

Practical Bayesian Design and Analysis for Drug and Device Clinical Trials

Brian P. Hobbs

Plan B Advisor: **Bradley P. Carlin**

brianho@biostat.umn.edu

Master of Science Plan B Seminar

March 7, 2007

Division of Biostatistics, School of Public Health, University of Minnesota

Basics of Bayesian inference

- Existing information is incorporated formally into the statistical analysis
- **Prior distributions** summarize our preexisting understanding/beliefs regarding unknown model parameters, $\boldsymbol{\theta} = (\theta_1, \dots, \theta_K)'$
- Inference is conducted on the **posterior distribution** of $\boldsymbol{\theta}$ given observed data $\mathbf{y} = (y_1, \dots, y_N)'$, defined by Bayes' Rule as:

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

Hierarchical modeling

- Suppose the prior distribution for θ depends on a vector of second-stage **hyperparameters** γ ,

$$\theta \sim p(\theta|\gamma)$$

- If γ is unknown, a third-stage prior, or **hyperprior**, $p(\gamma)$ may be chosen
- Hierarchical models permit “borrowing of strength” from prior distributions and across subgroups
- In clinical trials, $p(\gamma)$ is often determined using data from historical controls

Advantages of Bayesian designs

- Reduce time and expense
- Minimize unethical exposure of patients to inferior treatments
- Flexibility: better able to adapt to circumstances where randomization is unethical, or a trial needs to be stopped early.
- Natural approach for incorporating historical controls into the analysis
- Sample size calculations based on priors **uninformative**, **optimistic**, or **skeptical** with respect to the effectiveness of the intervention of interest, allow investigators to quantify the likely gain from incorporating prior knowledge, and assess the need for a new trial

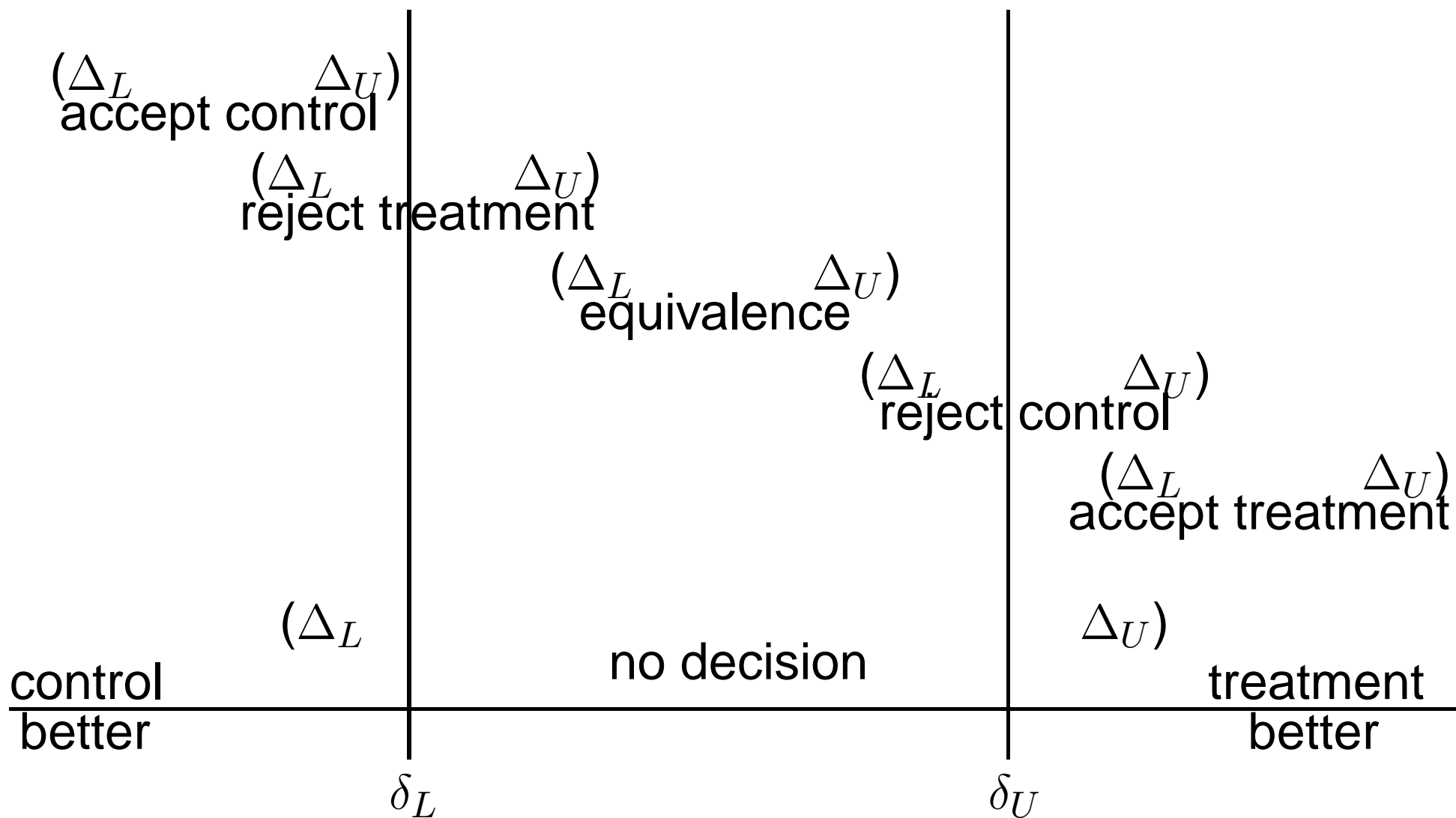
Bayesian approach

- In the case of medical device trials, the recently released FDA guidance document (U.S. Food and Drug Administration, 2006; <http://www.fda.gov/cdrh/osb/guidance/1601.html>) is a significant push for Bayesian methods, (roughly 10% of new device approvals, Berry 2006)
- I will present a practical approach for incorporating historical information into the design and analysis of clinical trials
- Demonstrate with sample size calculations in a drug trial motivated by FIRST (MacArthur et al., 2001)
- Classical concepts (sign. level, power, etc.) have analogues in defined relative to the **posterior distribution**

Designate decision-making paradigm

- Let Δ = effect difference between treatment and control (or intervention effect)
- We replace the point null $H_0 : \Delta = c$ with a range of null Δ 's, $\Delta \in [\delta_L, \delta_U]$ "indifference zone", proposed by Freedman et al. (1984)
- δ_U , represents the amount of improvement required by the intervention to suggest clinical superiority over control
- δ_L denotes the threshold below which the intervention would be considered clinically inferior
- Trial results will be based upon the location of the 95% posterior credible interval for Δ with respect to the indifference zone, (Spiegelhalter et al., 1994)

The six possible trial outcomes



Based on approach of Spiegelhalter et al. (1994)

Formulate the observed data likelihood

- Consider a binary endpoint trial where, $Y_i = 1$ if the i^{th} patient experiences progression of disease, $Y_i = 0$ otherwise, $i = 1, \dots, N$
- Then $Y_i \sim \text{Bernoulli}(\pi_i)$
- One possible model for π assumes

$$\text{logit}(\pi_i) = \log \left(\frac{\pi_i}{1 - \pi_i} \right) = \lambda_0 + \lambda_1 x_i$$

where λ_0 and λ_1 are random **hyperparameters**

- If $x_i = 0$ for control and 1 for treatment, then λ_1 captures the effect of intervention
- Let $\Delta = e^{\lambda_1}$, the odds ratio for disease progression

Analysis stage

- Evaluating the posterior distribution of Δ
- **Fitting priors** are used to compute or fit the **posterior distribution** for Δ using our statistical model given the observed trial data
- First consider distributional forms and hierarchical modeling schemes

$$\lambda_k \sim \text{Normal}(\mu_k, \sigma_k^2)$$

- Review historical evidence and meta-analysis construct a **clinical prior**
- Fit another posterior based solely on the observed likelihood using a **flat** or **noninformative prior**
- Comparing these posteriors allows us to measure the impact of incorporating historical evidence

Design stage

- For any fixed, “true” values of λ_0 and λ_1 , we can simulate the frequentist power of our Bayesian procedure
 1. compute the π_i from “true” λ_0 and λ_1
 2. generate fake data values Y_{ij}^* repeatedly from the binomial likelihood for $j = 1, \dots, N_{rep}$
 3. fit the posterior for Δ given Y_{ij}^*
 4. compare the 95% posterior credible interval for Δ to the indifference zone boundaries $[\delta_L, \delta_U]$
- Repeating this for N_{rep} datasets, we can compute the empirical probability of each of the six outcomes, and select N to correspond to our desired power

Fully Bayesian design

- A more fully Bayesian version of this procedure would be to replace the fixed, true values (λ_0, λ_1) with draws $\{(\lambda_{0j}^*, \lambda_{1j}^*), j = 1, \dots, N_{rep}\}$ from their prior distributions
- Here the uncertainty in these parameters acknowledged
- Priors used to generate the fake λ_{0j}^* and λ_{1j}^* are referred to as **design priors**

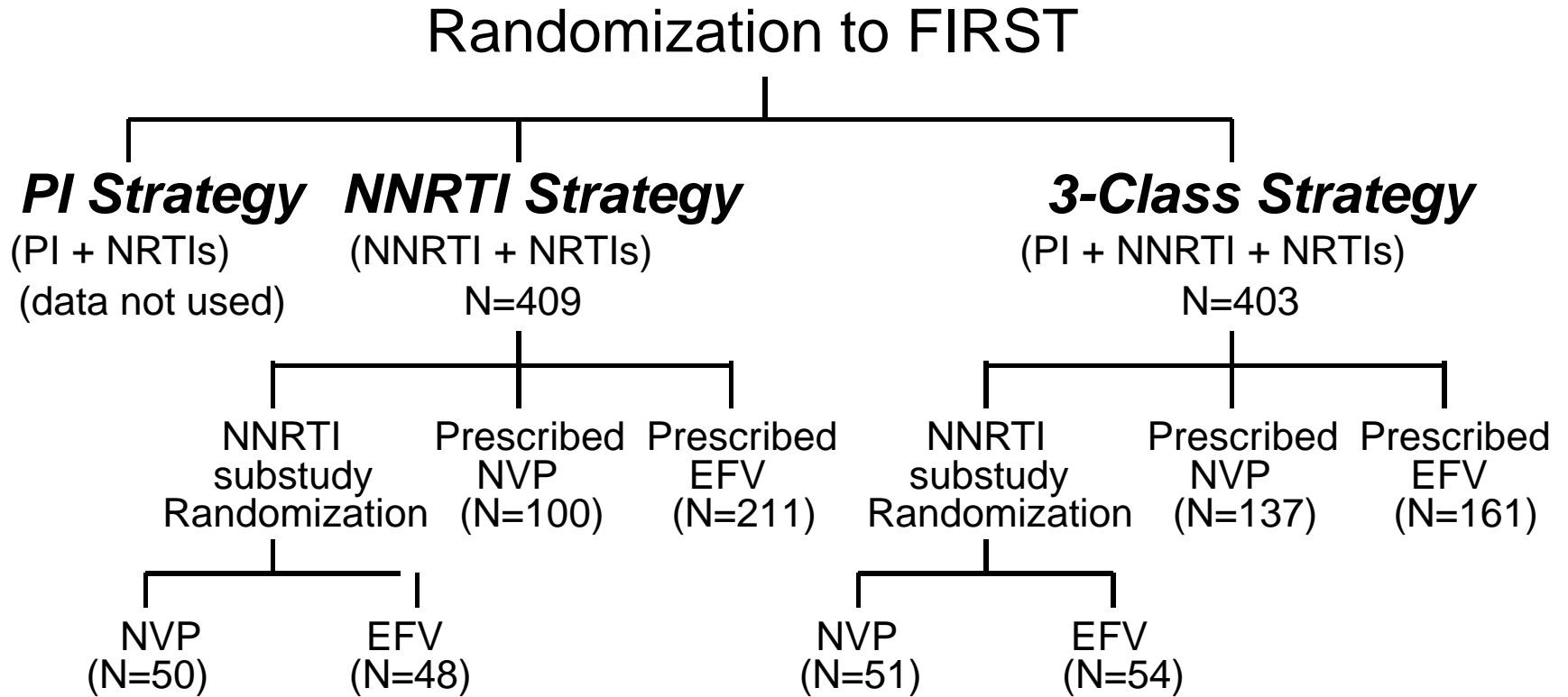
Bayesian design

- Thus our Bayesian sample size problem comes down to choosing a design (i.e., a sample size N and indifference zone) that delivers some prespecified acceptable frequentist properties
- The use of an informative **fitting prior** is likely to pay dividends (greater power) in cases where it is congruent with the “truth”
- The **design prior** need *not* be the same as the one used to fit the model
- Notice the posterior calculation for each fake data vector \mathbf{Y}_j^* may require MCMC methods. I used the new BRugs package, which permits the use of OpenBUGS commands within R (<http://www.r-project.org>)

Application in drug trials

- The first example concerns the Flexible Initial Retrovirus Suppressive Therapies (FIRST) trial (CPCRA 058).
- This was a large, long-term, randomized, prospective comparison of three different antiretroviral strategies in highly active antiretroviral therapy-naive, HIV-1-infected persons (MacArthur et al., 2001)
- We will use the Bernoulli-logit model to compare the probabilities of virological suppression (HIV RNA < 50 copies/mL) under two non-nucleoside reverse transcriptase inhibitors (NNRTIs) Efavirenz (EFV) and Nevirapine (NVP) at 32 weeks, $i = 1, \dots, 812$

Application in drug trials



Outline of FIRST design and randomization for eligible subjects (Berg-Wolf et al., 2006).

Application in drug trials

- Let $\beta = \lambda_1 - \lambda_2$
- The parameter of interest is $\Delta = e^\beta$

		percentiles		
		mean	2.5	97.5
EFV	λ_1	0.2870	-0.0763	0.6454
NVP	λ_2	-0.0644	-0.4344	0.3039
choose	λ_3	0.3195	-0.0068	0.6508
RNA < 10 ⁵	λ_4	0.5407	0.2465	0.8312
PI+NNRTI	λ_5	-0.2890	-0.5715	-0.0052
diff	β	0.3514	0.0595	0.6452
O.R.	Δ	1.4370	1.0610	1.9060

Using FIRST to design SECOND

- FIRST was the first prospective, randomized trial comparing EFV and NVP for VS of HIV-1 RNA when enrollment began in February of 1999
- Results from two other similar randomized studies have now appeared: the SENC (Spanish Efavirenz versus Nevirapine Comparison) trial reported by Núñez et al. (2002), and the 2NN (2 Non-Nucleoside reverse transcriptase inhibitors) study reported by van Leth et al. (2004)
- Contrary to FIRST (two-sided frequentist p -value 0.011 for $H_0 : \Delta = 1$), both SENC and 2NN suggested no significant difference between EFV and NVP at $\alpha = 0.05$
- Resolving the question of which NNRTI is more effective may require another randomized trial

Design prior specification

- Frequentist point **design priors** will be used for λ
- We need to model “true” Δ through λ_1 and λ_2
- Fix prior mean for λ_2 at -0.0644
- Increment the prior mean for λ_1 from -0.0644 by “true” β values of $\log(1)$, $\log(1.1)$, $\log(1.2)$, $\log(1.3)$, $\log(1.4)$, $\log(1.5)$, and $\log(1.6)$
- Therefore $\lambda_2 \sim N(-0.0644, 0)$, $\lambda_1 \sim N(\beta - 0.0644, 0)$
- Sample size calculations will correspond to fixed “true” Δ values of 1, 1.1, 1.2, 1.3, 1.4, 1.5, and 1.6.

Notice that when the true β equals $\log(1.4)$, this model assigns the prior means $(-0.0644, 0.2721)$ to (λ_1, λ_2) , roughly the same values seen in FIRST.

Fitting prior specification

- We will compare results under choices that are **optimistic**, **skeptical**, and **noninformative** with respect to the superiority of EFV for VS
- The **optimistic model** utilizes normal fitting priors with FIRST parameter estimates
- The model **skeptical** to the superiority of EFV assumes that the effects of EFV and NVP are *equivalent* by assigning mean -0.0644 to the fitting priors for *both* λ_1 and λ_2 , as well as a pooled variance estimate
- Define $\delta_U = 1 + \phi$ and $\delta_L = 1/(1 + \phi)$ for $\phi > 0$

Specificity of skeptical model

We ran the Gibbs sampler for 1500 iterations following a burn-in of 500 for each of $N_{rep} = 300$ replicate fake datasets

- Empirical probabilities for the six outcomes by n for the skeptical analysis model for true $\Delta = 1$, and $\phi = 0.30$

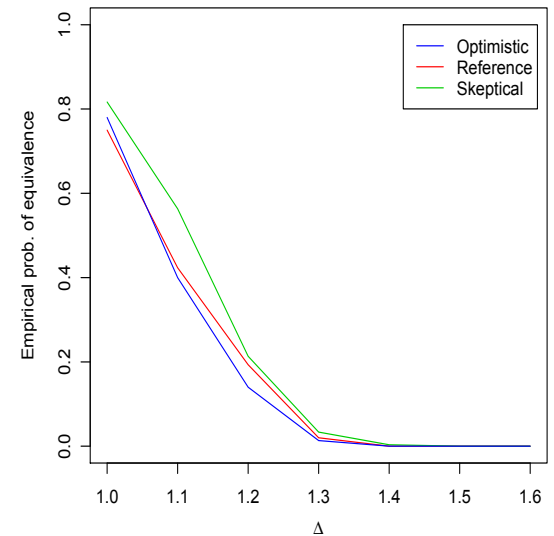
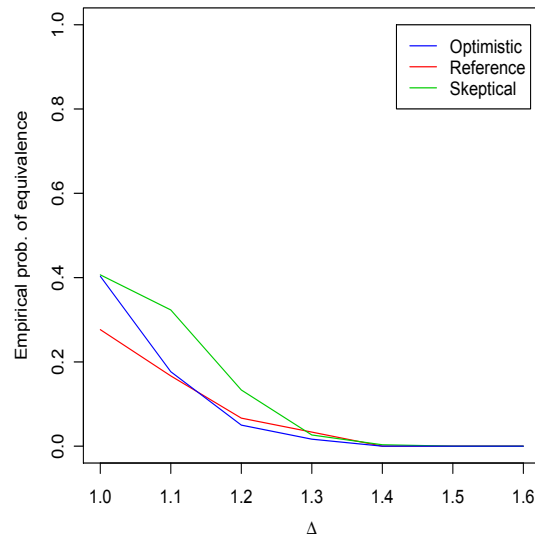
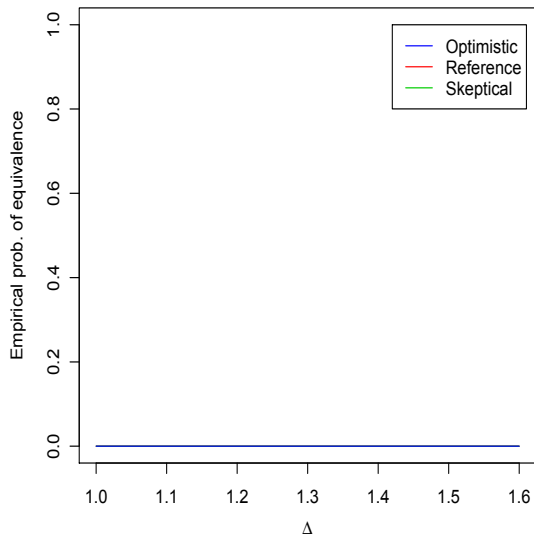
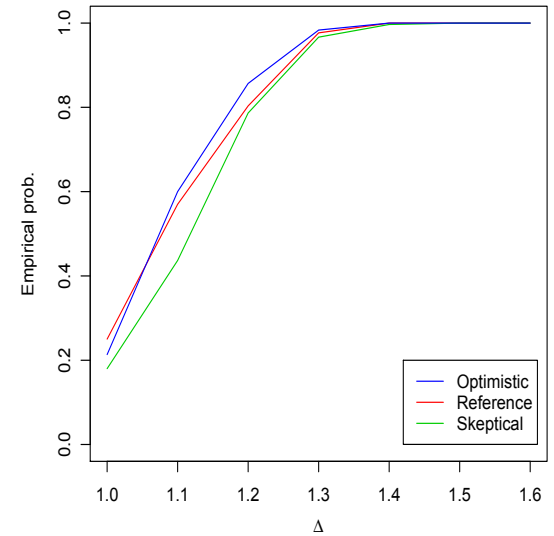
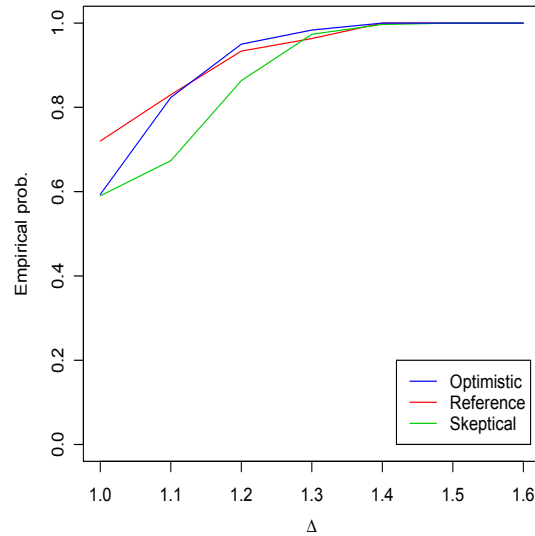
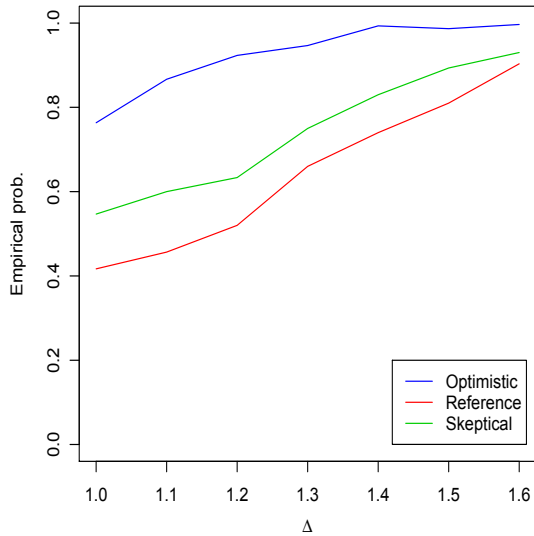
n	acc NVP	rej EFV	equiv	rej NVP	acc EFV	no dec
320	0.00	0.44	0.00	0.50	0.00	0.06
480	0.00	0.42	0.21	0.37	0.00	0.00
640	0.00	0.31	0.41	0.28	0.00	0.01
800	0.00	0.18	0.63	0.19	0.00	0.00
960	0.00	0.12	0.74	0.14	0.00	0.00
1120	0.00	0.08	0.82	0.10	0.00	0.00

Specificity of optimistic model

- Empirical probabilities for the six outcomes by n for the optimistic analysis model for true $\Delta = 1$ and $\phi = 0.30$

n	acc NVP	rej EFV	equiv	rej NVP	acc EFV	no dec
320	0.00	0.19	0.00	0.77	0.00	0.04
480	0.00	0.20	0.16	0.64	0.00	0.00
640	0.00	0.13	0.41	0.46	0.00	0.00
800	0.00	0.08	0.52	0.40	0.00	0.00
960	0.00	0.06	0.70	0.24	0.00	0.00
1120	0.00	0.05	0.78	0.17	0.00	0.00

Comparison of empirical probabilities



Closing thoughts

- Incorporating prior information through the use of a Bayesian design may save time and expense without sacrificing the specificity and sensitivity properties long prized by clinical trialists
- Bayesian designs account for uncertainty at every model stage and facilitate thorough reviews of contemporary knowledge
- The FDA has already acknowledged the advantages of being Bayesian in the device sector
- Learn BRugs: [Dr. Carlin's BRugs web-page](#)
- Visit my [software web-page](#)
- Thank you for coming to my talk