ADAPTIVE BAYESIAN COMPOUND DESIGNS FOR DOSE FINDING STUDIES

J. M. McGree[1], C. C. Drovandi[1], M. H. Thompson[1], J. A. Eccleston[2], S. B. Duffull[3], K. Mengersen[1], A. N. Pettitt[1] and T. Goggin[4]

[1] *Queensland University of Technology*
GPO Box 2434, Brisbane, QLD
Australia, 4001
james.mcgree@qut.edu.au

[2] *University of Queensland*
Hawken Dr, St. Lucia, QLD
Australia, 4072

[3] *University of Otago*
PO Box 913, Dunedin, Otago
New Zealand, 9001

[4] *Roche Pharmaceuticals*
F. Hoffmann-La Roche, Basel
Switzerland, CH-4070

Abstract

We consider the problem of how to efficiently and safely design dose finding studies. Both current and novel utility functions are explored using Bayesian adaptive design methodology for the estimation of a maximum tolerated dose (MTD). In particular, we explore widely adopted approaches such as the continual reassessment method and minimizing the variance of the estimate of an MTD. New utility functions are constructed in the Bayesian framework and are evaluated against current approaches. To reduce computing time, importance sampling is implemented to re-weight posterior samples thus avoiding the need to draw samples using Markov chain Monte Carlo techniques. Further, as such studies are generally first-in-man, the safety of patients is paramount. We therefore explore methods for the incorporation of safety considerations into utility functions to ensure that only safe and well-predicted doses are administered. The amalgamation of Bayesian methodology, adaptive design and compound utility functions is termed adaptive Bayesian compound design (ABCD). The performance of this amalgamation of methodology is investigated via the simulation of dose finding studies. The paper concludes with a discussion of results and extensions that could be included into our approach.

Key words: Adaptive design, Compound utility, Importance sampling, Markov chain

Monte Carlo, Optimal design, Safety, Utility functions.

1. Introduction

Early clinical trials in humans are generally interested in describing the probability of observing an adverse event for a range of doses, with the aim of estimating specific parameters such as the maximum tolerated dose (MTD). Questions about efficacious doses are not usually of interest but rather establishing a safe range of doses is important since these are brought forward in later trials. Indeed, an estimated MTD has important implications in drug development as this dose level may never be exceeded in subsequent trials. Generally, doses are administered to patients or cohorts as single or multiple ascending doses with all previous experimental information being available for selection of the following dose. The design problem then involves choosing the next dose to maximize the information about a parameter of interest (for example, an MTD) without compromising patient health.

One of the first papers in this area dates back to Tsutakawa (1980) who considered the estimation of a point on the logistic quantal response curve which could relate to, for example, the probability of observing an adverse event given the MTD was administered. A pragmatic approach to dose finding came from O'Quigley et al. (1990) who proposed the continual reassessment method (CRM). This approach concentrates experimental effort in the current best estimate of the area of interest. In terms of dosing finding trials, this relates to dosing subjects at the estimated MTD. One of the first Bayesian adaptive design procedures implemented within dose finding came from Whitehead and Brunier (1995). Priors induced through elicited data were included and doses were selected based on the gain in statistical information of an estimate. When a comparison of their approach and CRM was performed, limitations of the CRM were highlighted. Adaptive procedures were extended by Babb et al. (1998) who addressed the ethical need to control the probability of dosing over an MTD. Their idea was to approach dosing at the MTD whilst limiting the proportion of patients who receive an 'overdose'. Similar to the CRM, this methodology was developed with phase I cancer trials in mind. Haines et al. (2003) presented formal (Bayesian) optimal design theory based on c- and D-optimality to estimate an MTD. They too imposed constraints to keep a low probability of an administered dose exceeding some maximum acceptable dose level and also extended their approach to adaptive trials. Other notable research includes Whitehead and Williamson (1998); Whitehead et al. (2001, 2006a,b); Zhou et al. (2006, 2008); Thall et al. (2008). A review of recent methods for designing phase I clinical trials was given by Rosenberger and Haines (2002) and, more recently, a book on statistical methods for dose finding by Chevret (2006). The general focus of the above referenced material is either on the convergence to a quantile of interest or the minimization of the variance of a quantile estimate.

An important consideration in the design of dose finding trials is the safety of the patients. These trials are first-in-man, so there is the possibility of exposing patients to highly toxic doses, leading to the possible suspension of the trial. Current approaches

to design dose finding studies have been criticized for escalating too quickly (see Ratain et al. (1993), Korn et al. (1994) and Goodman et al. (1995)). A well-designed experiment should avoid, or at the very least minimize, the risk of administering a toxic dose. The issue of safety has been addressed recently by Zhou et al. (2006) and Zhou et al. (2008). In both papers, toxicity was controlled by bounding exposure to the drug, that is, not administering a dose if the estimated exposure is larger than a pre-defined tolerance level. Other approaches deal with the issue of safety by restricting the doses available for administration (as mentioned above, see Babb et al. (1998) and Haines et al. (2003)). Further examples include Dragalin et al. (2008), who developed a two-stage design approach accounting for both efficacy and safety by only selecting doses within a therapeutic range. Also, Biedermann et al. (2007) implemented (local) compound optimal designs for estimating multiple percentiles on the dose response curve where the design space was restricted due to ethical concerns over drug toxicity and/or efficacy.

Our approach to this design problem considers compound design criteria (or utility functions) in the context of Bayesian adaptive design for dual experimental goals of estimating the MTD and addressing the safety of subjects. We have termed this amalgamation of methodology adaptive Bayesian compound design (ABCD). Six utility functions, denoted as ψ_1 to ψ_6, for estimation of the MTD will be considered. Utility functions ψ_1 to ψ_3 are commonly found in the literature and are functions of the expectation of the 'new' data point. The other three utilities (ψ_4 to ψ_6) are evaluated by averaging over the predictive probability distribution of the 'new' data point. This requires additional computation, and importance sampling is used to substantially 'lighten' this burden. An indicator function is used to incorporate safety considerations into the utilities so that the 'best' dose for estimation is selected amongst those that are believed to be non-toxic (and therefore safe to administer). The sequential nature of the trial means that dose availability will adapt between participants, and larger doses will only be available if the probability of toxicity is well-predicted and relatively small. The performance of current utilities, our new approach to constructing designs, and the safety considerations will be evaluated via simulation studies under dose finding scenarios.

The paper is set out as follows. Initially, details regarding the dose finding studies are given in Section 2. This forms the basis for all utility functions and examples considered. The utilities are defined in Section 3, as well as the ABCD procedures that are implemented throughout the simulation studies. In Section 4, the results from the simulation studies are explored to compare utility functions and evaluate the safety provisions. The paper concludes with a discussion and suggestions for further research.

2. Dose finding studies

We consider a dose finding trial which could be, as given in Whitehead and Brunier (1995), a study of healthy volunteers for an experimental compound that is intended as a treatment for depression. The aim is to estimate the MTD denoted as D^* such that

$P(Y = 1|D^*) = p$, the probability of observing an adverse event.

Consider the following logistic generalized linear model (GLM) (see McCullagh and Nelder (1989)),

$$\begin{aligned} \pi &= g(\boldsymbol{\theta}, D) && (1) \\ &= \frac{\exp(\theta_0 + \theta_1 D)}{1 + \exp(\theta_0 + \theta_1 D)}, && (2) \end{aligned}$$

where π is the probability of observing an adverse event given drug dosage D and $\boldsymbol{\theta} = (\theta_0, \theta_1)'$.

We assume an uninformative prior for $\boldsymbol{\theta}$ such that

$$\begin{bmatrix} \theta_0 \\ \theta_1 \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 100 & 0 \\ 0 & 100 \end{bmatrix} \right).$$

We assume there is a discrete set of doses D available for selection in each study $D \in \{0.1, 0.2, \ldots, 2\}$.

In order to explore the adaptive design of dose finding studies, a true model must be assumed so that data can be generated for selected doses. Suppose the true value of the model parameter is represented by $\boldsymbol{\theta}_t = (\theta_{0t}, \theta_{1t})'$, then the MTD (denoted as D^*) can be found as follows

$$D^* = \frac{\log\left(\frac{p}{1-p}\right) - \theta_{0t}}{\theta_{1t}}. \quad (3)$$

Any dose which yields a probability of an adverse event that is less than or equal to (for example) $p = 0.2$ is considered safe. Similarly, doses which yield probabilities larger than p are deemed toxic.

To initialize the ABCD dose finding procedure (discussed in the next section), several 'pseudo' observations need to be assumed (here we use two). Suggestions for this initialization are given in Whitehead and Brunier (1995). In our implementation, two doses ($D_{-1} = 0.9$ and $D_0 = 2.0$) are selected and the probabilities of toxicity are assumed to be elicited. These elicited probabilities enter the likelihood directly (i.e. not as 0s or 1s). Consider the following likelihood for the pseudo data

$$p(\boldsymbol{\delta}|\boldsymbol{\theta}) = \Pi_{i=-1}^{0} \pi^{\delta_i} (1 - \pi)^{1-\delta_i}, \quad (4)$$

where π is a function of $\boldsymbol{\theta}$ and D (see equation (1)), and δ_i are the elicited data for $i = -1, 0$.

The induced prior for the study is compromised of two parts; the non-informative prior on the parameters $p(\boldsymbol{\theta})$ and the information obtained from the elicited data $p(\boldsymbol{\delta}|\boldsymbol{\theta})$. This induced prior is shown in Figure 1, with the predictions arising from the assumed

true model. The 90% credible interval for the model predictions essentially covers the entire interval $[0, 1]$ for doses less than one. For doses larger than one, there is movement away from the lower bound of zero.

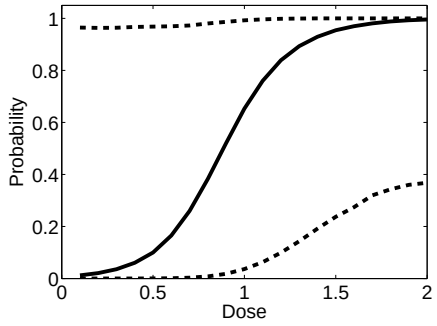

Figure 1: Model predictions (-) and 90% credible interval (- -) given the prior information and elicited data.

3. Adaptive Bayesian compound design (ABCD)

Adaptive design has received much attention in the literature. The idea is to choose the next dose (or design point) conditional on all current data. As more data become available, there is less uncertainty about the true values of the parameters, leading to a more efficient design. Chaudhuri and Mykland (1993) demonstrated this within a nonlinear context. Specifically in regard to binary response data, Dixon and Mood (1948) and Robbins and Monro (1951) presented the early work in this area. More recent developments have come from Haines et al. (2003) and Dror and Steinberg (2008).

The ABCD procedure used for all examples is outlined in this section. The procedure is based on that given in Whitehead and Brunier (1995) who consider a general adaptive design framework where each individual receives a single dose, and prior information is updated for dose selection after each individual. In their research, a conjugate-type prior was used for the experiment, and the expectation of the utility function was approximated by evaluating the utility at the expectation of the unknown parameters. Four key ingredients were discussed; the model, prior knowledge, set of possible actions and a utility function. The first three ingredients for our work were defined in Section 2, and will be considered in the examples that follow in Section 4. For the final ingredient, six utility functions will be explored, and are defined later in this section. Our work will extend the methodology of Whitehead and Brunier (1995) to fully Bayesian techniques using MCMC methods to accommodate a flexible choice of priors. Fully sequential simulation-based design has been explored previously by Müller et al. (2007). Their approach, however, is generally only suitable (i.e. practical) when the design set/decision set is small, such as whether to stop or continue the experiment.

Because of the lack of conjugacy of the model, we implemented a Metropolis-Hastings (MH) algorithm to sample from posterior distributions. In our MH-algorithm,

a proposal parameter, $\boldsymbol{\theta}^{(k)}$, was generated conditional on its current value, $\boldsymbol{\theta}^{(k-1)}$, from the proposal distribution $q(\boldsymbol{\theta}^{(k)}|\boldsymbol{\theta}^{(k-1)})$, for $k = 1, \ldots, nos$ where nos is the number of particles in the posterior sample. An independent proposal was used based on a bivariate normal distribution with mean and variance-covariance matrix determined by a maximum likelihood fit to the current data. The proposal was updated each time a new dose was selected, making it adaptive between patients.

Next, six utility functions are defined. The first three are functions of the expectation of a new data point and have appeared in the literature. We denote these as ψ_1 (Whitehead and Brunier, 1995), ψ_2 (O'Quigley et al., 1990) and ψ_3 (McGree et al., 2008). The remaining three utilities, denoted as ψ_4 (trimmed posterior variance), ψ_5 (posterior interquartile range) and ψ_6 (posterior variance-covariance), are evaluated by averaging over the predictive probability distribution of the new data point. For these latter three utilities, importance sampling was used to reduce the computational burden.

3.1. Variance of $\hat{D}^*$

Whitehead and Brunier (1995) proposed a utility function which aimed to minimize the variance of the estimate of the MTD. Let D^* denote the true MTD and $\hat{D}^*$ its estimate. Suppose we are interested in the optimal choice of dose for the ith individual. This choice will be based on $\boldsymbol{y}_{-1:i-1}$ which represents the elicited data and data up to individual $i-1$. Then under this utility, the next dose selected will be the D that minimizes:

$$\psi_1(D) \;=\; E_{\boldsymbol{\theta}|\boldsymbol{y}_{-1:i-1}}\{(\boldsymbol{\tau}'[\boldsymbol{M}(\boldsymbol{\theta}, D) + \boldsymbol{M}(\boldsymbol{\theta}, \boldsymbol{D_o})]^{-1}\boldsymbol{\tau})\},$$

where $\boldsymbol{M}(\boldsymbol{\theta}, D)$ represents the expected Fisher information from dose D, $\boldsymbol{M}(\boldsymbol{\theta}, \boldsymbol{D_o})$ represents the expected information from all previous doses $\boldsymbol{D_o}$, and $\boldsymbol{\tau}' = \partial D^*/\partial\boldsymbol{\theta}$. Here, $\boldsymbol{\tau}$ is a vector of partial derivatives of the functional expression for D^* with respect to $\boldsymbol{\theta}$ (defined in equation (3)). Hence, the utility is the minimization of the expectation of the asymptotic (Delta-method) variance of $\hat{D}^*$ over $\boldsymbol{\theta}$.

3.2. Continuous Reassessment Method (CRM)

The CRM was proposed by O'Quigley et al. (1990) and is one of the original utility functions for dose finding. The aim is to dose subjects at (or nearest to) the MTD. Then, the dose chosen for the next subject i will minimize:

$$\psi_2(D) \;=\; E_{\boldsymbol{\theta}|\boldsymbol{y}_{-1:i-1}}\{\text{abs}(g(\boldsymbol{\theta}, D) - p)\},$$

where $g(\boldsymbol{\theta}, D)$ is defined in equation (1).

Inherent in this utility is a safety consideration. Continual selection of the dose which is closest to the MTD ensures that patients are never (knowingly) exposed to highly toxic doses.

3.3. Coefficient of variation of $\hat{D}^*$

Similar to minimizing the variance of $\hat{D}^*$, this utility function considers minimizing the coefficient of variation of $\hat{D}^*$ given the posterior sample. A coefficient of variation as a utility function has been considered by McGree et al. (2008). The ψ_3-optimal design minimizes:

$$\psi_3(D) \;=\; E_{\boldsymbol{\theta}|\boldsymbol{y}_{-1:i-1}}\{(\boldsymbol{\tau}'[\boldsymbol{M}(\boldsymbol{\theta},D)+\boldsymbol{M}(\boldsymbol{\theta},\boldsymbol{D_o})]^{-1}\boldsymbol{\tau})^{1/2}/\hat{D}^*\}.$$

This utility weights the measure of variability by the magnitude of the estimate. Then, relatively large variances will be given less weight if the corresponding estimate of the MTD is large.

3.4. Algorithm One: Bayesian adaptive design procedure for ψ_1 to ψ_3

The Bayesian adaptive design procedure implemented for ψ_1 to ψ_3 is outlined in Algorithm 1. At the end of the trial, the median of the posterior distribution of D^* is taken as $\hat{D}^*$. This estimate is based (only) on observed data. That is, the elicited data has been removed.

3.5. Trimmed posterior variance

A Bayesian approach to adaptive design may seek to minimize the posterior variance of D^*. Equation (3) shows how D^* depends on $\boldsymbol{\theta}$. Notice that a ratio of the parameters needs to be computed in order to determine D^*. Consequently, when sampling from the posterior distribution of the parameters, outliers may be produced and inflate the estimate of the posterior variance of D^*. Therefore, a robust measure of variance is preferred. In this paper, we consider a trimmed variance based on the middle 98% of the data.

Suppose we have $\boldsymbol{y}_{-1:i-1}$, and need to choose an optimal dose for the next individual i. Given a sample from $p(D^*(\boldsymbol{\theta})|\boldsymbol{y}_{-1:i-1})$, we require to find the dose D that minimizes the expectation of the trimmed posterior variance of $D^*(\boldsymbol{\theta})$, that is

$$\psi_4(D) \;=\; \sum_{\tilde{y}=0}^{1} p_{D,\tilde{y}}\mathrm{VART}[D^*(\boldsymbol{\theta})|\boldsymbol{y}_{-1:i-1},\tilde{y},D)].$$

This utility estimates the average posterior variance of the MTD over the predictive probability distribution $p_{D,\tilde{y}}$. This is somewhat straightforward to perform in a binary setting as there are only two possible outcomes for a given dose (toxicity or no toxicity), $\tilde{y}=0,1$. As the proposals in the MH-algorithm are based on a maximum likelihood fit, the maximum likelihood estimates for $\boldsymbol{\theta}$ are available. These were used to estimate $p_{D,\tilde{y}}$.

Algorithm 1 General dose finding procedure

%% Loop over the number of subjects %%
for $i = 1 : N$ **do**
  if $i = 1$ **then**
    %% Initially choose lowest dose %%
    $D_{opt} = \boldsymbol{D_a}(i)$
  else
    %% All data up to individual $i - 1$ including elicited data %%
    Let $\boldsymbol{y}_{-1:i-1} = (\boldsymbol{y}'_{1:i-1}, y_0, y_{-1})'$
    %% Sample $\boldsymbol{\theta}$ from $p(\boldsymbol{\theta}|\boldsymbol{y}_{-1:i-1})$ %%
    Draw *nos* of $\boldsymbol{\theta} \sim p(\boldsymbol{\theta}|\boldsymbol{y}_{-1:i-1})$
    %% Loop over available doses %%
    for $j = 1 : \text{length}(\boldsymbol{D_a})$ **do**
      Let $D = \boldsymbol{D_a}(j)$
      %% Evaluate criterion value for D %%
      Evaluate $\psi(D)$ given $\boldsymbol{\theta}$
    end for
    %% Choose D that minimizes criterion %%
    $D_{opt} = \arg\min_{D \in \boldsymbol{D_a}} \psi(D)$
    %% Simulate data given D and true model %%
    Draw $y_i \sim \text{Bern}(g(\boldsymbol{\theta}_t, D_{opt}))$
  end if
end for
%% Sample $\boldsymbol{\theta}$ from $p(\boldsymbol{\theta}|\boldsymbol{y}_{1:N})$ %%
Draw $\boldsymbol{\theta} \sim p(\boldsymbol{\theta}|\boldsymbol{y}_{1:N})$
Find $p(D^*(\boldsymbol{\theta})|\boldsymbol{y}_{1:N})$
%% Estimate D^* as the median of the posterior distribution for D^* %%
$\hat{D}^* = \text{median}[D^*(\boldsymbol{\theta})|\boldsymbol{y}_{1:N})]$

3.6. Posterior interquartile range

As an alternative to minimizing the trimmed posterior variance, one may also minimize the interquartile range of the posterior distribution for D^*. When estimated by simulation, this alternative measure of spread is more robust to outliers than the variance. This utility function is also motivated by the shape of the posterior distribution of D^*. Despite $p(\boldsymbol{\theta}|D)$ appearing quite normal (stable), the posterior distribution of D^* may become skewed and longtailed due to D^* involving the ratio of two approximate normal random variables. Accordingly, the variance of such distributions may not be the ideal summary for spread. This utility is similarly defined (to ψ_4) as one that minimizes the following:

$$\psi_5(D) \quad = \quad \sum_{\tilde{y}=0}^{1} p_{D,\tilde{y}} \mathrm{IQR}[D^*(\boldsymbol{\theta})|\boldsymbol{y}_{-1:i-1}, \tilde{y}, D)].$$

3.7. Posterior variance-covariance

The final utility function considered is an extension of ψ_1 with the posterior variance-covariance inserted instead of the inverse of the Fisher information. Consider minimizing the following:

$$\psi_6(D) \quad = \quad \sum_{\tilde{y}=0}^{1} p_{D,\tilde{y}} \boldsymbol{\tau}' \boldsymbol{V}(\boldsymbol{\theta}, [\boldsymbol{y}_{-1:i-1}, \tilde{y}, D)])\boldsymbol{\tau},$$

where $\boldsymbol{V}(\boldsymbol{\theta}, [\boldsymbol{y}_{-1:i-1}, \tilde{y}, D)])$ denotes the posterior variance-covariance matrix given all previous data points, the supposed new data point $\tilde{y}$ and dose D.

3.8. Algorithm Two: Bayesian adaptive design procedure for ψ_4 to ψ_6

The utilities ψ_4 to ψ_6 can be implemented using Algorithm 1, but this can be computationally intensive as we are required to sample from a large number of posterior distributions. That is, each time a new dose is selected, a new sample from a posterior distribution must be found. Moreover, for each individual dose allocation, we need to re-compute the posterior distribution for each possible dose as part of the dose selection phases. We perform this using importance sampling, thereby avoiding a costly Metropolis-Hastings run for each possible dose. Importance sampling is typically used as an estimator based on independent, identically distributed samples. This has been extended to dependent samples (see Tierney (1994)).

Consider the pseudo data and data up to individual $i-1$, denoted as $\boldsymbol{y}_{-1:i-1}$, for which we have samples from the posterior distribution of the parameter, $p(\boldsymbol{\theta}|\boldsymbol{y}_{-1:i-1})$. We use this posterior as the importance distribution. The target distribution is $p(\boldsymbol{\theta}|\boldsymbol{y}_{-1:i-1}, \tilde{y})$, where $\tilde{y}$ is the new data point given dose D. By importance sampling, the current posterior samples can be weighted to reflect our target with unnormalized importance weights given by

$$
\begin{aligned}
w_k = \frac{p(\boldsymbol{\theta}^{(k)}|\boldsymbol{y}_{-1:i-1}, \tilde{y})}{p(\boldsymbol{\theta}^{(k)}|\boldsymbol{y}_{-1:i-1})} &\propto \frac{p(\boldsymbol{y}_{-1:i-1}, \tilde{y}|\boldsymbol{\theta}^{(k)})p(\boldsymbol{\theta}^{(k)})}{p(\boldsymbol{y}_{-1:i-1})|\boldsymbol{\theta}^{(k)})p(\boldsymbol{\theta}^{(k)})} \\
&= p(\tilde{y}|\boldsymbol{\theta}^{(k)}),
\end{aligned}
$$

where $\boldsymbol{\theta}^{(k)}$ is the kth sample from $p(\boldsymbol{\theta}|\boldsymbol{y}_{-1:i-1})$. Following a normalization of the weights, $W_k = w_k/\sum_j w_j$, the set $\{\boldsymbol{\theta}^{(k)}, W_k\}_{k=1}^{nos}$ represents a weighted sample from $p(\boldsymbol{\theta}|\boldsymbol{y}_{-1:i-1}, \tilde{y})$. We calculate the values of our utilities using this weighted sample. Whilst importance sampling was used in the dose selection stage, Metropolis-Hastings was still utilized to obtain the new posterior distribution after the new data point has been observed.

For illustrative purposes, consider ψ_4. To select the optimal dose for the ith individual, we have particles $\boldsymbol{\theta}^{(k)}$ $(k = 1, \ldots, nos)$ as a posterior sample from $p(\boldsymbol{\theta}|\boldsymbol{y}_{-1:i-1})$. Then, for all $D \in \boldsymbol{D_a}$

1. Let $p_{D,1} = g(\hat{\boldsymbol{\theta}}, D)$ and $p_{D,0} = 1 - p_{D,1}$
2. For $\tilde{y} = 0, 1$,
   - Find weights $w_{\tilde{y}}^{(k)} = p(\tilde{y}|\boldsymbol{\theta}^{(k)}, D)$ and normalize
   - For middle 98% of $p(D^*(\boldsymbol{\theta})|\boldsymbol{y}_{-1:i-1})$, compute

$$
\begin{aligned}
\bar{D}_{\tilde{y}}^*(\boldsymbol{\theta}, D) &= \sum_k w_{\tilde{y}}^{(k)} D^*(\boldsymbol{\theta}^{(k)}) \\
\psi_4(D, \tilde{y}) &= \sum_k w_{\tilde{y}}^{(k)} (D^*(\boldsymbol{\theta}^{(k)}) - \bar{D}_{\tilde{y}}^*(\boldsymbol{\theta}, D))^2
\end{aligned}
$$

3. Estimate expectation of utility given D with $\psi_4(D) = \sum_{\tilde{y}=0}^{1} p_{D,\tilde{y}} \psi_4(D, \tilde{y})$
4. Optimal dose minimizes $\psi_4(D)$

The algorithm for implementing utility functions ψ_4 to ψ_6 is outlined in Algorithm 2. As in Algorithm 1, the median of the posterior distribution of D^* is taken as $\hat{D}^*$, and this estimate is solely based on prior information $p(\boldsymbol{\theta})$ and data collected.

3.9. Safety

All utility functions discussed focus on the estimation of the MTD. With the exception of the CRM, no utility functions have any safety precautions incorporated into their specification, meaning that a highly toxic dose could potentially be given to a patient in instances where little information has become available. To include safety considerations into all utilities, we consider the posterior predictive distribution of the probability of observing a toxic event, given a specific dose. Given these distributions (one for each dose), we were only interested in the upper bounds, that is, how large the probability of toxicity could be. Hence, we only allow doses to be available for selection

Algorithm 2 Dose finding procedure with importance sampling

%% Loop over the number of subjects %%
for $i = 1 : N$ **do**
  if $i = 1$ **then**
    %% Initially choose lowest dose %%
    $D_{opt} = \boldsymbol{D_a}(i)$
  else
    %% All data up to individual $i - 1$ including elicited data %%
    Let $\boldsymbol{y}_{-1:i-1} = (\boldsymbol{y}'_{1:i-1}, y_0, y_{-1})'$
    %% Sample $\boldsymbol{\theta}$ from $p(\boldsymbol{\theta}|\boldsymbol{y}_{-1:i-1})$ %%
    Draw *nos* of $\boldsymbol{\theta} \sim p(\boldsymbol{\theta}|\boldsymbol{y}_{-1:i-1})$
    %% Find the maximum likelihood estimates (MLEs) of $\boldsymbol{\theta}$ %%
    Let $\hat{\boldsymbol{\theta}}$ be the MLEs given $\boldsymbol{y}$
    %% Loop over available doses %%
    for $j = 1 : \text{length}(\boldsymbol{D_a})$ **do**
      Let $D = \boldsymbol{D_a}(j)$
      %% Estimate the probability distribution of the data %%
      Let $p_{D,1} = g(\hat{\boldsymbol{\theta}}, D)$, $p_{D,0} = 1 - p_{D,1}$
      for $\tilde{y} = 0 : 1$ **do**
        %% Find 'new' weights and normalize %%
        $w_{\tilde{y}}^{(k)} = p(\tilde{y}|\boldsymbol{\theta}^{(k)}, D)$ and normalize
        %% Evaluate weighted criterion %%
        Evaluate $\psi(D, \tilde{y})$
      end for
      %% Average criterion value over the distribution of predicted data %%
      Estimate $\psi(D) = \sum_{\tilde{y}=0}^{1} p_{D,\tilde{y}} \psi(D, \tilde{y})$
    end for
    %% Choose D that minimizes (weighted) criterion %%
    $D_{opt} = \arg\min_{D \in \boldsymbol{D_a}} \psi(D)$
    %% Simulate data given D and true model %%
    Draw $y_i \sim \text{Bern}(g(\boldsymbol{\theta}_t, D_{opt}))$
  end if
end for
%% Sample $\boldsymbol{\theta}$ from $p(\boldsymbol{\theta}|\boldsymbol{y}_{1:N})$ %%
Draw $\boldsymbol{\theta} \sim p(\boldsymbol{\theta}|\boldsymbol{y}_{1:N})$
Find $p(D^*(\boldsymbol{\theta})|\boldsymbol{y}_{1:N})$
%% Estimate D^* as the median of the posterior distribution for D^* %%
$\hat{D}^* = \text{median}[D^*(\boldsymbol{\theta})|\boldsymbol{y}_{1:N})]$

if, say, the 95th percentile of the posterior predictive probability was less than a certain tolerance level *tol*. Importantly, as more information becomes available, a higher level of certainty regarding which doses can be administered will result. Consider the following decision rule:

$$I_D = \begin{cases} 1, & \text{if } \texttt{prctile}[g(\boldsymbol{\theta}|\boldsymbol{y}, D), \gamma] < tol \\ 0, & \text{otherwise,} \end{cases}$$

where $\texttt{prctile}[g(\boldsymbol{\theta}|\boldsymbol{y}, D), \gamma]$ denotes the 100γth percentile of the posterior predictive distribution of the probability of toxicity given dose D for $\gamma \in [0, 1]$.

I_D indicates if dose D is currently considered safe 'enough' to administer. That is, if $I_D = 0$ for a given dose, then it will not be available for administration. Importantly, this approach to safety incorporates the uncertainty regarding the predicted probability of an adverse event for a given dose.

4. Results

Given the dependence of optimal designs on prior information and, to a certain extent, simulated data, the examples provided throughout are intended to be illustrative, rather than comprehensive. In what follows, different dose finding scenarios are defined, and the performance of each of the six utility functions (Section 3) are explored. Two MTDs are considered that are motivated by the incidence of toxic events in past trials. For example, the incidence rate of cardiac events for rofecoxib is less than 5% (see Bresalier et al. (2005)). Further, Whitehead and Brunier (1995) considered drugs intended for the treatment of conditions of moderate seriousness such as ischaemic heart disease, depression and arthritis. Mild adverse reactions occurred with incidence of about 20%. As such, the results shown in this section are presented in two scenarios with $p = 0.05$ and 0.20 each with $N = 50$ patients. As a comparison, a simulation study also conducted for an incidence of 10% with less doses available ($D \in \{0.2, 0.4, 0.6, 0.8, 1.0, 1.2\}$). Given the results were similar for this incidence, these are presented in Appendix A.

For each incidence considered, three versions of the dose finding trial were simulated. The first is where estimation of the MTD is the sole purpose of the trial (that is, no safety considerations). The second version restricts dose availability. Only the current dose and doses one step up and one step down will be available for administration. Obviously, this is not possible at the boundaries of the design space, so only two doses will be considered when this occurs. This dosing regimen should be safer for patients as larger doses will only be administered once information about lower doses is available. Lastly, the third version considers the restricted dose availability with the approach to safety outlined in Section 3.9. The tolerance level for both scenarios was set to $tol = 0.50$. We note that this is quite flexible for Scenario 1 where $p = 0.05$ and considerably less so for Scenario 2 where $p = 0.20$. In practice, one would probably lower the tolerance level in Scenario 1, but for comparative purposes (between scenarios) this was kept fixed.

For each scenario, 500 dose finding studies were simulated for each utility function and the estimated MTD and selected doses were recorded. Binary responses for each patient were randomly generated given doses selected by the utility and the assumed true model. The MCMC chains were run with a burnin of 10,000 and 40,000 draws were kept as posterior samples. Given we used an efficient and independent proposal distributions within our Metropolis-Hastings step, low serial correlation was observed, hence, no thinning was necessary. Summary statistics (shown in Tables 1 and 2) were then used to evaluate and compare utility functions. Results for the additional run where $p = 0.1$ with fewer doses being available can be seen in Table 3.

4.1. Scenario 1: $p = 0.05$

In this scenario, we consider the estimation of the MTD which relates to $p = 0.05$. Under the assumption that $\boldsymbol{\theta}_t = (-5, 5)'$, the true MTD = 0.4111. The summary statistics for the sampling distribution of the estimate of D^* are given in Table 1.

In all versions of Scenario 1, it was found to be quite difficult to estimate the MTD. Indeed, all estimates were consistently larger than the true MTD. This is particularly evident when the dose availability is restricted and when safety measures are included. The CRM (ψ_2) seems to suffer the least from this overestimation problem. This is highlighted in Trial 1.2, where the mean and median of the estimates is noticeably less than those produced by all other criteria. This provides empirical evidence to support using the CRM (ψ_2). It is difficult to draw any conclusions from comparing the variability of the estimates as there does not appear to be any consistency across trials/versions. Alternatively, the MSE shows that utilities ψ_1 to ψ_3 perform well across all versions of the trial. Utility functions ψ_4 to ψ_6 seem to, on occasions, suffer from extreme estimates of D^* and consequently have inflated standard deviations.

4.2. Scenario 2: $p = 0.20$

In this scenario, we consider the estimation of the MTD which relates to $p = 0.20$. Again, under the assumption that $\boldsymbol{\theta}_t = (-5, 5)'$, the true MTD is 0.7227. The summary statistics for the sampling distribution of the estimate of D^* are given in Table 2.

Estimation of the location of the MTD via the mean and median appears to be similar across all utility functions within the three versions of the dose finding trial. The problem of overestimating the MTD as seen in Scenario 1 does not seem to be an issue here. In terms of variability of estimates, there are no discernible differences in the performance of utility functions in different versions of the trial. When the mean square error (MSE) is compared, however, ψ_1 to ψ_3 perform particularly and consistently well in comparison with ψ_4 to ψ_6. This difference in the MSE seems to be due to inflated standard deviations for ψ_4 to ψ_6 suggesting that some extreme estimates were obtained using these utilities.

As a comparison to the prior shown in Figure 1, this plot was reproduced for one of the simulation studies (after all 50 data points were collected) using utility ψ_1. The predicted probabilities (with 90% credible intervals) and the true model are shown in Figure 2. As can be seen, there is close agreement between the true and estimated

Table 1: Summary statistics for the sampling distribution of the estimates of D^* for Scenario 1 ($p = 0.05$, MTD $= 0.4111$, $N = 50$). The scenario is considered in three trials; No safety (Trial 1.1), restricted dose availability (Trial 1.2) and restricted dose availability with safety considerations (Trial 1.3) Here, $D \in \{0.1, 0.2, \ldots, 2.0\}$.

Scenario	Utility	Mean	Median	SD	IQR	MSE
Trial 1.1	ψ_1	0.4413	0.4558	0.1271	0.1586	0.0171
	ψ_2	0.4332	0.4412	0.1160	0.1493	0.0139
	ψ_3	0.4358	0.4477	0.1354	0.1593	0.0189
	ψ_4	0.4403	0.4558	0.1238	0.1723	0.0162
	ψ_5	0.4148	0.4369	0.1895	0.1577	0.0359
	ψ_6	0.4146	0.4376	0.1902	0.1573	0.0362
Trial 1.2	ψ_1	0.4938	0.4892	0.1507	0.1716	0.0296
	ψ_2	0.4293	0.4431	0.1240	0.1700	0.0157
	ψ_3	0.4945	0.4868	0.1551	0.2001	0.0310
	ψ_4	0.5005	0.4901	0.1695	0.2434	0.0367
	ψ_5	0.4837	0.4855	0.1499	0.1713	0.0277
	ψ_6	0.4998	0.4847	0.1276	0.1349	0.0242
Trial 1.3	ψ_1	0.4560	0.4665	0.1086	0.1396	0.0138
	ψ_2	0.4300	0.4417	0.1254	0.1733	0.0161
	ψ_3	0.4459	0.4572	0.1110	0.1505	0.0135
	ψ_4	0.5014	0.4920	0.1696	0.2446	0.0369
	ψ_5	0.4839	0.4847	0.1517	0.1799	0.0283
	ψ_6	0.4767	0.4754	0.0872	0.1121	0.0119

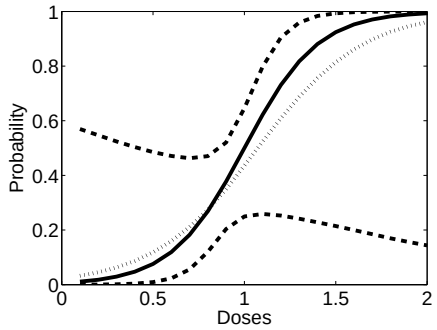

Figure 2: Model predictions under the true model (-) and the model predictions (·) and 90% credible intervals (- -) given data collected from 50 patients under utility function ψ_1.

probabilities close to the MTD. Further, there has been a substantial increase in the precision of the estimation of the MTD after 50 observations. The credible intervals in this region are tighter than that shown for higher doses highlighting that the utility

was focusing on estimation in this particular area.

Table 2: Summary statistics for the sampling distribution of the estimates of D^* for Scenario 2 ($p = 0.20$, MTD $= 0.7227$, $N = 50$). The scenario is considered in three trials; No safety (Trial 2.1), restricted dose availability (Trial 2.2) and restricted dose availability with safety considerations (Trial 2.3). Here, $D \in \{0.1, 0.2, \ldots, 2.0\}$.

Scenario	Utility	Mean	Median	SD	IQR	MSE
Trial 2.1	ψ_1	0.7304	0.7347	0.0766	0.0963	0.0059
	ψ_2	0.7239	0.7267	0.0766	0.0951	0.0059
	ψ_3	0.7308	0.7304	0.0721	0.1025	0.0053
	ψ_4	0.7311	0.7322	0.0778	0.1074	0.0061
	ψ_5	0.7261	0.7390	0.1123	0.1070	0.0126
	ψ_6	0.7208	0.7322	0.1184	0.1042	0.0140
Trial 2.2	ψ_1	0.7328	0.7275	0.0737	0.1013	0.0055
	ψ_2	0.7240	0.7293	0.0755	0.0999	0.0057
	ψ_3	0.7316	0.7313	0.0758	0.1038	0.0058
	ψ_4	0.7311	0.7322	0.0778	0.1074	0.0061
	ψ_5	0.7292	0.7369	0.0911	0.1085	0.0083
	ψ_6	0.7219	0.7304	0.0967	0.0961	0.0094
Trial 2.3	ψ_1	0.7189	0.7177	0.0750	0.0883	0.0056
	ψ_2	0.7165	0.7174	0.0813	0.0863	0.0066
	ψ_3	0.7174	0.7151	0.0805	0.1033	0.0065
	ψ_4	0.7135	0.7164	0.1012	0.1070	0.0103
	ψ_5	0.7167	0.7150	0.0873	0.1022	0.0077
	ψ_6	0.7073	0.7149	0.1053	0.1078	0.0113

4.3. Safety exploration

Our approach to addressing safety in dose-finding trials is evaluated in this section. Figures 3 and 4 show the median and 90% credible intervals for the doses selected across 500 trials for all patients where $p = 0.05$ and 0.20, respectively.

The three rows of Figure 3 relate to three utility functions (ψ_1, ψ_2 and ψ_6) while the three columns represent the different versions of the scenarios; no safety, restricted dose availability and restricted dose availability with safety considerations. Only results from three utilities are shown as the other results are similar. As expected, the CRM generally selects lower doses compared to the other utilities when safety is not included. The other utilities can often select doses much larger than the MTD when no safety considerations were incorporated. This is particularly noticeable for ψ_6 where the selection of extremely high doses is not uncommon. In fact, under the true model, these doses relate to a probability of toxicity greater than 0.9. When safety is included as a compound criterion, large doses are not common for any utility. This is demonstrated

by the upper bound for the 90% credible intervals of doses selected never exceeding $D = 1.0$ for any patient under any utility. This shows that, despite the tolerance level being much larger than the probability of toxicity relating to the MTD, there is a noticeable reduction in the selection of large doses when safety provisions are included. In fact, the selection of large doses in situations where safety is not considered may be a result of the estimation criteria preferring to observe some adverse events. Dosing near the MTD will only observe toxicity one in every twenty patients (on average).

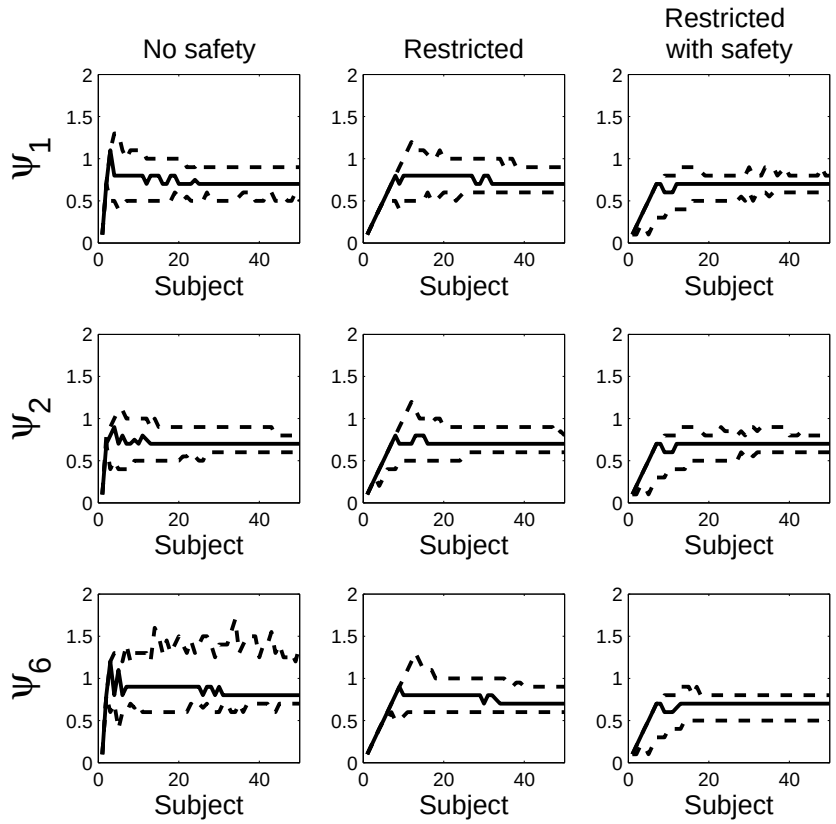

Figure 3: 5th, 50th and 95th percentiles of doses selected by utility functions ψ_1, ψ_2 and ψ_6 over 500 simulated dose trials for Scenario 1 where the MTD is 0.4111. The x-axis 'Subject' refers to the number of patients currently in the study.

Figure 4 is similar in presentation to Figure 3. In this scenario, the tolerance level is closer to the probability of toxicity relating to the MTD. Moreover, p is larger in this scenario in that we expect one in every five patients to observe an adverse event if dosed at the MTD. This may explain the decrease in the selection of large doses (in general) compared to Scenario 1. As before, when safety was included, the upper bound of the 90% credible interval for doses selected never exceeded $D = 1.0$. In fact, the dosing profiles appear to be similar across the utilities when safety is included. The percentages of doses selected above $D = 0.7$, 0.8 and 0.9 were approximately the same

across all utilities (when safety was included) at about 22%, 4% and 0.3%, respectively. In comparison, these percentages without safety were largest for ψ_6 which recorded 75%, 47% and 31%, respectively. The smallest percentages (without safety) were seen for the CRM (ψ_2) at 45%, 13% and 3%. This demonstrates how the inclusion of safety provisions into the CRM (ψ_2) may be beneficial. Furthermore, our safety measures yielded safe dosing regimens, despite the variety of dosing profiles that were given by different estimation criteria.

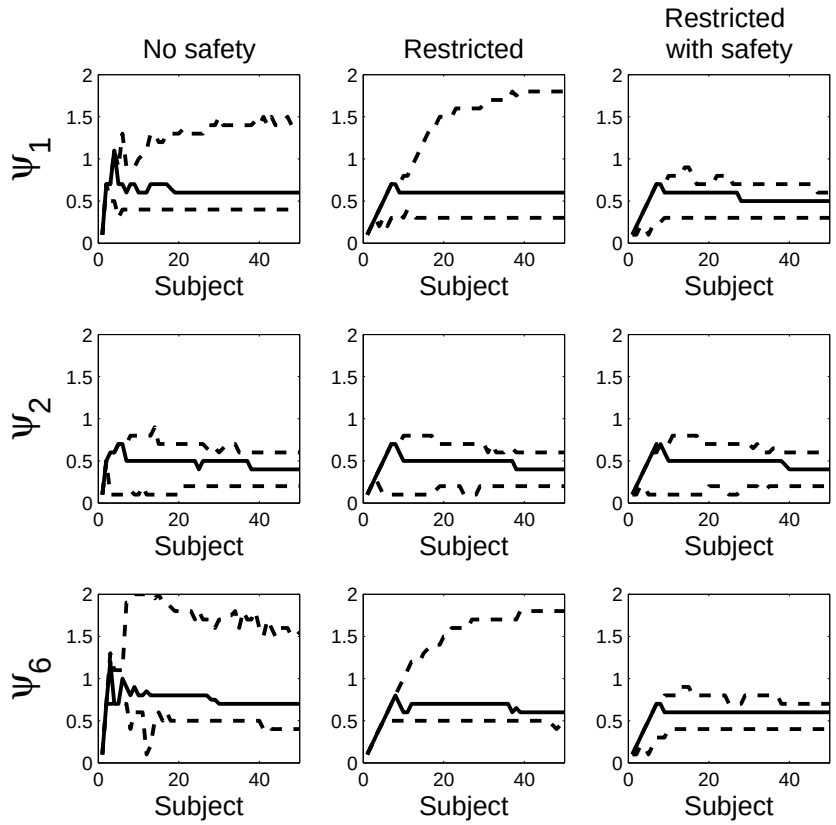

Figure 4: 5th, 50th and 95th percentiles of doses selected by utility functions ψ_1, ψ_2 and ψ_6 over 500 simulated dose trials for Scenario 2 where the MTD is 0.7227. The x-axis 'Subject' refers to the number of patients currently in the study.

5. Discussion

This paper considered methodology for Bayesian adaptive design with compound optimal design utilities. The application of our work was to dose finding studies and we explored novel design criteria for dose selection and the use of importance sampling for computational convenience. Various MTDs were considered, and scenarios were constructed to explore and compare our methodology.

Two of the examples considered a low ($p = 0.05$) and high ($p = 0.20$) incidence of observing an adverse event. The MTD for the high incidence example was generally well estimated by all utility functions, even when safety precautions were included. However, in the low incidence example, the MTD was generally overestimated with comparatively large uncertainty. It seems reasonable that an MTD relating to a low incidence of an adverse event will be more difficult to estimate than an MTD relating to a high incidence. This suggests that 'large' sample sizes may be required to obtain precise estimates of MTDs relating to a low incidence of adverse events.

Our approach to safety only allowed well-predicted doses to be available for administration. With this safety constraint, precise estimates of the MTD generally prevailed, highlighting the general applicability of our safety methodology. Placing restrictions on dose availability will generally affect the precision of the estimates, but this is compensated by the benefits of administering a safe dosing regimen. Moreover, the tolerance level in the safety provision allows the experimenter to control the level of safety of the study. Despite the CRM having inherent safety features, our approach to dealing with safety is still relevant and applicable. This is due to the fact that the CRM does not consider the uncertainty about the estimate of a probability of toxicity. In fact, plots of the doses selected over the 500 simulations demonstrate how the presence and absence of safety considerations can yield different designs for the CRM. The safe dosing regimen provided by incorporating a tolerance level ensures that doses higher than those already administered are only given to patients when sufficient information is provided regarding the probability of toxicity. This reduces the chance of exposing patients to highly toxic doses, and ensures safe dose finding studies.

This research included the development of utility functions, as well as the use of importance sampling to avoid the large computational burden that MCMC simulations present. There was little evidence to suggest using these new utility functions performed any 'better' than those in the literature. However, these utility functions may prove beneficial when expressions for the Fisher information are not available (or are based on the linearization of the model). This is not uncommon for mixed effects or hierarchical models, and our utilities offer a solution to this design problem.

The examples explored in this paper present a simplified dose finding study, and other concerns may be included. For example, the design space considered consisted of equally spaced doses. It may be more appropriate to assume a smaller dose increment at higher dose levels than at lower doses. For example, a modified Fibonacci sequence could be applied. Our assumptions regarding the design space are not a requirement for the methodology presented. In fact, any spacing between doses could be implemented within this framework.

An important consideration when designing dose finding trials is the possibility of early termination of the study due to the drug being unsafe or there being sufficient information for the location of the MTD. For example, O'Quigley and Reiner (1998) propose a binary tree approach to cease the trial once all possible outcomes for the remaining patients lead to no change in the current estimate of the MTD. Although stopping rules were not explored in this paper, this approach (and others) may read-

ily be incorporated. Indeed, after a patient's response is observed, a binary tree from O'Quigley and Reiner (1998) could be constructed, and a decision about the continuation of the study could be made. Alternatively, other stopping rules, such as those given by Zhou and Whitehead (2003), could be included.

Acknowledgements

J. M. McGree was supported by a grant from Roche Pharmaceuticals. The authors wish to thank both referees for their helpful comments. We also thank E. G. Ryan of Queensland University of Technology for proof reading our paper.

A. Addition scenario

Table 3: Summary statistics for the sampling distribution of the estimates of D^* for Scenario 3 ($p = 0.10$, MTD $= 0.5606$, $N = 50$). The scenario is considered in three trials; No safety (Trial 3.1), restricted dose availability (Trial 3.2) and restricted dose availability with safety considerations (Trial 3.3). Here, $D \in \{0.2, 0.4, 0.6, 0.8, 1.0, 1.2\}$.

Scenario	Utility	Mean	Median	SD	IQR	MSE
Trial 3.1	ψ_1	0.5740	0.5855	0.0993	0.1298	0.0100
	ψ_2	0.5667	0.5746	0.0994	0.1232	0.0099
	ψ_3	0.5738	0.5809	0.1001	0.1263	0.0102
	ψ_4	0.5752	0.5794	0.0973	0.1320	0.0097
	ψ_5	0.5851	0.5805	0.1342	0.1398	0.0186
	ψ_6	0.5779	0.5785	0.1355	0.1189	0.0186
Trial 3.2	ψ_1	0.5837	0.5889	0.0980	0.1167	0.0101
	ψ_2	0.5669	0.5807	0.1001	0.1304	0.0101
	ψ_3	0.5748	0.5800	0.0984	0.1179	0.0099
	ψ_4	0.6028	0.5980	0.1391	0.1799	0.0211
	ψ_5	0.5785	0.5868	0.0950	0.1157	0.0093
	ψ_6	0.5972	0.5942	0.1095	0.1089	0.0133
Trial 3.3	ψ_1	0.5658	0.5730	0.0933	0.1109	0.0087
	ψ_2	0.5579	0.5711	0.1030	0.1219	0.0106
	ψ_3	0.5624	0.5694	0.0978	0.1184	0.0096
	ψ_4	0.5579	0.5667	0.1027	0.1279	0.0105
	ψ_5	0.5646	0.5704	0.0983	0.1106	0.0097
	ψ_6	0.5646	0.5722	0.0953	0.1218	0.0091

References

Babb, J., Rogatko, A., Zacks, S., 1998. Cancer phase I clinical trials: Efficient dose escalation with overdose control. Statistics in Medicine 17, 1103–1400.

Biedermann, S., Dette, H., Zhu, W., 2007. Compound optimal designs for percentile estimation in dose-response models with restricted design intervals. Journal of Statistical Planning and Inference 137, 3838–3847.

Bresalier, R. S., Sandler, R. S., Quan, H., Bolognese, J. A., Oxenius, B., Horgan, K., Lines, C., Riddell, R., Morton, D., Lanas, A., Konstam, M. A., Baron, J. A., 2005. Cardiovascular events associated with Rofecoxib in a colorectal adenoma chemoprevention trial. New England Journal of Medicine 352, 1092–1102.

Chaudhuri, P., Mykland, P. A., 1993. Nonlinear experiments: Optimal design and inference based on likelihood. Journal of the American Statistical Association 88, 538–546.

Chevret, S., 2006. Statistical Methods for Dose-Finding Experiments. Wiley Statistics in Practice. John Wiley & Sons, England.

Dixon, W. J., Mood, A. M., 1948. A method for obtaining and analysing sensitivity data. Journal of the American Statistical Association 43, 109–126.

Dragalin, V., Fedorov, V. V., Wu, Y., 2008. Adaptive designs for selecting drug combinations based on efficacy-toxicity response. Journal of Statistical Planning and Inference 138, 352–373.

Dror, H. A., Steinberg, D. M., 2008. Sequential experimental designs for generalized linear models. Journal of the American Statistical Association 103, 288–298.

Goodman, S. N., Zahurak, M. L., Piantadosi, S., 1995. Some practical improvements in the continual reassessment method for phase I studies. Statistics in Medicine 14, 1149–1161.

Haines, L. M., Perevozskaya, I., Rosenberger, W. F., 2003. Bayesian optimal designs for phase I clinical trials. Biometrics 59, 591–600.

Korn, E., Midthune, D., Chen, T., Rubinstein, L., Christian, M. C., Simon, R., 1994. A comparison of two phase I trial designs. Statistics in Medicine 13, 1799–1806.

McCullagh, P., Nelder, J. A., 1989. Generalized Linear Models, 2nd Edition. Chapman and Hall.

McGree, J. M., Eccleston, J. A., Duffull, S. B., 2008. Compound optimal design criteria for nonlinear models. Journal of Biopharmaceutical Statistics 18, 646–661.

Müller, P., Berry, D. A., Grieve, A. P., Smith, M., Krams, M., 2007. Simulation-based sequential Bayesian design. Journal of Statistical Planning and Inference 137, 3140–3150.

O'Quigley, J., Pepe, M., Fisher, L., 1990. Continual reassessment method: a practical design for phase I clinical trials in cancer. Biometrics 46, 33–48.

O'Quigley, J., Reiner, E., 1998. A stopping rule for the continual reassessment method. Biometrika 85, 741–748.

Ratain, M., Mick, R., Schilsky, R. L., Siegler, M., 1993. Statistical and ethical issues in the design and conduct of phase i and ii clinical trials of new anti-cancer agents. Journal of the National Cancer Institute 85, 1637–1643.

Robbins, H., Monro, S., 1951. A stochastic approximation method. Annals of Mathematical Statistics 29, 400–407.

Rosenberger, W. F., Haines, L. M., 2002. Competing designs for phase I clinical trials: A review. Statistics in Medicine 21, 2757–2770.

Thall, P. F., Nguyen, H. Q., Estey, E. H., 2008. Patient-specific dose finding based on bivariate outcomes and covariates. Biometrics 64, 1126–1136.

Tierney, L., 1994. Markov chains for exploring posterior distributions. Annals of Statistics 22, 1701–1762, with discussion and a rejoinder by the author.

Tsutakawa, R. K., 1980. Selection of dose levels for estimating a percentage point of a logistic quantal response curve. Applied Statistics 29, 25–33.

Whitehead, J., Brunier, H., 1995. Bayesian decision procedures for dose determining experiments. Statistics in Medicine 14, 885–893.

Whitehead, J., Williamson, D., 1998. Bayesian decision procedures based on logistic regression models for dose-finding studies. Journal of Biopharmaceutical Statistics 8, 445–467.

Whitehead, J., Zhou, Y., Mander, A., Ritchie, S., Sabin, A., Wright, A., 2006a. An evaluation of Bayesian designs for dose-escalation studies in healthy volunteers. Statistics in Medicine 25, 433–445.

Whitehead, J., Zhou, Y., Patterson, S., Webber, D., Francis, S., 2001. Easy-to-implement Bayesian methods for dose-escalation studies in healthy volunteers. Biostatistics 2, 47–61.

Whitehead, J., Zhou, Y., Stevens, J., Blakey, G., Price, J., Leadbetter, J., 2006b. Bayesian decision procedures for dose-escalation based on evidence of undesirable events and therapeutic benefit. Statistics in Medicine 25, 37–53.

Zhou, Y., Whitehead, J., 2003. Practical implementation of Bayesian dose escalation procedure. Drug Information Journal 37, 45–59.

Zhou, Y., Whitehead, J., Bonvini, E., Stevens, J. W., 2006. Bayesian decision procedures for binary and continuous bivariate dose-escalation studies. Pharmaceutical Statistics 5, 125–133.

Zhou, Y., Whitehead, J., Korhonen, P., Mustonen, M., 2008. Implementation of a Bayesian design in a dose-escalation study of an experimental agent in healthy volunteers. Biometrics 64, 299–308.