Bayes Sequential Decision Theory in a Clinical Trial with Binary Measures

W. Qian*

MRC Clinical Trials Unit, London, NW1 2DA, UK

and

P. J. Brown

University of Kent, Canterbury, CT2 7NF, UK

SUMMARY. We apply Bayes sequential decision theory into a clinical trial with binary measures, in which the computational complexity could be overcome partly. Therefore the patients' horizon and treatment effect are considered in the loss functions in our procedures. These Bayesian sequential procedures are compared with frequentist procedures.

KEY WORDS: Group sequential clinical trials; Bayes sequential decision theory; Backward induction.

1. Introduction

Bayesian methods have been increasingly applied into group sequential clinical trials. They offer an approach for dealing with many difficult problems and have some practical advantages over frequentist methods (for example, see Freedman and Spiegelhalter 1989, 1993). Decision theory provides the framework for combining subjective distributions with action. However, Bayes sequential decision theory has not been widely introduced in clinical trials because of the computational complexity and the difficulty of specifying

* *email:* wq@ctu.mrc.ac.uk

loss and cost functions. Sylvester (1988) has used Bayes decision theory for a fixed sample phase II clinical trial in which the main outcome variable is binomially distributed. Berry and Ho (1988) have addressed one-sided sequential stopping boundaries for clinical trials from a decision-theoretic point of view. Lewis and Berry (1994), and Lewis (1996) have applied Bayes sequential decision theory into group sequential clinical trials with binomially distributed outcome measures and a piecewise continuous loss function and shown that there are Bayes groups sequential decision procedures with type I and type II error rates similar to those of the Pocock (1977) and the O'Brien-Fleming (1979) procedures but with smaller expected sample sizes and costs. The same results can also be found for clinical trials with normally distributed outcome measures (Qian 1997) and clinical trials with survival time (Qian and Brown 1999). Valchos and Gelfand (1998), and Carlin, Kadane and Gelfand (1998) also address the decision theory problem.

In this paper, the Bayes sequential decision theory is applied to a group sequential clinical trial with binary measures. After a brief description of the general procedure in Section 2, a specific procedure with a two-point prior distribution and the sample size calculation based on Bayes risk are developed in Section 3. This procedure is also compared with that of the frequentist methods. In Section 4, the Bayes sequential decision procedure with a beta prior distribution is explored. Finally, the discussion is given in Section 5.

2. Bayes Sequential Decision Theory in Group Sequential Clinical Trials

A clinical trial is designed to test a new drug response rate p. The conventional hypotheses are $H_0 : p \leq p_0$ (reject the new drug) vs $H_1 : p > p_0$ (accept the new drug), in which p_0 is the break-even value of the response rate.

Suppose the clinical trial is monitored at every n observations, $X_{ij}, i = 1, 2, ..., n, j = 1, 2, ..., l$, where X_{ij} are independent identically distributed Bernoulli random variables with unknown parameter p. The l is the maximum number of analyses. The clinical trial could be stopped earlier should accumulating evidence demonstrate the superiority of the new drug. Let Y_j be the sum of the group sample $X_{ij}, i = 1, 2, ..., n$. The sequence of random variables $Y_j, j = 1, 2, ..., l$, constitutes a classical sequential sample, and the distribution of Y_j is the binomial distribution $B(n, p)$, that is,

$$Y_j = \sum_{i=1}^{n} X_{ij} \sim B(n, p). \tag{1}$$

Since Y_j, $j = 1, 2, ..., l$, are sufficient statistics of the parameter p, the group sequential sample X_{ij}, $i = 1, 2, ..., n$, $j = 1, 2, ..., l$, can be replaced by the classical sequential sample Y_j, $j = 1, 2, ..., l$, without losing information on statistical inferences of the parameter p. Bayes sequential decision theory can be then applied. The basic elements of Bayes sequential decision procedure of the study are,

1) The parameter of interest is the new drug response rate p. The prior distribution $w(p)$ is assumed as a two-point prior distribution in Section 3 and beta prior distribution in Section 4.

2) The decision space of the study is $D = \{d_0$(reject the new drug),

d_1(accept the new drug)}.

3) the loss and cost functions

At each analysis $j, j = 1, 2, ..., l$, the loss function is defined as,

$$\begin{aligned} L(p, d_0) &= \begin{cases} 0 & p \leq p_0 \\ K(p - p_0)t & p > p_0, \end{cases} \\ L(p, d_1) &= \begin{cases} K(p_0 - p)t & p \leq p_0 \\ K(p_0 - p)t & p > p_0, \end{cases} \end{aligned} \tag{2}$$

where $p =$ the true underlying response rate of the new drug;

$p_0 =$ the break-even response rate;

$t =$ the patient horizon;

$K =$ the difference in cost(monetary or ethical) of further treatment between a patient who does not respond to the new drug and a patient who does respond($K > 0$).

The loss function reflects the expected new treatment benefit $p - p_0$ over a patient horizon t. If the response rate of new drug p is less than p_0, where the new drug should not be recommended, then there is no loss ($L(p, d_0) = 0$) in rejecting the new drug, and there is a loss $L(p, d_1) = K(p_0 - p)t$ in accepting the new drug. If the response rate of new drug p is greater than p_0, where the new drug could be recommended, then there is a loss $L(p, d_0) = K(p - p_0)t$ in rejecting the new drug, and there is a gain $L(p, d_1) = K(p_0 - p)t$ (< 0) in accepting the new drug.

Suppose that the unit of $K(> 0)$ is the cost of enrolling a patient into the trial. The cost function at analysis j, $j = 1, 2, ..., l$, is then

$$C_j = jn. \tag{3}$$

4) The sequential sample Y_j, $j = 1, 2, ..., l$, obtained by (1), are binomially distributed, that is, $Y_j \sim f(y|p) = B(n, p)$.

Backward induction (Berger 1985, Degroot 1970) is used to construct Bayes sequential decision procedure. At each interim analysis j, $j = 1, ..., l-1$, after observing $Y_1 = y_1$,$Y_2 = y_2$,..., $Y_j = y_j$, where the posterior distribution of p is denoted by $w^j = w(p|y_1, y_2, ..., y_j)$, we need to compare the risk from stopping the trial, denoted by $r_0(w^j, j)$, with the risk from the optimal continuation of the trial, denoted by $r_{l-j}(w^j, j)$, in which not more than $l-j$ groups of patients can be taken. If $r_0(w^j, j) \le r_{l-j}(w^j, j)$, then we make a decision without further observations; otherwise we continue sampling to observe the next group samples Y. Considering the constant cost function, the calculations of $r_{l-j}(w^j, j)$ can be obtained by the following recursive relationships,

$$r_0(w^j, j)=\min\{E_{p|y_1,...,y_j}L(p, d_0), E_{p|y_1,...,y_j}L(p, d_1)\}, \qquad j = 1, ..., l, \tag{4}$$

$$r_h(w^j, j)=\min\{r_0(w^j,j), E^*r_{h-1}(w^j(Y),j+1)\}+n,\ j, h=1, ..., l-1, j+h\le l. \tag{5}$$

The expected risk $E^*r_{h-1}(w^j(Y), j+1) = \int_0^\infty\int_0^\infty r_{h-1}(w^j(y), j+1)f(y|y_1, ..., y_j)dy$.

After stopping sampling, the decision $d \in D$ which achieves the Bayes risk as in (4) is chosen. The inferences on the response rate of new drug p are derived from the posterior distribution $w(p|y_1, y_2, ..., y_j)$.

3. Bayes Sequential Decision Procedure with Two-Point Prior Distribution

Suppose that the response rate of new drug p has the two-point prior distribution,

$$w(p) = \begin{cases} w_1 & p = p_1 \\ w_2 & p = p_2, \end{cases} \tag{6}$$

where $w_1 + w_2 = 1$, $0 < p_1 < p_0 < p_2 < 1$. Section 3.3 displays the choice of different p_1, p_2, w_1 and w_2.

At each analysis j, $j = 1, 2, ..., l$, after observing $Y_1 = y_1$, $Y_2 = y_2$,..., $Y_j = y_j$, the posterior distribution of p is

$$\begin{aligned} & w(p|y_1, y_2, ..., y_j) \\ & = \frac{p^{\sum_{i=1}^{j} y_i}(1-p)^{jn-\sum_{i=1}^{j} y_i}\ w(p)}{p_1^{\sum_{i=1}^{j} y_i}(1-p_1)^{jn-\sum_{i=1}^{j} y_i} w_1 + p_2^{\sum_{i=1}^{j} y_i}(1-p_2)^{jn-\sum_{i=1}^{j} y_i} w_2} \\ & = \begin{cases} \frac{p_1^{\sum_{i=1}^{j} y_i}(1-p_1)^{jn-\sum_{i=1}^{j} y_i} w_1}{p_1^{\sum_{i=1}^{j} y_i}(1-p_1)^{jn-\sum_{i=1}^{j} y_i} w_1+p_2^{\sum_{i=1}^{j} y_i}(1-p_2)^{jn-\sum_{i=1}^{j} y_i} w_2} & p = p_1 \\ \frac{p_2^{\sum_{i=1}^{j} y_i}(1-p_2)^{jn-\sum_{i=1}^{j} y_i} w_2}{p_1^{\sum_{i=1}^{j} y_i}(1-p_1)^{jn-\sum_{i=1}^{j} y_i} w_1+p_2^{\sum_{i=1}^{j} y_i}(1-p_2)^{jn-\sum_{i=1}^{j} y_i} w_2} & p = p_2. \end{cases} \end{aligned} \tag{7}$$

The predictive density function of Y is

$$\begin{aligned} & f(y|y_1, y_2, ..., y_j) \\ & = E_{p|y_1,y_2,...,y_j} f(y|p) \\ & = f(y|p_1)w(p_1|y_1, y_2, ..., y_j) + f(y|p_2)w(p_2|y_1, y_2, ..., y_j) \\ & = \frac{\binom{n}{y} p_1^{\sum_{i=1}^{j} y_i+y}(1-p_1)^{(j+1)n-(\sum_{i=1}^{j} y_i+y)} w_1 + \binom{n}{y} p_2^{\sum_{i=1}^{j} y_i+y}(1-p_2)^{(j+1)n-(\sum_{i=1}^{j} y_i+y)} w_2}{p_1^{\sum_{i=1}^{j} y_i}(1-p_1)^{jn-\sum_{i=1}^{j} y_i} w_1+p_2^{\sum_{i=1}^{j} y_i}(1-p_2)^{jn-\sum_{i=1}^{j} y_i} w_2}. \end{aligned} \tag{8}$$

3.1 *Bayes Sequential Decision Procedure*

The Stopping Rule

The stopping rule is first described in detail by using the clinical trial with a two-interim analysis. At the first interim analysis, $Y_1 = y_1$ has been observed. The next observation should be taken if, and only if, the risk from stopping the trial, $r_0(w^1, 1)$, is greater than the risk of continuing the trial with not more than two observations, $r_2(w^1, 1)$. The risk $r_0(w^1, 1)$ is

$$\begin{aligned} & r_0(w^1, 1) \\ & = \min_{d \in \{d_0, d_1\}} E_{p|y_1} L(p, d) \\ & = \begin{cases} K(p_2 - p_0)t\, w(p_2|y_1) & y_1 \le \frac{\log\left[\frac{p_0-p_1}{2(p_2-p_0)} \frac{w_1}{w_2} \left(\frac{1-p_1}{1-p_2}\right)^n\right]}{\log\left[\frac{p_2}{p_1} \frac{1-p_1}{1-p_2}\right]} \\ K(p_0 - p_1)t\, w(p_1|y_1) + K(p_0 - p_2)t\, w(p_2|y_1) & y_1 > \frac{\log\left[\frac{p_0-p_1}{2(p_2-p_0)} \frac{w_1}{w_2} \left(\frac{1-p_1}{1-p_2}\right)^n\right]}{\log\left[\frac{p_2}{p_1} \frac{1-p_1}{1-p_2}\right]}. \end{cases} \end{aligned} \tag{9}$$

Using the method of backward induction, the risk $r_2(w^1,1)$ is,

$$r_2(w^1,1) \quad = \quad \min\{r_0(w^1,1),\;\; E^* r_1(w^1(Y),2)+n\}, \tag{10}$$

where,

$$E^* r_1(w^1(Y),2) \quad = \quad \sum_{y=0}^{n} r_1(w^1(Y=y),2)\, f(y|y_1). \tag{11}$$

In (11) the predictive density function $f(y|y_1)$ is obtained by (8), and the calculation of $r_1(w^1(Y=y),2)$ is described as below.

From (5) it is obtained,

$$\begin{aligned} r_1(w^1,1) &= \min\{r_0(w^1,1),\; E^* r_0(w^1(Y'),2)+n\} \\ &= \begin{cases} \min\{E_{p|y_1}L(p,d_0),\; E^* r_0(w^1(Y'),2)+n\} & y_1 \leq \dfrac{\log\left[\frac{p_0-p_1}{2(p_2-p_0)}\frac{w_1}{w_2}\left(\frac{1-p_1}{1-p_2}\right)^n\right]}{\log\left[\frac{p_2}{p_1}\frac{1-p_1}{1-p_2}\right]} \\ \min\{E_{p|y_1}L(p,d_1),\; E^* r_0(w^1(Y'),2)+n\} & y_1 > \dfrac{\log\left[\frac{p_0-p_1}{2(p_2-p_0)}\frac{w_1}{w_2}\left(\frac{1-p_1}{1-p_2}\right)^n\right]}{\log\left[\frac{p_2}{p_1}\frac{1-p_1}{1-p_2}\right]}, \end{cases} \end{aligned}$$

where,

$$\begin{aligned} E^* r_0(w^1(Y'),2) \quad &= \quad \sum_{y'=0}^{n} r_0(w^1(Y'=y'),2)\, f(Y'=y'|y_1). \\ &= \quad \sum_{y'=0}^{M} K(p_2-p_0)t\, w(p_2|y_1,y')\, f(y'|y_1) \\ &+ \quad \sum_{y'=M+1}^{n} [K(p_0-p_1)t\, w(p_1|y_1,y') \\ &+ \quad K(p_0-p_2)t\, w(p_2|y_1,y')]\, f(y'|y_1), \end{aligned}$$

in which the M is the integer portion of value $\max\{\frac{\log\left[\frac{p_0-p_1}{2(p_2-p_0)}\frac{w_1}{w_2}\left(\frac{1-p_1}{1-p_2}\right)^{2n}\right]}{\log\left[\frac{p_2}{p_1}\frac{1-p_1}{1-p_2}\right]} - y_1,\ 0\}$. If $w^1 = w(p|y_1)$ is replaced by $w^1(y) = w(p|y_1,y)$, then $r_1(w^1(Y=y),2)$ in (11) is obtained, that is,

$$
\begin{aligned}
&r_1(w^1(Y=y),2)\\
&=\min\{r_0(w^1(Y=y)),E^*r_0(w^1(Y')|Y=y))+n\}\\
&=\begin{cases}\min\{E_{p|y_1,y}L(p,d_0),E^*r_0(w^1(Y')|Y=y))+n\} & y\le\frac{\log\left[\frac{p_0-p_1}{2(p_2-p_0)}\frac{w_1}{w_2}(\frac{1-p_1}{1-p_2})^{2n}\right]}{\log\left[\frac{p_2}{p_1}\frac{1-p_1}{1-p_2}\right]}-y_1\\ \min\{E_{p|y_1,y}L(p,d_1),E^*r_0(w^1(Y')|Y=y))+n\} & y>\frac{\log\left[\frac{p_0-p_1}{2(p_2-p_0)}\frac{w_1}{w_2}(\frac{1-p_1}{1-p_2})^{2n}\right]}{\log\left[\frac{p_2}{p_1}\frac{1-p_1}{1-p_2}\right]}-y_1,\end{cases}
\end{aligned}
$$

where,

$$
E^*r_0(w^1(Y')|Y=y))=\sum_{y'=0}^{M_1}K(p_2-p_0)t\,w(p_2|y_1,y,y')\,f(y'|y_1,y)
$$

$$
+\sum_{y'=M_1+1}^{n}\{K(p_0-p_1)t\,w(p_1|y_1,y,y')+Kt(p_0-p_2)\,w(p_2|y_1,y,y')\}f(y'|y_1,y).
$$

The M_1 is the integer portion of $\max\{\frac{\log\left[\frac{p_0-p_1}{2(p_2-p_0)}\frac{w_1}{w_2}(\frac{1-p_1}{1-p_2})^{3n}\right]}{\log\left[\frac{p_2}{p_1}\frac{1-p_1}{1-p_2},\right]}-(y_1+y),\ 0\}$.

Let

$$
D_1(y_1|n,K,t,w) \quad = \quad r_0(w^1,1)-r_2(w^1,1). \tag{12}
$$

If $D_1(y_1|n,K,t,w)\le 0$, then the clinical trial could be stopped; otherwise the trial needs to be continued.

At the second interim analysis, after $Y_1=y_1$ and $Y_2=y_2$ have been observed, the Bayes risk from stopping the trial is,

$$
\begin{aligned}
r_0(w^2,2) &= \min_{\{d_0,d_1\}} E_{p|y_1,y_2}L(p,d)\\
&= \min\{E_{p|y_1,y_2}L(p,d_0),\quad E_{p|y_1,y_2}L(p,d_1)\},
\end{aligned} \tag{13}
$$

where,

$$
\begin{aligned}
E_{p|y_1,y_2}L(p,d_0) &= K(p_2-p_0)t\,w(p_2|y_1,y_2),\\
E_{p|y_1,y_2}L(p,d_1) &= K(p_0-p_1)t\,w(p_1|y_1,y_2)+K(p_0-p_2)t\,w(p_2|y_1,y_2).
\end{aligned}
$$

The expected risk from observing the last observation is,

$$E^* r_0(w^2(Y), 3) = \sum_{y=0}^{n} r_0(w^2(Y = y), 3)\, f(y|y_1, y_2), \tag{14}$$

where,

$$\begin{aligned} r_0(w^2(Y = y), 3) &= \min_{d \in \{d_0, d_1\}} E_{p|y_1,y_2,y} L(p, d) \\ &= \min\{E_{p|y_1,y_2,y} L(p, d_0), \quad E_{p|y_1,y_2,y} L(p, d_1)\}. \end{aligned}$$

It is easy to show that the $r_0(w^2, 2)$ and $E^* r_0(w^2(Y), 3) + n$ are functions of $y_1 + y_2$ given the group sample size n, parameters of the loss function K, t and the prior distribution $w(p)$. Let

$$D_2(y_1 + y_2|n, K, t, w) = r_0(w^2, 2) - (E^* r_0(w^2(Y), 3) + n). \tag{15}$$

If $D_2(y_1 + y_2|n, K, t, w) \leq 0$, then the trial should be stopped; otherwise the trial needs to be continued.

The above stopping rule can be generalized to a clinical trial with more than two interim analyses. At each interim analysis j, $j = 1, 2, ..., l - 1$, after $Y_1 = y_1$, $Y_2 = y_2$, ..., $Y_j = y_j$ have been observed, the difference of the Bayes risk from stopping the trial, $r_0(w^j, j)$, and the expected risk of continuing the trial with not more than $l - j$ observations, $r_{l-j}(w^j, j)$, is the function of $\sum_{i=1}^{j} y_i$. Let

$$D_j(\sum_{i=1}^{j} y_i|n, K, t, w) = r_0(w^j, j) - r_{l-j}(w^j, j), \tag{16}$$

If $D_j(\sum_{i=1}^{j} y_i|n, K, t, w) \leq 0$, then the trial could be stopped; otherwise the trial needs to be continued. The stopping region with the form of statistics

$\sum_{i=1}^{j} Y_i$ is

$$\begin{aligned} &\{\sum_{i=1}^{j} y_i : D_j(\sum_{i=1}^{j} y_j|n,K,t,w) \leq 0\} = \\ &\quad \{\sum_{i=1}^{j} y_i : \sum_{i=1}^{j} y_i \leq c_1^j(n,K,t,w)\} \cup \{\sum_{i=1}^{j} y_i : \sum_{i=1}^{j} y_i \geq c_2^j(n,K,t,w)\}. \end{aligned} \tag{17}$$

where $c_1^j(n,K,t,w)$ and $c_2^j(n,K,t,w)$ may be the roots of the equation

$$D_j(\sum_{i=1}^{j} y_i|n,K,t,w) = 0.$$

The Decision Rule

At each analysis j, $j = 1,2,...,l$, after values $Y_1 = y_1$, $Y_2 = y_2$,..., $Y_j = y_j$ have been observed, if the clinical trial is terminated, the Bayes risk from stopping the trial, $r_0(w^j, j)$, is,

$$r_0(w^j,j) = \begin{cases} Kt(p_2-p_0)w(p_2|y_1,y_2,...,y_j) & \sum_{i=1}^{j} y_i \leq \frac{\log[\frac{p_0-p_1}{2(p_2-p_0)}\frac{w_1}{w_2}(\frac{1-p_1}{1-p_2})^{jn}]}{\log[\frac{p_2}{p_1}\frac{1-p_1}{1-p_2}]} \\ Kt(p_0-p_1)w(p_1|y_1,y_2,...,y_j) + & \sum_{i=1}^{j} y_i > \frac{\log[\frac{p_0-p_1}{2(p_2-p_0)}\frac{w_1}{w_2}(\frac{1-p_1}{1-p_2})^{jn}]}{\log[\frac{p_2}{p_1}\frac{1-p_1}{1-p_2}]}. \\ \quad Kt(p_0-p_2)w(p_2|y_1,y_2,...,y_j) & \end{cases} \tag{18}$$

The Bayes risk $r_0(w^j, j)$ is a function of $\sum_{i=1}^{j} y_i$. The decision rule after the trial be terminated is then

$$d = \begin{cases} d_0(\textit{reject the new drug}) & \sum_{i=1}^{j} y_i \leq \frac{\log[\frac{p_0-p_1}{2(p_2-p_0)}\frac{w_1}{w_2}(\frac{1-p_1}{1-p_2})^{jn}]}{\log[\frac{p_2}{p_1}\frac{1-p_1}{1-p_2}]} \\ d_1(\textit{accept the new drug}) & \sum_{i=1}^{j} y_i > \frac{\log[\frac{p_0-p_1}{2(p_2-p_0)}\frac{w_1}{w_2}(\frac{1-p_1}{1-p_2})^{jn}]}{\log[\frac{p_2}{p_1}\frac{1-p_1}{1-p_2}]}. \end{cases} \tag{19}$$

3.2 *Group Sample Size Calculation Based on Bayes risk*

Consider the group sequential clinical trial with the maximum number of analyses equal to l. At the final analysis, let $f(y_1, y_2, ..., y_l)$ be the joint probability density function of Y_1, Y_2,..., Y_l, that is,

$$\begin{aligned} f(y_1, y_2, ..., y_l) &= f(y_1, y_2, ..., y_l|p_1)\, w(p_1) + f(y_1, y_2, ..., y_l|p_2)\, w(p_2) \\ &= \prod_{i=1}^{l} \binom{n}{y_i} p_1^{\sum_{i=1}^{l} y_i} (1-p_1)^{nl-\sum_{i=1}^{l} y_i}\, w_1 \\ &+ \prod_{i=1}^{l} \binom{n}{y_i} p_2^{\sum_{i=1}^{l} y_i} (1-p_2)^{nl-\sum_{i=1}^{l} y_i}\, w_2. \end{aligned} \tag{20}$$

Since the Bayes risk $r_0(w^l, l)$ as in (18) with $j = l$ is a function of $\sum_{i=1}^{l} y_i$ where the constant cost function C_l as in (3) was ignored, the average Bayes risk is,

$$\begin{aligned} & E_{(Y_1,Y_2,\dots,Y_l)} r_0(w^l, l) + nl \\ &= \sum_{y_1, y_2, \dots, y_l} r_0(w^l, l)\, f(y_1, y_2, ..., y_l) + nl \\ &= \sum_{\sum y_i} r_0(w^l, l) f(\sum y_I) + nl \\ &= \sum_{\sum y_i = 0}^{nl} r_0(w^l, l) \left[\binom{nl}{\sum y_i} p_1^{\sum y_i} (1-p_1)^{nl-\sum y_i}\, w_1 \right. \\ &+ \left. \binom{nl}{\sum y_i} p_2^{\sum y_i} (1-p_2)^{nl-\sum y_i}\, w_2 \right] + nl \\ &= \sum_{\sum y_i = 0}^{M_0} E_{p|y_1, y_2, \dots, y_l} L(p, d_0) \left[\binom{nl}{\sum y_i} p_1^{\sum y_i} (1-p_1)^{nl-\sum y_i}\, w_1 \right. \\ &+ \left. \binom{nl}{\sum y_i} p_2^{\sum y_i} (1-p_2)^{nl-\sum y_i}\, w_2 \right] \\ &+ \sum_{\sum y_i = M_0+1}^{nl} E_{p|y_1, y_2, \dots, y_l} L(p, d_1) \left[\binom{nl}{\sum y_i} p_1^{\sum y_i} (1-p_1)^{nl-\sum y_i}\, w_1 \right. \\ &+ \left. \binom{nl}{\sum y_i} p_2^{\sum y_i} (1-p_2)^{nl-\sum y_i}\, w_2 \right] + nl \end{aligned} \tag{21}$$

where the M_0 is the integer portion of $\frac{\log\left[\frac{p_0 - p_1}{2(p_2 - p_0)} \frac{w_1}{w_2} (\frac{1-p_1}{1-p_2})^{nl}\right]}{\log\left[\frac{p_2}{p_1} \frac{1-p_1}{1-p_2}\right]}$.

It can been seen from above that given the prior distribution $w(p)$, the loss function $L(p,d)$, and the cost function $C_j, j = 1,2,...,l$, the average Bayes risk $E_{(Y_1,Y_2,...,Y_l)}r_0(w^l,l)$ is a function of nl, which is the maximum sample size of the group sequential clinical trial. Bayes inferences are independent of the number of analyses after the trial is terminated. Therefore in the view of Bayes sequential decision theory the maximum sample size $N = nl$ can be designed to reach the prespecified average Bayes risk level, say R, and

$$N \quad = \quad \min\{N : \ E_{(Y_1,Y_2,...,Y_l)}r_0(w^l,l) \leq R\}. \tag{22}$$

If the clinical trial is designed to have equal group size, then the group sample size is $n = \frac{N}{l}$.

3.3 *Comparison with Frequentist Methods*

The most popular frequentist procedures in group sequential clinical trials, which are the procedures of Pocock and O'Brien-Fleming, are used to compare with the Bayes sequential decision procedure. The comparisons are based on type I error rates α, type II error rates β, and their corresponding expected sample sizes, *Mean* N_α, *Mean* N_β, respectively,

Suppose the break-even value of response rate of the new drug $p_0 = 0.20$, the new drug is rejected if the response rate of the drug $p = p_1 = 0.01$, 0.05, and 0.10, and the new drug is accepted if the response rate of the drug $p = p_2 = 0.25$. Let the two-point prior distribution of p in (6) have $w_1 = 0.3$, 0.5, 0.7, and 0.9. The group sample sizes n are designed to be the same as those used in the procedures of Pocock and O'Brien-Fleming with type I error rate $\alpha = 0.05$ and type II error rate $\beta = 0.20$ and 0.10. Assume that the patient horizon $t = 500$, 1000, 5000, 10000. The two interim analyses procedure is used as an example. The Table 1 and Table 2 are the results of the type I error rate α and its expected sample size *Mean* N_α, and the type II error rate β and its expected sample size *Mean* N_β, respectively, with the group sample size designed by the procedure of Pocock with type I error rate $\alpha = 0.05$ and type II error rate $\beta = 0.2$.

Tables 1 and 2 about here

It has been shown that there are Bayes sequential decision procedures which have the type I error rates α, type II error rates β close to the procedures of Pocock, but their corresponding expected sample sizes are smaller than those of Pocock's procedures, which are in bold print in the tables. The more prior belief in $p = p_1$ of the new drug being ineffective, that is, the larger value of w_1, then the lower type I error rate α and the smaller expected sample size Mean N_α but the higher type II error rate β and the bigger expected sample size Mean N_β. The less prior belief in $p = p_1$ or the more prior belief in $p = p_2$, that is, the larger the value of $w_2 = 1 - w_1$, then the smaller type II error rate β is. The larger the patient horizon t is, the larger the expected sample size is needed to be able to make a decision with greater accuracy(that is, both type I and type II error rates are small). The bigger the difference of response rate(that is, the value of p_1 is smaller because the value of p_2 is fixed to be 0.25.) is, the earlier the trial is stopped since the expected sample size is smaller. The same results are found when the type II error rate $\beta = 0.1$, and when the Bayes sequential decision procedures are compared with the procedures of O'Brien-Fleming as well.

4. Bayes Sequential Decision Procedure with Beta Prior Distribution

For computational simplicity, we assumed that the response rate p has a two-point prior in Section 3. This is a very strong assumption. In practice, it would be more proper to assume a beta prior distribution for p, which is explored as follows.

Suppose that the response rate of new drug p has a beta prior distribution with parameters $u(u > 0)$ and $v(v > 0)$,

$$p \sim w(p) \quad = \quad \frac{1}{\beta(u, v)} p^{u-1}(1-p)^{v-1}. \tag{23}$$

At each analysis j, $j = 1, 2, ..., l$, after observing $Y_1 = y_1$, $Y_2 = y_2$, ..., $Y_j = y_j$, the posterior distribution of p is still a beta distribution with parameters $\sum_{i=1}^{j} y_i + u$ and $jn - \sum_{i=1}^{j} y_i + v$, that is,

$$w(p|y_1, y_2, ..., y_j) \quad = \quad \frac{p^{\sum_{i=1}^{j} y_i + u - 1}(1-p)^{jn - \sum_{i=1}^{j} y_i + v - 1}}{\beta(\sum_{i=1}^{j} y_i + u,\ jn - \sum_{i=1}^{j} y_i + v)}. \tag{24}$$

The predictive density function of Y is

$$f(y|y_1, y_2, ..., y_j) \ = \ \binom{n}{y} \frac{\beta(\sum_{i=1}^{j} y_i + y + u,\ (j+1)n - \sum_{i=1}^{j} y_i - y + v)}{\beta(\sum_{i=1}^{j} y_i + u,\ jn - \sum_{i=1}^{j} y_i + v)}. \tag{25}$$

4.1 *Bayes Sequential Decision Procedure*

Unlike the Bayes sequential decision procedure with a two-point prior, there are no closed forms for the risk from stopping the trial $r_0(w^j, j)$ or the optimal continuation of the trial $r_{l-j}(w^j, j)$ in the procedures with a beta prior. Their computations have to rely on the approximation or numerical approaches. As an example, a one-interim analysis procedure is used to describe the Bayes sequential decision procedure.

At the interim analysis, suppose that the value $Y_1 = y_1$ has been observed, the Bayes risk from stopping the trial, $r_0(w^1, 1)$, is,

$$\begin{aligned} r_0(w^1, 1) &= \min_{d \in \{d_0, d_1\}} E_{p|y_1} L(p, d) \\ &= \min\{E_{p|y_1} L(p, d_0),\ E_{p|y_1} L(p, d_1)\}, \end{aligned} \tag{26}$$

where

$$\begin{aligned} E_{p|y_1} L(p, d_0) &= \int_0^1 L(p, d_0)\, w(p|y_1)\, dp \\ &= \int_{p_0}^1 K(p - p_0) t \frac{p^{y_1+u-1}(1-p)^{n-y_1+v-1}}{\beta(y_1+u,\ n-y_1+v)}\, dp \\ &= Kt \left(\frac{y_1+u}{n+u+v} \int_{p_0}^1 \frac{p^{y_1+u+1-1}(1-p)^{n-y_1+v-1}}{\beta(y_1+u+1,\ n-y_1+v)}\, dp \right. \\ &\qquad \left. - p_0 \int_{p_0}^1 \frac{p^{y_1+u-1}(1-p)^{n-y_1+v-1}}{\beta(y_1+u,\ n-y_1+v)}\, dp \right), \\ E_{p|y_1} L(p, d_1) &= \int_0^1 L(p, d_1)\, w(p|y_1)\, dp \\ &= \int_0^1 K(p_0 - p) t \frac{p^{y_1+u-1}(1-p)^{n-y_1+v-1}}{\beta(y_1+u,\ n-y_1+v)}\, dp \\ &= Ktp_0 - Kt \frac{y_1+u}{n+u+v}. \end{aligned}$$

The expected risk from observing the next observation Y is,

$$\begin{aligned} E^* r_0(w^1(Y), 2) &= \sum_{y=0}^n r_0(w^1(Y=y), 2)\, f(y|y_1) \\ &= \sum_{y=0}^n \left(\min_{d \in \{d_0, d_1\}} E_{p|y_1, y} L(p, d) \right) f(y|y_1) \\ &= \sum_{y=0}^n \left(\min_{d \in \{d_0, d_1\}} \int_0^1 L(p, d)\, w(p|y_1, y)\, dp \right) f(y|y_1), \end{aligned} \tag{27}$$

where the posterior density function $w(p|y_1, y)$ and the predictive density function $f(y_1|y)$ are obtained by (24) and (25), respectively. The numerical methods are used to obtain the integrals in (26) and (27).

If the Bayes risk from stopping the trial $r_0(w^1, 1)$ is less than $E^* r_0(w^1(Y), 2)+n$, the clinical trial could be stopped; otherwise the trial needs to be continued.

At each analysis j, $j = 1, 2, ..., l$, in the study, after values $Y_1 = y_1$, $Y_2 = y_2$,..., $Y_j = y_j$ have been observed, if the trial is terminated, then the Bayes risk from stopping the trial $r_0(w^j, j)$ is,

$$\begin{aligned} r_0(w^j, j) &= \min_{d \in \{d_0, d_1\}} E_{p|y_1, y_2, \ldots, y_j} L(p, d) \\ &= \min\{E_{p|y_1, y_2, \ldots, y_j} L(p, d_0), \quad E_{p|y_1, y_2, \ldots, y_j} L(p, d_1)\}. \end{aligned} \tag{28}$$

Since the $E_{p|y_1, y_2, \ldots, y_j} L(p, d_0)$ and the $E_{p|y_1, y_2, \ldots, y_j} L(p, d_1)$ are functions of $\sum_{i=1}^{j} y_i$, and the drug is not accepted if the response rate is not high enough, that is, the $\sum_{i=1}^{j} y_i$ is small. The decision rule after the trial being terminated has the form,

$$d = \begin{cases} d_0(\mathit{reject\ the\ new\ drug}) & \sum_{i=1}^{j} y_i \leq k_1 \\ d_1(\mathit{accept\ the\ new\ drug}) & \sum_{i=1}^{j} y_i \geq k_2 \end{cases} \tag{29}$$

where $0 \leq k_1 \leq k_2 \leq jn$.

4.2 *Prior Information*

The expectation and variance of the prior distribution of p in (23) are,

$$\begin{aligned} E(p) &= \frac{u}{u+v}. \\ Var(p) &= \frac{uv}{(u+v)^2(u+v+1)} \\ &= \left(\frac{u}{u+v}\right)\left(1-\frac{u}{u+v}\right)\left(\frac{1}{u+v+1}\right) \\ &= E(p)(1-E(p))\left(\frac{1}{u+v+1}\right). \end{aligned}$$

The variance $Var(p)$ is a function of $u+v$ if the expectation $E(p)$ is fixed. Therefore the value of $u+v$ may be regarded as a measure of prior information.

If $E(p) < p_0$, then there is prior belief that the new drug is not effective. The larger value of $u+v$ (or the more prior information), the more likely that the new drug would be rejected, and hence the smaller type I error rate α would be. When $E(p) > p_0$, it is assumed that the new drug is effective by the prior information. The larger value of $u+v$, that is, the more prior information, the more chance that the new drug be accepted, and hence the smaller type II error rate β would be. These are useful in defining a beta prior distribution when summarsing prior information.

4.3 *Comparison with the Two-point Prior Distribution*

The comparisons are based on type I, type II error rates, their corresponding expected sample sizes, and average Bayes risks. The values of parameters u, v of the beta prior distribution are selected to have the same expectation and variance as those of the two-point prior distributions in Table 1 and Table 2. The loss and cost functions are also same as those of the two-point prior distributions in Table 1 and Table 2. The results of the type I error rate α and its corresponding expected sample size *Mean* N_α, and the type II error rate β and its corresponding expected sample size *Mean* N_β on the beta prior distribution are listed in Table 3 and Table 4. The expected Bayes risk *Mean Risk* of two-point prior distribution and beta prior distribution are listed in Table 5. The first column of *Mean Risk* under each patient horizon t is corresponding to two-point priors and the second to the beta priors.

Comparing Table 1 with Table 3, and Table 2 with Table 4, it can been seen that the type I error rates α of Bayes sequential decision procedures with the beta prior distribution are smaller than those of Bayes sequential decision procedures with the two-point prior distribution; the type II error rates β of Bayes sequential decision procedures with the beta prior distribution are larger than those of Bayes sequential decision procedures with two-points prior distribution, but can still be smaller than those of procedures of Pocock in some situations. From Table 5, it can been seen that the average Bayes risks of Bayes sequential decision procedures with beta prior distributions are bigger than those with two-point prior procedures. This is because the two-point prior distribution could be assumed when we have very strong prior information.

Tables 3, 4, 5 about here

5. Discussion

As we all know that the difficult in using the decision theory is the computational complexity and the difficulty of specifying loss and cost functions. For a clinical trial with binary measures, the group sequential samples, $X_{ij}, i = 1, 2, ..., n, j = 1, 2, ..., l$, could be replaced by classical sequential sample $Y_j, j = 1, 2, ..., l$ with binomial distribution $B(n, p)$, where p is the parameter of interest. The set of all possible values of Y_j is $\{0, 1, 2, ..., n\}$, which is a finite set. Consequently the computational difficulties of Bayesian inferences on the parameter p may be partly overcome. Hence we introduced the loss function in which the patient horizon and treatment effect are considered. The simulation results have shown that there are Bayes sequential decision procedures which have the type I error and type II error rates close to the frequentist procedures with smaller expected sample sizes. We are very interested in applying our methods in practice and develop the corresponding software to make it feasible.

The issue of health economics in clinical trials has been raised increasingly. The Bayesian decision procedure will be a very useful technique in this field. For example, the sample size calculation for a clinical trial developed can be used to design a clinical trial in which the cost effect is considered. We believe that Bayes decision theory will be useful in clinical trials.

References

Berger, J. (1985). *Statistical Decision Theory and Bayesian Analysis.* Springer, Berlin.

Carlin, B., Kadane, J. and Gelfand, A. (1998). Approaches for optimal sequential decision analysis in clinical trials. *Biometrics* **54**, 964–975.

DeGroot, M. (1970). *Optimal Statistical Decisions.* McGraw-Hill, New York.

Freedman, L. and Spiegelhalter, D. (1989). Comparison of Bayesian with group sequential methods for monitoring clinical trials. *Controlled Clin. Trials* **10**, 357–367.

Freedman, L. and Spiegelhalter, D. (1993). Application of Bayesian statistics to decision making during a clinical trial. In *Yearbook of Medical Informatics*, pages 353–365.

Lewis, R. (1996). Bayesian hypothesis testing: Interim analysis of a clinical trial evaluating phenytoin for the prophylaxis of early post-traumatic seizures in children. In Berry, D. and Stangl, D., editors, *Bayesian Biostatistics*, pages 279–296. Marrel Dekker, New York.

Lewis, R. and Berry, D. (1994). Group sequential clinical trials - A classical evaluation of Bayesian decision theoretic designs. *J. Amer. Statist. Assoc.* **89**, 1528–1534.

O'Brien, P. and Fleming, T. (1979). A multiple testing procedure for clinical trials. *Biometrics* **35**, 549–556.

Pocock, S. (1977). Group sequential methods in the design and analysis of clinical trials. *Biometrika* **64**, 191–199.

Qian, W. (1997). *Bayes Methods in Group Sequential Clinical Trials.* PhD Thesis,University of Kent.

Qian, W. and Brown, P. J. (1999). Bayes sequential decision theory in clinical trials. In Bernardo, J. M., Berger, J. O., Dawid, A. P. and Smith, A. F. M., editors, *Bayesian Statistics 6*, pages 829–838. Oxford University Press.

Sylvester, R. (1988). A Bayesian approach to the design of phase II clinical trials. *Biometrics* **44**, 823–836.

Vlachos, P. and Gelfand, A. (1998). MCMC implementation of multistep backward induction. Technical report, Department of Statistics, University of Connecticut.

Table 1 Type I error rate α and expected sample size Mean N_α for two-point prior

	t=500		t=1000		t=5000		t=10000	
w_1	α	Mean N_α	α	Mean N_α	α	Mean N_α	α	Mean N_α
Pocock's procedure: $p_1 = 0.01$, Max N = 39, Mean N = 17.5								
Bayes sequential decision procedure:								
0.3	**0.0082**	**14.7**	**0.0082**	**14.7**	0.0085	28.8	0.0024	29.0
0.5	**0.0082**	**14.7**	**0.0082**	**14.7**	**0.0007**	**14.9**	0.0008	26.4
0.7	**0.0072**	**13**	**0.0082**	**14.7**	**0.0007**	**14.9**	**0.0007**	**14.9**
0.9	**0.0004**	**13.1**	**0.0004**	**13.1**	**0.0007**	**14.9**	**0.0007**	**14.9**
Pocock's procedure: $p_1 = 0.05$, Max N = 63, Mean N = 28								
Bayes sequential decision procedure:								
0.3	0.0849	21	0.0898	25.4	0.0277	37.2	0.0149	37.8
0.5	0.0849	21	**0.0205**	**22.5**	**0.0080**	**27.8**	0.0066	37.9
0.7	**0.0189**	**21**	**0.0205**	**22.5**	**0.0061**	**27.8**	0.0038	27.9
0.9	0.0032	21	0.0036	21.4	0.0011	23.0	0.0012	23.4
Pocock's procedure: $p_1 = 0.10$, Max N = 129, Mean N = 55.5								
Bayes sequential decision procedure:								
0.3	0.1332	43	0.1333	43	0.0728	60.6	0.0373	65.0
0.5	0.0607	43	0.0607	43	**0.0318**	**55.0**	0.0162	64.2
0.7	0.0607	43	0.0607	43	**0.0124**	**49.4**	0.0067	56.7
0.9	0.0087	43	0.0087	43	**0.0044**	**45.9**	0.0024	49.4

Table 2 Type II error rate β and expected sample size $MeanN_\beta$ for two-point prior

	t=500		t=1000		t=5000		t=10000	
w_1	β	Mean N_β	β	Mean N_β	β	Mean N_β	β	Mean N_β
Pocock's procedure: $p_1 = 0.01$, Max N = 39, Mean N = 20.5								
Bayes sequential decision procedure:								
0.3	**0.0275**	**14.5**	**0.0275**	**14.5**	0.0042	14.9	0.0048	17.7
0.5	**0.0275**	**14.5**	**0.0275**	**14.5**	**0.0292**	**17.8**	0.0092	18.2
0.7	**0.1267**	**13**	**0.0275**	**14.5**	**0.0292**	**17.8**	**0.0292**	**17.8**
0.9	**0.1321**	**16.0**	**0.1321**	**16.0**	**0.0292**	**17.8**	**0.0292**	**17.8**
Pocock's procedure: $p_1 = 0.05$, Max N = 63, Mean N = 33.5								
Bayes sequential decision procedure:								
0.3	0.0745	21	0.0236	22.3	0.0080	26.0	0.0085	30.0
0.5	0.0745	21	**0.0843**	**23.7**	**0.0313**	**29.3**	0.0151	31.1
0.7	**0.1917**	**21**	**0.0843**	**23.7**	**0.0323**	**30.4**	0.0329	34.9
0.9	0.3674	21	0.2063	25.1	0.0976	34.6	0.0913	34.8
Pocock's procedure: $p_1 = 0.10$, Max N = 129, Mean N = 66								
Bayes sequential decision procedure:								
0.3	0.0612	43	0.0612	43	0.0161	48.6	0.0147	53.8
0.5	0.1237	43	0.1237	43	**0.0339**	**52.6**	0.0217	60.5
0.7	0.1237	43	0.1237	43	**0.0740**	**57.1**	0.0420	67.6
0.9	0.3390	43	0.3390	43	**0.1411**	**61.0**	0.0916	73.4

Table 3 Type I error rate α and expected sample size Mean N_α for beta prior

		t=500		t=1000		t=5000		t=10000	
u	v	α	MeanN_α	α	MeanN_α	α	MeanN_α	α	MeanN_α
Pocock's procedure: $p_1 = 0.01$, Max N = 39, Mean N = 17.5									
Bayes sequential decision procedure:									
1.98	9.12	.0003	13	.0003	13.1	$<$ .0001	14.6	$<$ .0001	14.6
0.89	5.96	.0003	13	$<$ .0001	13.0	$<$ .0001	13.1	$<$ .0001	14.6
0.43	4.79	.0003	13	$<$ .0001	13.0	$<$ .0001	13.1	$<$ .0001	14.6
0.18	5.15	$<$ .0001	13	$<$ .0001	13.0	$<$ .0001	13.1	$<$ .0001	13.1
Pocock's procedure: $p_1 = 0.05$, Max N = 63, Mean N = 28									
Bayes sequential decision procedure:									
1.98	9.12	.0189	21	.0189	21	.0006	22.8	.0005	27.0
0.89	5.96	.0032	21	.0032	21	.0005	22.8	.0001	22.9
0.43	4.79	.0032	21	.0032	21	.0001	21.4	.0001	22.9
0.18	5.15	.0004	21	.0004	21	$<$ .0001	21.1	$<$ .0001	21.4
Pocock's procedure: $p_1 = 0.10$, Max N = 129, Mean N = 55.5									
Bayes sequential decision procedure:									
1.98	9.12	.0607	43	.0607	43	.0283	47.7	.0117	54.3
0.89	5.96	.0244	43	.0244	43	.0105	45.2	.0041	48.9
0.43	4.79	.0087	43	.0087	43	.0091	43.7	.0032	45.6
0.18	5.15	.0008	43	.0008	43	.0002	43.1	$<$ .0001	43.2

Table 4 Type II error rate β and expected sample size Mean N_β for beta prior

		t=500		t=1000		t=5000		t=10000	
u	v	β	MeanN_β	β	MeanN_β	β	MeanN_β	β	MeanN_β
Pocock's procedure: $p_1 = 0.01$, Max N = 39, Mean N = 20.5									
Bayes sequential decision procedure:									
1.98	9.12	0.3326	13	0.1952	15.7	0.2175	28.3	0.2175	28.3
0.89	5.96	0.3326	13	0.3817	16.9	0.2622	26.2	0.2175	28.3
0.43	4.79	0.3326	13	0.3817	16.9	0.2622	26.2	0.2175	28.3
0.18	5.15	0.5843	13	0.4163	16.3	0.2927	25.2	0.2686	30.1
Pocock's procedure: $p_1 = 0.05$, Max N = 63, Mean N = 33.5									
Bayes sequential decision procedure:									
1.98	9.12	0.1917	21	0.1917	21	0.1479	34.1	0.1261	37.7
0.89	5.96	0.3674	21	0.3674	21	0.2120	34.9	0.1990	40.4
0.43	4.79	0.3674	21	0.3674	21	0.2845	35.8	0.2077	42.8
0.18	5.15	0.5666	21	0.5666	21	0.4607	34.7	0.3769	42.2
Pocock's procedure: $p_1 = 0.10$, Max N = 129, Mean N = 66									
Bayes sequential decision procedure:									
1.98	9.12	0.1237	43	0.1237	43	0.0864	49.7	0.0632	58.6
0.89	5.96	0.2175	43	0.2175	43	0.1591	52.3	0.1158	63.8
0.43	4.79	0.3390	43	0.3390	43	0.2439	48.2	0.1794	61.3
0.18	5.15	0.6145	43	0.6145	43	0.6273	53.5	0.5177	73.7

Table 5 Expected Bayes Risk for the Two-point and Beta Priors

			t=500		t=1000		t=5000		t=10000	
w_1	u	v	Mean Risk		Mean Risk		Mean Risk		Mean Risk	
$p_1 = 0.01$, Max N = 39, Mean $N_\alpha = 20.5$, Mean N_β=23.0										
0.3	1.98	9.12	-13.0	-7.6	-26.1	-18.0	-136.9	-110.5	-276.2	-221.1
0.5	0.89	5.96	-9.1	-5.7	-18.3	-13.3	-93.9	-79.8	-195.8	-165.4
0.7	0.43	4.79	-4.1	-3.4	-10.5	-8.2	-56.1	-48.5	-112.3	-100.2
$p_1 = 0.05$, Max N = 63, Mean $N_\alpha = 33.5$, Mean N_β=26.5										
0.3	1.98	9.12	-10.4	-7.7	-23.4	-15.3	-132.8	-99.8	-269.9	-206.5
0.5	0.89	5.96	-6.0	-4.8	-15.4	-9.6	-91.3	-67.9	-190.0	-140.3
0.7	0.43	4.79	-2.9	-2.1	-8.3	-4.2	-53.6	-32.6	-108.9	-71.8
$p_1 = 0.10$, Max N = 129, Mean $N_\alpha = 66$, Mean N_β=72										
0.3	1.98	9.12	-10.7	-8.3	-21.4	-16.5	-126.8	-91.4	-262.8	-195.7
0.5	0.89	5.96	-6.3	-4.7	-12.6	-9.4	-86.9	-55.0	-184.8	-122.0
0.7	0.43	4.79	-2.8	-1.5	-5.6	-3.0	-47.7	-19.0	-106.2	-48.9