Case Study of a Maximally Flexible Bayesian Design in Biologics Research with a Skeptical Prior and Cox Model-based Analysis

Frank E Harrell Jr[1] Nate Mercaldo[1]
David Apelian[2] Timothy C. Rodell[2]

[1]Department of Biostatistics
Vanderbilt University School of Medicine
Nashville TN

[2]GlobeImmune, Inc.
Louisville CO

2^{nd} Bayesian Biostatistics Conference—MD Anderson 28 Jan 2009

Goals

- Reliable inference about therapeutic efficacy
- Appropriate trade-off between flexibility and carefulness
- Infinitely many looks at the data
- Extend an equivocal study without penalty

Options for Statistical Planning & Inference

- Frequentist
- Randomization/permutation inference
- Non-Bayesian non-frequentist likelihood
- "Relative" Bayes—Bayes factors
- Predictive Bayes
- Bayes using posterior probabilities of efficacy

Advantages of Bayes

Frequentist	discover an event whose probability is easy to compute
Bayesian	discover how to compute the probability of the event of interest

- Bayes result in a direct forward probability model
- Adheres to *likelihood principle*
- Does not dwell on sample space of contemplated experiments
- Uses only the sample that *has arisen*

Harrell and Shih [2001],Emerson [1995]

Advantages of Bayes, *continued*

No penalty for multiple looks or extending the study

- no down-weighting of previously collected data
- interval estimates coherent with inferential probabilities

Cost and speed: $\downarrow$ average sample size

- early termination for futility, efficacy, or harm
- extending the **right** study

Using the latest information because ∞ looks at data are OK

Only ways to cheat

- suppressing new information
- changing the prior to tilt results
- the usual suspects

Prior Distribution

- Use the prior to encapsulate carefulness/skepticism instead of tightening criteria on the posterior
  - early looks discount observed results; prior loses influence as information $\uparrow$
  - very early termination difficult
  - if terminate early, skepticism pulls point estimate towards zero (unlike commonly used frequentist methods)
- Skeptical prior placing small probability τ of a large treatment effect δ
- $N(0, \sigma^2), \sigma^2 = [\frac{\delta}{\Phi^{-1}(1-\tau)}]^2$
- $\tau = 0.025, \delta = \log(\frac{1}{3}), \sigma = 0.5606$
- $\Pr(\lambda < \frac{1}{3}) = 0.025$

Spiegelhalter et al. [2004], section 5.5

Data Model: Cox PH

- Let T=time to event
- $P[T > t] = S_0(t)^{exp(X\beta)}$, S_0 unspecified
- $Y_1, Y_2, \ldots, Y_n$ = event/censoring times
- X_i = treatment assignment for i^{th} subject
- log partial likelihood is
  $\log L(\beta) = \sum_{Y_i \text{ uncensored}} \{X_i\beta - \log[\sum_{Y_j \geq Y_i} \exp(X_j\beta)]\}$

Exact Bayesian Calculations

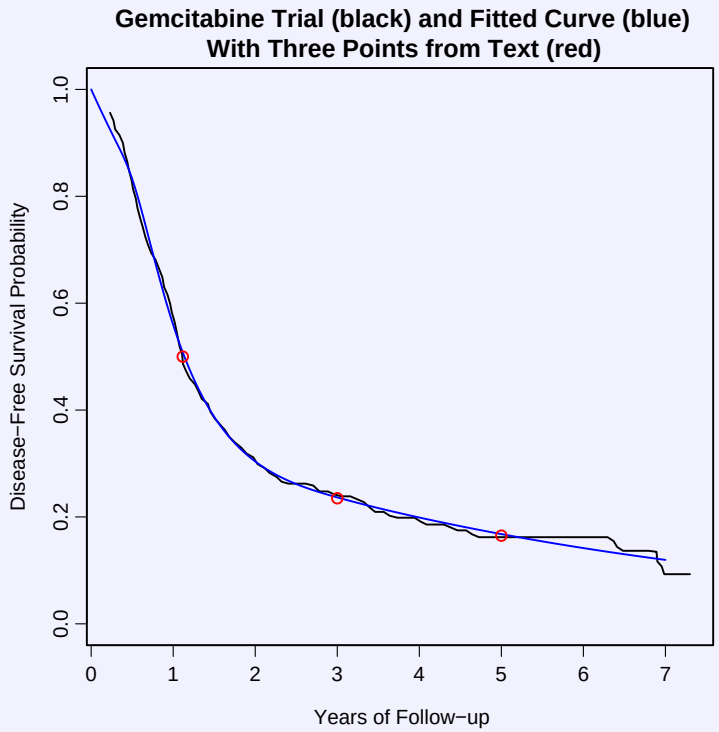
- "Poisson-multinomial trick" described in the BUGS manual
- Model suitable for BUGS-type Markov chain Monte Carlo (MCMC) simulation
- Counting process increments $dN_i(t)$ are distributed as Poisson$(I_i(t))$, where $I_i(t) = \lambda_0(t)e^{X_i\beta}$ if subject i is observed at time t or zero otherwise, and λ_0 is the underlying nonparametric hazard function
- Assume independent increments in Λ are failure times whose logarithms are given "non-informative" priors
- Used `JAGS` with R package `rjags` (both by Martyn Plummer)

Thomas et al. [1992] and p. 61 of `http://www.mrc-bsu.cam.ac.uk/bugs/documentation/Download/eg05vol1.pdf`; `http://calvin.iarc.fr/~martyn/software/jags`

Clinical Setting and Data used for Control $S(t)$

- Ongoing Phase 2 RCT in newly diagnosed, resected pancreas cancer comparing gemcitabine + therapeutic vaccine (GI-4000) vs. gemcitabine alone
- $\hat{S}(t)$ for control arm from gencitabine trial $n = 186$
- Fit restricted cubic spline in t to $\hat{\Lambda}(t)$ without intercept

Only used for simulating operating characteristics



Control survival estimates from Oettle et al. [2007], accrual 1998-2004

Criteria for Efficacy and Extension

- Efficacy: $\Pr(\lambda < 1) \geq 0.95$
- Extension of follow-up and recruitment: $0.7 \leq \Pr(\lambda < 1) < 0.95$

Simulations

- Excellent approximation to posterior for $\log \hat{\lambda} = \hat{\beta}$ assuming ordinary Cox partial likelihood is Gaussian
- Use estimated variance $\hat{v}$ from the partial likelihood
- Posterior for $\hat{\beta}$ is $\approx N(\mu, \Sigma)$

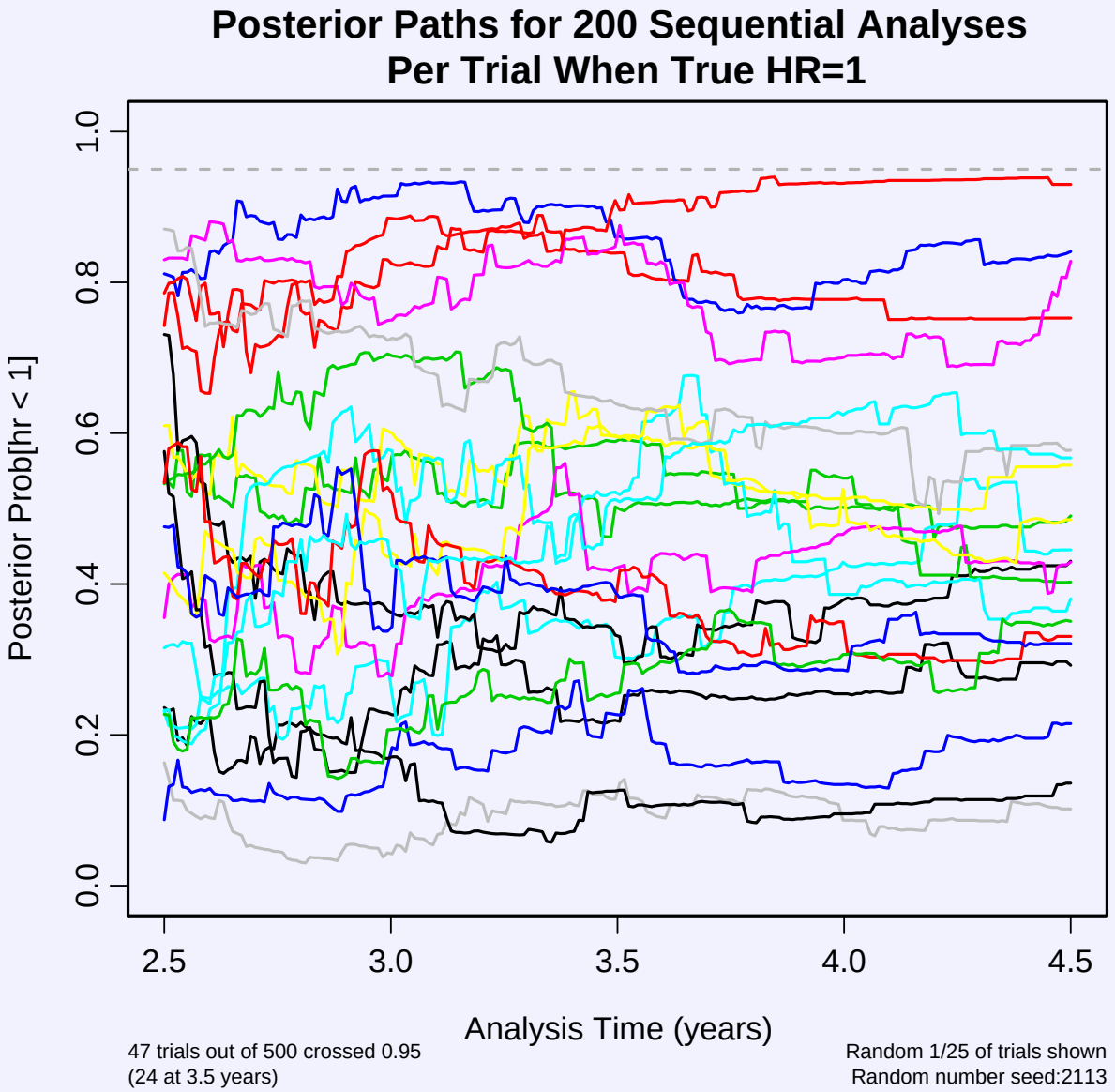
$$\Sigma = \frac{1}{\frac{1}{\sigma^2} + \frac{1}{\hat{v}}}$$

$$\mu = \frac{\Sigma}{\hat{v}} \hat{\beta}$$

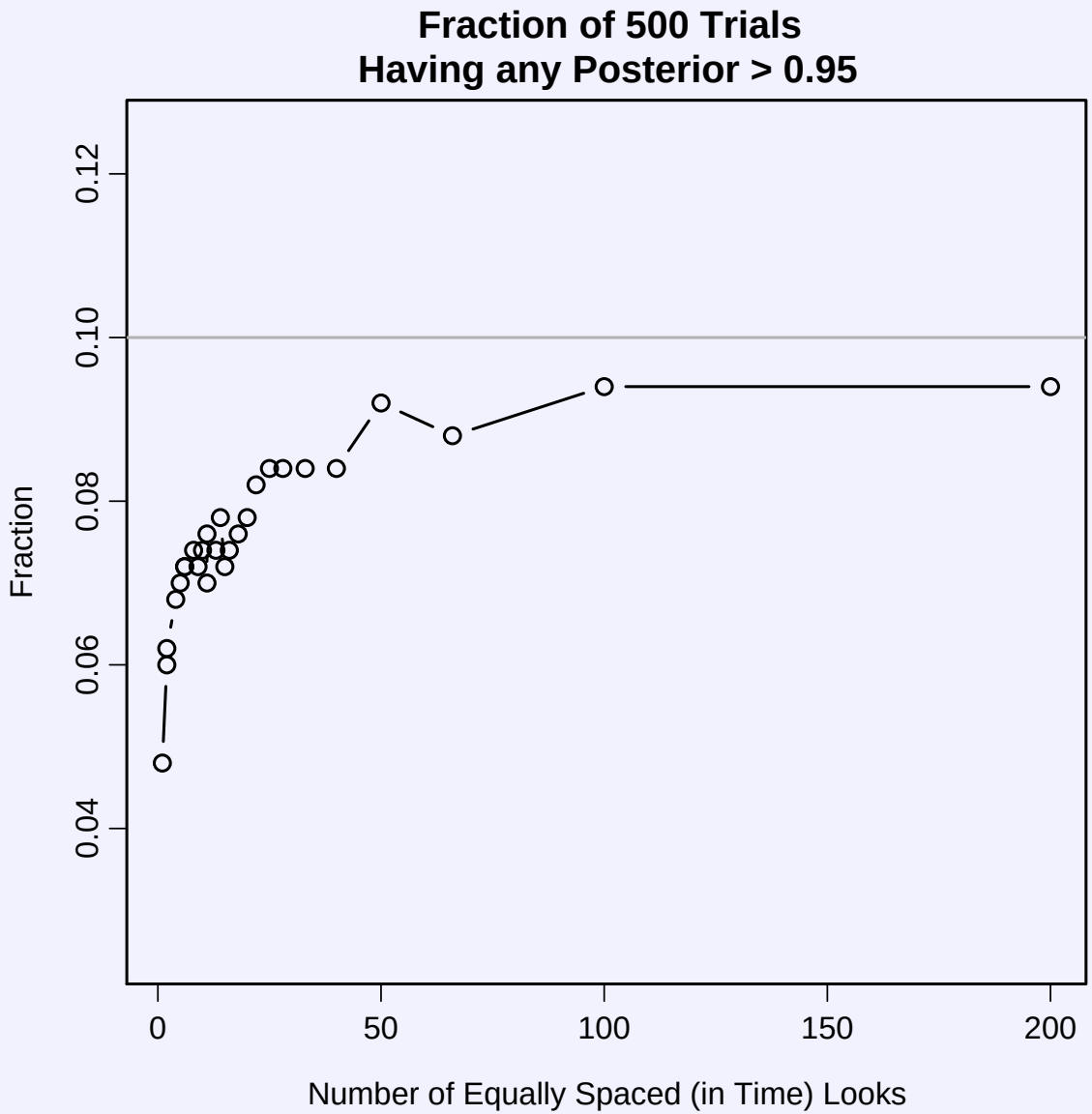
Simulations, *cont.*

- Simulate 500 clinical trials from fitted distribution for control and distributions shifted by a hazard ratio
- Approximate ∞ looks by doing more analyses (200) than patients (100)
- Uniform accrual uniform over 29 months
- Uniform analysis times from 30 to 54 months after first patient accrued
- Simulate all failure times once per trial, revealing progressively more data as analyses progress
- Toughest test: assume $\lambda = 1$
- Misleading evidence: posterior probability ≥ 0.95
- Compute proportion of trials in which posterior crosses 0.95
  - at any time
  - at 42 months (3.5y)

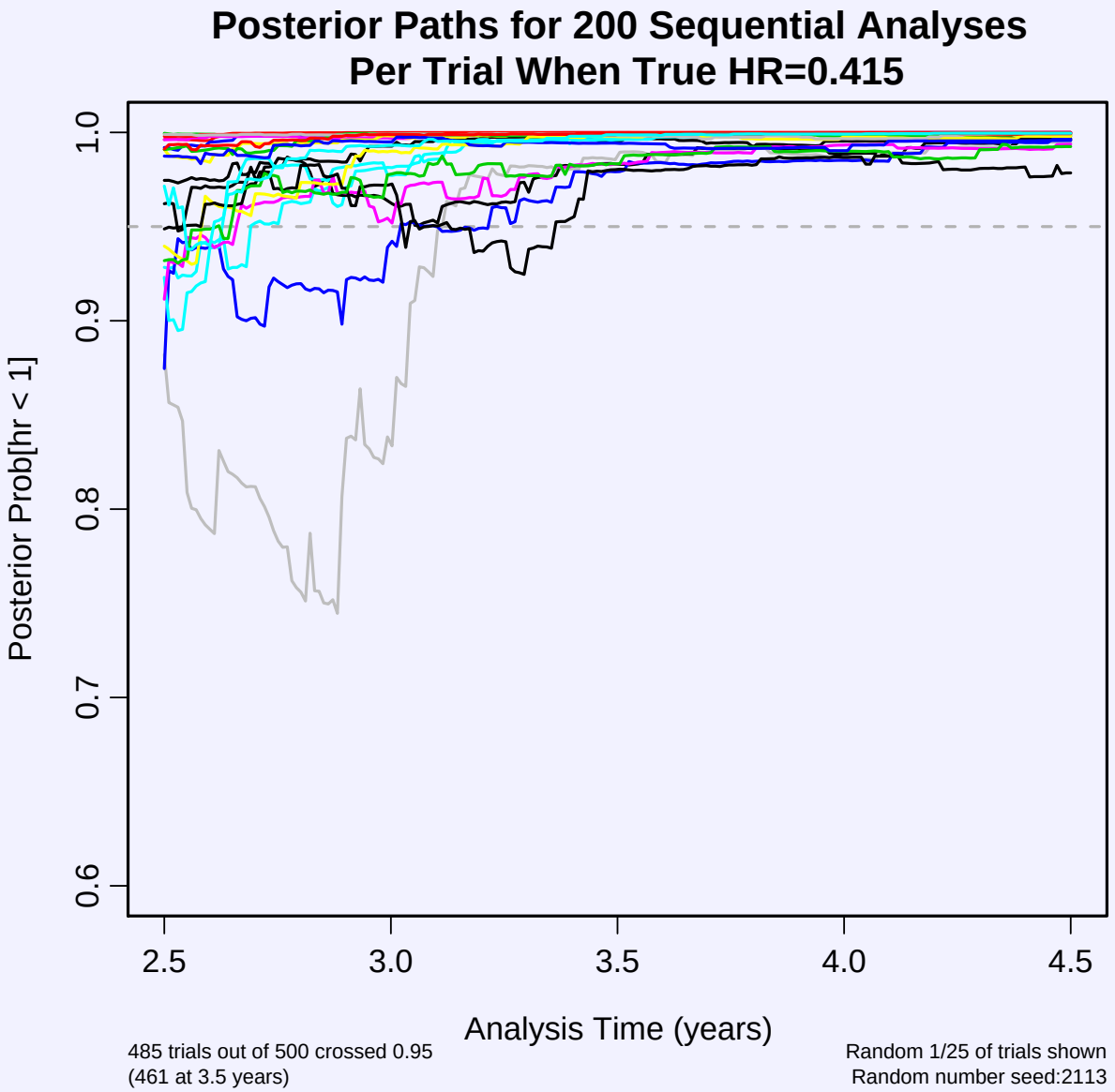
Simulation Results: Null Case



Null Case, *continued*



Simulation Results: $\lambda = 0.415$



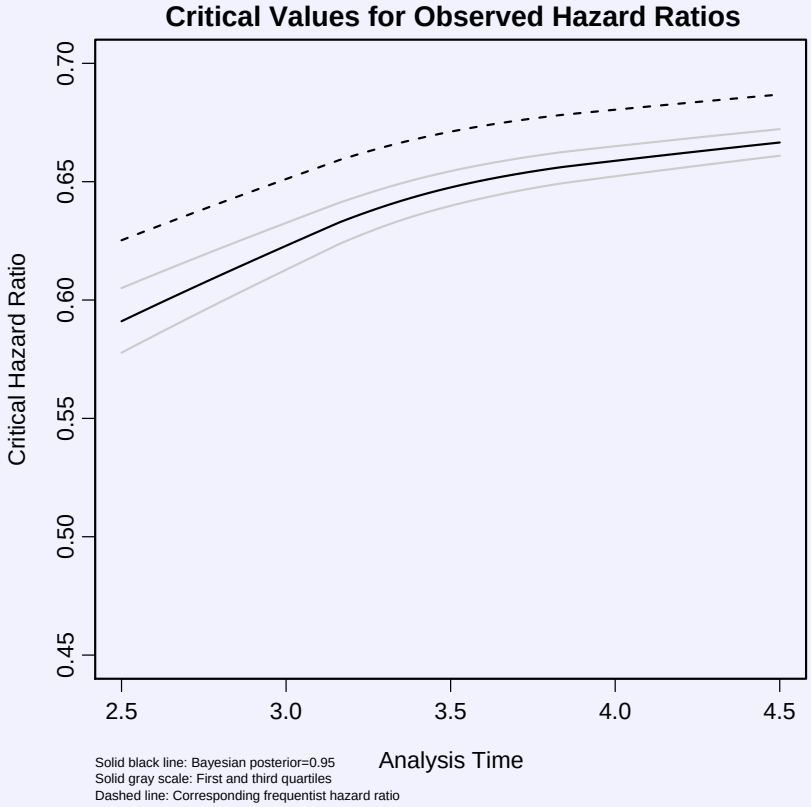
Critical Values for $\hat{\lambda}$

- Useful to show clinical researchers and reviewers the values of $\hat{\lambda}$ that would result in a posterior probability of 0.95 at a given analysis time
- This can be compared to the corresponding frequentist $\hat{\lambda}$
- Will depend on true value of λ as there are more events with λ closer to 1.0
- Critical value $\approx \exp[-\Phi^{-1}(0.95)\frac{\hat{\nu}}{\sqrt{\Sigma}}]$

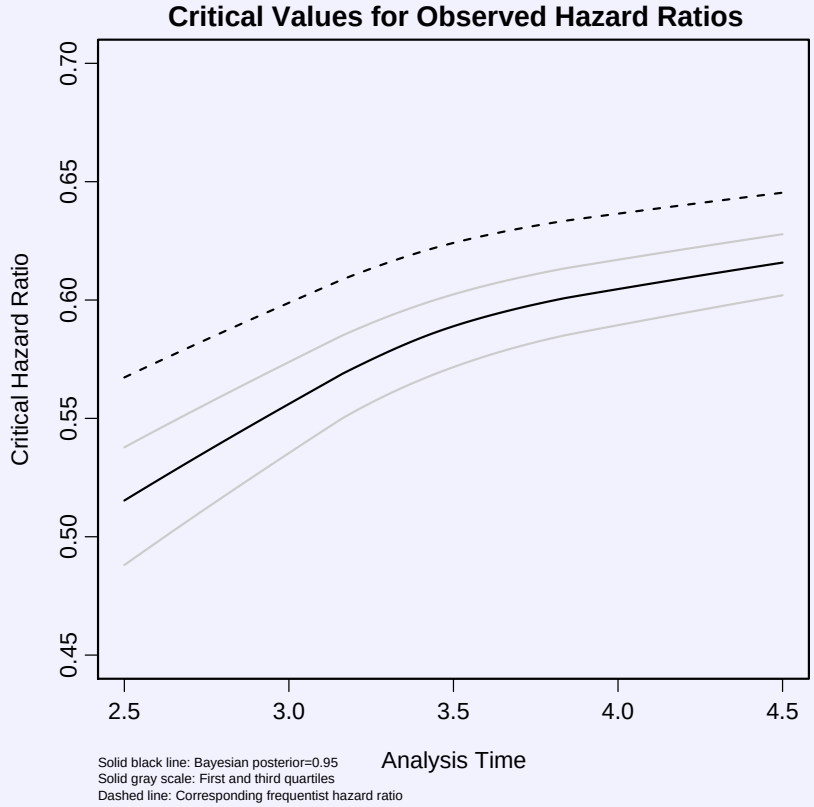
See Emerson [2006] regarding utility of clinically scaled critical values

Bayesian and Nominal Frequentist Critical Values of $\hat{\lambda}$

$\lambda = 1$



$\lambda = 0.415$



Gray-scale bands indicate lowess smooths of outer quartiles of observed approx. Bayesian critical values.

Experience Working with FDA

- More flexibility in trial design/analysis in the Phase 2 setting
- An excellent opportunity to engage FDA and your own organization
- Acceptability of Bayesian design for pivotal registration trial to be discussed
- FDA statistical reviewer wants highly annotated simulation code
- Major points of discussion:
  - FDA's wish for complete specification of criteria for study extension
  - Changing original endpoint from 18m binary event to time-to-event

Difficulties with FDA CDRH Bayesian Guidance Document

- Emphasis on *the* frequentist operating characteristics
- No such thing as *the*
- *The* depends on intentions, conduct of DMC meetings, ...
- Conditions on the future and on the treatment effect being **exactly** zero
- Creates work
- Perpetuates *P*-values and arbitrarily chosen multiplicity adjustments
- Fails to recognize that Pr(efficacy) at any moment is fully meaningful no matter what the previous analysis path has been
- E.g., veracity of an estimate of Pr(heads) after 100 coin flips does not depend on the veracity of Pr(heads) after 50
- Why not concentrate on Bayesian operating characteristics of Bayesian procedures (sensitivity analysis re: prior, etc.)?

References

D. A. Berry. Bayesian clinical trials. *Nat Rev*, 5:27–36, 2006. Editorial, p. 3.

S. S. Emerson. Issues in the use of adaptive clinical trial designs. *Stat Med*, 25:3270–3296, 2006.

S. S. Emerson. Stopping a clinical trial very early based on unplanned interim analysis: A group sequential approach. *Biometrics*, 51:1152–1162, 1995.

F. E. Harrell and T. Shih. Using full probability models to compute probabilities of actual interest to decision-makers. *International Journal of Technology Assessment in Health Care*, 17:17–26, 2001.

H. Oettle, S. Post, P. Neuhaus, et al. Adjuvant chemotherapy with gemcitabine vs observation in patients undergoing curative-intent resection of pancreatic cancer. *JAMA*, 297(3):267–277, 2007.

D. J. Spiegelhalter, K. R. Abrams, and J. P. Myles. *Bayesian Approaches to Clinical Trials and Health-Care Evaluation*. Wiley, New York, 2004.

A. Thomas, D. J. Spiegelhalter, and W. R. Gilks. Bugs: A program to perform Bayesian inference using Gibbs sampling. In J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith, editors, *Bayesian Statistics*, volume 4, pages 837–842. Clarendon Press, Oxford, UK, 1992. `http://www.mrc-bsu.cam.ac.uk/bugs`.

Case Study of a Maximally Flexible Bayesian Design in Biologics Research with a Skeptical Prior and Cox Model-based Analysis

Frank Harrell and Nate Mercaldo
Dept of Biostatistics, Vanderbilt University School of Medicine
David Apelian and Timothy Rodell
GlobeImmune, Inc.

This presentation will cover the development of a Bayesian flexible design for a biologic agent in an ongoing Phase 2 trial in pancreatic cancer. The design allows for infinitely many looks at the data and for possible study extension and conversion to adaptive allocation. Unlike frequentist sample size re-estimation procedures, the Bayesian procedure does not require penalizing for the final analysis for having done earlier analyses nor does it require down-weighting of data collected before the decision to extend the study. The study is easily extended if results obtained at the originally planned study termination are equivocal. The final analysis uses the same analysis procedure as used at the initial analysis, whereas there is no consensus in the frequentist world for how to analyze an extended study.

Our primary analysis is based on a Bayesian Cox proportional hazards model using a skeptical prior distribution. The endpoint is time to cancer recurrence or death. Evidence for efficacy is taken to be a posterior probability of efficacy ≥ 0.95 at any analysis time, where "efficacy" means a true hazard ratio < 1.0. The planned rule for extending the study is a probability of efficacy ≥ 0.7 at the last pre-planned analysis. Results of simulations to study the properties of the design will be presented, and experience in presenting the design to the FDA will be briefly discussed.

The traditional frequentist approach, in order to compute the probability of getting a result as or more impressive than that observed if there is truly no treatment effect (the P-value), requires contemplating experiments that were never carried out and analyses that were never done. Great simplicity is had when only "forward" probabilities (Bayesian posterior probabilities) are computed, conditioning only on what has been observed up to the time of analysis and not conditioning on unknowable information such as the true population treatment effect.