bayesian CI”) for θ .

- Though not necessarily narrowest, this **equal tail** interval is easy to compute.
- Unlike frequentist CIs, interpretation of Bayesian CIs is direct: **“The probability that θ lies in (q_L, q_U) is $(1 - \alpha)$.”**

Bayesian hypothesis testing

- Classical approach bases accept/reject decision on

$$\text{p-value} = P\{T(\mathbf{Y}) \text{ more "extreme" than } T(\mathbf{y}_{obs}) | \boldsymbol{\theta}, H_0\},$$

where “extremeness” is in the direction of H_A

- Bayesian approach: for two models, a commonly used summary historically is the **Bayes factor**,

$$BF = \frac{P(M_1 | \mathbf{y}) / P(M_2 | \mathbf{y})}{P(M_1) / P(M_2)} = \frac{p(\mathbf{y} | M_1)}{p(\mathbf{y} | M_2)},$$

i.e., the likelihood ratio if both hypotheses are simple

- **Problem:** If $\pi_i(\boldsymbol{\theta}_i)$ is **improper**, then $p(\mathbf{y} | M_i)$ necessarily is as well \implies **BF is not well-defined!**...

Bayesian hypothesis testing via DIC

- A generalization of the Akaike Information Criterion (AIC) to the case of hierarchical models based on the posterior distribution of the deviance statistic,

$$D(\boldsymbol{\theta}) = -2 \log f(\mathbf{y}|\boldsymbol{\theta}) + 2 \log h(\mathbf{y}) ,$$

where $f(\mathbf{y}|\boldsymbol{\theta})$ is the likelihood and $h(\mathbf{y})$ is any standardizing function of the data alone

- Summarize the fit of a model by the posterior expectation of the deviance, $\bar{D} = E_{\theta|y}[D]$
- Summarize the complexity of a model by the effective number of parameters,

$$p_D = E_{\theta|y}[D] - D(E_{\theta|y}[\boldsymbol{\theta}]) = \bar{D} - D(\bar{\boldsymbol{\theta}}) .$$

Bayesian hypothesis testing via DIC

- The *Deviance Information Criterion* (DIC) is then

$$DIC = \bar{D} + p_D = 2\bar{D} - D(\bar{\theta}) ,$$

with **smaller** values indicating **preferred** models.

- Both building blocks of DIC and p_D , that is, $E_{\theta|y}[D]$ and $D(E_{\theta|y}[\theta])$, are easily estimated via MCMC methods, and in fact are automatic within **WinBUGS**.
- While p_D has a scale (effective model size), DIC does **not**, so only **differences** in DIC across models matter.
- DIC can be sensitive to **parametrization** and **“focus”** (i.e., what is considered to be part of the likelihood)
 - $f(\mathbf{y}|\theta)$: “focused on θ ”
 - $p(\mathbf{y}|\eta) = \int f(\mathbf{y}|\theta)p(\theta|\eta)d\theta$: “focused on η ”
- Like AIC, DIC tends to select “bigger” models

Software for Bayesian CTs I

Expensive but incredibly cool commercial software: **FACTS** (Fixed and Adaptive Clinical Trials Simulator) software for many Phase I and II trial designs

- permits dose finding in the presence of both safety and efficacy endpoints; stopping for success or futility
- supports continuous, dichotomous, or time-to-event (TITE) endpoints; also longitudinal data (e.g. biomarkers)
- joint venture between Berry Consultants (statistics, algorithms) and Tessella Technology and Consulting (software interface)

Website:

<http://www.smarterclinicaltrials.com/what-we-offer/facts/>

– probably best for large companies with ongoing commitments to Bayesian CT development

Software for Bayesian CTs II

Noncommercial but still very professional software: freely available from the M.D. Anderson Cancer Center Department of Biostatistics Software page, <https://biostatistics.mdanderson.org/SoftwareDownload/>

- All are stand-alone packages, free for download and local install
- easy to learn: menu- and dialog box-driven, nice graphics where appropriate
- accompanied by extensive tutorials, typically with guidelines, exercises, and solutions
- remarkably broad coverage of areas from all phases of the regulatory process...

Software for Bayesian CTs II

Partial list of MD Anderson Phase I software packages:

- **CRMSimulator**: Simulates power and Type I error of Continual Reassessment Method (CRM) dose-finding designs, offering improvements in power and/or sample size over “3 + 3” designs; see [BCLM Section 3.2](#).
- **bCRM**: handles *bivariate* dose-finding with two competing outcomes (say, toxicity and progression) and a single agent; see [BCLM Section 3.4.2](#).
- **EffTox**: finds a best dose when efficacy must be traded off against toxicity, both assumed increasing in dose; see [BCLM Section 3.3](#).
- **ToxFinder**: for [combination therapy](#), i.e., two *drugs* being administered in combination, with only one *outcome* (usually toxicity); see [BCLM Section 3.4.4](#).

Software for Bayesian CTs II

Partial list of MD Anderson Phase II software packages:

- **Phase II PP Design:** computes stopping boundaries for a single-arm Phase II predictive probability design with a binary endpoint; see [BCLM Section 4.2](#).
- **MultcLean:** for monitoring toxicity and efficacy in single arm phase II clinical trials with binary data; see [BCLM Section 4.3.2](#).
- **Adaptive Randomization (AR):** for designing and simulating outcome-adaptive randomized trials with up to 10 arms, using binary or time-to-event (TITE) outcomes; more patients are treated with the better treatment while retaining the benefits of randomization; see [BCLM Section 4.4](#).

Software for Bayesian CTs III

Noncommercial and not-all-that-professional software, but with R and BUGS source code freely available: from the **BCLM book's data and software page**, <http://www.biostat.umn.edu/~brad/data3.html>

- organized by chapter in the book
- similar range of models/problems as the MDACC software
- continuing to grow
- Mostly written in R, but those that require MCMC typically call *OpenBUGS* using the **BRugs** package, whose installation and exemplification are given here: <http://www.biostat.umn.edu/~brad/software/BRugs/>

Software for Bayesian CTs III

Partial list of BCLM Phase I software packages:

- **betabinHM.R**: elementary BRugs **metaanalysis** program for a single success proportion; see **BCLM Section 2.4**.
- **Power_BRugs.txt**: slightly more advanced BRugs power and Type I error-simulating program, Weibull survival model; see **BCLM Section 2.5.4**.
- **titecrm.R**: a basic R implementation of the TITE-CRM method; see **BCLM Section 3.2.3**.
- **324.R**: an R program for dose-finding based on toxicity *intervals* (rather than fixed target levels); see **BCLM Section 3.2.4**.
- **354.R**: a basic R implementation of the 2-agent combination therapy dose-escalation method (similar to **ToxFinder**); see **BCLM Section 3.4.4**.

Software for Bayesian CTs III

Partial list of BCLM Phase II-III software packages:

- **431.R**: an R program for binary stopping for futility, efficacy, or toxicity (similar to **MultcLean**); see **BCLM Section 4.3.2**.
- **443.R**: R code for outcome adaptive randomization with delayed survival response ; see **BCLM Section 4.4.4**.
- **adapt.R**: R code to compute the simulated Type I error and other operating characteristics of a basic one-arm binary response confirmatory trial; see **BCLM Section 5.2.1**.
- **example5.4.R**: an R program to simulate operating characteristics of the basic confirmatory trial with delayed outcomes; see **BCLM Section 5.2.3**.

BUGS Example: Survival Modeling

- Our data arises from a clinical trial comparing two treatments for *Mycobacterium avium complex (MAC)*, a disease common in late stage HIV-infected persons. Eleven clinical centers (“units”) have enrolled a total of 69 patients in the trial, of which 18 have died.
- For $j = 1, \dots, n_i$ and $i = 1, \dots, k$, let
 - t_{ij} = time to death or censoring
 - x_{ij} = treatment indicator for subject j in stratum i
- Next page gives survival times (in half-days) from the MAC treatment trial, where “+” indicates a censored observation...

MAC Survival Data

unit	drug	time	unit	drug	time	unit	drug	time
A	1	74+	E	1	214	H	1	74+
A	2	248	E	2	228+	H	1	88+
A	1	272+	E	2	262	H	1	148+
A	2	344				H	2	162
			F	1	6			
B	2	4+	F	2	16+	I	2	8
B	1	156+	F	1	76	I	2	16+
			F	2	80	I	2	40
C	2	100+	F	2	202	I	1	120+
			F	1	258+	I	1	168+
D	2	20+	F	1	268+	I	2	174+
D	2	64	F	2	368+	I	1	268+
D	2	88	F	1	380+	I	2	276
D	2	148+	F	1	424+	I	1	286+
...
						K	2	106+

MAC Survival Data

- With **proportional hazards** and a **Weibull** baseline hazard, stratum i 's hazard is

$$\begin{aligned}h(t_{ij}; x_{ij}) &= h_0(t_{ij})\omega_i \exp(\beta_0 + \beta_1 x_{ij}) \\ &= \rho_i t_{ij}^{\rho_i - 1} \exp(\beta_0 + \beta_1 x_{ij} + W_i) ,\end{aligned}$$

where $\rho_i > 0$, $\beta = (\beta_0, \beta_1)' \in \mathbb{R}^2$, and $W_i = \log \omega_i$ is a clinic-specific **frailty** term.

- The W_i capture overall differences among the clinics, while the ρ_i allow differing baseline hazards which either increase ($\rho_i > 1$) or decrease ($\rho_i < 1$) over time. We assume i.i.d. specifications for these random effects,

$$W_i \stackrel{iid}{\sim} N(0, 1/\tau) \quad \text{and} \quad \rho_i \stackrel{iid}{\sim} G(\alpha, \alpha) .$$

MAC Survival Data

- As in the `mice` example (WinBUGS Examples Vol 1),

$$\mu_{ij} = \exp(\beta_0 + \beta_1 x_{ij} + W_i) ,$$

so that

$$t_{ij} \sim Weibull(\rho_i, \mu_{ij}) .$$

- We **recode** the drug covariate from (1,2) to (-1,1) (i.e., set $x_{ij} = 2drug_{ij} - 3$) to **ease collinearity** between the slope β_1 and the intercept β_0 .
- We place vague priors on β_0 and β_1 , a moderately informative $G(1, 1)$ prior on τ , and set $\alpha = 10$.
- **Data:** www.biostat.umn.edu/~brad/data/MAC.dat
Code:
www.biostat.umn.edu/~brad/data/MACfrailty_BUGS.txt

MAC Survival Results

node (unit)	mean	sd	MC error	2.5%	median	97.5%
W_1 (A)	-0.04912	0.835	0.02103	-1.775	-0.04596	1.639
W_3 (C)	-0.1829	0.9173	0.01782	-2.2	-0.1358	1.52
W_5 (E)	-0.03198	0.8107	0.03193	-1.682	-0.02653	1.572
W_6 (F)	0.4173	0.8277	0.04065	-1.066	0.3593	2.227
W_9 (I)	0.2546	0.7969	0.03694	-1.241	0.2164	1.968
W_{11} (K)	-0.1945	0.9093	0.02093	-2.139	-0.1638	1.502
ρ_1 (A)	1.086	0.1922	0.007168	0.7044	1.083	1.474
ρ_3 (C)	0.9008	0.2487	0.006311	0.4663	0.8824	1.431
ρ_5 (E)	1.143	0.1887	0.00958	0.7904	1.139	1.521
ρ_6 (F)	0.935	0.1597	0.008364	0.6321	0.931	1.265
ρ_9 (I)	0.9788	0.1683	0.008735	0.6652	0.9705	1.339
ρ_{11} (K)	0.8807	0.2392	0.01034	0.4558	0.8612	1.394
τ	1.733	1.181	0.03723	0.3042	1.468	4.819
β_0	-7.111	0.689	0.04474	-8.552	-7.073	-5.874
β_1	0.596	0.2964	0.01048	0.06099	0.5783	1.245
RR	3.98	2.951	0.1122	1.13	3.179	12.05

MAC Survival Results

- Units A and E have moderate overall risk ($W_i \approx 0$) but increasing hazards ($\rho > 1$): few deaths, but they occur late
 - Units F and I have high overall risk ($W_i > 0$) but decreasing hazards ($\rho < 1$): several early deaths, many long-term survivors
 - Units C and K have low overall risk ($W_i < 0$) and decreasing hazards ($\rho < 1$): no deaths at all; a few survivors
 - Drugs differ significantly: CI for β_1 (RR) excludes 0 (1)
-
- **Note:** This has all been for two sets of random effects (W_i and ρ_i), called “Model 2” in the BUGS code. You will also see models having three (adding β_{1i}), one (deleting ρ_i), or zero sets of random effects!

BRugs Example: Clinical Trial Design

- Following our MAC 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

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

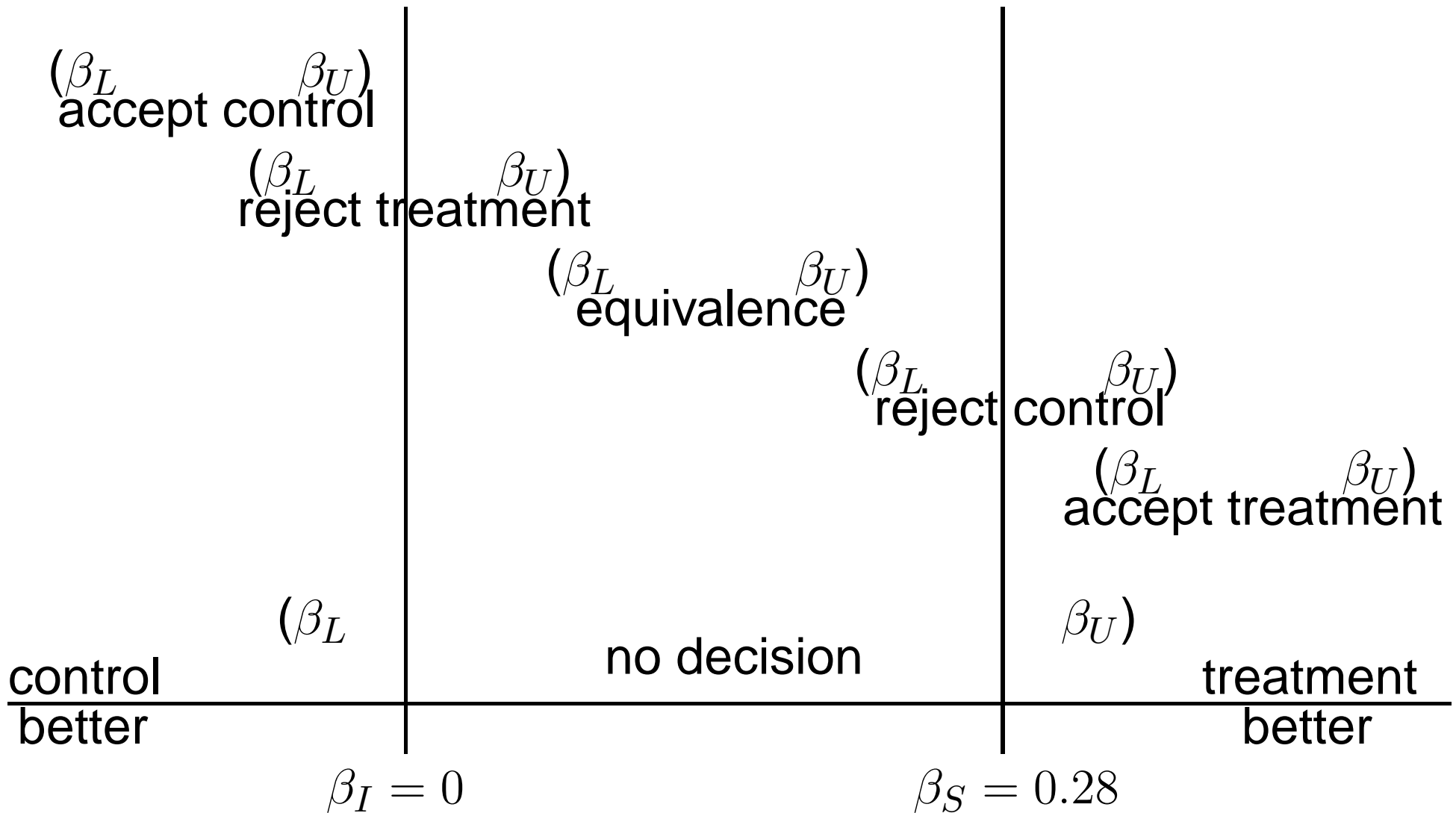
- The value of β_1 corresponding to a **15% increase in median survival in the treatment group** satisfies

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

Range of equivalence

- The range of β_1 values within which we are **indifferent** as to use of treatment or control
- 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**
 - One that believes the treatment is likely no better than control (as might be believed by the FDA)
- **Enthusiastic (or Clinical) Prior**
 - One that believes the treatment will succeed (typical of the clinicians running the trial)
- **Reference (or Noninformative) Prior**
 - One that expresses no particular opinion about the treatment's merit
 - Often a **improper uniform** (“flat”) prior is permissible

MCMC-based Bayesian design

To simulate power or other operating characteristics:

- Sample “true” β values from an assumed **design 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) by **calling OpenBUGS from R** using the **analysis prior** (may be different from the design prior)
- 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

Software for doing this: **BRugs**

- **BRugs** is a suite of R routines for calling `OpenBUGS` from R, originally written by `WinBUGS` head programmer Andrew Thomas, and refined and maintained by Uwe Ligges
- All necessary programs and instructions can be downloaded from www.biostat.umn.edu/~brad/software/BRugs
- Note that we will now have **two** text files with code:
 - an **R** program that organizes the dataset, contains all the `BRugs` commands, and summarizes the output
 - a piece of **BUGS** code that is sent by R to `OpenBUGS` (must be saved in working directory)
- `BRugs` code for our Weibull model is available as **Example 3** at www.biostat.umn.edu/~brad/software/BRugs

Running BRugs

- Using BRugs, a minimum of four commands are needed to run a Bayesian model:
 - `library(BRugs)`
 - `bugsData`
 - `bugsInits`
 - `BRugsFit`
- Then the MCMC samples for the parameters specified in the `BRugsFit` command are available for analysis with both BUGS commands **and** ordinary R commands
- To get a trace plot of the MCMC chain for new study 10, use either:
 - `samplesHistory("p[10]")` or
 - `plot(samplesSample("p[10]"), type='l')`

Results from Power_BRugs.txt

- 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”

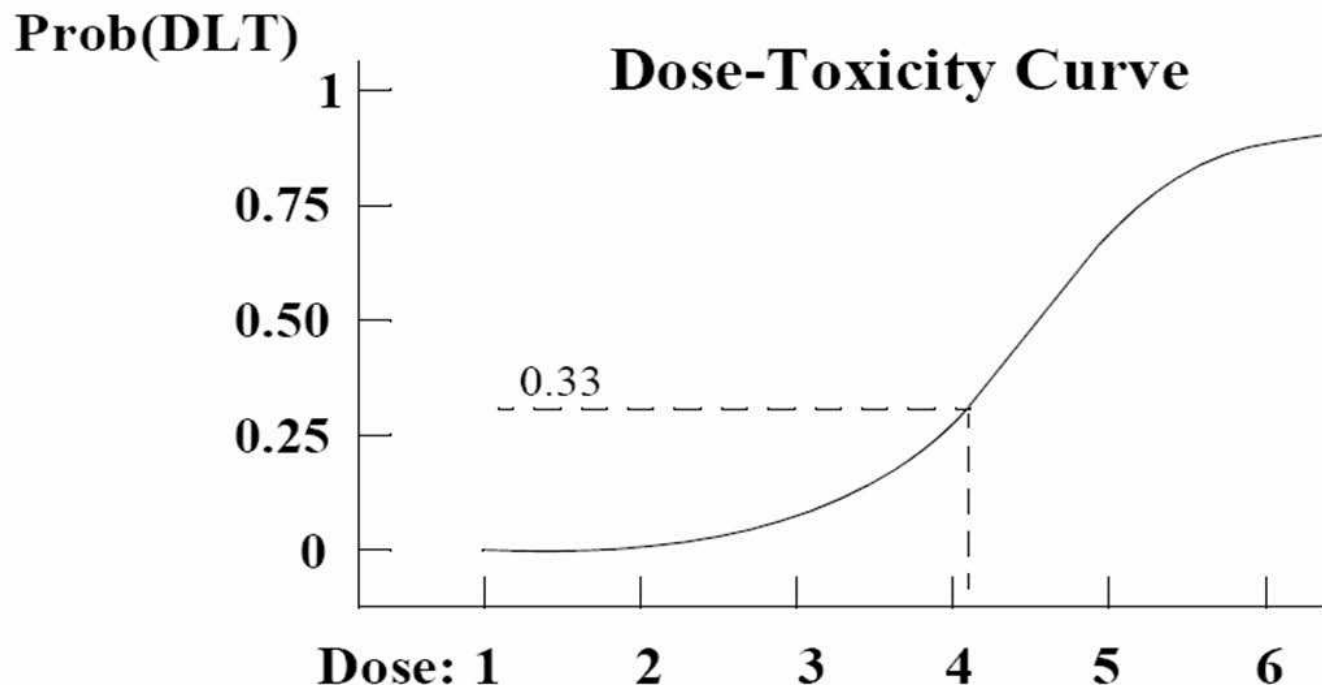
We obtain the following output from $Nrep = 100$ reps:

- 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

Ch 3: Phase I studies

- The first application of a new drug to humans
- Typically small (20-50 patients)
- **Main goal:** to establish the **safety** of a proposed drug, often through determining an appropriate dosing schedule (*dose-finding*)
- For cytotoxic agents (cancer), we assume the drug benefit (as well as the severity of its toxicity and other side effects) increases with dose, and thus seek the **maximum tolerated dose (MTD)**
- Key elements:
 - a starting dose (often $\frac{1}{10}LD_{10,mice}$)
 - a definition of **dose limiting toxicity (DLT)**
 - a **target toxicity level (TTL)** (say, 20-30%)
 - a dose escalation scheme

Sec 3.2: Model-based MTD designs



- Here we assume there is a monotonic relationship between dose and $P(\text{DLT})$, typically in the form of a simple one- or two-parameter model.
- In the figure, for a target toxicity level (TTL) of 33%, dose level 4 is the MTD

Continual Reassessment Method (CRM)

- The first Bayesian model-based design introduced in the literature (O'Quigley et al., 1990).
- Often characterizes the dose-toxicity relationship via simple one-parameter parametric models. That is, letting $p(d) = P(DLT|dose = d)$, 3 possible models are:

Hyperbolic tangent:
$$p(d) = \left[\frac{\exp(d)}{\exp(d) + \exp(-d)} \right]^a$$

Logistic:
$$p(d) = \frac{\exp(3 + ad)}{1 + \exp(3 + ad)}$$

Power:
$$p(d) = d^{\exp(a)}$$

The Bayesian posterior distribution of a induces a posterior for $p(d)$, and hence that of the MTD for any given TTL!

CRM algorithm

1. Assume a vague or fully non-informative prior for a .
2. Treat 1 patient at the level closest to the current estimate of the MTD, and observe the toxicity outcome
3. Update the **posterior distribution of a** , proportional to the prior for a times the likelihood,

$$L(a; \mathbf{d}, \mathbf{y}) \propto \prod_{i=1}^n p(d_i)^{y_i} [1 - p(d_i)]^{1-y_i},$$

where d_i is the dose level for patient i , and where $y_i = 1$ if a DLT is observed for i and $y_i = 0$ if not.

4. Treat the next patient at the level closest to the updated estimate of MTD based on the posterior distribution of a .
5. Repeat these steps until a sufficiently precise estimate of a is achieved or the maximum sample size is reached

Properties of the CRM

● Advantages

- Model-based method with a clearly defined objective
- Treats more patients at close to the target MTD level, hence reduces the number of patients treated at low or ineffective dose levels
- Uses all the data to model the dose-toxicity curve

● Disadvantages

- The dose assignment may be too aggressive
- Success depends on a proper choice of the dose-toxicity curve and the prior distribution on a
- Need special a computer program to implement the design

In order to address some of these safety concerns...

Modified CRM

c.f. Faries (1994, *J. Biopharm. Stat.*); Korn et al. (1994, *Statist. in Med.*); Goodman et al. (1995, *Statist. in Med.*)

Use CRM but with the following **modifications**:

- Start at the lowest dose level, and do not skip doses
- Treat patients in **cohorts** of size greater than 1 (say, 3)
- No dose escalation to new doses until all treatment in patients from the previous doses are completed
- Use asymmetric metrics to determine the current MTD, e.g., the level closest to **but no higher than** the TTL
- Possibly use a more conservative stopping rule, e.g., no more than 2 of 6 developed DLT at any given dose level

Software for Modified CRM

1. **MD Anderson option:** The `CRMSimulator` package from biostatistics.mdanderson.org/SoftwareDownload/ can do simulations for the power model...

2. **BCLM book-related option:** The `phaseIsim.R` function from www.biostat.umn.edu/~brad/software/BCLM_ch3.html

The following code corresponds to the example on the next 2 pages!

```
p.tox0 <- c(.05, .15, .3, .45, .6)
s.dose  <- log(p.tox0/(1-p.tox0)) - 3
phaseIsim(nsim=10000, npat=30, sdose=s.dose, prob.tox=p.tox0,
  design='3+3', outfile='3plus3.txt') # 3+3 results

phaseIsim(nsim=100, npat=30, sdose=s.dose, prob.tox=p.tox0,
  outfile='CRM1.txt') # CRM1 results

phaseIsim(nsim=100, npat=30, sdose=s.dose, prob.tox=p.tox0,
  crm.group.size=3, outfile='CRM3.txt') # CRM3 results
```

Example using phaseIsim.R

- Suppose that in developing a new agent, we have five potential dose levels, and our target toxicity level is 30%. We wish to simulate and compare the operating characteristics of the 3+3 and two CRM designs using 10,000 simulated trials.
- Suppose the true probabilities of DLT at dose levels 1 to 5 are 0.05, 0.15, 0.30, 0.45, and 0.60, respectively, so that dose level 3 is the true MTD.
- The table on the next page shows that the CRM design with a cohort size of 1 (**CRM 1**) treats more patients at the true MTD level, but also more patients at dose levels above the MTD. The overall percent of DLT for the 3+3 and CRM 1 designs are 21.1 and 27.0, respectively.
- By increasing the CRM cohort size from 1 to 3 (**CRM 3**), we treat **fewer** patients at levels above the MTD.

Modified CRM outperforms 3+3

		Dose					Ave	%
		1	2	3	4	5	N	DLT
<i>Scenario 1</i>	P(DLT):	0.05	0.15	0.30	0.45	0.60		
3+3	% patients	26.0	32.5	27.2	12.1	2.3	15.2	21.1
	% MTD	20.5	42.7	27.5	5.7	0		
CRM 1	% patients	15.6	24.1	34.7	19.0	6.7	18.5	27.0
	% MTD	1.0	21.4	52.4	23.0	2.2		
CRM 3	% patients	21.3	31.4	29.1	15.8	2.5	19.0	23.3
	% MTD	1.5	22.6	49.8	23.7	2.4		

- CRM designs are much more likely to find the true MTD
- CRM 3 offers protection for just a bit higher Ave N
- CRM also beats 3+3 when the assigned doses are less and more toxic than anticipated, respectively.

Escalation w/ Overdose Control (EWOC)

Same as CRM, except when choosing the next dose use the α^{th} **quantile** of the MTD's posterior, instead of the mean

- For dose x , $P(DLT|x = MTD) \equiv \theta$, the TTL
- Start at the lowest dose level, i.e., set $x_1 = d_1$
- For any patient k , let $\pi_k(\gamma)$ be the MTD posterior cdf,

$$\pi_k(\gamma) = P(MTD \leq \gamma | \mathbf{y}_k)$$

where \mathbf{y}_k is the data currently available

- Ideally, select the next dose level x_k such that

$$\pi_k(x_k) = \alpha$$

To restrict to our dose levels $\{d_1, \dots, d_r\}$, take

$$x_k^* = \max\{d_1, \dots, d_r : d_i - x_k \leq T_1 \text{ and } \pi_k(x_k) - \alpha \leq T_2\}$$

for prespecified **tolerances** $T_1, T_2 > 0$

Bayesian EWOC

The EWOC doses x_k **minimize risk** with respect to the **asymmetric loss function**,

$$L(x, \gamma) = \begin{cases} \alpha(\gamma - x) & \text{for } x \leq \gamma \text{ (i.e., } x \text{ is an underdose)} \\ (1 - \alpha)(x - \gamma) & \text{for } x > \gamma \text{ (i.e., } x \text{ is an overdose)} \end{cases}$$

- Choosing the feasibility bound $\alpha < 0.5$ corresponds to placing a **higher penalty** on **overdosing** than on **underdosing**
- Choosing $\alpha = 0.5$ implies a **symmetric** loss function, and produces the MTD posterior median as the new dose
- When $\alpha \ll 0.5$, the final dose recommended for Phase II study (e.g., MTD posterior median) may be significantly larger than the dose **any** Phase I patient has received \Rightarrow use a **varying** feasibility bound?

EWOC Implementation

Consider EWOC under the logistic model,

$$Prob(DLT|dose = x) \equiv p(x) = \frac{\exp(\beta_0 + \beta_1 x)}{1 + \exp(\beta_0 + \beta_1 x)}.$$

To ease prior specification, reparametrize from (β_0, β_1) to (ρ_0, γ) , where $\rho_0 = p(X_{min})$, the probability of DLT at the minimum dose, X_{min} , and γ is the MTD. Then

$$\beta_0 = \frac{1}{\gamma - X_{min}} [\gamma \text{logit}(\rho_0) - X_{min} \text{logit}(\theta)]$$

$$\text{and } \beta_1 = \frac{1}{\gamma - X_{min}} [\text{logit}(\theta) - \text{logit}(\rho_0)].$$

Note: We assume that $\gamma \in (X_{min}, X_{max})$ with probability 1; we would typically take the starting dose $d_1 = X_{min}$.

EWOC Code and Extensions

- www.biostat.umn.edu/~brad/data3.html offers WinBUGS EWOC code, using the model above and adopting independent uniform priors on γ and ρ_0 on the ranges (X_{min}, X_{max}) and $(0, \theta)$, respectively.
- sisyphus.emory.edu/software_ewoc.php is “the” EWOC website, featuring papers and relevant programs
- Covariate adjustment = “individualized patient dosing”

Example: a phase I non-small-cell lung cancer trial of **PNU-214936**, a monoclonal antibody, with a covariate (**anti-SEA**) previously shown to have a neutralizing effect on PNU. A convenient **dose-toxicity model**:

$$P(DLT|x, c) = \frac{\exp[\beta_0 + \beta_1 \log(x) + \beta_2 \log(c)]}{1 + \exp[\beta_0 + \beta_1 \log(x) + \beta_2 \log(c)]}$$

where c denotes anti-SEA level.

EWOC Extensions

- Recent work by Zabor (2010) considers the case of **two** covariates, one categorical and one continuous

Example: Trial of **852A**, an “agonist” that can enhance the tumor-inhibiting and immune response-boosting properties of other oncologic agents. Let z be 0 for male, 1 for female, and let $c \in [39, 80]$ be the patient’s age in years. Then our model for $P(DLT)$ is

$$P(DLT|x, c, z) = \frac{\exp[\beta_0 + \beta_1 x + \beta_2 c + \beta_3 z]}{1 + \exp[\beta_0 + \beta_1 x + \beta_2 c + \beta_3 z]} .$$

- Reparametrize to $\gamma_{max} = \gamma(c = 80, z = 0)$ and

$$\rho_1 = P(DLT|X = X_{min}, C = 39, Z = 0)$$

$$\rho_2 = P(DLT|X = X_{min}, C = 80, Z = 0)$$

$$\rho_3 = P(DLT|X = X_{min}, C = 39, Z = 1)$$

EWOC Results

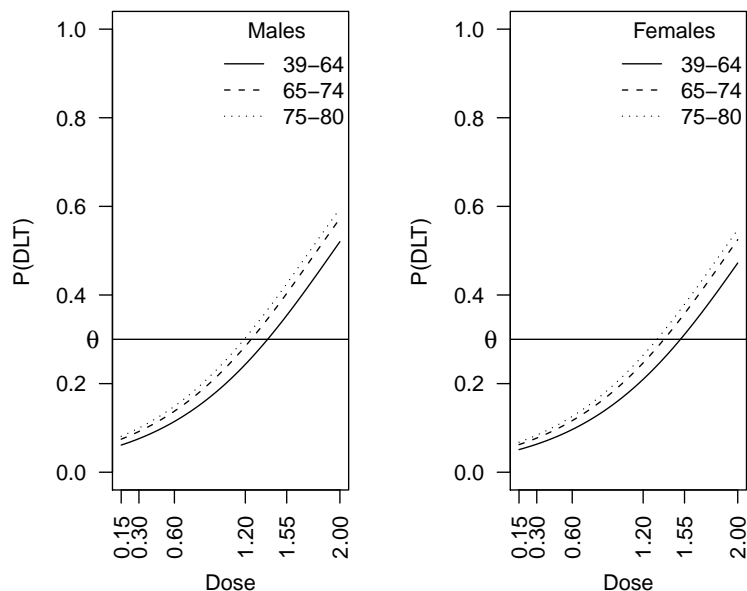


Figure 1: Probability of DLT by covariate group and dose

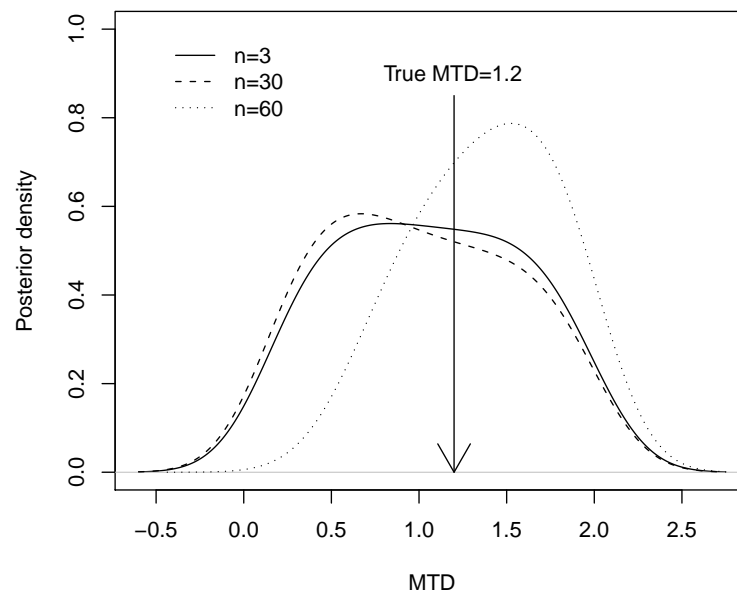


Figure 2: The marginal posterior distribution of the maximum tolerated dose (MTD) for an 80 year old male as data accumulates

- Left plot shows true dose-response curves by gender and grouping ages into 3 clinically meaningful groups
- Simulated trials used a feasibility bound of $\alpha = 0.25$, entered up to 20 cohorts of size 3, and stopped early if 95% Bayesian CI had width < 1.2
- **BRugs** results for posterior of $\gamma(80, 0)$ in one trial (right plot) reveal Bayesian learning as data accumulates!

Section 4.4: Adaptive Randomization

“Bayesians don’t need/like randomization”

- plays **no role** in calculating posterior probabilities (whereas crucial for frequentist inference)
- we can control for prognosis-related covariates anyway
- ethically difficult for physicians
- patients willing to be randomized are inherently different

Our take: Randomization **IS** still essential:

- ensures pt prognosis is uncorrelated with trt assigned
- balances trt assignment within patient subgroups
- we can’t control for **unknown/unmeasured** covariates!

BUT: Note that we needn’t randomize patients **with equal probabilities to all arms...**

Principles of Adaptive Randomization

- By “adaptive,” we mean a procedure that alters something **based on the results of the trial so far** \implies implications for Type I error!
- Here we focus on **outcome-adaptive** designs, not **covariate-adaptive** designs that seek to balance covariates across treatments

Basic idea: Treatment arms A_k having response probabilities θ_j , $j = 1, \dots, m$. Given data \mathbf{y} , randomize to treatment A_k with probability

$$r_k(\mathbf{y}) \propto \{p(\theta_k = \max_j \theta_j \mid \mathbf{y})\}^c \quad \text{for some } c \geq 0 \quad (1)$$

- $c = 0 \implies$ equal randomization
- might take $c = n/2N$ where n is number of currently enrolled pts and N is maximum enrollment

MDACC software package: AR

- Windows application for trials having up to 10 arms
- outcomes may be either binary or time-to-event (TITE), though latter case currently restricted to exponential survival with a conjugate prior
- easy-to-read user's guide
- **583** registered downloads between 2005 and 2009

Binary case: Assign $Beta(\alpha_k, \beta_k)$ priors to the θ_k by

- choosing the (α_k, β_k) pairs directly
- specifying either two quantiles **or** the mean and the variance; AR can then “back out” (α_k, β_k)

Assuming x_k positive and $n_k - x_k$ negative (independent) responses $\implies \theta_k | \mathbf{y} \sim Beta(x_k + \alpha_k, n_k - x_k + \beta_k)$ as usual \implies used to define a variety of stopping rules...

Algorithm used by AR

Step 1. Early loser: If for some prespecified probability p_L ,

$$P(\theta_k > \theta_{j \neq k} | \mathbf{y}) < p_L ,$$

then arm k is declared a **loser** and is **suspended**.

- Normally p_L is fairly small (say, 0.10 or less)
- AR permits an arm to return later

Step 2. Early winner: If for some prespecified probability p_U ,

$$P(\theta_k > \theta_{j \neq k} | \mathbf{y}) > p_U ,$$

then arm k is declared the **winner** and the trial is stopped early.

- Normally p_U is fairly large (say, $1 - p_L$ for a two-arm trial)

Algorithm used by AR (cont'd)

Step 3. Final winner: If, *after all patients have been evaluated*, for some prespecified probability p_U^* ,

$$P(\theta_k > \theta_{j \neq k} | \mathbf{y}) > p_U^* ,$$

then arm k is declared the **winner**.

- If **no** treatment arm can meet this criterion, AR does not make a final selection.
- Normally $p_U^* < p_U$ (say, between 0.70 and 0.90)

Step 4. Futility: If for some θ_{min} and some prespecified p_L^* ,

$$P(\theta_k > \theta_{min} | \mathbf{y}) < p_L^* ,$$

then arm k is declared **futile** and its accrual is stopped.

- Reactivation of a futile arm is **not** permitted
- Normally p_L^* is quite small (say, 0.10 or less)

Example: Sensitizer Trial

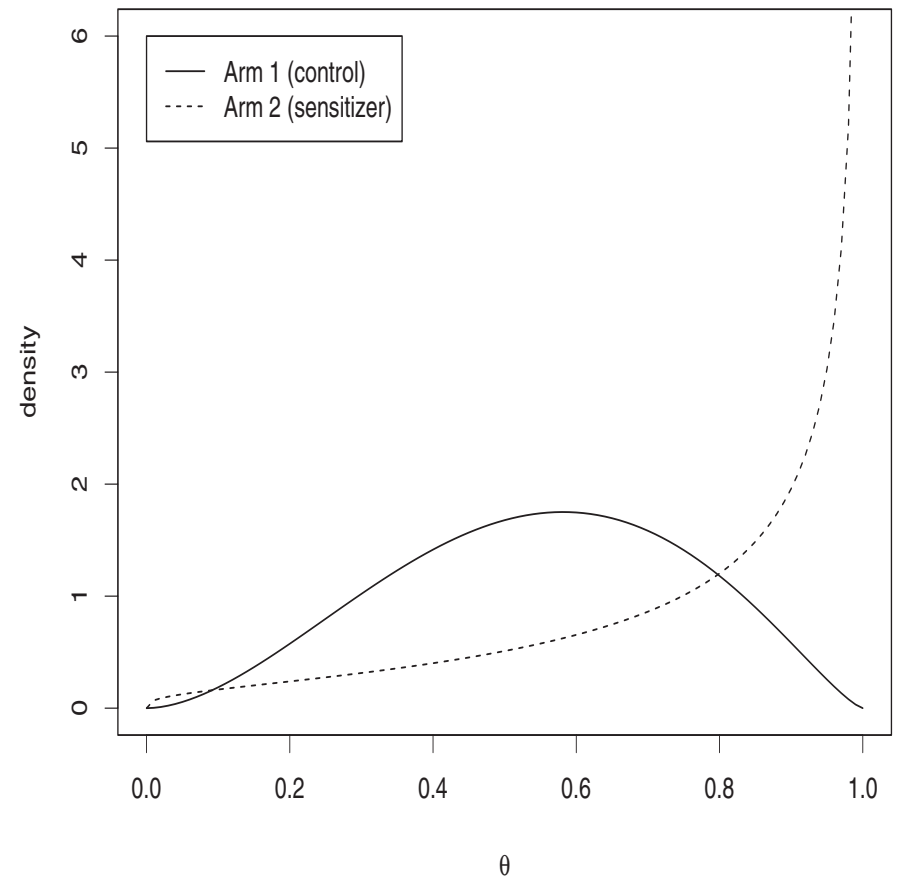
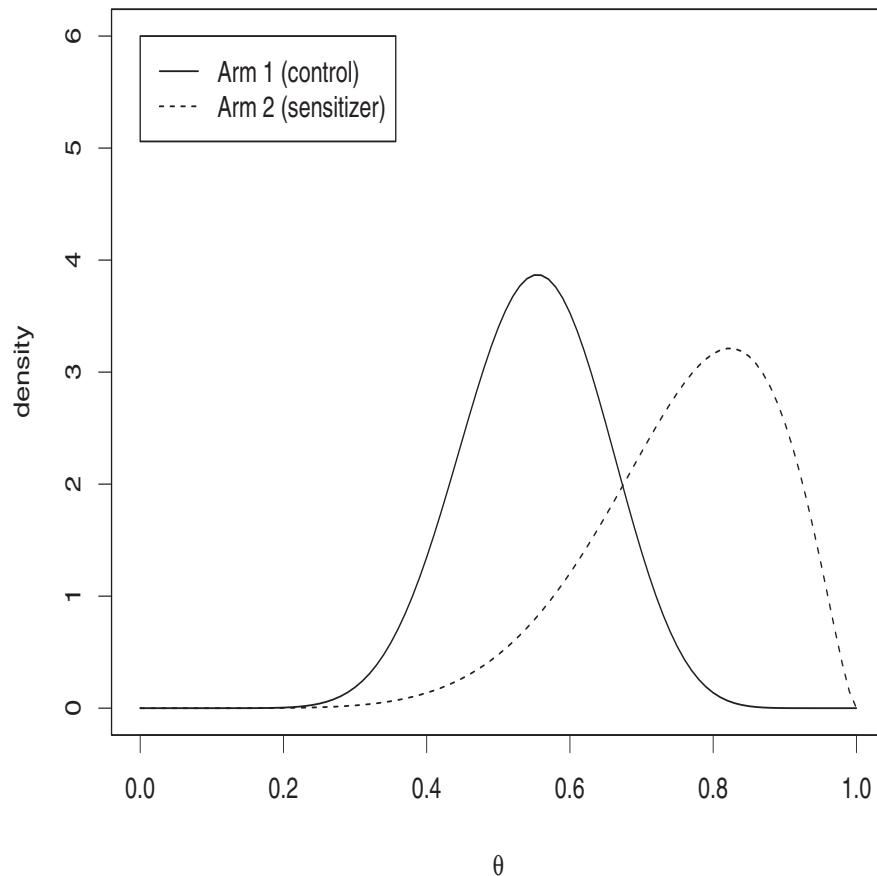
Goal: Evaluate ability of a “sensitizer” (given concurrently with another chemotherapeutic agent) to produce complete remission (CR) at 28 days post-treatment

- Classical two-arm comparison of drug-plus-sensitizer vs. drug alone \Rightarrow sample size near 100 – too large for our accrual rate of just 30/year!

Instead, use some prior information with a Bayesian AR design having

- maximum patient accrual = 60
- minimum randomization probability = 0.10
- first 14 patients randomized fairly (7 to each arm) before AR begins
- AR tuning parameter $c = 1$ (i.e., modest deviation from equal randomization after the first 14 patients)

Priors for the Sensitizer Trial



- Left: “standard” priors; right, “conservative” priors (both sd’s doubled)
- Both Arm 1 priors have mean 0.55; both Arm 2 priors have mean 0.75

Control Parameters for Sensitizer Trial

Begin with a “standard” rule that sets

- early loser selection probability $p_L = 0.025$
- early winner selection probability $p_U = 0.975$
- final winner selection probability $p_U^* = 0.90$
- futility parameters $\theta_{min} = 0.50$ and $p_L^* = 0.05$

Use AR to compare results from three different scenarios:

- Scenario 1: true response rates of .55 in both groups (the “null” scenario),
- Scenario 2: true response rates of .55 control, .70 sensitizer (the “most likely” scenario), and
- Scenario 3: true response rates of .55 control, .80 sensitizer (the “optimistic” scenario).

100 Simulated Trials, Standard Prior

Scenario 1 (Average Trial Length: **22.5** months)

Arm	True Pr (success)	Pr (select)	Pr(select early)	Pr(stop early)	# Patients (2.5%, 97.5%)
Arm1	0.55	0.01	0	0.11	19.6 (5, 38)
Arm2	0.55	0.16	0.11	0	35.6 (8, 53)

Scenario 2 (Average Trial Length: **16.4** months)

Arm1	0.55	0	0	0.55	10.1 (4, 22)
Arm2	0.7	0.74	0.55	0	30.8 (4, 51)

Scenario 3 (Average Trial Length: **10.8** months)

Arm1	0.55	0	0	0.89	7.01 (4, 16)
Arm2	0.8	0.96	0.89	0	20.1 (4, 51)

- **Good** Type I error (17% total) and power (74%, 96%)
- But **fairly high** sample sizes & **fairly long** trial lengths

100 Simulated Trials, Conservative Prior

Scenario 1 (Average Trial Length: 21.0 months)

Arm	True Pr (success)	Pr (select)	Pr(select early)	Pr(stop early)	# Patients (2.5%, 97.5%)
Arm1	0.55	0.19	0.05	0.15	26.2 (5, 45)
Arm2	0.55	0.16	0.13	0.09	25.4 (4, 47)

Scenario 2 (Average Trial Length: 18.1 months)

Arm1	0.55	0.04	0.02	0.39	15.2 (4, 44)
Arm2	0.7	0.52	0.39	0.02	29.4 (4, 49)

Scenario 3 (Average Trial Length: 14.3 months)

Arm1	0.55	0	0	0.63	10.6 (2, 26)
Arm2	0.8	0.80	0.63	0	25.6 (3, 50)

- is more conservative; higher Type I error, lower power
- still fairly low average sample sizes (36.2, 44.6)

Summary

- MD Anderson software available at <https://biostatistics.mdanderson.org/SoftwareDownload/>
 - BCLM text-related software available at: <http://www.biostat.umn.edu/~brad/data3.html>
 - BRugs package installation and further examples: <http://www.biostat.umn.edu/~brad/software/BRugs/>
 - Related design site for binary and Cox PH models: www.biostat.umn.edu/~brianho/papers/2007/JBS/prac_bayes_design.html
-
- *Thanks for your attention!*