Paper 4692-2020

Bayesian Sequential Monitoring of Clinical Trials Using SAS®

Matthew A. Psioda, Department of Biostatistics,
University of North Carolina at Chapel Hill

ABSTRACT

In this paper, we provide an overview of Bayesian sequential monitoring for clinical trials. In such trials, patients are continually enrolled and their data are analyzed as often as is desired or feasible until a hypothesis has been proven or disproven, or until the allocated resources for the trial have been exhausted (that is, the maximum sample size or study duration has been reached). Such an approach is particularly appealing in the case of difficult-to-enroll populations such as pediatric populations or for rare disease populations. A Bayesian sequentially monitored trial does not require a pre-specified sample size or number of analyses. For proving efficacy in a sequentially monitored trial, the Bayesian collects data until the evidence in favor of the investigational treatment is substantial from the perspective of an a priori skeptical judge who doubts the possibility of large treatment effects. The Bayesian approach naturally allows for the incorporation of prior information when determining when to stop patient accrual and ultimately in evidence evaluation once the complete data are available. We give easy-to-understand examples for how Bayesian methods can be applied in the setting of pediatric trials where it is of interest to extrapolate information from adult trials. We discuss SAS/IML® software that can be used to efficiently perform design simulations without high-powered computing infrastructure.

INTRODUCTION

The use of Bayesian methods for trial design has the potential to make the drug development process more efficient. Both the Prescription Drug User Fee Amendments of 2017 (PDUFA VI) and the 21st Century Cures Act contain language that is designed to make the use of Complex Innovative Trial Designs (CIDs), for which Bayesian approaches are commonly proposed, easier to use in drug development programs and in decision-making by regulatory authorities. The use of CIDs perhaps has the greatest promise in rare diseases and in pediatric indications where it is difficult to enroll patients and therefore important to use methods that allow for early and frequent analyses of accumulating data. In this setting, use of data from outside of a trial (henceforth, *historical data*) in the evaluation of treatment effectiveness may be valuable and the Bayesian paradigm offers a useful calculus for doing this.

MOTIVATING EXAMPLE

In this paper, we consider the design of a single arm pediatric trial undertaken after two phase III pivotal adult trials were completed for the same disease indication. Specifically, we will focus on redesigning the first phase of the T72 pediatric trial "A Study of the Safety and Efficacy of Infliximab (REMICADE) in Pediatric Patients with Moderately to Severely Active Ulcerative Colitis (UC)" (Hyman et al., 2012) using Bayesian sequential monitoring methods. The T72 trial also included a second phase to evaluate maintenance regimens for the investigational treatment. For the first phase of the T72 trial, the investigational product led to a response rate equal to 73.3% (44 of 60). Data previously collected from adults in the ACT1 and ACT2 trials (Rutgeerts et al., 2005), which used the same endpoint and the same period of follow-up from baseline to primary outcome assessment, could have been used prospectively in the pediatric trial's design and analysis. For the combined ACT1 and

ACT2 trials, there were 242 adult patients who received the dose investigated in the pediatric trial and a 67% responder rate was observed. The T72 trial primary analysis was designed to prove that the response rate for pediatric patients was greater than 40% (derived from the upper 95% confidence limit of the placebo responder rate from the adult trials).

SEQUENTIAL MONITORING

SEQUENTIAL MONITORING PHILOSOPHY

To frame our discussion in this paper, following the T72 trial, we consider a simple single arm trial with binary primary endpoint based on the response probability θ. We consider the one-sided hypotheses $H_0: \theta \leq \theta_0$ versus $H_1: \theta > \theta_0$ where $\theta_0 \in [0,1]$ is some fixed, prespecified constant. Summarizing Spiegelhalter et al. (1993) and others: A Bayesian may monitor data continually and stop collection when any of the following criteria have been met:

- A skeptical observer is convinced H_1 is true.
- An enthusiastic observer is convinced H_1 is false or that the benefit of treatment is not likely to be what was expected.
- The probability of *eventually* proving that H_1 is true is sufficiently low.
- The resources allocated for the trial have been exhausted (e.g., maximum sample size reached).

Assuming one can give satisfactory definitions for what it means to be a skeptical and enthusiastic observer and for what constitutes substantial evidence in favor of a hypothesis, it is difficult against the intuitiveness and interpretability of these criteria.

DEFINING SUBSTANTIAL EVIDENCE

It is common practice in Bayesian inference for analysts to reject a null hypothesis (e.g., $\theta \leq \theta_0$) when the event defining the alternative hypothesis has a high posterior probability. For example, for observed data $\boldsymbol{D}$, it is common to reject the null hypothesis when $P(\theta > \theta_0 | \boldsymbol{D}) > 0.975$. Note that the posterior probability $P(\theta > \theta_0 | \boldsymbol{D})$ depends on both the data $\boldsymbol{D}$ as well as the prior used for analysis. Thus, when we say a posterior probability exceeding 0.975 constitutes substantial evidence, it should be clear that the totality of evidence is comprised of the information in the data as well as the information in the prior. If one were to observe a posterior probability exactly equal to 0.975, you might say they would be *all but convinced* that the corresponding claim is true (e.g., $\theta > \theta_0$). The notion of being *all but convinced* of a claim is central to how we choose to define skeptical and enthusiastic priors.

SKEPTICAL & ENTHUSIASTIC CONJUGATE PRIORS

Suppose an effect θ_1 is thought to be highly clinically relevant and plausible given available data. Based on the hypotheses previously specified, it should be clear that $\theta_1 > \theta_0$. Given choices for θ_0 and θ_1, it is straightforward to construct skeptical and enthusiastic priors.

We define the belief of a skeptical observer (e.g., a skeptical prior) as one that satisfies the following properties:

1. The most likely value of θ is equal to θ_0, and
2. The probability that θ is equal to or exceeds θ_1 is 0.025.

Thus, the skeptical observer is all but convinced that the actual treatment effect is less than the hypothesized value(i.e.., θ_1). Note that skepticism here is a statement about the magnitude of the treatment effect and not the probability that the null hypothesis is true. To complete a *conjugate* skeptical prior (for the case of binary data), one must recognize that

the beta family is conjugate for the binomial likelihood and thus solve the system of non-linear equations:

1. $(\alpha - 1)/(\alpha + \beta - 2) = \theta_0$, and
2. $F(\theta_1|\alpha,\beta) = 0.975$,

where α and β are the shape paramters of a beta density and $F(\theta_1|\alpha,\beta)$ is the cumulative distribution function. The solution to this system of equations does not have a closed form but an adequate solution is easy to obtain using a grid search algorithm. The following source code provides an implementation for the skeptical prior using The IML Procedure.

```
proc iml;
      theta0 = 0.40; theta1 = 0.67; evidence_crit = 0.975;

      start calc_beta(alpha,mode);
            return ((1-mode)*alpha + 2*mode - 1)/mode;
      finish;

      delta = 0.01;
      do alpha = 1+delta to 10 by delta;
            beta = calc_beta(alpha,theta0);
            prob = cdf('beta',theta1,alpha,beta);
                  if (beta>1) &
                  ((evidence_crit-0.00005)<prob) &
                  (prob<(evidence_crit+0.00005)) then do;
                        print alpha beta prob evidence_crit;
                        alpha = 1e10;
                  end;
      end;
quit;
```

The source code assumes $\theta_0 = 0.40$, $\theta_1 = 0.67$, and an error tolerance for the error calculation that is equal to 0.00005. The shape parameters identified by the grid search algorithm are $\alpha = 5.830$ and $\beta = 8.245$. The skeptical beta prior associated with these shape parameters is plotted in Figure 1.

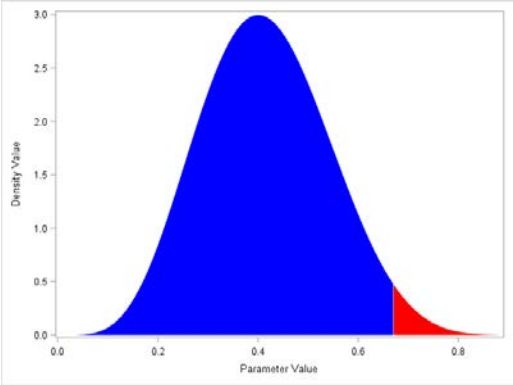

Figure 1: Skeptical Prior

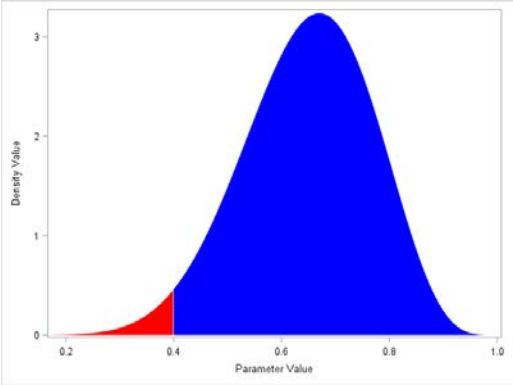

Figure 2: Enthusiastic Prior

We define the belief of an enthusiastic observer (e.g., an enthusiastic prior) as one that satisfies the following two properties:

1. The most likely value of θ is equal to θ_1.
2. The probability that θ is equal to θ_0 or less than θ_0 is 0.025.

Thus, the enthusiastic observer is all but convinced that the alternative is true. Minor modifications of the source code above allow once to identify the beta prior that meets these criteria. The shape parameters identified by the grid search algorithm are $\alpha = 9.790$ and $\beta = 5.329$. The enthusiastic beta prior associated with these shape parameters is plotted in Figure 2.

Constructing conjugate skeptical and enthusiastic priors is only possible in simple models but using such priors provides computational advantages and also results in a useful interpretation. For example, the skeptical prior can be viewed as being equivalent to observing $\alpha + \beta = 14.075$ patient's worth of data that is perfectly consistent with $\theta = \theta_0$ (i.e, the MLE $\hat{\theta}$ is equal to θ_0). The enthusiastic prior can be viewed as being equivalent to observing $\alpha + \beta = 15.119$ patient's worth of data that is perfectly consistent with $\theta = \theta_1$. Using a conjugate prior is this case is quite natural because it allows the skeptical (or enthusiastic prior) to be translated into hypothetical observed data that would lead to the posterior belief that is reflected by the prior.

SEQUENTIAL MONITORING PROCESS

Among other things, there are four important considerations for a sequentially monitored trial:

1. The minimum sample size $N_{\min}$ that corresponds to the number of patients that will be enrolled prior to commencing sequential monitoring. In most cases, regulatory authorities may require a minimum sample size to ensure some degree of safety assessment is possible.
2. The maximum sample size $N_{\max}$ that corresponds to the maximum number of patients that can be enrolled in the trial. Through in principle a sequentially monitored trial does not require a maximum sample size, for logistical reasons this constraint will almost always exist due to financial considerations.
3. The criteria for determining when there is sufficient evidence in favor or treatment efficacy such that further enrollment of patients in the trial is not warranted for proving efficacy.
4. The criteria for determining when there is sufficient evidence of futility (a null treatment effect or a treatment effect that is sufficiently small) such that further study of the investigational product in the trial may not be warranted.

Posterior Probabilities of Efficacy and Futility

Based on observed data $\boldsymbol{D}$, early stopping of enrollment may be justified when $P_{(S)}(\theta > \theta|\boldsymbol{D}) > 0.975$. The subscript (s) in the posterior probability is used to indicate that the analysis is performed using the skeptical prior. Note that we are careful to use the term stopping of enrollment in the above discussion as when data demonstrate proof of efficacy while some patients are still ongoing in the trial, it is likely that those patients will be followed through until the time of outcome ascertainment. If a substantial number of patients are ongoing in the trial (relative to the number already completed), the finding based on the interim data may not be consistent with the finding derived from the complete data, once it is observed.

For futility, early stopping of the trial may be justified when $P_{(E)}(\theta < \theta_{\mathrm{m}}|\boldsymbol{D}) > 0.975$ where $\theta_{\mathrm{m}} \in [\theta_0, \theta_1]$. For example, we consider $\theta_{\mathrm{m}} = 0.5\theta_0 + 0.5\theta_1$ in our investigations in this paper. In words, when the observe data provide substantial evidence that the treatment effect is less than that expected from the perspective of an a priori enthusiastic observer, stopping the trial for futility may be justified. If a substantial number of patients are ongoing in the trial (relative to the number completed), the finding based on the interim data may not be

consistent with the finding derived from the complete data, should all enrolled patients be followed through until the time of outcome ascertainment.

Probability of Sustained Substantial Evidence

In any sequentially monitored trial, the observed data at an interim analysis will be different from the final data that would be observed if all patients already enrolled were followed through until the time of outcome ascertainment. For example, at some point in the trial $n = 30$ outcomes may be ascertained but an additional $n_{\text{miss}} = 5$ patients may already be enrolled and it may be the case that substantial evidence of a claim at an interim analysis is no longer present at the time of the final analysis. In such cases it may be advantages to incorporate the probability that substantial evidence will be sustained if the remaining n_{miss} outcomes are ascertained. This criteria is formalized by requiring that

$$E\{1[P(\theta > \theta_0|\boldsymbol{D}_{\text{obs}}, \boldsymbol{D}_{\text{miss}}) > 0.975]|\boldsymbol{D}_{\text{obs}}\} > \psi_{\text{PSEE}}$$

and

$$E\{1[P(\theta < \theta_{\text{m}}|\boldsymbol{D}_{\text{obs}}, \boldsymbol{D}_{\text{miss}}) > 0.975]|\boldsymbol{D}_{\text{obs}}\} > \psi_{\text{PSEF}},$$

where $\boldsymbol{D}_{\text{obs}} = \{y_{\text{obs}}, n_{\text{obs}}\}$ is the observed data at the time of the interim analysis, $\boldsymbol{D}_{\text{miss}} = \{y_{\text{miss}}, n_{\text{miss}}\}$ is the data associated with outcomes yet to be obtained for patients already enrolled, ψ_{PSEE} is a threshold for the probability that the evidence remains substantial in favor of efficacy after the missing data are observed, and ψ_{PSEF} is a threshold for the probability that the evidence remains substantial in favor of futility after the missing data are observed. Using the criteria above instead of using the posterior probabilities derived from the observed data at the time of the interim analysis will become particularly useful when the proportion of enrolled patients with outcomes not yet ascertained increases. In this paper we perform efficacy and futility monitoring based on $\psi_{\text{PSEE}} = 0.975$ and $\psi_{\text{PSEF}} = 0.80$.

Probability of Ultimately Obtaining Substantial Evidence

It will also be useful to know, based on the current set of n_{obs} outcomes, if the maximum number of patients were enrolled and followed-up, whether the probability of eventually obtaining substantial evidence in favor of efficacy is high. If that is not the case, further data collection may not be warranted. This criteria is formalized by requiring that

$$E\{1[P(\theta > \theta_0|\boldsymbol{D}_{\text{obs}}, \boldsymbol{D}_{\text{rem}}) > 0.975]|\boldsymbol{D}_{\text{obs}}\} > \psi_{\text{PUSE}},$$

where $\boldsymbol{D}_{\text{rem}} = \{y_{\text{rem}}, n_{\text{rem}}\}$ is the data yet-to-be observed assuming the maximum enrollment is reached and ψ_{PUSE} is a threshold for the probability that the evidence will ultimately be substantial in favor of efficacy should the maximum sample size be reached. In this paper we perform futility monitoring based on $\psi_{\text{PUSE}} = 0.10$.

USING EXTERNAL DATA PROSPECTIVELY IN MONITORING

As noted in the motivating example, data will sometimes be available that could directly inform the monitoring of accumulating data in a sequentially monitored trial.

SKEPTICAL POWER PRIORS

Such data can be incorporated into the monitoring process by using a skeptical power prior − a power prior (Ibrahim and Chen, 2000) where the skeptical prior previously mentioned serves as the initial prior. Formally, the skeptical power prior is given as follows:

$$\pi(\theta) \propto L(\theta|\boldsymbol{D}_0)^{a_0}\pi_s(\theta),$$

where $L(\theta|\boldsymbol{D}_0)$ is the likelihood for the existing data (e.g., the adult trial data), $a_0 \in [0,1]$ is a borrowing parameter that discounts the existing data's influence and $\pi_s(\theta)$ is the skeptical prior mentioned in previous sections.

ESTIMATING A VALUE FOR THE BORROWING PARAMETER

In principle the value of a_0 can be prespecified but in practice we prefer to assess the compatibility of the two data sources to obtain an estimate for it. One intuitive strategy is to choose a_0 so that it weights the existing data based on the compatibility of the new trial data with respect to its prior predictive distribution based on the posterior distribution from the existing data which we denote by $\pi(\theta|\boldsymbol{D}_0) \propto L(\theta|\boldsymbol{D}_0)\pi_0(\theta)$ where $\pi_0(\theta)$ is some weakly informative initial prior. For the example application, we might take $\pi_0(\theta) = \text{Beta}(0.5,0.5)$ which has negligible impact on the posterior distribution.

For example, based on the adult trial data discussed above from the combined ACT1 and ACT2 trials, the adult trial posterior distribution would be a beta distribution with shape parameters 162.5 and 80.5, respectively. One obtains the prior predictive distribution for the pediatric data as follows:

$$p(y|n,\boldsymbol{D}_0) = \int p(y|n,\theta)\,\pi(\theta|\boldsymbol{D}_0)d\theta$$

where $p(y|n,\theta)$ is the binomial probability mass function for number of responses y for the pediatric trial data based on a sample size of n. It is well known that $p(y|n,\boldsymbol{D}_0)$ has a closed-form in this case – the beta-binomial distribution. One can then compute Box's p-value (Evans and Moshonov, 2006) which is a measure of how extreme the observed data y_{obs} are relative to what would be expected if the pediatric data were generated from the predictive distribution based on the existing data. Formally, we define c_0, c_1, and c_2 as follows:

$$c_1 = \sum_{y=0}^{n} p(y|n,\boldsymbol{D}_0) \times 1\{p(y|n,\boldsymbol{D}_0) \le p(y_{\text{obs}}|n,\boldsymbol{D}_0)\},$$

$$c_2 = \textstyle\sum_{y=0}^{n} p(y|n,\theta_0) \times 1\{p(y|n,\theta_0) \le p(y_{\text{obs}}|n,\theta_0)\}, \text{ and}$$

$$c_0 = \max(\min(c_1 - c_2, 1), 0),$$

where c_1 is Box's p-value, c_2 is the p-value assuming $\theta = \theta_0$, and $c_0 \in [0,1]$ is the difference reflecting how much more likely the observed pediatric data are under the predictive distribution based on the adult data compared to an assumed null distribution. If we then define ρ as the number of adult patients that can be borrowed for each new pediatric patient, then we compute a_0 as follows:

$$a_0 = \min(1, c_0 \cdot \rho n)$$

It is important to note that when computing the probability of sustained substantial evidence as described in the previous section, one must compute the value of a_0 for each possible value of y_{miss} or y_{rem} which can be computationally demanding in large scale simulation studies.

COMPUTATIONS USING SAS IML

SIMULATION OF DATA

For the proposed design, simulation of data using the IML Procedure is relatively straightforward given specified values for several inputs:

- Distribution for enrollment times (we assume exponentially distributed interarrival times with mean PRM1_ENR)

- Distribution of outcome ascertainment times (we assume a normal distribution with mean and standard deviation, PRM1_ASC and PRM2_ASC, respectively)
- True probability of response (TRUE_PI)

The following IML source code can be used to both generate the patient-level data for MAXN patients and order the data according to the sequence in which the outcomes are ascertained.

```
/** Generate the complete hypothetical dataset **/
r    = J(M,1,0);
w    = J(maxN,1,0);
y    = J(maxN,1,0);

/** Generate enrollment times via a Poisson Process **/
call randgen(r,'exponential',prm1_enr);
r = cusum(r);

/** Generate outcome ascertainment times via a normal distribution **/
call randgen(w,'normal',prm1_asc,prm2_asc);

/** Calculated time from enrollment to outcome ascertainment **/
e = r + w;

/** Generate outcomes **/
call randgen(y,'bernoulli',true_pi);

/** Sort dataset in ascending order of outcome ascertainment **/
dat = r||w||e||y;
call sort(dat,3);

r = dat[,1];
w = dat[,2];
e = dat[,3];
y = dat[,4];

free dat;
```

ACCUMULATION OF DATA AT A GIVEN TIME POINT

For the sequential monitoring process, one needs to accumulate the number of responses up to the point where the appropriate number of outcomes are ascertained for analysis. The IML source code below illustrates how this can be done. Note that the variable NBY specifies how many additional outcomes are ascertained prior to the next analysis (N is initialized to zero). The TIME_OF_ANALYSIS vector is a two dimensional vector that stores the time of the interim analysis (first component) and final analysis (second component, should additional follow-up occur).

```
/** increment number of ascertained outcomes **/
n = n + nBy;

/** increment the analysis number **/
if n >= minN then analysis = analysis + 1;

/** identify time of current analysis (overwrite element 1) **/
time_of_analysis[1] = e[n];

/** accumulate the outcome data into sufficient statistics **/
y0 = sum((y=0)[1:n]);
y1 = sum((y=1)[1:n]);
```

```
/** determine how many enrolled patients are ongoing **/
nMiss = nrow(y[loc(r<time_of_analysis[1])])-n;

/** determine how many outcomes could still be ascertained **/
nLeft = max(0,maxN - n);
```

The variables NMISS and NLEFT store the number of patients enrolled with outcomes not yet ascertained and the number of patients that could have ascertained outcomes (that do not currently) if the maximum sample size (i.e., MAXN) were reached.

PRECOMPUTATION OF THE BORROWING PARAMETER

The following source code defines an IML function for the beta-binomial probability mass function:

```
start BetaBin (y,n) global(alpha,beta);
  return exp(lcomb(n,y)+logbeta(y+alpha,n-y+beta)-logBeta(alpha,beta));
finish;
```

Using the beta-binomial probability mass function, the following IML source code computes the quantity c_0 for all possible data values (y,n) based on the specified maximum sample size MAXN. This is done one time for an entire set of simulation studies and subsequently these values can simply be accessed from the matrix in which they are stored. This vastly reduces the expense of computations required for simulation studies.

```
c0_matrix = J(maxN,maxN+1,0);
c1_matrix = J(maxN,maxN+1,0);
c2_matrix = J(maxN,maxN+1,0);
a0_matrix = J(maxN,maxN+1,0);

do n  = 1 to maxN;
do y1 = 0 to n;
     r              = n;
     c              = y1+1;

     alpha          = h_y1 + 0.5;
     beta           = h_y0 + 0.5;
     PredProb       = betaBin(t(do(0,n,1)),n);
     obsProb        = betaBin(y1,n);
     c1_matrix[r,c] = ((PredProb<=obsProb)#PredProb)[+];

     PredProb       = pdf('binomial',t(do(0,n,1)),theta0,n);
     obsProb        = pdf('binomial',y1,theta0,n);
     c2_matrix[r,c] = ((PredProb<=obsProb)#PredProb)[+];

     c0_matrix      = c1_matrix - c2_matrix;
     c0_matrix      = c0_matrix >< 1;
     c0_matrix      = c0_matrix <> 0;

     a0_matrix[r,c] = min(1,c0_matrix[r,c]*rho*n);
end;
end;
```

In the above, the shape parameters ALPHA and BETA make use of the number of successes and failures from the adult trial data, denoted by H_Y1 and H_Y0, respectively. Having the possible values of c_0 precomputed allows for determination of each value of a_0 as well.

PRECOMPUTATION OF POSTERIOR PROBABILITIES

The following IML source code computes the required posterior probabilities for all possible data values (y,n) based on the specified maximum sample size MAXN. This also is done one time for an entire set of simulation studies and subsequently these values can simply be accessed from the matrix in which they are stored.

```
pp_skept_matrix_borrow   = J(maxN,maxN+1,0);
pp_skept_matrix_noborrow = J(maxN,maxN+1,0);
pp_enthu_matrix          = J(maxN,maxN+1,0);

/** Precomputation of posterior probabilities **/
do n  = 1 to maxN;
do y1 = 0 to n;
     y0                          = n-y1;
      r                          = n;
      c                          = y1+1;

      alpha                           = 5.830 + y1 + a0_matrix[r,c]*h_y1;
      beta                            = 8.245 + y0 + a0_matrix[r,c]*h_y0;
      pp_skept_matrix_borrow[r,c]     = sdf('beta',theta0,alpha,beta);

      alpha                           = 5.830 + y1;
      beta                            = 8.245 + y0;
      pp_skept_matrix_noborrow[r,c] = sdf('beta',theta0,alpha,beta);

      alpha                           = 9.790 + y1;
      beta                            = 5.329 + y0;
      pp_enthu_matrix[r,c]            = cdf('beta',thetam,alpha,beta);
end;
end;
```

Posterior probabilities are computed for both the skeptical prior and the skeptical power prior.

COMPUTATION OF PROBABILITES FOR SUSTAINED SUBSTANTIAL EVIDENCE

The following source IML code illustrates how probabilities of sustained substantial evidence can be computed. The code assumes the existence of a binary variable (BORROW=0 or 1) that determines whether the skeptical prior or skeptical power prior is used for analysis. To avoid unnecessary computation, calculations are only performed when the number of ascertained outcomes reaches the minimum (MINN).

```
if (n>= minN) then do;

      /** identify borrowing parameter **/
      row = n;
      col = y1+1;
      c0  = c0_matrix[row,col];
      a0  = a0_matrix[row,col]*borrow;

      /** identify posterior probability for efficacy **/
      postprob_skept = pp_skept_matrix_borrow[row,col]*(borrow=1)
                     + pp_skept_matrix_noborrow[row,col]*(borrow=0);

      /** identify posterior probability for futility **/
      postprob_enthu = pp_enthu_matrix[row,col];
```

```
      /** compute PSSE for skeptical & enthusiastic priors **/
      if nMiss = 0 then do;
            PSE_skept = postprob_skept;
            PSE_enthu = postprob_enthu;
      end;
      else do;

      nMissTot = n + nMiss;
      nLeftTot = n + nLeft;

            /** Compute distribution for missing outcomes **/
            alpha        = 5.830 + y1 + a0*h_y1;
            beta         = 8.245 + y0 + a0*h_y0;
            MissPredProb = betaBin(t(do(0,nMiss,1)),nMiss);

            /** Compute the probability of sustained evidence **/
            PSE_skept = 0;
            do    y1Miss = 0 to nMiss;
                  y0Miss = nMiss - y1Miss;
                  y1MissTot = y1 + y1Miss;
                  y0MissTot = nMissTot - y1MissTot;

                  row  = nMissTot;
                  col  = y1MissTot+1;
                  pps  = pp_skept_matrix_borrow[row,col]*(borrow=1)
                       + pp_skept_matrix_noborrow[row,col]*(borrow=0);

                  PSE_skept = PSE_skept
                            + (pps>evidence_crit)*MissPredProb[y1Miss+1];
            end;

            /** Compute distribution for possible remaining outcomes **/
            alpha        = 5.830 + y1 + a0*h_y1;
            beta         = 8.245 + y0 + a0*h_y0;
            LeftPredProb = betaBin(t(do(0,nLeft,1)),nLeft);

            /** Compute the probability of sustained evidence **/
            PSE_skept2 = 0;
            do    y1Left = 0 to nLeft;
                  y0Left = nLeft - y1Left;
                  y1LeftTot = y1 + y1Left;
                  y0LeftTot = nLeftTot - y1LeftTot;

                  row  = nLeftTot;
                  col  = y1LeftTot+1;
                  pps2 = pp_skept_matrix_borrow[row,col]*(borrow=1)
                       + pp_skept_matrix_noborrow[row,col]*(borrow=0);

                  PSE_skept2 = PSE_skept2
                             + (pps2>evidence_crit)*LeftPredProb[y1Left+1];
            end;
            .
            .
            .
      end;
end;
```

Note that the source code associated with computing the probability of sustained substantial evidence of futility is excluded for brevity (represented by the … section).

DETERMINATION OF EARLY STOPPING

The following IML source code is used to apply the early stopping criteria. Note that all three criteria discussed above (PSSE, PSSF, and PUSE) are checked as well as whether the maximum sample size has been reached.

```
/** Stop for efficacy **/
if    (PSE_skept>=pse_eff_crit  )   then do;
      stop_trial = 1;
      Eff[1+final_analysis] = 1;
end;

/** Stop for futility **/
if    (PSE_enthu >= pse_fut_crit1  )   then do;
      stop_trial = 1;
      FUT1[1+final_analysis] = 1;
end;
if  (PSE_skept2 < pse_fut_crit2 )   then do;
      stop_trial = 1;
      FUT2[1+final_analysis] = 1;
end;

FUT3 = (FUT1 + FUT2)><1;

/** Stop for maximum sample size **/
if ( n>= maxN )   then stop_trial = 1;
```

Note that the STOP_TRIAL and FINAL_ANALYSIS variables are initialized to zero. When STOP_TRIAL is set to 1 for early stopping due to efficacy, one more analysis occurs that includes the outcomes for all patients that were already enrolled. The above source code would also only be executed when the minimum number of outcomes has been ascertained.

OVERALL LOOPING STRUCTURE

The main source IML code is setup to run a large scale set of simulation studies to examine the operating characteristics of the sequential trial. Simulations are performed to investigate an array of possible true values for θ and to evaluate properties of the design when external data are incorporated as well as when they are not. The source IML code below provides the loop structure. The vector TPV contains the true θ values to be investigated in the simulations.

```
tpv = do(0.40,0.76,0.02);
do borrow = 0 to 1;
do t = 1 to ncol(tpv);
      true_pi = tpv[t]; results = J(nSims,26,0);
      do sim = 1 to nSims;

            [SOURCE IML CODE TO INITIALIZE VARIABLES (e.g. STOP_TRIAL)]
            [SOURCE IML CODE TO SIMULATE DATA]
            /** Simulate the sequentially monitored trial **/
            do until(stop_trial=1 & final_analysis=1);
                  [SOURCE IML CODE TO ACCUMULATE AND ANALYZE DATA]
            end;
            [SOURCE IML CODE TO ACCUMULATE SIMULATED STUDY RESULTS]
      end;
end;
end;
```

With MAXN=60, NBY=2, and NSIMS=10000, the above code takes <5 minutes to complete.

DESIGN SIMULATION RESULTS FOR MOTIVATING EXAMPLE

Figure 3 shows the estimated power curve for both the sequential monitoring designs with (red line) and without (blue line) borrowing information prospectively from the adult trial.

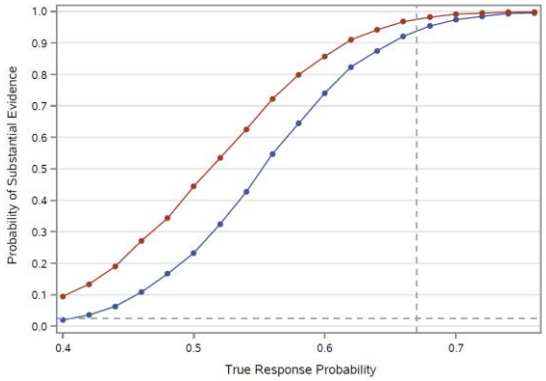

Figure 3: Statistical Power/Type I Error Rate

Of note, the power for both designs exceeds 90% when the pediatric response probability matches the adult trial MLE and is above 80% when the response probability is at least 0.58 and 0.62 for the borrowing and non-borrowing designs, respectively. The non-borrowing design attains a 2.5% type I error rate when $\theta = 0.40$ whereas the borrowing design has a type I error rate of approximately 10%. This should not be viewed as problematic for the borrowing design since, as was discussed in earlier sections, the strategy for borrowing information is entirely appropriate if one believes the adult data are pertinent and accounts for the fact that the null hypothesis may be true by reducing borrowing in accordance with how well the observed data are supported by a null predictive distribution.

Figure 4 shows the distribution for the final sample size for both the sequential monitoring designs with (red) and without (blue) borrowing information prospectively from the adult trial. The expected sample size is represented by a diamond.

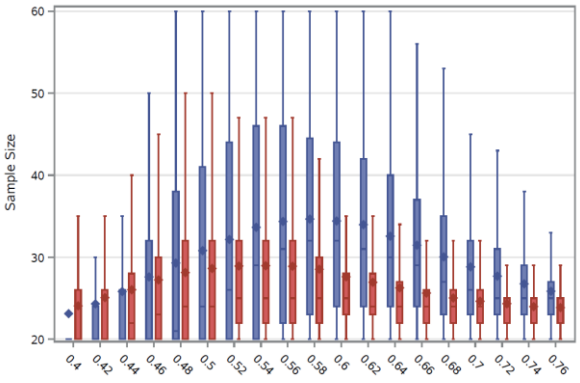

Figure 4: Distribution of Final Sample Size

One can see that the information borrowing design has must less variable sample size at the cost of a slightly higher expected sample size when the null hypothesis is true.

CONCLUSION

In this paper we have discussed a SAS implementation for design simulations for a sequentially monitoring trial using the IML Procedure. All code for this paper can be found at https://github.com/psioda/SUGI-Sequential-Monitoring.

REFERENCES

Hyams J, Damaraju L, Blank M, et al. 2012. "Induction and maintenance therapy with infliximab for children with moderate to severe ulcerative colitis." *Clinical Gastroenterology and Hepatology*. 10(4):391-399.e1.

Rutgeerts P, Sandborn WJ, Feagan BG, et al. 2005. "Infliximab for induction and maintenance therapy for ulcerative colitis." *New England Journal of Medicine*. 353(23):2462-2476.

Spiegelhalter DJ, Freedman LS, Parmar MKB. 1993. "Applying Bayesian ideas in drug development and clinical trials." *Statistics in Medicine*. 12(15-16):1501-1511.

Ibrahim JG, Chen M-H. 2000. "Power prior distributions for regression models." *Statistical Science*. 15(1):46-60.

Evans M, Moshonov H. 2006. "Checking for prior-data conflict." *Bayesian Anal*ysis. 1(4):893-914.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Matthew A. Psioda

Collaborative Studies Coordinating Center

UNC Department of Biostatistics

matt_psioda@unc.edu