Statistics in Medicine

WILEY

Incorporating adverse event relatedness into dose-finding clinical trial designs

Research Article

Incorporating adverse event relatedness into dose-finding clinical trial designs

Darsy Darssan, Mery H. Thompson and Anthony N. Pettitt*

Dose-finding designs estimate the dose level of a drug based on observed adverse events. Relatedness of the adverse event to the drug has been generally ignored in all proposed design methodologies. These designs assume that the adverse events observed during a trial are definitely related to the drug, which can lead to flawed dose level estimation. We incorporate adverse event relatedness into the so called continual reassessment method. Adverse events which have 'doubtful' or 'possible' relationships to the drug are modeled using a two-parameter logistic model with an additive probability mass. Adverse events 'probably' or 'definitely' related to the drug are modeled using a cumulative logistic model. To search for the maximum tolerated dose, we use the maximum estimated toxicity probability of these two adverse event relatedness categories. We conduct a simulation study which illustrates the characteristics of the design under various scenarios. This article demonstrates adverse event relatedness is important for improved dose estimation. It opens up further research pathways into continual reassessment design methodologies.
Copyright © 2013 John Wiley & Sons, Ltd.

Keywords: adverse event relatedness; continual reassessment method; cumulative logistic; dose-finding; ordinal toxicity; phase I

1. Background

Dose-finding trials are designed to determine the tolerable dose of an investigational new drug (IND) known as the maximum tolerated dose (MTD). Designs using statistical models, including the continual reassessment method (CRM) [1], have been proven to be efficient in determining the MTD [2]. The CRM was invented over two decades ago and has generated extensive literature, much of which involves modifications made to adopt the CRM methodology under various circumstances [3]. Since the invention of the CRM, several model-based designs have been proposed. These include the Bayesian decision-theoretic designs by Whitehead and Brunier [4], escalation with overdose control by Babb *et al.* [5], isotonic design by Leung and Wang [6], and Bayesian c-optimal design by Haines *et al.* [7]. All proposed design methodologies, including the CRM, only take into account the adverse events which are definitely related to the IND. Adverse events which have possible or doubtful relationships to the IND are either ignored or wrongly included in the design as definitely related to the IND, which implies the estimated MTD is potentially inaccurate.

Mathematical Sciences School, Science and Engineering Faculty, Queensland University of Technology, Brisbane, Queensland, Australia
* Correspondence to: Anthony N. Pettitt, Professor of Statistics, Mathematical Sciences School, 2 George Street, Brisbane, QLD-4000, Australia.
*E-mail: a.pettitt@qut.edu.au

Statistics
in Medicine

If an adverse event is observed during a clinical trial then the relationship of that adverse event to the IND must be thoroughly identified [8, 9]. A phase I trial conducted in developing a new drug for rheumatoid arthritis at Roche pharmaceutical company reports adverse event relatedness on a 0 to 3 ordered scale, with a relatedness score of 0 being 'unrelated', 1 is 'remotely', 2 is 'possibly' and 3 is 'probably' related to the IND. This phase I trial motivates the need for incorporating adverse event relatedness into design methodology. A conventional classification of adverse event relatedness is 'doubtful', 'possible', 'probable', and 'definite' [9]. Incorporating this classification into dose-finding design methodology would yield better MTD estimation.

Adverse events which have possible or doubtful relationships to the IND may occur due to an underlying disease, environmental or toxic factors, or other drug or therapy. These background adverse reactions are produced at some level when the dose of the IND is zero. Statistical models used to describe dose-toxicity relationships based on possibly or doubtfully related adverse events should account for such background adverse reactions. Possibly or doubtfully related adverse events have a vague relationship to the IND. Since additional severity information about these adverse events is superfluous, they can be modelled as binary outcomes: presence or absence of dose-limiting toxicity (DLT). Generally CRM type designs describe dose-toxicity relationships using a one-parameter model [10]. However, two-parameter logistic models are flexible and parsimonious. The CRM is a sequential design in which the overall concept is not model fitting, but efficient MTD estimation [11]. Piantadosi *et al.* [12] describe a CRM type dose-finding design using a two-parameter logistic model. Their design requires minimal prior data, and therefore, toxicity information of a low and a high toxic dose level. They use maximum likelihood estimation to estimate model parameters. Their design is efficient, flexible and pragmatic. Incorporating adverse event relatedness into their design would be appropriate whenever toxicity is identified as a binary outcome.

Probably or definitely related adverse events are closely associated with the IND. Incorporating severity of these adverse events into the dose-toxicity model will bring forth more information about low and high grade toxicities. If toxicity is recorded as an ordinal outcome a two-parameter logistic model is insufficient to incorporate more toxicity information. Several attempts have been made to incorporate ordinal toxicity outcomes into design methodology. Notably, Bekele and Thall [13] use a multivariate ordinal probit regression model and Yuan *et al.* [14] use a quasi-Bernoulli likelihood to incorporate ordinal toxicity outcomes. These designs are complicated, computationally challenging and barely comply with dose-finding designs in practice. Recently, Iasonos *et al.* [15] present a multistage CRM type design which incorporates low grade toxicity into the CRM design methodology. Their approach amalgamates higher grade toxicities and therefore disaggregated information on those toxicities is lost. Meter *et al.* [16] use a cumulative logistic model to describe the relationship between dose and ordinal toxicity. A cumulative logistic model with different intercept parameters corresponding to each toxicity grade is capable of providing additional information on low and high grade toxicities. In addition, a simple binary DLT cut-off (e.g., toxicity grade of 3 or more is a DLT) can be used to estimate dose levels.

1.1. A Motivating Trial for the New Design

This motivating trial is based on the dose-finding trial of AZD8330, conducted in patients with advanced malignancies [17]. The primary objective of this trial was to estimate the MTD of AZD8330 while assessing the safety profile of AZD8330. Ten test dose levels were used in the trial. The highest dose level was selected as the MTD using some up-and-down rules with a sample size of 82 patients. This dose-finding trial does not define toxicity quantitatively and neither prior data nor statistical models were used. Furthermore, adverse event relatedness was collected during the trial but not included in the design methodology. Based on the information published by Cohen *et al.* [17], we discuss the need to incorporate adverse event relatedness in the design of the AZD8330 trial.

This trial does not incorporate adverse event relatedness in the design methodology, as with all other trials currently in practice. A dose-limiting toxicity was defined as any toxicity possibly related to the IND or any toxicity identified as significant to the IND by the investigator. Such a shallow definition discards adverse events which have a doubtful

Darssan, Thompson, and Pettitt

relationship to AZD8330. Furthermore, possibly and probably related adverse events are used in the trial with equal consideration. The trial was conducted on 82 patients, of whom 72 patients experienced adverse events. However, only 47 patients had adverse events related to the IND. If the trial had categorised adverse events into doubtful, possible, probable, and definite, would investigators identify more adverse events as related to the the IND?

The AZD8330 trial reports the severity of adverse events as a binary outcome; grade 3 or 4, and all other grades. Out of 72 patients who experienced adverse events, only 32 patients had grade 3 or 4 adverse events. However, only four of these adverse events were considered to be related to AZD8330. Cohen *et al.* [17] mention that ten patients dropped out of the trial due to adverse events. Patient drop outs might have been omitted if adverse events were specified using toxicity probability and severity of adverse events were kept as ordinal outcomes. Severity of adverse events would provide more information on low and high grade adverse events and, therefore, it is unlikely that higher toxic dose levels would be assigned to the ten patients.

The IND is a unique mitogen-activated, extracellular signal-regulated kinase half (MEK 1/2) inhibitor. Cohen *et al.* [17] mention that AZD8330 inhibits approximately 90% of tumour growth in xenograft models at a tolerated once daily dose level of 1.0 mg/kg. Therefore, toxicity produced at the dose level of 1.0 mg/kg can be used to determine prior toxicity data of a low and a high toxic dose level. If the prior data had been constructed then a feasible starting dose would have been estimated. In addition, when enrolling patients sequentially, observed data is insufficient to estimate a dose for the first few cohort of patients. Prior data would have played a key role in estimating a dose for the first few cohort of patients, and later their role might have been diminished using prior data weights.

There were no statistical considerations in designing the trail. The adverse event relatedness was recorded but not included in the trial. The severity of adverse events were recorded but amalgamated into two categories. Adequate prior data were readily available to be used in the trial but ignored. A new design methodology with incorporated adverse event relatedness, severity of toxicity, and prior toxicity information is indispensable for a better MTD estimation of the dose-finding trial of AZD8330.

In this paper we present a novel dose-finding design, namely an adverse event relatedness continual reassessment method (Aerd-CRM), which incorporates adverse event relatedness to the IND into the design methodology. To describe the dose-toxicity relationship generated by doubtfully or possibly related adverse events, we use a two-parameter logistic model with a probability mass for background adverse reactions which causes those doubtfully or possibly related adverse events. We use a cumulative logistic model to describe the dose-toxicity relationship which is generated by probably or doubtfully related adverse events. Section 2 describes the Aerd-CRM design methodology, the binary CRM and the ordinal CRM. In Section 3, we present a simulation study which shows the characteristics of Aerd-CRM and a comparison made with the binary CRM and the ordinal CRM. Section 4 shows a single Aerd-CRM trial example which illustrates step by step MTD estimation. We conclude with our findings and limitations of Aerd-CRM in Section 5.

2. Methods

Relatedness of an adverse event to the investigational new drug (IND) is generally collected in dose-finding trials conducted in chronic diseases. However, adverse event relatedness has been ignored in dose-finding design methodology. Although several dose-finding design methodologies have been invented, the so called CRM has generated an extensive literature, proven to be robust in estimating MTD and is in the era of converting to practical circumstances [18]. This section presents an extension of CRM methodology which incorporates adverse event relatedness.

Design methodologies described in this section share the following initial design parameters: target toxicity probability ϕ, number of test dose levels K and sample size N. Values selected for these parameters should satisfy the sample size constraint given by Cheung [19],

$$\frac{N - 3(K - 2)}{2} > \frac{1}{\phi}. \tag{1}$$

Statistics
in Medicine
Darssan, Thompson, and Pettitt

The two crucial design components, 'prior data' in toxicity and 'dose-toxicity model' vary for each design methodology presented in this paper. In this paper the term prior data refers to data points that represent clinicians' expert belief or information obtained from pre-clinical studies. For instance, we might posit that two doses, dose = (200 mg, 3000 mg) produce toxicity outcome, toxicity = (0.10, 0.90). Throughout this paper 'pure likelihood' based CRM-type designs, as described by Piantadosi *et al.* [12], are discussed. Therefore, the term prior data should not be confused with the term 'prior distribution' often used as information for statistical model parameters under Bayesian CRM. Prior data of a low and a high dose can be obtained from pre-clinical studies or drugs which have a toxicity profile similar to the IND. If there is a reason to have prior toxicity data of low and high doses then the amount of information prior data carries as the trial progresses should be reasonably diminished. Section 2.4 discusses weighting schemes for prior data. Specification of prior toxicity data is described under each design methodology section in detail.

In this paper we use logistic models to describe the dose-toxicity relationship. Dose-toxicity model(s) used are shown under each design methodology section. We use maximum likelihood estimation to estimate model parameters. The CRM is a sequential design in which model parameter values are updated after toxicity outcomes of newly entered patients are known. Based on a fitted model, the dose level which has toxicity probability closer to the target toxicity probability is selected for the following cohort of patients. To ensure patient safety, Section 2.5 describes a safety constraint and a stopping rule.

If toxicity is observed as the binary outcome presence or absence of DLT we use binary (Bin) CRM. The Bin-CRM described in Section 2.1 is identical to the CRM type design described by Piantadosi *et al.* [12]. When toxicity is observed as an ordinal outcome, Section 2.2 describes Ordinal (Ord) CRM. The Ord-CRM is similar to the design described by Meter *et al.* [16]. Section 2.3 describes a design methodology which incorporates adverse event relatedness into the dose-finding design methodology.

2.1. Binary CRM

Adverse event U is observed as a binary (0,1) outcome, where 0 is a 'non DLT' and 1 is a 'DLT'. The design methodology for binary CRM is as follows.

1. Determine the initial CRM design parameters: ϕ, K and N, and verify the sample size constraint in equation (1).
2. Obtain prior toxicity data of a low and a high dose level. For instance, 200 mg produces 5% DLTs and 3000 mg produces 85% DLTs. Decide on the amount of information prior data carries as the trial progresses (Section 2.4).
3. Specify the dose-toxicity model in equation (2) and estimate the model parameters α and β using maximum likelihood estimation.

$$\psi(d_i, \alpha, \beta) = \Pr(U = 1 | d_i) = \frac{\exp(\alpha + \beta d_i)}{1 + \exp(\alpha + \beta d_i)}. \tag{2}$$

4. Implement the safety constraint and the stopping rule in Section 2.5. Use the inequality in (3) to find the dose level (d_l) which has toxicity probability closest to the target toxicity probability.

$$|\psi(d_l, \hat{\alpha}, \hat{\beta}) - \phi| < |\psi(d_i, \hat{\alpha}, \hat{\beta}) - \phi|, \quad i = 1, \ldots, K, \ d_i \neq d_l. \tag{3}$$

5. Update the values of $\hat{\alpha}$ and $\hat{\beta}$ once toxicity outcomes of the newly entered cohort of patients are known.
6. The dose level selected for the final cohort of patients is the MTD. The MTD is a dose with estimated $\Pr(U = 1)$ closest to the target toxicity probability.

2.2. Ordinal CRM

We assume that adverse event U is observed on a 1 to 4 ordered scale, with adverse event scores of 1 to 4 being 'mild', 'moderate, 'severe' and 'life-threatening' adverse events, respectively. The design of an ordinal CRM trial proceeds as follows.

Statistics
in Medicine

Darssan, Thompson, and Pettitt

1. Determine the initial CRM design parameters: ϕ, K and N, and verify the sample size constraint in equation (1).

2. Decide on adverse event scores which are 'dose-limiting'. For instance, adverse event scores of 3 and 4 are 'dose-limiting'.

3. Obtain prior toxicity data of a low and a high dose level. For instance, 200 mg is 10% toxic with 75% of these toxic events being mild, 20% moderate, 4% severe and 1% life-threatening adverse events. Similarly, 3000 mg is 90% toxic with 5% of these toxic events being mild, 10% moderate, 35% severe and 50% life-threatening adverse events. Decide on the amount of information prior data carries as the trial progresses (Section 2.4).

4. The cumulative logistic model in (4) describes the dose-toxicity relationship. Estimate the model parameters γ_j and δ using maximum likelihood estimation.

$$\psi(d_i, \gamma_j, \delta) = \Pr(U \geq j | d_i) = \frac{\exp(\gamma_j + \delta d_i)}{1 + \exp(\gamma_j + \delta d_i)}, \quad j = 2, 3, 4. \tag{4}$$

5. Implement the safety constraint and the stopping rule in Section 2.5. Use the inequality in (5) to find the dose level (d_l) which has toxicity probability closest to the target toxicity probability.

$$|\psi(d_l, \hat{\gamma_3}, \hat{\delta}) - \phi| < |\psi(d_i, \hat{\gamma_3}, \hat{\delta}) - \phi|, \quad i = 1, \ldots, K, \ d_i \neq d_l. \tag{5}$$

6. Update the values of $\hat{\gamma_j}$ ($j = 2, 3, 4$) and $\hat{\delta}$ once toxicity outcomes of the newly entered cohort of patients are known.

7. The dose level selected for the final cohort of patients is the MTD. The MTD is a dose with estimated $\Pr(U \geq 3)$ closest to the target toxicity probability.

2.3. Adverse Event Relatedness CRM

Adverse event relatedness CRM (Aerd-CRM) incorporates relatedness of adverse events into the CRM. As in Section 2.2, adverse event U is observed on an ordinal scale of 1 to 4, where adverse event scores of 1 to 4 correspond to 'mild', 'moderate', 'severe' and 'life-threatening' adverse events, respectively. Adverse event relatedness V is categorised into two groups, where a score of 1 is an adverse event doubtfully or possibly related to the IND and 2 is an adverse event probably or definitely related to the IND. The distribution of adverse events has two dimensional structure. Grids in Figure 1 can be looked at as a visual guide. The Aerd-CRM design is implemented as follows.

1. Determine the initial CRM design parameters: ϕ, K and N and verify the sample size constraint in equation (1).

2. Decide on adverse events which are dose-limiting. For instance, severe and life-threatening adverse events which are probably or definitely, and life-threatening adverse events which are doubtfully or possibly, related to the IND are dose-limiting.

3. Obtain prior toxicity data of a low and a high toxic dose level. For instance, 200 mg is 10% toxic with 75% of these toxic events being mild, 20% moderate, 4% severe and 1% life-threatening adverse events. Similarly, 3000 mg is 90% toxic with 5% of these toxic event being mild, 10% moderate, 35% severe and 50% life-threatening adverse events. In addition, specify prior data about the proportion of adverse events doubtfully or possibly and probably or definitely related to the IND. For instance, 35% of adverse events are doubtfully or possibly, and 65% of adverse events are probably or definitely, related to the IND. Decide on the amount of information prior data carries as the trial progresses (Section 2.4).

4. Decide on the proportion of toxicities, p, which background adverse reactions accounts for, for doubtful or possible adverse events. For instance, doubtfully or possibly related adverse events carry 5% of toxicity from background adverse reactions.

Statistics
in Medicine

5. Specify the dose-toxicity models in (6) and (7). Estimate the model parameters using maximum likelihood estimation.

$$\psi_1(d_i, \alpha, \beta) = \Pr(U = 4 | V = 1, d_i) = p + \left(\frac{\exp(\alpha + \beta d_i)}{1 + \exp(\alpha + \beta d_i)} \right) \times (1 - p), \tag{6}$$

$$\psi_2(d_i, \gamma_j, \delta) = \Pr(U \geq j | V = 2, d_i) = \frac{\exp(\gamma_j + \delta d_i)}{1 + \exp(\gamma_j + \delta d_i)}, \quad j = 2, 3, 4. \tag{7}$$

6. Implement the safety constraint and the stopping rule in Section 2.5. Use the inequality (8) to find the dose level (d_l) which has toxicity probability closest to the target toxicity probability.

$$| \max\{\psi_1(d_l, \hat{\alpha}, \hat{\beta}), \psi_2(d_l, \hat{\gamma_3}, \hat{\delta})\} - \phi | < | \max\{\psi_1(d_i, \hat{\alpha}, \hat{\beta}), \psi_2(d_i, \hat{\gamma_3}, \delta)\} - \phi |, \quad i = 1, \ldots, K, \ d_i \neq d_l. \tag{8}$$

7. Update the estimated parameter values once toxicity outcomes of the newly entered cohort of patients are known.
8. The dose level selected for the final cohort of patients is the MTD. The MTD is a dose with maximum estimated $\Pr(U = 4 | V = 1)$ and $\Pr(U \geq 3 | V = 2)$ closest to the target toxicity probability.

The Aerd-CRM requires prior data in the toxicity probability of a low and a high dose level, severity of toxicity of those dose levels and percent toxicity relatedness. A toxicity probability of a low and a high dose level can be obtained from pre-clinical studies or from drugs which have similar pharmacological effects similar to the IND. These dose levels need not be the test dose levels. Prior information on severity of toxicity for these dose levels may be obtained based on clinicians' expert knowledge. Prior information on proportion of toxicity relatedness can simply be determined from adverse event relatedness recorded for drugs similar to the IND. If clinicians are not confident with the prior information then smaller weight should be given to the prior data as the trial progresses.

The dose-toxicity relationship in the Aerd-CRM is established using a conditional model for $\Pr(U|V)$ through equations (6) and (7). The model in equation (6) describes adverse events which are doubtfully or possibly related to the IND. The probability p added to this model accounts for the toxicity probability arising from background adverse reactions. Toxicity probability of background adverse reactions can be determined based on any common underlying disease most enrolling subjects carry, environmental or toxic factors, and other drugs or therapies given to the study subjects. The probability p is not estimated rather decided in advance. We suggest that a reasonable value for p is greater than zero and less than 0.1.

The maximum estimated ψ_1 and ψ_2 is used to select a dose level (inequality (8)). That is, for each dose level the maximum of estimated toxicity probability based on doubtfully or possibly related adverse events and probably or definitely related adverse events should be calculated in the first instance. The dose level with maximum toxicity probability closest to the target toxicity probability is selected. Instead of the maximum a weighted average or a linear combination of ψ_1 and ψ_2 could be used. However, the maximum criterion provides a conservative solution in selecting dose levels when the proportion of doubtfully or possibly, and probably or definitely related adverse events of the IND for each dose level is unknown.

2.4. Managing Prior Data

Prior data in adverse event toxicity, severity and relatedness are used to estimate the starting dose and to proceed the trial with a small amount of adverse event responses. Contribution of the prior data throughout the design is crucial and reasonable, because phase I trials are conducted in small sample sizes. However, it can be down weighted or eliminated as the trial progresses. It should be noted that if no heterogeneous outcomes are present in the current data and the prior data are completely weighted out then the dose toxicity model cannot be fitted. For instance, if prior data is eliminated before a patient is observed with a doubtfully or possibly related adverse event, then the parameters in equation (6) cannot

Darssan, Thompson, and Pettitt

be estimated. An appropriate weight for prior data can be obtained by conducting a simulation study under different weighting schemes.

Table 1. A prior data weight distribution

Cohort No.	2:1-Scheme		1:1-Scheme	
	Prior data	Observed data	Prior data	Observed data
1	66.67	33.33	50.00	50.00
2	50.00	50.00	33.33	66.67
3	40.00	60.00	25.00	75.00
4	33.33	66.67	20.00	80.00
5	28.57	71.43	16.67	83.33
6	25.00	75.00	14.29	85.71
7	22.22	77.78	12.50	87.50
8	20.00	80.00	11.11	88.89
9	18.18	81.82	10.00	90.00
10	16.67	83.33	9.09	90.91
11	15.38	84.62	8.33	91.67

Table 1 shows an example of two different weighting schemes for prior data. These weights are created with cohort of size four. The first weighting scheme, namely the '2:1-Scheme', uses two times the cohort size as the prior data weight. For instance, once the first cohort of four patients are entered into the trial the prior data has double the amount, eight. So the total weight prior data carry is equal to $\frac{8}{12}$x100%. This weighting scheme starts with over 65% weight on prior data and is gradually reduced to 25% when the sixth cohort of patients is entered into the trial. The second weighting scheme is named the '1:1-Scheme', which ensures the prior data weight is equal to the cohort size. It starts with 50% and ends with just over 8% wight for prior data.

2.5. Safety Constraint and Stopping Rule

The dose for the next patient is selected so that the maximum estimated toxicity probabilities (inequality (8)) at that dose is close to the target DLT. A safety constraint set up selects the adjacent lower dose level of an estimated dose level only if the estimated toxicity at the dose level is over c_t% higher than the target toxicity probability. This stops the selection of a dose level for the next patient that has toxicity probability larger than the target toxicity probability plus the margin c_t%. It selects the closest dose level even if it has a maximum estimated toxicity probability higher than the target toxicity probability. This constraint aids in identifying dose levels that are within a range of the target toxicity probability during the early stages of the trial.

A higher level of safety can be implemented by changing this constraint level c_t% to be equal to 0%. A high level for the safety constraint will tend to select adjacent lower dose levels more frequently and, therefore, if the estimated dose level is the lowest dose level then the trial will stop prematurely. Clinicians should decide on the toxicity level above the target toxicity probability at which patients can be treated. A reasonable value for c_t is between zero and ten. Towards the end of a trial, a higher level of safety could be implemented by changing the value of c_t to zero after a certain number of patients.

Dose-limiting adverse events are produced whenever patients are treated at high toxic dose levels. These dose-levels are given to patients based on observed toxicity data and a fraction of prior data. Considering patients' safety, a stopping rule can be set to stop the trial if recorded dose-limiting adverse events are over s_t% of all observed outcomes. If N is the planned total number of enrolled patients then the trial will stop as soon as the number of observed dose-limiting toxicities are greater than $s_t \times N$. If a small cohort size is chosen this stopping rule will stop the trial prematurely. A stopping criterion is entirely the clinician's choice. However, caution should be taken because as the trial progress, there is less possibility of stopping the trial.

Statistics
in Medicine

3. Simulation Study

We analyse characteristics of the Aerd-CRM design using a simulation study. As described in Section 2.3, determining the initial CRM design parameters, ϕ, K and N is the first step. The convention for target toxicity probability ϕ is a value between 0.20 and 0.30. A value outside this range should be verified by clinicians. In this simulation study we set ϕ to be 0.25. Generally, the number K of test dose levels is determined based on dose increments. The convention for dose increments is the modified Fibonacci sequence defined by multiplications of $2.00, 1.67, 1.50, 1.33, 1.33, \ldots$. Modification of the Fibonacci number sequence is neither verified in the literature nor is there a biological basis for its use [20]. However, testing close dose levels is unnecessary as they will show similar pharmacological properties during the later phases of the drug development process, and therefore one of them will be abandoned. In this simulation study, we use dose levels tested for a rheumatoid arthritis trial at the Roche pharmaceutical company. Dose levels are: 50 mg, 150 mg, 500 mg, 1200 mg, 1500 mg, 2000 mg, 2400 mg, and 2600 mg. The required sample size N has a monotonic relationship with the number K of test dose levels. We use 44 patients which satisfies sample size constraint (1). Patients are entered into the trial in a cohort of size four. We assume that 5% of toxicities come from background adverse reactions.

3.1. Simulation Results of Binary Adverse Event Relatedness CRM

The concept of Aerd-CRM is complicated to grasp with the presence of ordinal toxicity outcome. We show the simulation results of the simple version of the Aerd-CRM design, namely Bin-Aerd-CRM. That is, adverse events are observed as a binary outcome. Doubtfully or possibly related adverse events are modelled using (6). Probably or definitely related adverse events are modelled using a two parameter logistic model without the additive probability mass.

Suppose prior toxicity data of a low (200 mg) and a high (3000 mg) dose are known. Assume that the prior data for adverse events relating to a 200 mg dose yields 5% DLT. Similarly, 3000 mg yields 85% DLT. The relatedness score is recorded as 1 or 2, where 1 is doubtfully or possibly, and 2 is probably or definitely related to the IND. Relatedness is independent of doses, but it dependents on the IND. Assume that adverse events are 35% doubtfully or possibly, and 65% probably or definitely, related to the IND.

Table 2 summarizes the operating characteristics of the Bin-Aerd-CRM. Under each scenario, the first two rows of the table represent the true probabilities of toxicities. These probabilities are monotonically increasing with the dose. However, due to rounding error, they may look equal at some consecutive dose levels. The third row corresponds to the selection percentage of each dose level using the Bin-Aerd-CRM. The fourth row is the number of patients treated at each dose level of completed trials averaged over 10000 simulated trials. Furthermore, this table shows percentage of inconclusive trials (denoted as 'Early Stop') due to safety constraint with $c_t = 10\%$ and stopping rule with $s_t = 75\%$ as explained in Section 2.5. Prior data were diminished using the 1:1 weighting scheme explained in Section 2.4.

In scenario 1, toxicity probabilities for doubtful or possible adverse events are low in the first three dose levels. It jumps to around 30% at 1200 mg where it dramatically increases to just over 85% at 2600 mg. Toxicity probability of probable or definite adverse events are 15% at 500 mg with a jump to just under 50% at 1200 mg, followed by a steady increase to 96% at 2600 mg. Over 63% of the simulated trials select the targeted dose, 500 mg, as the MTD. In each simulated trial, on average, over 20 patients were treated at this dose level.

In scenario 2, 1200 mg is the targeted dose level. Under this scenario, over 46% of trials correctly select this dose level as the MTD. However, around 16%-17% of trials select 1500 mg as the MTD, which is the adjacent higher dose level to the targeted dose level of 1200 mg. However, only 7 patients were treated at 1500 mg, whereas more than twice (18 patients) this were treated per simulated trial at the targeted dose level of 1200 mg.

Scenario 3 targets a dose level of 1500 mg. Probably or definitely related adverse events have the highest toxicity probability (0.21) at the targeted dose level. Just under 60% of simulated trials selects 1500 mg at the MTD. A dose level of 1200 mg is the adjacent lower toxic dose level to the targeted MTD. Over 15% of trials select this dose level as the MTD. About 20 patients per trial were treated at 1500 mg, where as only 10 patients per trial were treated at the adjacent

Darssan, Thompson, and Pettitt

Table 2. Percent selecting a dose level as MTD: Bin-Aerd-CRM

	50 mg	150 mg	500mg	1200 mg	1500 mg	2000 mg	2400 mg	2600 mg	Early Stop
Scenario 1									
$\Pr(U=1\|V=1)$	0.08	0.09	0.12	0.29	0.42	0.66	0.82	0.87	
$\Pr(U=1\|V=2)$	0.06	0.07	0.15	0.48	0.65	0.85	0.94	0.96	
Selected %	6.60	8.25	**63.21**	7.33	0.12	0.00	0.00	0.00	14.49
Patients per trial	4.79	3.56	20.13	9.20	0.93	0.50	0.00	0.00	
Scenario 2									
$\Pr(U=1\|V=1)$	0.08	0.09	0.11	0.28	0.42	0.67	0.82	0.88	
$\Pr(U=1\|V=2)$	0.00	0.00	0.01	0.16	0.40	0.85	0.97	0.99	
Selected %	5.00	0.41	11.41	**46.15**	16.83	0.07	0.00	0.00	20.13
Patients per trial	5.02	0.30	4.56	18.00	7.87	2.04	0.00	0.00	
Scenario 3									
$\Pr(U=1\|V=1)$	0.05	0.05	0.06	0.08	0.12	0.26	0.48	0.60	
$\Pr(U=1\|V=2)$	0.00	0.00	0.01	0.05	0.21	0.80	0.97	0.99	
Selected %	6.51	0.03	13.72	16.17	**58.82**	0.88	0.00	0.00	3.87
Patients per trial	2.72	0.02	5.36	10.84	19.68	4.10	0.04	0.00	
Scenario 4									
$\Pr(U=1\|V=1)$	0.04	0.05	0.06	0.08	0.11	0.22	0.39	0.49	
$\Pr(U=1\|V=2)$	0.00	0.00	0.01	0.04	0.08	0.25	0.50	0.64	
Selected %	6.73	0.01	9.25	8.44	20.88	**46.93**	2.84	0.12	4.80
Patients per trial	2.86	0.01	4.15	5.88	8.96	17.57	2.84	0.24	
Scenario 5									
$\Pr(U=1\|V=1)$	0.05	0.05	0.05	0.07	0.08	0.12	0.19	0.23	
$\Pr(U=1\|V=2)$	0.00	0.00	0.01	0.02	0.04	0.11	0.24	0.34	
Selected %	5.19	0.00	13.54	8.30	1.88	28.28	**28.90**	11.54	2.37
Patients per trial	2.19	0.00	5.59	5.05	2.69	13.75	9.89	4.11	
Scenario 6									
$\Pr(U=1\|V=1)$	0.07	0.07	0.07	0.10	0.11	0.15	0.19	0.22	
$\Pr(U=1\|V=2)$	0.00	0.00	0.00	0.02	0.03	0.08	0.18	0.25	
Selected %	6.87	0.00	8.98	7.99	1.12	18.31	27.34	**25.72**	3.67
Patients per trial	2.79	0.01	4.18	4.99	2.22	11.62	10.06	7.00	

lower dose level of 1200 mg.

Scenario 4 targets 2000 mg, in which $\Pr(U=1|V=1)$ slowly increases to 11% at 1500 mg, followed by a doubling at 2000 mg and with a steady increase to 49% at 2600 mg. Toxicity probability of probably or definitely related adverse events has similar pattern. Around 47% of simulated trials select the targeted dose as the MTD. At this dose level, about 17 patients per trial were treated. The adjacent lower dose level of 1500 mg was selected in 20% of trials, in which just 8 patients per trial were treated.

Scenarios 5 and 6 have the target at dose levels of 2400 mg and 2600 mg respectively. In these two scenarios Bin-Aerd-CRM selects the adjacent lower dose levels with a similar probability as the target dose levels. Scenario 5 selects 2000 mg and 2400 mg in just over 28% of trials. More patients (13) were treated at 2000 mg than at the target dose level of 2400 mg. In scenario 6, the adjacent lower dose level to the target dose is selected in just over 27% of trials, where as, over 25% of trials select the target dose level. These two scenarios shows that if low dose levels are ineffective and the true MTD is at the upper end of high dose levels then it is highly likely that the design would select adjacent lower dose level of the true MTD.

The stopping rule and the safety constraint stopped less than 5% of trials in scenarios 3, 4, 5 and 6. In scenarios 1 and 2, about 15% and over 20% trials are stopped early respectively. These scenarios target third and fouth dose levels respectively, and therefore, due to safety constraint, more trials were stopped early.

The scenarios presented above demonstrate that adverse event relatedness incorporated into the design methodology with binary toxicity outcome often correctly selects the targeted dose level. The design's performance could be increased if probably or definitely related adverse events are incorporated in an ordinal fashion.

Statistics
in Medicine
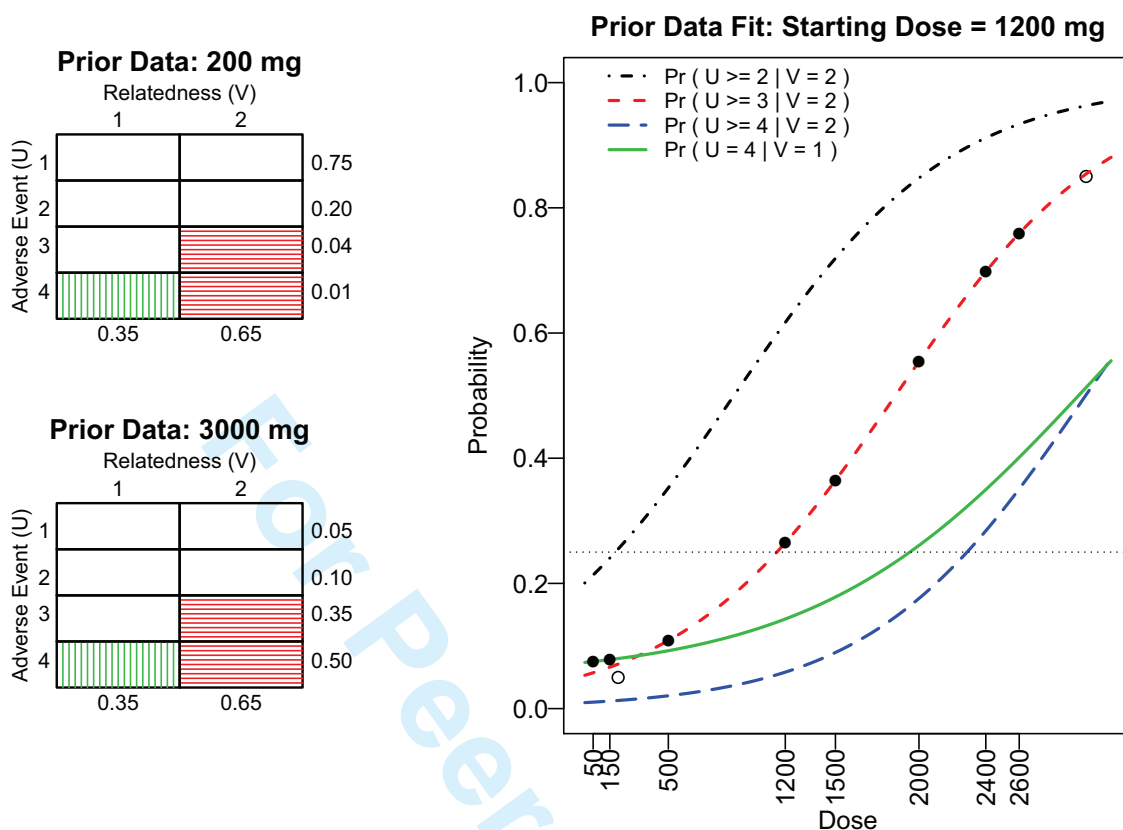
Darssan, Thompson, and Pettitt

Figure 1. From left to right: prior data of a low toxic (200 mg) and a high toxic (3000 mg) dose level, and prior data fit with starting dose of 1200 mg.

3.2. Simulation Setup for Adverse Event Relatedness CRM

We present a simulation study for Aerd-CRM, which highlights the inclusion of ordinal toxicity outcome into the deign methodology. Figure 1 shows a prior toxicity data of a low toxic (200 mg) and a high toxic (3000 mg) dose level, and the models (6) and (7) fitted to the prior data. Adverse event responses are denoted by U. Severity of an adverse event is observed on an ordinal scale of 1 to 4, where adverse event scores of 1 to 4 correspond to mild, moderate, severe and life-threatening adverse events, respectively. The expected toxicity percent is shown at the right side of each grid of Figure 1. The prior data for adverse events relating to a 200 mg dose level yields 75% mild, 20% moderate, 4% severe and 1% life-threatening adverse events. Similarly, 3000 mg yields 5% mild, 10% moderate, 35% severe and 50% life-threatening adverse events.

The adverse event relatedness is denoted by V. The relatedness is recorded as 1 or 2, where 1 is an adverse event which is doubtfully or possibly related to the IND, and 2 is an adverse event which is probably or definitely related to the IND. Prior data for relatedness percent is shown at the bottom of each grid of Figure 1. Relatedness is independent of dose levels, but it depends on the IND. We assume that 35% of adverse events are 'doubtfully or possibly' and 65% are probably or definitely related to the IND. The decision on dose-limiting for this simulation is shown as shaded boxes. The adverse event scores of 3 and 4 with relatedness score of 2, and adverse event score of 4 with related score of 1, are defined as dose-limiting. In other words, severe and life-threatening adverse events which are probably or definitely related to the IND, and life-threatening adverse events which are doubtfully or possibly related to the IND, are defined as dose-limiting.

A dose level estimated based on these prior data is shown in the rightmost of Figure 1. Models (6) with $p = 0.05$ and (7) are fitted with $\alpha = -3.664552$, $\beta = 0.001205$, $\gamma_2 = -1.381882$, $\gamma_3 = -2.878874$, $\gamma_4 = -4.641482$ and $\delta = 0.001549$.

Darssan, Thompson, and Pettitt

The maximums of the estimated toxicity probabilities, $\Pr(U = 4|V = 1)$ and $\Pr(U \geq 3|V = 2)$, for each dose level are shown as solid circles. We implement the safety constraint discussed in Section 2.5 with a 10% margin. This means that if the estimated toxicity at the starting dose level is 10% higher than the target DLT then the adjacent lower dose level is selected as the starting dose. Based on the prior data fit, the first cohort of patient will be treated at 1200 mg.

Figure 2 shows six scenarios set for simulation. Scenario 7 in the top left panel targets 500 mg as the MTD. In this scenario, the toxicity probability of adverse events doubtfully or possibly related to the IND is higher in first two dose levels than probable or doubtful adverse events. Scenario 8 in the right top panel targets 1200 mg as the MTD in which adverse events doubtfully or possibly related to the IND play the key role in estimating the targeted MTD. The graph in the middle left panel shows scenario 9, which targets 1500 mg as the MTD. This scenario is set with the toxicity probability of adverse events probably or definitely related to the IND being lower than adverse events doubtfully or possibly related to the IND in the first four dose levels, and the converse is true in last four dose levels. Scenario 10 is set to observe the properties of Aerd-CRM in selecting a high toxic dose level (2000 mg) as the MTD. Scenarios 11, which targets 2400 mg, is created such that first five dose levels have low toxic probabilities. Scenario 12 is created such that all dose levels have toxic probabilities less than the target toxicity probability. Scenario 11 and 12 are created to observe the characteristics of Aerd-CRM whenever some or all of the selected dose levels are ineffective.

We have simulated ten thousand trials for each scenario under each weighting scheme discussed in Section 2.4. Every trial in each scenario uses the initial parameters discussed in this section, begins with the same starting dose, and is subject to the safety constraint with $c_t = 10\%$ and stopping rule with $s_t = 75\%$, as discussed in Section 2.5. In addition, for each scenario using the data simulated for Aerd-CRM we simulate Ord-CRM in Section 2.2 and Bin-CRM in Section 2.1. This will bring forth the importance of incorporating adverse event relatedness into the CRM. The Ord-CRM and the Bin-CRM ignore adverse event relatedness to the IND, which implies that the data simulated for Aerd-CRM should be combined by ignoring the relatedness score.

For Ord-CRM, simulated life-threatening adverse events which are doubtfully or possibly related to the IND were recoded as severe adverse events. All other doubtfully or possibly related adverse events were recoded as mild adverse events. Similarly, a dose-limiting toxicity in Aerd-CRM (shaded cells in the grids of Figure 1) is treated as dose-limiting in Bin-CRM. Therefore, the DLT criteria in Aerd-CRM is identical to both Ord-CRM and Bin-CRM, which allows comparison.

Since the Ord-CRM and the Bin-CRM ignore adverse event relatedness to the IND in each scenario, the targeted MTD of those designs may be higher than the targeted MTD of Aerd-CRM. Sensitivity of scenario MTD to adverse event relatedness ($\Pr(V = 1)$ and $\Pr(V = 2)$) can be determined using inequality (9). Let $\tilde{U} \in \{1, \ldots, 4\}$ be the adverse event score of Ord-CRM. Inequality (5), which selects dose d_l for the following cohort of patients in the Ord-CRM, will become

$$|\Pr(\tilde{U} \geq 3|d_l) - \phi| < |\Pr(\tilde{U} \geq 3|d_i) - \phi|, \quad i = 1, \ldots, K, \; d_i \neq d_l, \tag{9}$$

$$\text{where,} \;\; \Pr(\tilde{U} \geq 3|d_i) = \Pr(U \geq 3|V = 2, d_i) \times \Pr(V = 2) + \Pr(U \geq 4|V = 1, d_i) \times \Pr(V = 1).$$

Because inequality (9) is averaged over V, the toxicity probability of dose levels are lower than those calculated in Aerd-CRM using inequality (8). However, toxicity probability of the targeted MTD calculated using both inequality (8) and inequality (9) are closer to the target toxicity probability for each scenario shown in this paper. That is, the target MTD of each scenario for Aerd-CRM and Ord-CRM remains unchanged. Similarly, the target MTD for Bin-CRM and Aerd-CRM is identical for those scenarios shown in this paper.

3.3. Simulation Results of Adverse Event Relatedness CRM

Table 3 displays the percentage of trials selecting a dose level as the MTD based on 10000 simulated trials for each design and weighting scheme. The Ord-CRM and Bin-CRM, which ignore adverse event relatedness, were also simulated under each of the prior data weighting schemes. The percent of trials of a particular dose level being selected for targeted MTD

Statistics
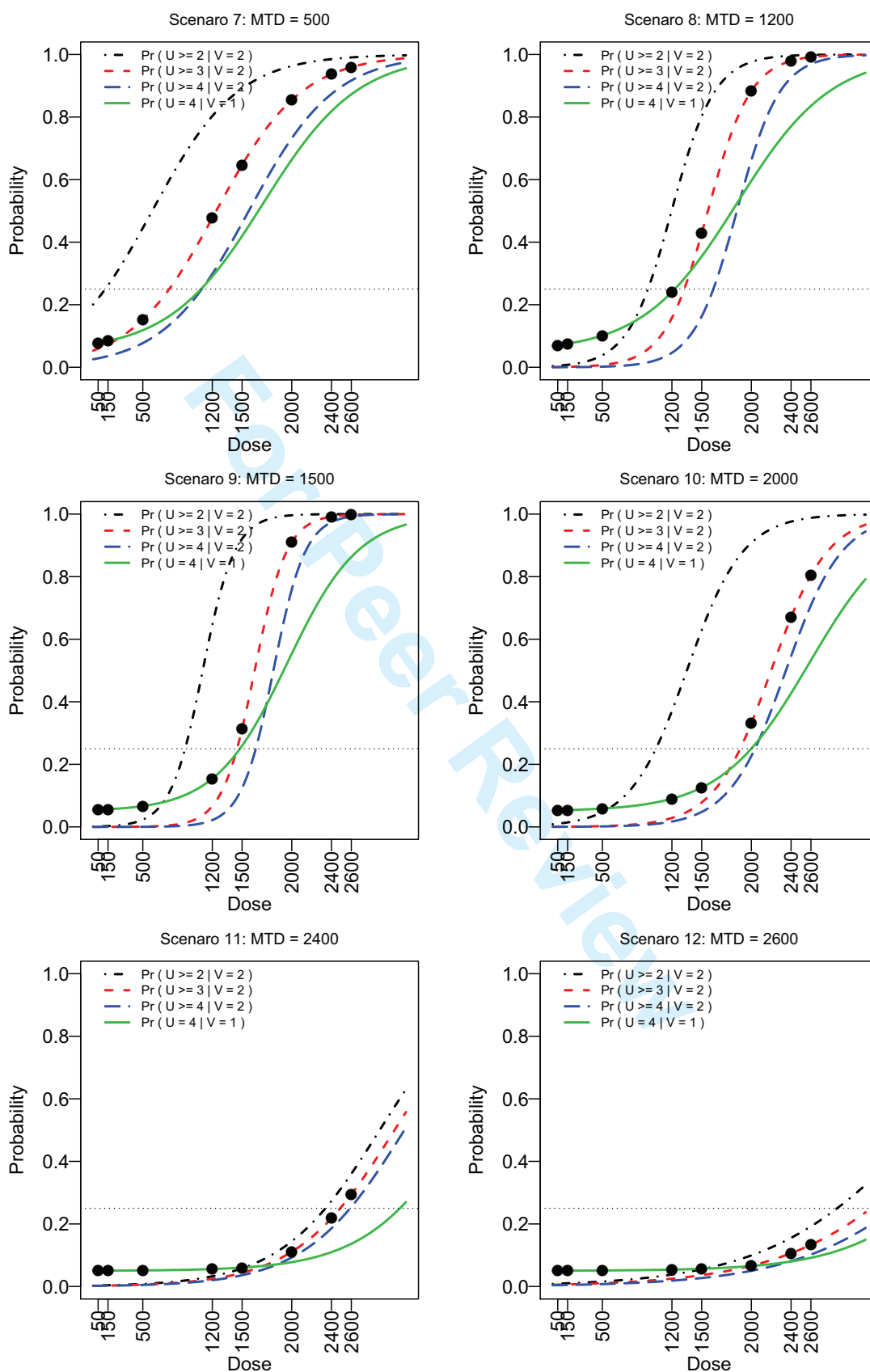# in Medicine

Darssan, Thompson, and Pettitt

Figure 2. Scenarios created for the purpose of simulation of Aerd-CRM

Darssan, Thompson, and Pettitt

of each design under each scenario is shown in bold numbers. Dose levels 50 mg and 150 mg are very low toxic dose levels. In all scenarios, these dose levels were selected by Aerd-CRM a few times. These dose levels may not be necessary in a trial if it is believed 200 mg produces only 5% DLT (Figure 1). Both dose levels of 2400 mg and 2600 mg are very high toxic doses in scenarios 7, 8, 9 and 10. In these scenarios, none of the simulated Aerd-CRM trials select these dose levels as the MTD. Both Bin-CRM and Ord-CRM did select these extremely high does levels in scenarios 7 to 10, under either one or both weighting schemes. In contrast, scenarios 11 and 12 are set such that, dose levels of 2400 mg and 2600 mg have the toxicity probabilities closer to the target toxicity probabilities.

Scenario 7 targets a dose level of 500 mg as the MTD. Over 65% of simulated Aerd-CRM trials correctly select this dose level as the MTD compared to less than 20% of Bin-CRM and Ord-CRM trials. Additionally, under scenario 7, Bin-CRM and Ord-CRM incorrectly select dose levels higher than the targeted MTD more often than Aerd-CRM. For example, over 50% of Bin-CRM and Ord-CRM trials incorrectly select 1200 mg as the MTD under the 2:1 weighting scheme compared to less than 25% for Aerd-CRM. Doses of 1500 mg and 2000 mg are highly toxic doses (probability of toxicity $> 60\%$, see Figure 2). In scenario 7, Bin-CRM and Ord-CRM incorrectly select 1500 mg as the MTD in over 30% of trials under the 1:1 weighting scheme compared to less than 1% for Aerd-CRM. These results suggest that designs which ignore adverse event relatedness under scenario 7 circumstances select a higher MTD, putting patients at a higher risk.

In scenario 8, 1200 mg is the targeted dose level. Under this scenario, over 69% of Aerd-CRM trials correctly select this dose level as the MTD compared to less than 43% of Bin-CRM and Ord-CRM trials. As with scenario 7, Bin-CRM and Ord-CRM incorrectly select dose levels higher than the targeted MTD more often than Aerd-CRM, indicating that Aerd-CRM is, again, the superior design and is more effective in minimising patient risk. Note, however, that under scenario 8, around 11%-12% of Aerd-CRM trials select 500 mg as the MTD, which is below the targeted dose level of 1200 mg, compared to less than 1% of Bin-CRM and Ord-CRM trials. This suggests that Aerd-CRM may be too risk-aversive in a small percentage of trials.

Scenario 9 targets a high toxic dose level of 1500 mg as the MTD. Under this scenario, Bin-CRM and Ord-CRM outperform Aerd-CRM under the 2:1 weighting scheme. Here, over 75% of Bin-CRM and Ord-CRM trials correctly select 1500 mg as the MTD compared to around 58% of Aerd-CRM trials. Because the toxicity probability of the next higher dose level of 2000 mg is extremely high (90% probability of toxicity, see Figure 2), both Bin-CRM and Ord-CRM select the adjacent lower dose level of 1500 mg, which is the targeted MTD, more often. Also, the toxicity probability of the targeted MTD for this scenario is 0.31. Since it is higher than the target toxicity probability of 0.25, Aerd-CRM selects the next lower dose level of 1200 mg as the MTD more often than Bin-CRM and Ord-CRM.

Scenario 10 targets 2000 mg, which has toxicity probability of 0.33 (Figure 2), as the MTD. Over 47% of Aerd-CRM trials select this dose level as the MTD. A dose level of 1500 mg is the adjacent lower toxic dose level to the targeted MTD. Over 35% of Aerd-CRM trials select this dose level as the MTD, compared to less than 19% for Bin-CRM and Ord-CRM. This indicates that Bin-CRM and Ord-CRM designs are not as conservative as Aerd-CRM in selecting the adjacent lower dose level to the targeted MTD.

In scenario 11, the first five dose levels have low toxic probabilities. This scenario observes the characteristics of Aerd-CRM whenever the selected low dose levels are ineffective. Under the 2:1 weighting scheme, over 45% of Aerd-CRM trials selects the targeted dose level of 2400 mg as the MTD, compared to over 59% of Ord-CRM trials. In addition, over 36% of Aerd-CRM trials select the adjacent lower dose level of 2000 mg, where as only 23% of Ord-CRM trials and 14% of Bin-CRM trials selects this dose level as the MTD.

Scenario 12 is set with low toxic dose levels. Since the dose-toxicity relationship is monotonically increasing, the target dose is the last dose level of 2600 mg. It is unfortunate to have all the test dose levels with toxicity probability less than the target toxicity probability. However, the majority of Aerd-CRM trials select the highest toxic dose level. Under this scenario Bin-CRM outperforms Aerd-CRM under both weighting schemes.

The results of the simulation study were, in some cases, dependent on the weights attributed to prior data, which is why we considered both a 2:1 and a 1:1 weighting schemes. There is very little difference between the two prior data weighting

Statistics
in Medicine

Darssan, Thompson, and Pettitt

Table 3. Percent selecting a dose level as MTD: Aerd-CRM

	Weight	50 mg	150 mg	500mg	1200 mg	1500 mg	2000 mg	2400 mg	2600 mg	Early Stop
Scenario 7										
Aerd-CRM	2:1	0.71	2.37	**68.93**	24.39	0.10	0.00	0.00	0.00	3.50
Ord-CRM	2:1	0.00	0.01	**6.95**	61.22	27.82	0.50	0.00	0.00	3.50
Bin-CRM	2:1	0.00	0.00	**9.07**	53.21	33.62	0.60	0.00	0.00	3.50
Aerd-CRM	1:1	0.43	1.83	**67.64**	17.91	0.08	0.00	0.00	0.00	12.11
Ord-CRM	1:1	0.00	0.01	**7.31**	44.58	31.92	3.81	0.17	0.09	12.11
Bin-CRM	1:1	0.00	0.00	**16.17**	32.61	35.84	3.20	0.16	0.02	12.00
Scenario 8										
Aerd-CRM	2:1	0.01	0.75	10.99	**75.55**	12.21	0.00	0.00	0.00	0.49
Ord-CRM	2:1	0.00	0.00	0.32	**42.48**	55.27	1.44	0.00	0.00	0.49
Bin-CRM	2:1	0.00	0.00	0.29	**25.09**	69.54	4.58	0.01	0.00	0.49
Aerd-CRM	1:1	0.38	0.40	12.09	**69.99**	13.37	0.00	0.00	0.00	3.77
Ord-CRM	1:1	0.00	0.00	0.26	**28.16**	59.69	7.66	0.61	0.37	3.25
Bin-CRM	1:1	0.00	0.00	0.71	**12.79**	59.84	21.99	1.09	0.33	3.25
Scenario 9										
Aerd-CRM	2:1	0.19	0.21	3.16	37.69	**58.54**	0.15	0.00	0.00	0.06
Ord-CRM	2:1	0.00	0.00	0.03	10.69	**81.86**	7.33	0.03	0.00	0.06
Bin-CRM	2:1	0.00	0.00	0.00	3.30	**75.94**	20.58	0.12	0.00	0.06
Aerd-CRM	1:1	0.10	0.05	1.92	29.53	**61.04**	0.30	0.00	0.00	7.06
Ord-CRM	1:1	0.00	0.00	0.01	4.00	**70.81**	17.36	1.14	0.42	6.26
Bin-CRM	1:1	0.00	0.00	0.02	1.07	**31.84**	56.42	3.83	0.56	6.26
Scenario 10										
Aerd-CRM	2:1	0.01	0.00	0.15	2.66	44.24	**52.66**	0.28	0.00	0.00
Ord-CRM	2:1	0.00	0.00	0.00	0.03	18.35	**79.06**	2.52	0.04	0.00
Bin-CRM	2:1	0.00	0.00	0.00	4.35	82.94	12.39	0.32	0.00	
Aerd-CRM	1:1	0.02	0.02	0.20	1.95	35.45	**47.21**	0.81	0.00	14.34
Ord-CRM	1:1	0.01	0.00	0.00	0.01	4.28	**66.61**	12.73	3.57	12.79
Bin-CRM	1:1	0.00	0.00	0.00	0.00	1.10	**30.18**	39.65	16.28	12.79
Scenario 11										
Aerd-CRM	2:1	0.00	0.00	0.00	0.00	1.49	36.13	**45.03**	17.10	0.25
Ord-CRM	2:1	0.00	0.00	0.00	0.00	0.05	23.08	**59.45**	17.41	0.01
Bin-CRM	2:1	0.00	0.00	0.00	0.00	0.06	14.44	**46.75**	38.74	0.01
Aerd-CRM	1:1	0.16	0.00	0.00	0.00	0.38	8.67	**22.64**	19.81	48.34
Ord-CRM	1:1	0.00	0.00	0.00	0.00	0.00	3.20	**19.18**	35.46	42.16
Bin-CRM	1:1	0.00	0.00	0.00	0.00	0.02	1.75	**5.78**	50.29	42.16
Scenario 12										
Aerd-CRM	2:1	0.00	0.00	0.00	0.01	0.42	12.60	29.98	**56.72**	0.27
Ord-CRM	2:1	0.00	0.00	0.00	0.00	0.01	5.81	40.61	**53.53**	0.04
Bin-CRM	2:1	0.00	0.00	0.00	0.00	0.01	3.99	27.35	**68.61**	0.04
Aerd-CRM	1:1	2.84	0.00	0.00	0.00	0.06	1.07	5.11	**36.91**	54.01
Ord-CRM	1:1	0.00	0.00	0.00	0.00	0.00	0.26	3.01	**51.18**	45.55
Bin-CRM	1:1	0.00	0.00	0.00	0.00	0.00	0.28	1.69	**52.48**	45.55

schemes for Aerd-CRM under scenarios 7, 8, 9 and 10. However, for Bin-CRM and Ord-CRM, the percentage of trials selected at the targeted MTD is lower under the 1:1 weighting scheme than the 2:1 weighting scheme for those scenarios.

In this simulation study early stopping may occur due to the safety constraint and stopping rule discussed in Section 2.5. In scenario 7, over 3% of trials with the 2:1 weighting scheme stopped early. The percent of trials stopped early slightly increased to just over 12% due to the heterogeneity issue explained in Section 2.4. No trials stopped early in scenario 10 with the 2:1 weighting scheme. In scenarios 11 and 12, majority of the trials stops under 2:1 weighting scheme. Overall, Table 3 shows that Aerd-CRM trials select the targeted MTD more often, whereas both Ord-CRM and Bin-CRM select highly toxic dose levels. As shown in scenarios 9 and 10, if the adjacent higher toxic dose level is extremely toxic then both Ord-CRM and Bin-CRM outperform Aerd-CRM.

Statistics
in Medicine

Darssan, Thompson, and Pettitt

4. A Single Trial Example

We show a single Aerd-CRM trial associated with a simulated set of data. This example illustrates, step by step, the MTD estimation using Aerd-CRM. In addition, it provides some guidelines for constructing prior data. Suppose the number of patients to be enrolled is $N = 32$ with the target toxicity probability $\phi = 0.25$. The sample size constraint in inequality (1) suggests that $K < 10$, meaning the sample size is insufficient to test 10 or more dose levels. If we choose five test dose levels then the MTD can be reached using the resources: $N = 32$, $\phi = 0.25$, $K = 5$. The original CRM does not use the actual doses that are administered [1]. Although actual doses are used in the Aerd-CRM, they carry no biological meaning to the algorithm. This example is set with five fictitious but reasonably selected doses with mg/kg as the unit of measurement. Suppose the test dose levels are: 0.25 mg/kg, 0.55 mg/kg, 0.85 mg/kg, 1.2 mg/kg, and 1.6 mg/kg.

Adverse event U is observed on an ordinal scale of 1 to 4, where adverse event scores of 1 to 4 correspond to 'mild', 'moderate', 'severe' and 'life-threatening' adverse events, respectively. Adverse event relatedness V is categorised into two groups, where a score of 1 is an adverse event doubtfully or possibly related to the IND and 2 is an adverse event probably or definitely related to the IND. Dose-limiting toxicity is defined as severe and life-threatening adverse events which are probably or definitely, and life-threatening adverse events which are doubtfully or possibly, related to the IND.

In designing real trials, prior data is constructed by asking questions to clinicians. The first question would be: what dose is likely to yield 10%-15% chance of DLT? Suppose that, 0.15 mg/kg is most likely to produce 15% DLT. The follow-up question would be: what is the expected percent of severity of adverse events this dose level produces? Suppose 0.15 mg/kg is expected to produce 65% mild, 20% moderate, 10% severe and 5% life-threatening adverse events. Similar questions can be asked to gather information of a higher toxic dose level. Suppose 2 mg/kg is most likely to produce 85% DLT with 5% mild, 10% moderate, 35% severe and 50% life-threatening adverse events. If this information is believed to be conservative, additional information of a dose which is likely to produce 50% chance of DLT can be obtained. However, this example is kept simple by ignoring this extra dose level information.

We next ask: what proportion of probably or definitely related adverse events are expected during the trial? The second part of the prior data is about the proportion of adverse events doubtfully or possibly and probably or definitely related to the IND. The study subjects' previous medical history would revel the expected adverse events due to the underlying disease and other treatments. In addition, drugs similar to the IND tested on similar study subjects would provide hints for this question. For this example, assume 70% of adverse events are probably or definitely related, and therefore, 30% are doubtfully or possibly related to the IND. Suppose the proportion p of background adverse reactions accounted for, for doubtful or possible adverse events, is 3%.

The toxicity percentage each (u, v) combination accounts for, is calculated by assuming independence between adverse events and its relatedness. The grids in Figure 1 can be referred to as a visual guide. Models in equations (6) and (7) are fitted to these prior data points. The fitted curve is equivalent to the skeleton mentioned in the original CRM [1]. Based on the fitted curve the starting dose is 0.55 mg/kg.

In this example, patients are enrolled in a cohort of size four. Whenever adverse event and relatedness data are observed, reliability of these prior data need to be identified. Suppose clinicians believe this prior information is only 60% reliable. So the first cohort of observed data is set to carry 40% weight. The prior data should be diminished as the trial progress. In this example, once 50% of the trial progressed the prior data information is diminished to one half, and when the trial ends this information is down to one fourth.

The safety constraint and the stopping rules discussed in Section 2.5 are set with $c_t = 5\%$ and $s_t = 70\%$ respectively. The first cohort of four patients are treated at the starting dose. The observed data simulated for this example and the estimated subsequent dose levels are shown in Table 4. For each cohort, at the estimated dose level, the estimated toxicity probability based on doubtfully or possibly related adverse events and probably or definitely related adverse events is shown in columns three and four. The Aerd-CRM estimates a dose level by using the maximum of these two estimated toxicity probabilities. The amount of information prior data carries is shown in the last column of the table.

Statistics
in Medicine

Darssan, Thompson, and Pettitt

Table 4. Step by step MTD estimation using Aerd-CRM

Cohort No.	Adverse Events (u, v)	Prob. $\{U=4\|V=1\}$	Prob. $\{U\geq3\|V=2\}$	Estimated Dose	Prior Data Weight
1	(2,2), (3,2), (3,1), (4,1)	0.28	0.24	0.25 mg/kg	60%
2	(1,2), (1,1), (2,2), (2,1)	0.22	0.29	0.55 mg/kg	43%
3	(3,1), (1,2), (2,2), (2,2)	0.19	0.18	0.55 mg/kg	34%
4	(1,2), (2,1), (1,1), (2,2)	0.19	0.28	0.85 mg/kg	28%
5	(2,2), (1,2), (3,1), (2,2)	0.17	0.19	0.85 mg/kg	24%
6	(1,2), (2,2), (1,2), (1,1)	0.22	0.28	1.20 mg/kg	21%
7	(3,2), (1,2), (4,1), (2,1)	0.19	0.19	0.85 mg/kg	18%
8	(4,2), (2,2), (2,2), (3,1)	0.18	0.20	**0.85 mg/kg**	15%

Two dose-limiting toxicities were observed in the first cohort of patients. One is probably or definitely related to the severe adverse event and the other is doubtfully or possibly related to the life-threatening adverse event. The estimated dose is 0.25 mg/kg which was decided using the curve fitted to doubtfully or possibly related adverse events. No dose-limiting toxicities were observed from cohort two to cohort six. The estimated dose levels were slowly increased to 1.20 mg/kg. The toxicity probability of these estimated dose levels is higher in doubtfully or possibly related adverse events for cohort three. The converse is true for cohorts two, four, five and six. Cohort seven is administered a dose level of 1.20 mg/kg. Two dose-limiting adverse events, a severe form $\{V=2\}$, and a life-threatening form $\{V=1\}$ category are observed at this dose level. The estimated dose level is 0.85 mg/kg, where the toxicity probability is slightly higher in $\{V=1\}$ category. It is not visible in the table, because probabilities shown in the table are rounded to two digits. The last cohort of patients were administered to 0.85 mg/kg, in which a life-threatening DLT of $\{V=2\}$ category is observed. The global MTD is 0.85 mg/kg. The prior data weight for the first cohort of patients is 60%, where it is steadily reduced by one half at cohort four. From cohort four to the last cohort the amount of information prior data carries is diminished slowly to 15%.

5. Discussion

In this article, we assessed a design methodology which incorporates adverse event relatedness. Patients administered low toxic doses could experience adverse events which are doubtfully or possibly related to the IND and, therefore, this toxic information is modeled using a two-parameter logistic model with an additive probability mass. The probability mass accounts for the background exposure of these adverse events. Patients who receive high toxic doses are most likely to experience adverse events which are probably or definitely related to the IND. Including this information means that information on each toxicity grade can be made available using a cumulative logistic model.

The design that incorporates adverse event relatedness, namely Aerd-CRM, selects the targeted dose efficiently and gathers more information about adverse events which are vaguely related to the IND. Whenever the adjacent higher dose level is extremely toxic this design tends to select the adjacent lower toxic dose level. In such circumstances designs which do not incorporate adverse event relatedness appear to be better than the Aerd-CRM. The Aerd-CRM is an extension of CRM which uses the cumulative logistic model introduced by Meter *et al.* [16]. This means it has the key characteristics of CRM with additional information on ordinal toxicity outcome and adverse event relatedness. The Aerd-CRM estimates dose levels using the maximum of toxicity probabilities calculated based on adverse event categories. Therefore, the design methodology can be interpreted as a design for categorised adverse events. Our simulation study shows that ignoring adverse event relatedness categories results in estimating higher toxic or delusive dose levels as the MTD.

We have shown that dose-finding design methodology should incorporate adverse event relatedness. The CRM is proven to be efficient and in an era of spreading to clinicians and statisticians whom have no professional knowledge of the CRM

Darssan, Thompson, and Pettitt

design methodology [21]. Incorporating adverse event relatedness into CRM opens up further investigation into the CRM design methodology. For example, the bivariate CRM by Braun [22] can be extended to include adverse event relatedness and efficacy relatedness. Incorporating the toxicity and efficacy due to any background information would yield better dose estimation. Combining the Aerd-CRM concepts with the TITE-CRM introduced by Cheung *et al.* [23] would further improve the CRM with late-onset toxicities. If a dose-finding trial is set with little data and the objective is to estimate a dose level based on binary toxicity outcome then the Bin-Aerd-CRM can be used.

The Aerd-CRM discussed in this paper has a few restrictions in addition to the limitations of the CRM. The prior data required for Aerd-CRM is complicated. In addition to the prior data in toxicity and severity required by the CRM, the Aerd-CRM needs prior data in adverse event relatedness. Prior data in adverse event relatedness can only be obtained from drugs previously tested which are similar to the IND. However, clinicians' expert knowledge can also be included as prior data in adverse event relatedness. As shown in the simulation study, if prior data is down weighted with caution then the choice of the prior data does not affect the characteristics of the Aerd-CRM. The performance of the Aerd-CRM would be extreme if test dose levels of the IND are highly toxic. If a very low toxic dose level is administered and there are no observed toxicities, then this means the adverse event relatedness is unknown. In such circumstances the non-toxic outcome should be included into the design as a mild toxicity based on the prior data on adverse event relatedness.

Clinicians are required to thoroughly identify the relatedness of each adverse event. In practice, sometimes it is difficult to identify the relatedness of an adverse event to the IND. If an adverse event is identified as related, then it is more challenging to classify the relatedness into categories. Zohar *et al.* [24] investigate the sensitivity of dose-finding designs to two types of observation errors, identifying a drug related toxicity as unrelated toxicity, and an unrelated toxicity as drug related toxicity. They claim that the CRM is less affected by errors of the former type, identifying a drug related toxicity as unrelated toxicity, compared to other rule-based dose-finding designs. Iasonos *et al.* [25] support this claim, as their simulation study proves that the CRM is robust to both types of errors, compared to the standard 3+3 design. Therefore, incorporating adverse event relatedness into the CRM bring a new dimension to this design methodology. However, the Aerd-CRM demands clinician to accurately classify the drug related adverse events into two categories 'doubtful or possible', and 'probable or definite'. A wrong classification can lead to an inaccurate MTD estimation.

Acknowledgements

References

1. O'Quigly J, Pepe M, Fisher L. Continual reassessment method: a practical design for phase I clinical trials in cancer. *Biometrics* 1990; **46**:33 – 48.
2. Onar A, Kocak M, Boyett JM. Continual Reassessment Method vs. Traditional Empirically Based Design: Modifications Motivated by Phase I Trials in Pediatric Oncology by the Pediatric Brain Tumor Consortium. *Journal of Biopharmaceutical Statistics* 2009; **19**:437 – 455.
3. O'Quigly J, Conaway M. Continual reassessment method and related dose-finding designs. *Statistical Science* 2010; **25**:202 – 216.
4. Whitehead J, Brunier H. Bayesian decision procedures for dose determining experiments. *Statistics in Medicine* 1995; **14**:885 – 893.
5. Babb J, Rogatko A, Zacks S. Cancer phse I clinical trials: Efficient dose escalation with overdose congrol. *Statistics in Medicine* 1998; **17**:1103 – 1120.
6. Leung DHY, Wang YG. Isotonic Designs for Phase I Trials. *Controlled Clinical Trials* 2001; **22**:126 – 138.
7. Haines L, Perevozskaya I, Rosenberger W. Bayesian Optimal Designs for Phase I Clinical Trials. *Biometrics* 2003; **59**:591 – 600.
8. Hughes WT. Postulates for the Evaluation of Adverse Reactions to Drugs. *Clinical Infectious Diseases* 1995; **20**:179 – 182.
9. Karch FE, Lasagna L. Adverse Drug Reactions. *The Journal of the American Medical Association* 1975; **234**:1236 – 1241.
10. Rosenberger WF, Haines LM. Competing designs for phase I clinical trials: a review. *Statistics in Medicine* 2002; **21**:2757 – 2770.

Statistics
in Medicine

11. Shen LZ, O'Quigly J. Consistency of continual reassessment method under model misspecification. *Biometrika* 1996; **83**:395 – 405.

12. Piantadosi S, Fisher JD, Grossman S. Practical implementation of a modified continual reassessment method for dose-finding trials. *Cancer Chemotheraphy Pharmacology* 1998; **41**:429 – 436.

13. Bekele BN, Thall PF. Dose-Finding Based on Multiple Toxicities in a Soft Tissue Sarcoma Trial. *Journal of the American Statistical Association* 2004; **99**:26 – 35.

14. Yuan Z, Chappell R, Bailey H. The Continual Reassessment Method for Multiple Toxicity Grades: A Bayesian Quasi-Likelihood Approach. *Biometrics* 2007; **63**:173 – 179.

15. Iasonos A, Zohar S, O'Quigley J. Incorporating lower grade toxiciy information into dose finding designs. *Clinical Trials* 2011; **8**:370 – 379.

16. Van Meter EM, Garrett-Mayer E, Bandyopadhyay D. Proportional odds model for dose-finding clinical trial designs with ordinal toxicity grading. *Statistics in Medicine* 2011; **30**:2070 – 2080.

17. Cohen RB, Aamdal S, Nyakas M, Cavallin M, Green D, Learoyd M, Smith I and Kurzrock R. A phase I dose-finding, safety and tolerability study of AZD8330 in patients with advanced malignancies. *European Journal of Cancer* 2013; DOI: 10.1016/j.ejca.2013.01.013

18. Iasonos A, O'Quigley J. Continual reassessment and related designs in dose finding studies. *Statistics in Medicine* 2011; **30**:2057 – 2061.

19. Cheung YK. 2011. *Dose Finding by the Continual Reassessment Method*. Chapman and Hall/CRC Press, New York.

20. Omura GA. Modified Fibonacci Search. *Journal of Clinical Oncology* 2003; **21**:3177.

21. Garrett-Mayer E. The continual reassessment method for dose-finding studies: a tutorial. *Clinical Trials* 2006; **3**:57 – 71.

22. Braun TM. The bivariate continual reassessment method: extending the CRM to phase I trials of two competing outcomes. *Controlled Clinical Trials* 2002; **23**:240 – 256.

23. Cheung YK, Chappell R. Sequential Designs for Phase I Clinical Trials with Late-Onset Toxicities. *Biometrics* 2000; **56**:1177 – 1182.

24. Zohar S, O'Quigly J. Sensitivity of dose-finding studies to observation errors. *Contemporary Clinical Trials* 2009; **30**:523 – 530.

25. Iasonos A, Gounder M, Spriggs DR, Gerecitano JF, Hyman DM, Zohar S, and O'Quigley J. The Impact of Non–Drug-Related Toxicities on the Estimation of the Maximum Tolerated Dose in Phase I Trials. *Statistics in Clinical Cancer Research* 2012; **18**:5179 – 5187.