Towards Structured Use of Bayesian Sequential Monitoring in Clinical Trials

Evan Kwiatkowski[†], Eugenio Andraca-Carrera[‡],
Mat Soukup[‡], Matthew A. Psioda[†*]

† Department of Biostatistics, University of North Carolina,
McGavran-Greenberg Hall, CB#7420,
Chapel Hill, North Carolina, U.S.A.

‡ Division of Biometrics VII, Office of Biostatistics
Center for Drug Evaluation and Research,
US Food and Drug Administration,
Silver Spring, Maryland, USA

June 28, 2019

Abstract

The text of your abstract. 200 or fewer words.

Keywords: 3 to 6 keywords, that do not appear in the title

[*]The authors gratefully acknowledge *please remember to list all relevant funding sources in the unblinded version*

1 Introduction

Things to discuss:

- 21st Century Cures Act (MATT)
- PDUFA VI reauthorization (MATT)
- Expansive work already done on sequential monitoring (EVAN – draft on 6/21)
    1. Foundational Bayesian sequential monitoring Cornfield (1966*a*) Cornfield (1966*b*) Neyman & Greenhouse (1967).
    2. Further development Freedman & Spiegelhalter (1989) Freedman & Spiegelhalter (1992) Spiegelhalter et al. (1993) Spiegelhalter et al. (1994) Fayers et al. (1997).
    3. Need to go further, and mention work on 3-arm Bayesian trials.
- Our majors contribution (EVAN – as early as possible in introduction without having the flow appear weird – draft on 6/21)
- Outline for the remaining section of the paper (EVAN – draft on 6/21)

2 Methods

As you introduce ideas that come from or extend other ideas in the literature, cite the relevant literature.

2.1 Monitoring versus Estimation Priors (EVAN – draft on 6/21)

- Define generally in terms of $\boldsymbol{\theta} = (\gamma, \boldsymbol{\psi})$ where γ is a parameter of interest and $\boldsymbol{\psi}$ is a nuisance parameter (possible vector valued).
- Define *Monitoring* Priors and *Inference* Priors.
- Make connection between Inference priors and two-part mixture prior and BMA.
- Define *Skeptical* and *Enthusiastic* monitoring priors and how each would be used.
- I would have a generic graphic to illustrate the types of priors and the mixture.

2.1.1 Bayesian hypothesis testing based on posterior probabilities

The Bayesian formulation allows direct inference on a parameter of interest through specification of a model for the data generating mechanism and prior distributions for unknown quantities. Let $\mathbf{D}$ be a random variable representing the data collected in the trial with density $p(\mathbf{D}|\theta, \psi)$ where θ and ψ are the unknown quantities. Let θ be the parameter of interest and ψ be the unknown quantities that are not of research interest (i.e. "nuisance parameters"). Define the sample spaces for the unknown quantities as $\theta \in \Theta$ and $\psi \in \Psi$. Suppose the hypotheses under consideration are $H_0 : \theta \in \Theta_{H_0}$ versus $H_1 : \theta \in \Theta_{H_1}$, where $\Theta = \Theta_{H_0} \cup \Theta_{H_1}$ and $\Theta_{H_0} \cap \Theta_{H_1} = \emptyset$. These hypotheses are adjudicated based on posterior probabilities of θ by evaluating $P(\theta \in \Theta_{H_i}|\mathbf{D}) = \int_{\Theta_{H_i}} p(\theta|\mathbf{D})d\theta$ for $i \in \{0, 1\}$, where $p(\theta|\mathbf{D}) = \int_{\Psi} p(\theta, \psi|\mathbf{D})d\psi$. The posterior distribution of θ depends on the choice of prior distribution $\pi(\theta, \psi)$ since $p(\theta, \psi|\mathbf{D}) = p(\mathbf{D}|\theta, \psi)\pi(\theta, \psi)/p(\mathbf{D})$ by Bayes rule. The specification of the prior distribution depends on the research objective. An *inference prior* is a prior that is used when the research objective is to make final analysis after data collection

is complete. A *monitoring prior* is a prior that is used when the research objective is monitor enrollment with an eye towards early termination.

It has been said that 'the purpose of a trial is to collect data that bring to conclusive consensus at termination opinions that had been diverse and indecisive at the outset' (Kass and Greenhouse (1989)). These opinions manifest as priors $\pi(\theta, \psi)$ for which their relation to $P(\theta \in \Theta_i | \pi(\theta, \psi))$, $i \in \{0, 1\}$ is examined (note this quantity does not depend on the data $\mathbf{D}$ and therefore reflect a-priori opinion). An inference prior is often non-informative or objective in the sense that it does not show preference to H_0 or H_1. This prior $\pi_I(\theta, \psi) \equiv \pi_I$ has the property that $P(\theta \in \Theta_0 | \pi_I) \approx P(\theta \in \Theta_1 | \pi_I)$. A skeptical prior is an informative or subjective prior that gives preference to H_0. This prior $\pi_S(\theta, \psi) \equiv \pi_S$ has the property that $P(\theta \in \Theta_0 | \pi_S) > 1 - \epsilon_S$ (equivalently $P(\theta \in \Theta_1 | \pi_S) < \epsilon_S$) for some $\epsilon_S > 0$ whereby $1 - \epsilon_S$ signifies a high level of certainty. For example, if $\epsilon_S = 0.05$, then this choice of skeptical prior places 95% prior probability that $\theta \in \Theta_0$. An enthuastic prior $\pi_E(\theta, \psi) \equiv \pi_E$ similarly gives preference to H_1 through the property that $P(\theta \in \Theta_1 | \pi_E) > 1 - \epsilon_E$ (equivalently $P(\theta \in \Theta_0 | \pi_E) < \epsilon_E$) for some $\epsilon_E > 0$.

Early termination of the trial is only appriopriate if diverse prior opinions about θ would be in agreement given the interm data. It is then reasonable to stop data collection when a sufficiently *skeptical person* is convinced that H_1 is true or a sufficiently *enthuastic person* is convinced that H_1 is false. The use of monitoring based on changing the opinion of skeptical and enthuastic priors has been described as overcoming a handicap (Freedman & Spiegelhalter (1989)) and providing a brake (Fayers et al. (1997)) on the premature termination of trials, or constructing "an adversary who will need to be disilusioned by the data to stop further experimentation" (Spiegelhalter et al. (1994)). Let $\delta > 0$ be chosen such that $1 - \delta$ is a threshold for "convinced" that will be used for the posterior probability

of θ at an interm analysis. It is reasonable to stop the trial for efficacy if, once presented with the data, the skeptical viewpoint as updated via Bayes rule becomes convinced of the alternative hypothesis, that is, if $P(\theta \in \Theta_1|\mathbf{D}, \pi_S) > 1 - \delta_S$. For example, if $\delta_S = 0.1$, then the trial can be stopped if the skeptical viewpoint now is 90% convinced of treatment efficacy. Similarly, it is reasponable to stop the trial for futility if $P(\theta \in \Theta_0|\mathbf{D}, \pi_E) > 1-\delta_E$.

Final inference on θ is made once enrollment is stopped based on the monitoring priors or at the planned end of the trial. We propose use of a mixture prior constructed from the monitoring process as the inference prior. Define $p(\mathbf{D}|\pi) = \int L(\theta|\mathbf{D})\pi(\theta)d\theta$ to be the marginal likelihood for the data given the prior π. One can average the posterior means from the analyses using the skeptical and enthuastic priors, such that,

$$E(\theta|\mathbf{D}, \pi_{Inference}) = \omega \times E(\theta|\mathbf{D}, \pi_S) + (1 - \omega) \times E(\theta|\mathbf{D}, \pi_E)$$

for $\omega \in [0, 1]$. Choosing $\omega = 1/2$ corresponds to an agnostic inference prior. Choosing ω based on posterior model probabilities of the null and alternative hypothesis yields $\omega = p(\mathbf{D}|\pi_S/(p(\mathbf{D}|\pi_S) + p(\mathbf{D}|\pi_E))$.

2.1.2 Default parameterization of monitoring priors for common designs

Define prior distribution as $\pi(\theta|\lambda)$ where λ is a vector of hyperparameters.

Skeptical prior might be centered around the clinical inferiority boundary. Enthuastic prior might be centered around clinical superiority boundry. Variance based on tail area considerations, or by matching the skeptical prior variance.

Reference prior attempts to express no particular opinion about the treatment's merit.

2.2 Futility Monitoring Using Probability of Success (EVAN – draft on 6/21)

- Futility monitoring using POS is about stopping early when their is a high likelihood of a study being inconclusive at the end of the study.
- Since the final analysis uses the *Inference* prior, POS should be based on the inference prior.
- Develop the framework for POS and show how it is a weighted average POS based on the skeptical and enthusiastic priors.

As an alternative strategy to futility analysis, one can monitor the probability of success (POS) for the trial. Let $\pi(\theta|D) \equiv \pi(\theta|D, \pi_S)$ denote the posterior distribution for θ based on the skeptical prior π_S and the current data D. Let $\psi \equiv \psi(D, \pi_S)$ denote the POS which is given as follows:

$$\psi = E[1\{P(\theta > 0.20|D_1, D, \pi_S) \geq 0.95\}]$$

where the expectation is taken with respect to the posterior predictive distribution $p(D_1)$ for future data D_1 (ongoing + subjects yet to enroll).

$$p(D_1) = \int p(D_1|\theta) \times \pi(\theta|D) d\theta$$

One may stop the enrollment if ψ is sufficiently small (i.e., $\psi < 0.05$).

Stochastic curtailment in frequentist setting.

3 Examples – (EVAN)

3.1 Single-Arm Oncology Proof-of-Activity Trial w/ Binary Endpoint

For example, consider testing the hypothesis $H_0 : \theta \leq \theta_0$ versus $H_1 : \theta > \theta_0$ where θ is a treatment effect of interest. Suppose an effect $\theta_1 > \theta_0$ is thought to be highly clinically relevant and plausible given what limited data is available. Consider a clinical trial with a single analysis and fixed sample size chosen that there is a high probability of proving H_1 when $\theta = \theta_1$. A standard Bayesian decision rule would reject H_0 when $P(\theta > \theta_0|D) \geq 0.95$ which will result in a type one error rate of 0.05 (approximately) if $\theta = \theta_0$ when the analysis prior is non-informative (a so-called reference or flat prior).

3.2 Parallel Two-Group Superiority Trial /w Continuous Binary Endpoint

3.3 Three-Arm, Placebo Controlled Non-Inferiority Trial w/ Continuous Endpoint

4 Discussion – (MATT/EVAN)

SUPPLEMENTARY MATERIAL

5 BibTeX

References

Cornfield, J. (1966*a*), 'A Bayesian Test of Some Classical Hypotheses, with Applications to Sequential Clinical Trials', *Journal of the American Statistical Association* **61**(315), 577.
URL: *https://www.jstor.org/stable/2282772?origin=crossref*

Cornfield, J. (1966*b*), 'Sequential Trials, Sequential Analysis and the Likelihood Principle', *The American Statistician* **20**(2), 18.
URL: *https://www.jstor.org/stable/2682711?origin=crossref*

Fayers, P. M., Ashby, D. & Parmar, M. K. B. (1997), 'Tutorial in Biostatistics: Bayesian Data Monitoring in Clinical Trials', *Statistics in Medicine* **16**(12), 1413–1430.
URL: *http://doi.wiley.com/10.1002/%28SICI%291097-0258%2819970630%2916%3A12%3C1413%3A%*
SIM578%3E3.0.CO%3B2-U

Freedman, L. S. & Spiegelhalter, D. J. (1989), 'Comparison of Bayesian with group sequential methods for monitoring clinical trials', *Controlled Clinical Trials* **10**(4), 357–367.
URL: *https://www.sciencedirect.com/science/article/pii/0197245689900019?via%3Dihub*

Freedman, L. S. & Spiegelhalter, D. J. (1992), 'Application of bayesian statistics to decision making during a clinical trial', *Statistics in Medicine* **11**(1), 23–35.
URL: *http://doi.wiley.com/10.1002/sim.4780110105*

Neyman, J. & Greenhouse, S. W. (1967), *Proceedings of the Berkeley Symposium on Mathematical Statistics and Probability.*, University of California Press.
URL: *https://projecteuclid.org/euclid.bsmsp/1200513830*

Spiegelhalter, D. J., Freedman, L. S. & Parmar, M. K. B. (1993), 'Applying Bayesian ideas in drug development and clinical trials', *Statistics in Medicine* **12**(15-16), 1501–1511.
URL: *http://doi.wiley.com/10.1002/sim.4780121516*

Spiegelhalter, D. J., Freedman, L. S. & Parmar, M. K. B. (1994), 'Bayesian Approaches to Randomized Trials', *Journal of the Royal Statistical Society. Series A (Statistics in Society)* **157**(3), 357.
URL: *https://www.jstor.org/stable/10.2307/2983527?origin=crossref*