Communications for Statistical Applications and Methods
2022, Vol. 29, No. 4, 421–439

https://doi.org/10.29220/CSAM.2022.29.4.421
Print ISSN 2287-7843 / Online ISSN 2383-4757

An R package *UnifiedDoseFinding* for continuous and ordinal outcomes in Phase I dose-finding trials

Haitao Pan[1,a], Rongji Mu[b], Chia-Wei Hsu[a], Shouhao Zhou[c]

[a]Department of Biostatistics, St Jude Children's Research Hospital, USA;
[b]Clinical Research Center, Shanghai Jiao Tong University School of Medicine, China;
[c]Department of Public Health Sciences, Pennsylvania State University College of Medicine, USA

Abstract

Phase I dose-finding trials are essential in drug development. By finding the maximum tolerated dose (MTD) of a new drug or treatment, a Phase I trial establishes the recommended doses for later-phase testing. The primary toxicity endpoint of interest is often a binary variable, which describes an event of a patient who experiences dose-limiting toxicity. However, there is a growing interest in dose-finding studies regarding non-binary outcomes, defined by either the weighted sum of rates of various toxicity grades or a continuous outcome. Although several novel methods have been proposed in the literature, accessible software is still lacking to implement these methods. This study introduces a newly developed R package, `UnifiedDoseFinding`, which implements three phase I dose-finding methods with non-binary outcomes (Quasi- and Robust Quasi-CRM designs by Yuan *et al.* (2007) and Pan *et al.* (2014), gBOIN design by Mu *et al.* (2019), and by a method by Ivanova and Kim (2009)). For each of the methods, `UnifiedDoseFinding` provides corresponding functions that begin with `next_` that determines the dose for the next cohort of patients, `select_`, which selects the MTD defined by the non-binary toxicity endpoint when the trial is completed, and `get_oc`, which obtains the operating characteristics. Three real examples are provided to help practitioners use these methods. The R package `UnifiedDoseFinding`, which is accessible in R CRAN, provides a user-friendly tool to facilitate the implementation of innovative dose-finding studies with nonbinary outcomes.

Keywords: Phase I dose-finding, toxicity grades, quasi-likelihood, continual reassessment method Bayesian optimal interval design

1. Introduction

Phase I trials are conducted to seek a new drug's toxic effect on patients through the identification of an optimal dose, called the maximum tolerated dose (MTD), which maximizes its therapeutic effect while maintaining a tolerable toxic effect. The estimated MTD and the administered schedule of a new drug or treatment determined in a phase I clinical trial will then be employed in phase II and III clinical trials for efficacy and therapeutic effects to be assessed. There are three categories of designs for phase I clinical trials; namely, algorithm-based (e.g., 3+3 design in Storer, 1989), model-based (e.g., the continual reassessment method (CRM) design by O'Quigley *et al.*, 1990), and model-assisted designs (for instance, the Bayesian optimal interval (BOIN) design in Liu and Yuan, 2015).

In all the above-mentioned designs, a binary indicator of dose-limiting toxicity (DLT) (if a DLT occurs during the observation window for toxicity assessment) is adopted to describe the toxicity

[1] Corresponding author: Department of Biostatistics, St Jude Children's Research Hospital, Memphis, TN 38105, USA.
E-mail: Haitao.Pan@stjude.org

Published 31 July 2022 / journal homepage: http://csam.or.kr

outcomes. In spite of its ease of application, there are unavoidable limitations in using DLT as a binary indicator. First, patients often have multiple toxicities, but in using DLT as an assessment, all other toxicities of a patient are ignored except for the worst one. For example, the National Cancer Institute-Common Terminology Criteria for Adverse Events (NCI-CTCAE) (2003) is often used as an instrument for grading adverse events. For each adverse event, the severity is graded on a scale from 0 to 5, with grade 0 being no toxicity and grade 5 being death. In real cancer trials, a tremendous amount of toxicity data is collected and reported for each patient. The binary toxicity summaries using DLTs disregard lower grade toxic effects, which individually are less severe than dose-limiting, but in aggregate they can be concerning. Second, binary toxicity summaries also do not differentiate between types of toxic effects. Third, the definition of DLT varies by study (Dent *et al.*, 1996). Last, there is an additional impetus for the non-binary toxicity response due to the emergence of molecularly targeted agents and immunotherapies that has changed the landscape of many oncology drug developments. These new therapeutic agents appear more likely to induce multiple low or moderate grade toxicities rather than DLTs (Brahmer *et al.*, 2010). To account accurately for the side effects of these agents, it is also important to incorporate the grade of toxicity in dose-finding and decision making. For some of these agents, the side effects are often less frequent and severe; and the dose of interest in this type of dose-finding trial may aim to evaluate the biological activity of an agent, often measured as a continuous variable, rather than the binary toxicity. Because DLT might not be reached, a continuous biomarker endpoint might be a better primary endpoint in a phase I trial that uses a cytostatic agent. Examples include a measure of target inhibition or pharmacokinetic endpoints such as plasma drug concentrations that correlate with biological activity (Le Tourneau *et al.*, 2009) or the percentage inhibition of an enzyme (Plummer *et al.*, 2008).

Several methods for incorporating grade toxicity or a non-binary toxicity endpoint have been proposed in recent years for phase I dose-finding trials, e.g., Ivanova and Kim (2009), Bekele and Thall (2004), Yuan *et al.* (2007), Meter *et al.* (2011), and Mu *et al.* (2019). Among them, Bekele and Thall (BT method for short) applied severity weights to a soft tissue sarcoma trial using five types of DLTs. Physicians assigned a severity weight on a common numeric scale for each type of toxicity to each observed patient, and the sum of these weights over the five toxicity types was called the total toxicity burden (TTB). The authors then considered a hypothetical collection of cohorts with a variety of different possible outcomes. They Quasi-CRM design, proposed by Yuan *et al.* (2007), also adopted severity weights to convert toxicity grades to numerical scores and incorporated these scores into the CRM design. The recommended dose for the next patient is the dose level with an estimated score (the equivalent toxicity score) closest to the target score, which was obtained from a pre-specified toxicity profile at the MTD. This Quasi-CRM method has been demonstrated to be superior to the BT method, and it has a higher probability for recommending the optimal dose in further studies. Pan *et al.* (2014) further extended the Quasi-CRM and proposed the Robust Quasi-CRM by utilizing a parallel of skeletons. Without specifying any skeleton, Ivanova and Kim (2009) presented a different model-based design approach in which the target dose is defined as the dose at which the outcome of interest is equal to some pre-specified value. By minimizing the normalized difference between the current dose and the target dose, this new design was applied to trials with continuous outcomes, ordinal toxicity outcomes, and binary outcomes. In the school of model-assisted designs, Mu *et al.* (2019) generalized the BOIN design and proposed a unified approach that can deal with binary and non-binary outcomes. Additional measures to incorporate ordinal toxicity grades into dose-finding also include the toxicity burden score (TBS) by Lee *et al.* (2012), which summarizes toxicity by using a weighted sum, where the severity weights were estimated via regression using historical data, and the total toxicity profile (TTP) by Ezzalfani *et al.* (2013), which is computed as

the Euclidean norm of the severity weights.

Although these novel designs for non-binary outcomes have been developed in recent years, there has been very limited usage of these methods in real practices. One obstacle is the lack of accessible software. For instance, among the above-introduced methods, only the design by Emily *et al.* (2011), which incorporated toxicity grades using a continuation ratio (CR) model in the likelihood-based CRM, has been developed into an R package `ordcrm` for implementing the associated method. Therefore, there is a need to include various representative methods in one freely available and user-friendly package that is available for trial design and design comparison, which is the aim of this manuscript.

In the `UnifiedDoseFinding` package, three methods are included: the Quasi-CRM and Robust Quasi-CRM designs (Yuan *et al.*, 2007, Pan *et al.*, 2014), the gBOIN design – a model-assisted unified design (Mu *et al.*, 2019), and the model-based unified approach in Ivanova and Kim (2009).

As pointed out by one reviewer, for capturing a more comprehensive profile of toxicity in a dose-finding study, with the exception of the above approach by re-defining the binary-based DLT, there are other ways that exist for capture, for example, the late-onset toxicity and accumulative toxicity. For example, one work by Zhang *et al.* (2018) mentions the concept of relative dose intensity as another alternative endpoint. Another way is to refine the DLT beyond the 1st cycle, and Paoletti *et al.* (2014) provides some examples, especially when there is a developed R package `phase1RMD` that takes into consideration the longitudinal toxicity endpoint in multiple cycles. Furthermore, health-related quality of life as an endpoint in oncology phase I trials has been reviewed by Fiteni *et al.* (2019) since some oncology diseases can be 'cured', e.g., some types of pediatric ALL/AML. Finally, the current paradigm of phase I methods ignores the PK information, though PK also contains or is correlated with toxicity per se, for example, as in a recent trial shown in Muehz (2017).

The paper is organized as follows. We concisely introduce these three designs in Section 2. In Section 3, we show how to use the package to implement the trial in practice and conduct simulations for protocol development or research purposes. Section 4 provides a real trial use of this package. The conclusion is found at the end in Section 5.

2. Methods

This section briefly reviews the phase I design methods proposed by Yuan *et al.* (2007) and Pan *et al.* (2014), Mu *et al.* (2017), and Ivanova and Kim (2009), respectively. These designs were proposed for dose-finding trials, which use non-binary toxicity outcomes.

2.1. Quasi-CRM method

Yuan *et al.* (2007) proposed the concept of "equivalent toxicity score" (ETS) to measure the relative severity by incorporating different toxicity grades in the dose-finding procedure. Rather than using a conventional 0–1 rule for a severe adverse event (AE) or above (i.e., AE grade ≥ 3) to define a binary toxicity outcome, an ordinal score is assigned to AEs at different grades. At each dose level, DLT is assessed by the weighted average score using a corresponding toxicity profile that is defined by an AE grade. For example, a medical oncologist can pre-assign the scores of 0, 0.5, 1, and 1.5 to AE grades 0/1, 2, 3, and 4, respectively. In the meantime, if the following toxicity profile is considered tolerable for the MTD, such that 49% grade 0 and grade 1, 18% grade 2, 23% grade 3, and 10% grade 4, the target ET score for dose-finding is then obtained by the weighted sum over all the grades, that is,

$$R_0 = 0.49 \times 0 + 0.18 \times 0.5 + 0.23 \times 1.0 + 0.10 \times 1.5 = 0.47.$$

For the dose-finding trial, the above target ET score replaces the conventional DLT target. Indeed, the conventional definition of DLT often just considers ≥ 3 grades toxicity. For the toxicity profile using the ET score, it is essentially a more familiar concept for medical oncologists and a more consistent target for toxicity.

From the above, we can define a normalized ET score as the continuous endpoint (more precisely we can call it quasi-binary, which is explained in Section 3.2) used for the Quasi-CRM design. Suppose that n patients have been tested sequentially at dose levels $d(1), d(2), \ldots, d(n)$ with corresponding ET scores $s(1), s(2), \ldots, s(n)$. The normalized ET score is then defined as,

$$s^{\star}(i) = \frac{s(i)}{s_{\max}}, \quad i = 1, 2, \ldots, n,$$

where $s_{\max} < \infty$ ensures $s^{\star}(i) \in [0, 1]$, $i = 1, 2, \ldots, n$.

The normalized scores can be viewed as fractional events and modeled using the Quasi-Bernoulli likelihood (Papke and Wooldrige, 1996). If the dose-toxicity model is correctly specified, the quasi-maximum likelihood estimate (QMLE) will be strongly consistent because the Bernoulli distribution belongs to the binomial family. If a functional dose-score curve is not assumed, the QMLE will be equal to the observed average ET score at each dose level. Thereby, the goal of the study is to find the MTD d_0 that is the highest dose level such that the projected normalized ET score is $p_s^{\star}(d_0) \leq p_0/s_{\max}$.

The Quasi-Bernoulli likelihood provides a simple way to incorporate ordinal grades into parametric models. Yuan *et al.* (2007) successfully applied it with the CRM for developing the Quasi-CRM. However, the Quasi-CRM model suffers from the same problem as the CRM model in that the choice of the skeleton may dramatically affect the performance of the model operating characteristics. Pan *et al.* (2014) utilized the Bayesian model selection approach and proposed the Robust Quasi-CRM model, which inherits the BMA-CRM model proposed by Yin and Yuan (2009) and considers a parallel of skeletons for the Quasi-CRM. The superior performance of the Robust Quasi-CRM model was also demonstrated by extensive simulation studies conducted by Pan *et al.* (2014).

Specifically, let $(M_1, \ldots, M_K)$ be the K models corresponding to each set of prior guesses of toxicity probabilities $\{(p_{11}, \ldots, p_{1J}), \ldots, (p_{K1}, \ldots, p_{KJ})\}$. The probability of toxicity at dose j in model $M_k (k = 1, \ldots, K)$ is given by,

$$\pi_{kj}(\alpha_k) = p_{kj}^{\exp(\alpha)}, \quad j = 1, \ldots, J$$

which is based on the k^{th} skeleton $(p_{k1}, \ldots, p_{kJ})$. Let $\mathrm{pr}(M_k)$ be the prior probability and model M_k is the true model; that is, the probability that the k^{th} skeleton $(p_{k1}, \ldots, p_{kJ})$ matches the true dose-toxicity curve. If there is no preference a priori for any single model in the CRM case, then one can assign equal weights to the different skeletons by simply setting $\mathrm{pr}(M_k) = 1/K$. At a certain stage of the trial, which is based on the observed data $D = \{(n_j, y_j), j = 1, \ldots, J\}$, denotes the quasi-Bernouli likelihood function under model M_k by $L(D|\alpha_k, M_k)$.

The posterior model probability for M_k is given by,

$$\mathrm{pr}(M_k|D) = \frac{L(D|M_k)\mathrm{pr}(M_k)}{\sum_1^K L(D|M_i)\mathrm{pr}(M_i)},$$

where $L(D|M_k)$ is the marginal likelihood of model M_k, $L(D|M_k) = \int L(D|\alpha_k, M_k) f(\alpha_k|M_k) d\alpha_k$, and α_k is the power parameter in the CRM associated with model M_k, and $f(\alpha_k|M_k)$ is the prior distribution of α_k under model M_k.

As the posterior model probability transformed the prior opinion for each model M_k through consideration of data D, a Bayesian model selection approach can be naturally developed to estimate toxicity probabilities and base decision on dose assignment. Specifically, at each point of the decision making for dose assignment, we select the model with the highest posterior probability, i.e., model,

$$k^\star = \arg \max_{k \in 1,\ldots,K} (\text{pr}(M_K|D)),$$

and use this model to make inference and dose assignment.

Then, after n patients, the quasi-posterior estimation of the toxicity probability for dose j under the *kth* skeleton will be updated by

$$\hat{\pi}_{kj} = \int p_{kj}^{\exp(\alpha)} \frac{\text{L}(D|\alpha_k, M_k) f(\alpha_k|M_k)}{\int \text{L}(D|\alpha_k, M_k) f(\alpha_k|M_k) d\alpha_k} d\alpha_k. \tag{2.1}$$

We require early termination of a trial if the lowest dose is too toxic, as noted by

$$\text{pr}\left(\pi_{k^\star 1}(\alpha_{k^\star}) > \phi | M_{k^\star}, D\right) > 90\%.$$

The functions `get_oc_RQ_CRM()`, `next_RQ_CRM()`, and `select_mtd_RQ_CRM()` in the `Unified DoseFinding` package implement the dose-finding methods described in this section. When a single skeleton is used, these functions apply the Quasi-CRM design while a multiple-skeletons setting applies the Robust Quasi-CRM approach.

2.2. gBOIN method

The gBOIN design by Mu *et al.* (2019) is a model-assisted design to generalize the BOIN design by Liu and Yuan (2015), which accommodates various existing toxicity grade scoring systems.

Assume there are J specified doses $d_1 < \cdots < d_J$ under investigation. Let y denote the toxicity outcome, which is either binary or quasi-binary (e.g., DLT or ETS) or continuous (e.g., TTB, TBS or TTP). Define $\mu = \text{E}(y)$ and $\mu_j = \text{E}(y|d_j)$. Given the dose d_j, the distribution of y belongs to the exponential family,

$$f(y|d_j) = h(y) \exp\left\{\eta(\theta_j) T(y) - A(\theta_j)\right\},$$

where,

- $\theta_j = \mu_j$, $\eta(\theta_j) = \log\{\mu_j/(1-\mu_j)\}$, $A(\theta_j) = -\log(1-\mu_j)$, $T(y) = y$, and $h(y) = 1$, if y follows a binomial distribution;
- $\theta_j = (\mu_j, \sigma_j^2)$, $\eta(\theta_j) = \mu_j/\sigma_j^2$, $A(\theta_j) = \mu_j^2/(2\sigma_j^2)$, $T(y) = y$, and $h(y) = (1/\sqrt{\pi}\sigma) \exp\{-y^2/(2\sigma_j^2)\}$, if y follows a normal distribution.

Let ϕ_0 denote the target value of μ for dose finding. Specifically, for binary or quasi-binary toxicity endpoints, ϕ_0 is the target DLT probability; for continuous endpoints, ϕ_0 is the targeted value of TTB, TBS or TTP. Denote $\hat{\mu}_i$ the sample mean of the observed toxicity data at dose level d_i. For the interval-based design, dose transition decisions are made by comparing $\hat{\mu}_j$ with the decision boundaries, $\lambda_e(d_j, n_j, \phi_0)$ and $\lambda_d(d_j, n_j, \phi_0)$. Specifically, if $\hat{\mu}_j < \lambda_e(d_j, n_j, \phi_0)$, then escalate to the higher dose level $j+1$, and if $\hat{\mu}_j > \lambda_d(d_j, n_j, \phi_0)$, de-escalate to th thene lower dose level $j-1$; otherwise retain the same dose level j. The selection of the decision boundaries $\lambda_e(d_j, n_j, \phi_0)$ and $\lambda_d(d_j, n_j, \phi_0)$

is critical because these two parameters essentially determine operating characteristics of a design. Let the decisions of retainment, escalation and de-escalation (each based on the current dose level) be denoted as $\mathcal{R}$, $\mathcal{E}$ and $\mathcal{D}$, respectively and let $\overline{\mathcal{R}}$ denote the decisions that are complementary to $\mathcal{R}$ (i.e., $\overline{\mathcal{R}}$ includes $\mathcal{E}$ and $\mathcal{D}$); and $\overline{\mathcal{E}}$ and $\overline{\mathcal{D}}$ denote the decisions that are complementary to $\mathcal{E}$ and $\mathcal{D}$, respectively. Following the same rule in Liu and Yuan (2015) that obtains optimal decision boundaries under some criteria, the gBOIN considers three point hypotheses $H_0 : \mu_j = \phi_0$, $H_1 : \mu_j = \phi_1$, and $H_2 : \mu_j = \phi_2$; and minimizes an incorrect decision probability α,

$$\alpha = \text{Prob}(H_0)\text{Prob}\left(\overline{\mathcal{R}}|H_0\right) + \text{Prob}\left(H_1\right)\text{Prob}\left(\overline{\mathcal{E}}\,|H_1\right) + \text{Prob}\left(H_2\right)\text{Prob}\left(\overline{\mathcal{D}}|H_2\right), \tag{2.2}$$

where ϕ_1 is a value deemed subtherapeutic such that dose escalation is warranted, and ϕ_2 is a value deemed overly toxic such that dose de-escalation is required. Taking a noninformative prior, i.e., $\text{Prob}(H_0) = \text{Prob}(H_1) = \text{Prob}(H_2) = 1/3$, and minimizing the incorrect decision probability α in equation (2.2), the decision boundaries can be obtained as,

$$\lambda_e^* = \frac{A(\vartheta_1) - A(\vartheta_0)}{\eta(\vartheta_1) - \eta(\vartheta_0)}, \qquad \lambda_d^* = \frac{A(\vartheta_2) - A(\vartheta_0)}{\eta(\vartheta_2) - \eta(\vartheta_0)}.$$

Specifically, when y follows a Bernoulli or quasi-Bernoulli distribution, we have $\vartheta_k = \phi_k$, $A(\vartheta_k) = -\log(1-\phi_k)$, $\eta(\vartheta_k) = \log\{\phi_k/(1-\phi_k)\}$. Then,

$$\lambda_e^* = \frac{\log\frac{1-\phi_1}{1-\phi_0}}{\log\frac{\phi_0(1-\phi_1)}{(1-\phi_0)\phi_1}}, \qquad \lambda_d^* = \frac{\log\frac{1-\phi_0}{1-\phi_2}}{\log\frac{\phi_2(1-\phi_0)}{(1-\phi_2)\phi_0}}, \tag{2.3}$$

which are exactly the same as boundaries provided by the original BOIN design (Liu and Yuan, 2015). When y follows a normal distribution, we have $\vartheta_k = (\phi_k, \sigma_j^2)$, $A(\vartheta_k) = \phi_k^2/(2\sigma_j^2)$, $\eta(\vartheta_k) = \phi_k/\sigma_j^2$. Then,

$$\lambda_e^* = \frac{\phi_0 + \phi_1}{2}, \qquad \lambda_d^* = \frac{\phi_0 + \phi_2}{2}. \tag{2.4}$$

Based on the above decision boundaries, the gBOIN design is summarized as follows:

(a) Patients in the first cohort are treated at the lowest dose level or at a prespecified dose level.

(b) At the current dose level j, a dose is assigned to the next cohort of patients,

- if $\hat{\mu}_j \le \lambda_e^*$, then escalate the dose level to $j+1$,
- if $\hat{\mu}_j \ge \lambda_d^*$, then de-escalate the dose level to $j-1$,
- otherwise, if, i.e., $\lambda_e^* < \hat{\mu}_j < \lambda_d^*$, then retain the same dose level, j.

(c) This process is continued until the maximum sample size is reached or the trial is terminated because of excessive toxicities.

The optimal decision boundaries $(\lambda_e^*, \lambda_d^*)$ are free of d_j and n_j, which means that the same pair of boundaries can be used throughout the trial no matter which dose is the current dose or how many patients have been treated at the current dose.

The functions `get_oc_gBOIN_Continuous()`, `next_gBOIN_Continuous()`, and `select_mtd_gBOIN_Continuous()` in the `UnifiedDoseFinding` package implement the gBOIN design with continuous outcomes.

2.3. Ivanova design

Ivanova and Kim (2009) proposed a unified dose-finding approach for studies with the target dose being defined as the dose at which the outcome of interest is equal to some specified value. This design can apply to trials with a variety of response types, such as, continuous outcomes, ordinal toxicity outcomes, and binary outcomes.

Let $D = \{d_1, \ldots, d_K\}$ denote the set of ordered dose levels selected for a trial. A subject's response at d_k has a distribution function $F(\cdot; \mu_k, \sigma_k^2)$, where $(\mu_1, \ldots, \mu_K$ and $(\sigma_1^2, \ldots, \sigma_K^2)$ are vectors of means and variances corresponding to D. This design assumed that the mean response was monotone in dose. The goal is to find dose $d_m \in D$ such that $\mu_m = \mu^*$. If there is no such dose, the goal is to find dose d_m with μ_m closest to μ^*, where μ^* is the target value and d_m is the target dose. Subjects are assigned sequentially starting with the lowest dose. The total number of subjects is equal to M. Let $\boldsymbol{n}(t) = (n_1(t), \ldots, n_K(t))$ be the number of subjects at each of the K doses right after subject $t, t \leq M$, has been assigned, that is, $n_1(t) + \cdots + n_K(t) = t$. Let Y_{ji} be the observation from the ith subject assigned to dose $d_j, i = 1, \ldots, n_j(t)$. Let $\bar{Y}_j(n_j(t)) = \sum_{i=1}^{n_j(t)} Y_{ji}/n_j(t)$ and $s_j^2(n_j(t) = \sum_{i=1}^{n_j(t)} \{Y_{ji} - \bar{Y}_j(n_j(t))\}^2/(n_j(t) - 1)$ be the sample mean and variance computed from all available observations at $d_j, n_j(t) = 2, 3, \ldots$. Define $T_j(n_j(t)), n_j(t) = 2, 3, \ldots$, to be the t-statistic

$$T_j(n_j(t)) = \frac{\bar{Y}_j(n_j(t)) - \mu^*}{s_j(n_j(t))/\sqrt{n_j(t)}}.$$

Suppose the most recent subject t was assigned to dose d_j. The next subject is assigned as follows:

(i) if $T_j(n_j(t)) \leq -\Delta$, the next subject is assigned to dose d_{j+1};

(ii) if $T_j(n_j(t)) \geq \Delta$, the next subject is assigned to dose d_{j-1};

(iii) if $-\Delta < T_j(n_j(t)) < \Delta$, the next subject is assigned to dose d_j.

Here, $\Delta > 0$ is called the design parameter. The isotonic estimates of the mean response at the end of a trial was used for target dose selection.

The authors recommend a start-up rule of using the above algorithm with at least two subjects to any untried dose before the dose can be escalated. The authors also introduced how to compute the distribution for subject allocation $\boldsymbol{n}(M)$, which is the number of subjects assigned to each dose by the time M subjects have been assigned. If F is the normal distribution function, the distribution of $\boldsymbol{n}(M)$ can be computed based on the joint distribution of sequential t-statistics.

The last step is to choose the design parameter Δ. If Δ is small, the current dose is repeated if the average response is very close to the target and changed otherwise. A narrow window $(-\Delta, \Delta)$ results in a small probability of repeating a dose even if the true mean value of the quantity of interest at that dose is equal μ^*. The optimal value for Δ is to maximize the number of subjects assigned to the target dose. Though the optimal Δ may depend on a dose-response assumption, the authors recommend $\Delta = 1$ as a reasonable choice used in simulation studies; that is, the design parameters can be chosen before the trial, which is similar to the BOIN design. However, the authors suggest it might be worth fine-tuning the design parameters during the trial as information about the spacing between doses and the variance of the outcomes becomes available, especially for continuous outcomes. If the optimal criterion is to maximize the proportion of trials that select the target dose correctly, the extensive simulations show that the same, or nearly the same parameter Δ works.

The functions `get_oc_Ivanova_continuous()`, `next_Ivanova_continuous()`, and `select_mtd_Ivanova_continuous()` in the `UnifiedDoseFinding` package implement the Ivanova dose-finding method for the continuous outcome described in this section. Similar functions ending with `_binary` can be used for the dose-finding with binary endpoints.

3. R functions

Package `UnifiedDoseFinding` implements all the dose-finding methods described in Section 2. For an ongoing study, the accumulated patient data can be used in the functions beginning with `next_···` to determine the next recommended dose. At the end of a trial, the functions beginning with `select_···` will report the dose level for the target score. All functions beginning with `get_oc_···` can be used for simulating various scenarios to develop clinical protocols or for research purposes.

The package `UnifiedDoseFinding` is available on the CRAN archive and can be installed through the URL `https://cran.r-project.org/web/packages/UnifiedDoseFinding`. Once the package is installed, it can be loaded with the command:

```
> library(UnifiedDoseFinding)
```

In the remainder of this section, we present the R functions available in the package along with the required input parameters and examples. A more extensive demonstration and documentation can be accessed from the on-line user manual on the CRAN server by installing the package or by accessing it directly from within the R console.

3.1. Determine the dose for the next cohort of new patients

The functions beginning with `next_···` give the recommended dose to administer to the next cohort of patients or the final estimated MTD if applied at the end of the trial.

(1) Next dose for the Quasi-CRM design

```
> next_RQ_CRM(target, n, y, dose.curr, score, skeleton)
```

This function applies the Quasi-CRM and Robust-Quasi-CRM designs for the ETS-defined MTD target, which is essentially a quasi-binary endpoint essentially. Among the input arguments: `target` is for the toxicity target score, `n` is for the number of patients treated at each dose level, `y` is the toxicity score at each dose level, `dose.curr` is the current dose level, score is the weighted vector for ordinal toxicity levels introduced in the Quasi-CRM design section, `skeleton` is a skeleton for the Quasi-CRM design or a matrix with multiple-skeletons for the Robust-Quasi-CRM design.

An example for using this function is shown below:

```
> ### Implement Robust-Quasi-CRM design (Pan {\it et al.} 2014) with pre-specifying 3 skeletons
> target <- 0.47
> score <- c(0, 0.5, 1, 1.5)
> p1 <- c(0.11, 0.25, 0.40, 0.55, 0.75, 0.85)
> p2 <- c(0.05, 0.10, 0.15, 0.25, 0.40, 0.65)
> p3 <- c(0.20, 0.40, 0.60, 0.75, 0.85, 0.95)
> skeletons <- rbind(p1, p2, p3)
> n <- c(3, 3, 3, 9, 3, 0)
> y <- c(0, 0, 1, 1.333333, 3, 0)
>
> ## Example to get the ET score 1 on dose 3
> ## Assume three patients their corresponding score on the dose 3 is
> ## 0.5, 0.5 and 0.5. Then we calculate ET score as this:
> ## (0.5 + 0.5 + 0.5) / 1.5 = 1
>
> ## Example to get the ET score 1.333333 on dose 4
```

```
> ## Assume nine patients their corresponding score on the dose 4 is
> ## 0, 0, 0, 0, 0, 0, 0.5, 0.5 and 1. Then we calculate ET score as this:
> ## (0 + 0 + 0 + 0 + 0 + 0 + 0.5 + 0.5 + 1) / 1.5 = 1.333333
>
> next_RQ_CRM(target = target, n = n, y = y, dose.curr = 5,
+             score = score, skeleton = skeletons)
[1] 4
```

(2) Next dose for the gBOIN design

```
> next_gBOIN_continuous(target, n, y, d)
```

This function applies to the gBOIN design for the continuous endpoints for quasi-binary endpoints. Among the input arguments, `target` is for the toxicity target score, `n` is the number of patients treated at each dose level, `y` is the toxicity score at each dose level, and `d` is the current dose.

An example for using this function is shown below:

```
> target <- 1.47
> n <- c(3, 3, 3, 9, 0, 0)
> y <- c(0.1951265, 1.5434317, 2.1967343, 13.9266838, 0, 0)
> d <- 4
> next_gBOIN_Continuous(target = target, n = n, y = y, d = d)
[1] 4
```

There are also two functions, `next_gBOIN_TB()` and `next_QuasiBOIN()`, for the gBOIN design that accept the continuous endpoints (e.g., TBS) or quasi-binary endpoints (e.g., ETS) as input parameters, respectively. For example:

```
> target <- 3.344
> n <- c(3, 9, 6, 0, 0, 0, 0, 0, 0, 0)
> y <- c(5.5, 26.95, 25.3, 0, 0, 0, 0, 0, 0, 0)
> d <- 2
> next_gBOIN_TB(target = target, n = n, y = y, d = d)
[1] 2

> target <- 0.47 / 1.5
> n <- c(3, 3, 6, 3, 3, 0)
> y <- c(0, 0, 1.333333, 0, 1, 0)
> d <- 5
> next_QuasiBOIN(target = target, n = n, y = y, d = d)
[1] 5
```

(3) Next dose for the Ivanova design

```
> next_Ivanova_continuous(target, eps, c_resp, n, d)
```

Given the relatively inferior performance of the Ivanova design for ordinal and binary outcomes, this function applies the Ivanova design for the continuous endpoints. Among the input arguments: `target` is for the target toxicity score, `eps` is for the decision criterion, that is, the design parameter Δ in section 2.3, `c_resp` is for the observed continuous values for each dose level, `n` is for the number of patients enrolled at each dose level, and `d` is the current dose level.

Now, we show an example of using this function below:

```
> target <- 1.47
> eps <- 1
> c_resp <- list(c(0, 0.05475884, 0.12446843, 0.10131912),
+                c(0, 0.4716962, 0.2792428, 0.3296575),
+                c(0, 0.3931168, 1.6116607, 0.1642561),
+                c(0, 0.9410027, 1.6021326, 1.6115235,
+                  1.1735981, 2.5575655, 1.6513679, 1.4269044,
+                  0.8983843, 2.2209587),
+                0,
+                0)
```

```
> n <- c(3, 3, 3, 9, 0, 0)
> d <- 4
> next_Ivanova_continuous(target = target, eps = eps, c_resp = c_resp,
+                         n = n, d = d)
[1] 4
```

The above example shows that a trial has 6 dose levels, the continuous target is 1.47, and at that moment, there were 3, 3, 3 and 9 patients treated at dose levels 1 ,2, 3, and 4, respectively, with the current dose level being 4. The recommended design parameter $\Delta = 1$ was used and when executing to `next_Ivanova_continuous()` function, the Ivanova design recommended de-escalating to dose level 4 for the next cohort of patients.

3.2. Select the target dose when the trial is completed

The functions beginning with `select_···` are used to choose the target dose defined by the continuous (e.g., TTB) or quasi-binary endpoints (e.g., ETS) at the end of the trial.

(1) Select the target dose for the Quasi-CRM design

```
> select_mtd_RQ_CRM(target, n, y, score, skeleton)
```

This function applies Quasi-CRM and Robust-Quasi-CRM designs for the ETS-defined MTD target. Among the input arguments: `target` is for the toxicity target score, `n` is for the number of patients treated at each dose level, `y` is the toxicity score at each dose level, `score` is the weighted vector for ordinal toxicity levels introduced in the Quasi-CRM design section, `skeleton` is a skeleton for the Quasi-CRM design or a matrix with multiple-skeletons for the Robust-Quasi-CRM design.

An example for using this function is shown below:

```
> target <- 0.47
> score <- c(0, 0.5, 1, 1.5)
> p1 <- c(0.11, 0.25, 0.40, 0.55, 0.75, 0.85)
> p2 <- c(0.05, 0.10, 0.15, 0.25, 0.40, 0.65)
> p3 <- c(0.20, 0.40, 0.60, 0.75, 0.85, 0.95)
> skeletons <- rbind(p1, p2, p3)
> n <- c(3, 3, 3, 9, 3, 0)
> y <- c(0, 0, 1, 1.333333, 3, 0)
>
> ## Example to get the ET score 1 on dose 3
> ## Assume three patients their corresponding score on the dose 3 is
> ## 0.5, 0.5 and 0.5. Then we calculate ET score as this:
> ## (0.5 + 0.5 + 0.5) / 1.5 = 1
>
> ## Example to get the ET score 1.333333 on dose 4
> ## Assume nine patients their corresponding score on the dose 4 is
> ## 0, 0, 0, 0, 0, 0, 0.5, 0.5 and 1. Then we calculate ET score as this:
> ## (0 + 0 + 0 + 0 + 0 + 0 + 0.5 + 0.5 + 1) / 1.5 = 1.333333
>
> select_mtd_RQ_CRM(target = target, n = n, y = y, score = score,
+                   skeleton = skeletons)
[1] 4
```

(2) Select the target dose for the gBOIN design

The following function applies to the continuous outcomes of the gBOIN design.

```
> select_mtd_gBOIN_continuous(target, npts, ntox)
```

Among the input arguments: `target` is for the toxicity target score, `npts` is for the number of patients treated at each dose level, `ntox` is the toxicity score at each dose level.

An example for using this function is shown below:

```
> target <- 1.47
> n <- c(3, 3, 3, 9, 0, 0)
> y <- c(0.1951265, 1.5434317, 2.1967343, 13.9266838, 0, 0)
> select_mtd_gBOIN_continuous(target = target, npts = n, ntox = y)
[1] 4
```

There are also two functions, `select_mtd_gBOIN_TB()` and `select_mtd_QuasiBOIN()`, for the gBOIN design that accept the continuous (e.g., ETS) as input parameters. The examples are not shown here, but can be found in the package.

(3) Select the target dose for the Ivanova design

```
> select_mtd_Ivanova_continuous(target, c_resp, n)
```

This function applies the Ivanova design for continuous outcomes. The input arguments are similar to the corresponding `next_` function: `target` is for the target toxicity score, `c_resp` is for observed continuous values for each dose level, `n` is for the number of patients treated at each dose level.

Now, we show an example of using this function below:

```
> target <- 1.47
> c_resp <- list(c(0, 0.05475884, 0.12446843, 0.10131912),
+                c(0, 0.4716962, 0.2792428, 0.3296575),
+                c(0, 0.3931168, 1.6116607, 0.1642561),
+                c(0, 0.9410027, 1.6021326, 1.6115235,
+                  1.1735981, 2.5575655, 1.6513679, 1.4269044,
+                  0.8983843, 2.2209587),
+                0,
+                0)
> n <- c(3, 3, 3, 9, 0, 0)
> select_mtd_Ivanova_continuous(target = target, c_resp = c_resp, n = n)
$dselect
[1] 4

$n
[1] 3 3 3 9 0 0
```

3.3. Generate operating characteristics

The functions beginning with `get_oc_···` are used to obtain the operating characteristics of the dose-finding design. This function should be used to assess trial performance for the design of clinical studies.

(1) Obtain the OC for the Quasi-CRM design

```
> get_oc_RQ_CRM(ptox, skeletons, target, score, cohortsize, ncohort,
start.dose = 1, seed = 100)
```

This function applies Quasi-CRM and Robust-Quasi-CRM designs for the equivalent toxicity score defined MTD target. Among the input arguments, `ptox` is true toxicity probability at each dose level, `target` is for the toxicity target score, `cohortsize` is the cohort size, `ncohort` is the number of cohorts, `score` is the weighted vector for ordinal toxicity levels introduced in the Quasi-CRM design section, `skeleton` is a skeleton for the Quasi-CRM design or a matrix with multiple-skeletons for the Robust Quasi-CRM design, `start.dose` is the starting dose level, and `seed` is the seed setting for replicating the simulating results.

An example of using this function is shown as below:

```
> ### Scenario 1 in Yuan {\it et al.} (2007) and Pan {\it et al.} (2014)
> target <- 0.47
> score <- c(0, 0.5, 1, 1.5)
> cohortsize <- 3
```

```
> ncohort <- 10
> ntrial <- 10
>
> ptox <- matrix(nrow = 4, ncol = 6)
> ptox[1,] <- c(0.83, 0.75, 0.62, 0.51, 0.34, 0.19)
> ptox[2,] <- c(0.12, 0.15, 0.18, 0.19, 0.16, 0.11)
> ptox[3,] <- c(0.04, 0.07, 0.11, 0.14, 0.15, 0.11)
> ptox[4,] <- c(0.01, 0.03, 0.09, 0.16, 0.35, 0.59)
> ### specify one skeleton (Quasi-CRM design)
> p1 <- c(0.11, 0.25, 0.40, 0.55, 0.75, 0.85)
>
> get_oc_RQ_CRM(ptox = ptox, skeletons = p1, target = target,
+               score = score, cohortsize = cohortsize,
+               ncohort = ncohort, ntrial = ntrial)
$selpercent
[1]  0  0 40 60  0  0

$nptsdose
[1]  3.3  4.8 10.5 10.5  0.9  0.0
> ### specify three skeletons (Quasi-CRM design)
> p1 <- c(0.11, 0.25, 0.40, 0.55, 0.75, 0.85)
> p2 <- c(0.05, 0.10, 0.15, 0.25, 0.40, 0.65)
> p3 <- c(0.20, 0.40, 0.60, 0.75, 0.85, 0.95)
> skeletons <- rbind(p1, p2, p3)
>
> get_oc_RQ_CRM(ptox = ptox, skeletons = skeletons, target = target,
+               score = score, cohortsize = cohortsize,
+               ncohort = ncohort, ntrial = ntrial)
$selpercent
[1]  0  0 30 60 10  0

$nptsdose
[1]  3.3  4.8  7.5 12.6  1.8  0.0
```

(2) Obtain the OC for the gBOIN design

The following function is for a continuous outcome in the gBOIN design.

```
> get_oc_gBOIN_continuous(target, c_true, ncohort, cohortsize, ntrial, startdose = 1)
```

Among the input arguments: `target` is for the toxicity target score, `c_true` is for the true mean value of the continuous measure, `ncohort` is the number of cohorts, the `cohort size` is the cohort size, `ntrial` is the number of simulated trials, `startdose` is the starting dose level, and, `seed` is the seed setting for replicating the simulating results.

An example for using this function is shown below:

```
> target <- 1.47
> c_true <- c(0.11, 0.25, 0.94, 1.47, 2.38, 2.40)
> ncohort <- 10
> cohortsize <- 3
> ntrial <- 4000
> get_oc_gBOIN_continuous(target = target, c_true = c_true,
+                         ncohort = ncohort, cohortsize = cohortsize,
+                         ntrial = ntrial)
$selpercent
[1]  0.000  0.000 13.550 79.125  6.725  0.600

$nptsdose
[1]  3.00000  3.22050  8.35650 12.29325  2.91000  0.21975
```

Using the gBOIN design, there are also two functions, `get_oc_gBOIN_TB()` and `ge_oc_QuasiBOIN()`, for the gBOIN design to accept the continuous outcomes or ordinal-converted, quasi-binary outcomes as input parameters. We omitted the examples here, but they can be found in the package by running the sample code.

(3) Obtain the OC for the Ivanova design

```
> get_oc_Ivanova_continuous(target, ptox, ncohort, cohortsize,
ntrial, startdose = 1)
```

This function applies the Ivanova design for continuous outcomes. The input arguments: `target` is for the target toxicity score, `ptox` is the true mean value of the continuous measure, `ncohort` is the number of cohorts, `cohortsize` is the cohort size, `ntrial` is the number of simulated trials, `startdose` is the starting dose level, and `seed` is the seed setting for replicating the simulating results.

Now, we show an example of using this function below:

```
> target <- 1.47
> ptox <- c(0.11, 0.25, 0.94, 1.47, 2.38, 2.40)
> ncohort <- 10
> cohortsize <- 3
> ntrial <- 4000
> get_oc_Ivanova_continuous(target = target, ptox = ptox, ncohort = ncohort,
+                           cohortsize = cohortsize, ntrial = ntrial)
$selpercent
[1]  0.000  0.000 16.625 72.450  9.625  1.300

$nptsdose
[1]  3.0000  3.2055  8.3700 12.4770  2.7165  0.2310
```

This function returns the operating characteristics of the Ivanova design as a list that includes: (1) selection percentage at each dose level ($selpercent), (2) the number of patients treated at each dose level ($nptsdose).

4. A trial example

Example 1. In this illustrative example, we used the `UnifiedDoseFinding` package to apply the gBOIN design in the setting of an on-going phase I/II study at St.Jude Children's Research Hospital. In its phase I part, the primary objective is to evaluate the safety of combining intravenous atezolizumab, every three weeks with daily oral cyclophosphamide, pharmacokinetic (PK)-guided sorafenib, and IV bevacizumab once every 3 weeks in children, adolescents and young adults (AYA) with relapsed or refractory solid malignancies. Instead of a conventional dose-finding trial for the MTD, pharmacokinetic outcomes (PK) of this study will be obtained and the targeted dose is defined to be the area under the curve (AUC) between 20 and 55 hrg/mL until day 21 of the course. Phase II used a two-stage design and the endpoint is the response rate. Here, we show how to employ the gBOIN design to design the phase I part of this study.

From the derived boundaries of (2.4) in Section 2.2, it is known that we can use recommended default ϕ_1 and ϕ_2, or choose values for them based on the specific clinical scenario.

In this study, the AUC target has been specified by a fixed interval. Though different from the setting in the original gBOIN design, we adopted the following procedures so that the gBOIN design method can still be properly applied. First, we will take two targets $\phi_0^\ell = 20$ hrg/mL and $\phi_0^r = 55$ hrg/mL that correspond to the left and right endpoints of this interval, respectively. Second, we can take, $H_1 : \phi_1 = 0.8 * \phi_0^\ell$ hrg/mL and $H_2 : \phi_2 = 1.2 * \phi_0^r$ hrg/mL. In other words, if the value of AUC is less than ϕ_1 hrg/mL, the dose would be safe, and if AUC is greater than ϕ_2 hrg/mL, the dose would be over-toxic; otherwise, the dose is very likely to fall into the target interval. Therefore, the decision

Example of a trial with continuous outcomes with subjects assigned in cohorts of three. The target AGT activity is 5 fmol/mg protein. Data were resampled from Friedman et al. (1998). The new design with $\Delta = 1$ is used.

	Y_j	t-statistic	Decision
Dose 1, 40 mg/m^2			
$(y_{11}, y_{12}, y_{13}) = (26.35, 42.00, 15.00)$	27.78	2.91	Increase the dose to 60 mg/m^2
Dose 2, 60 mg/m^2			
$(y_{21}, y_{22}, y_{23}) = (23.00, 13.50, 10.83)$	15.78	2.92	Increase the dose to 80 mg/m^2
Dose 3, 80 mg/m^2			
$(y_{31}, y_{32}, y_{33}) = (11.70, 9.03, 5.00)$	8.58	1.84	Increase the dose to 100 mg/m^2
Dose 4, 100 mg/m^2			
$(y_{41}, y_{42}, y_{43}) = (4.07, 5.00, 8.70)$	5.92	0.65	Repeat the dose at 100 mg/m^2
$(y_{41}, y_{45}, y_{46}) = (2.50, 4.07, 6.13)$	5.08[a]	0.09[a]	Repeat the dose at 100 mg/m^2
$(y_{47}, y_{48}, y_{49}) = (3.60, 5.00, 5.00)$	4.90[a]	−0.18[a]	Repeat the dose at 100 mg/m^2
$(y_{410}, y_{411}) = (6.80, 6.60)$	5.22[a]	0.43[a]	Repeat the dose at 100 mg/m^2

[a]Calculated based on the combined sample.

Figure 1: *Table 1 from Ivanova and Kim (2009).*

boundaries can be calculated accordingly,

$$\lambda_e^* = \frac{\phi_0^\ell + \phi_1}{2} = \frac{20 + 16}{2} = 18,$$
$$\lambda_d^* = \frac{\phi_0^r + \phi_2}{2} = \frac{55 + 66}{2} = 60.5.$$

By calculating $\hat{\mu}_j$, the averaged AUC at the current treated dose level j, we can assign a dose to the next cohort of patients as follows:

- if $\hat{\mu}_j \le \lambda_e^*$, then escalate the dose level to $j + 1$,
- if $\hat{\mu}_j \ge \lambda_d^*$, then escalate the dose level to $j - 1$,
- otherwise, if, i.e., $\hat{\lambda}_e^* < \hat{\mu}_j < \lambda_d^*$, retain the current dose level j.

The decision of dose escalation and de-escalation involves only a simple comparison of the sample mean of the AUC with two pre-specified dose escalation and de-escalation boundaries. Although the gBOIN design with the non-binary DLT endpoint was eventually not used for this study, it provides an example of how to implement the gBOIN design approach for a continuous outcome.

Example 2. Now we show how to use the Ivanova method to conduct a real study. Data of this study was originally from Friedman *et al.* (1998), and we use Figure 1 (Table 1 from Ivanova and Kim (2009)) which shows the data of AGT activity and t-statistic that is used to make decisions for each dose cohort, and show how this trial was conduct. We see that there are four dose levels: 40 mg/m^2, 60 mg/m^2, 80 mg/m^2, and 100 mg/m^2. The goal is to find a dose with AGT activity equal to 5 fmol/mg of protein, which has a target continuous value of 5.

It should be noted that in this study, the AGT activity was believed to be decreasing with dose, and therefore, in the following codes, we take the negative sign for the target and response values. This is because the decision rules based on calculated t-statistics by Ivanova's method are to escalate if the t-statistic is small while to de-escalate if the t-statistic is large for this example.

We now show how to use the replicate, `next_Ivanova_continuous`, to replicate the decisions shown in Figure 1 (Table 1 from Ivanova and Kim (2009)). At a Dose 1 of 40 mg/m^2, the observed ACG activity measurements from the first three patients are $(y_{11}, y_{12}, y_{13}) = (26.35, 42.00, 15.00)$.

From the below output, we see that the next recommended dose is Dose 2, and thus, the decision is to escalate.

```
> target<- -5
> eps <- 1
> c_resp<-list(c(-26.35,-42,-15))
> n <- c(3,0,0,0)
> d <- 1
> next_Ivanova_continuous(target = target, eps = eps, c_resp = c_resp,
+                         n = n, d = d)
[1] 2
```

Then, the study enrolled three patients at a Dose 2 of 60 mg/m^2. If the observed ACG activity measurements from these three patients are $(y_{21}, y_{22}, y_{23}) = (23.00, 13.50, 10.83)$, then we see from the below output, the next recommended dose is Dose 3, and thus, the decision is to escalate.

```
> c_resp<-list(c(-26.35,-42,-15),c(-23,-13.5,-10.83))
> n <- c(3,3,0,0)
> d <- 2
> next_Ivanova_continuous(target = target, eps = eps, c_resp = c_resp,
+                         n = n, d = d)
[1] 3
```

The rest of decisions made from Figure 1 can be see evidently by the following codes and outputs.

Enroll three patients at a Dose 3 of 80 mg/m^2 with the observed three ACG activity measurements being $(y_{31}, y_{32}, y_{33}) = (11.70, 9.03, 5.00)$, and thus, the decision is to escalate to Dose 4.

```
> c_resp<-list(c(-26.35,-42,-15),c(-23,-13.5,-10.83),c(-11.7,-9.03,-5))
> n <- c(3,3,3,0)
> d <- 3
> next_Ivanova_continuous(target = target, eps = eps, c_resp = c_resp,
+                         n = n, d = d)
[1] 4
```

Enroll three patients at a Dose 4 of 100 mg/m^2 with the observed three ACG activity measurements being $(y_{41}, y_{42}, y_{43}) = (4.07, 5.00, 8.70)$, and thus, the decision is to stay at Dose 4.

```
> c_resp<-list(c(-26.35,-42,-15),c(-23,-13.5,-10.83),c(-11.7,-9.03,-5),c(-4.07,-5,-8.7))
> n <- c(3,3,3,3)
> d <- 4
> next_Ivanova_continuous(target = target, eps = eps, c_resp = c_resp,
+                         n = n, d = d)
[1] 4
```

Continue to enroll the three patients at Dose 4 and the observed three ACG activity measurements are $(y_{44}, y_{44}, y_{44}) = (2..50, 4.07, 6.13)$, then the decision is to still stay at Dose 4, etc.

```
> c_resp<-list(c(-26.35,-42,-15),c(-23,-13.5,-10.83),c(-11.7,-9.03,-5),
+              c(-4.07,-5,-8.7,-2.50,-4.07,-6.13))
> n <- c(3,3,3,6)
> d <- 4
> next_Ivanova_continuous(target = target, eps = eps, c_resp = c_resp,
+                         n = n, d = d)
[1] 4
```

Example 3. We show how to use the Quasi-CRM method for a phase I trial of a pre-surgical gemcitabine with an external beam radiation (EBR) for patients with soft tissue sarcoma in a real trial background. This trial was introduced in Bekele and Thall (2004). In this study, each patient received a fixed dose of 50 cGy external beam radiation and 1 of 10 doses of gemcitabine, 100, 200, ... or 1,000 mg/m^2. There were five types of toxicity considered: myelosuppression, dermatitis, liver, nausea/vomiting, and fatigue. In this example, we only focused on the myelosuppression toxicity as an

example. As noted in Bekele and Thall (2004), only the first 11 patients were real; the remaining 25 were hypothetical. Therefore, we also hypothesized the observed ET scores for the patients.

To be specific in this trial, for the myelosuppression toxicity with fever or not, there were five grades from grade 0 to 5. Grades 0 and 1 are combined into one grade and denoted as grade 01. From discussions with the clinical investigator, the targeted toxicity profile for grade 01, 2, 3, and 4 is 0.39, 0.28, 0.20, and 0.13, respectively. If we assign the toxicity scores of 0, 0.5, 1, and 1.5 to the four toxicity grades, we can obtain the target ET score for this study as: $0.39 \times 0 + 0.28 \times 0.5 + 0.20 \times 1 + 0.13 \times 1.5 = 0.535$. This target ET score is equivalent to a target DLT rate of 33% in a conventional binary toxicity case if the investigator will dichotomize the multiple toxicities and only attribute $\geq$ grade 3 as the DLT.

To use the R package, we first elicit a set of 3 skeletons using the `getprior(.)` function from the `dfcrm` package as shown below.

```
> library(dfcrm)
> p1 <- getprior(halfwidth = 0.10, target = 0.33, nu = 4, nlevel=6)
> p2 <- getprior(halfwidth = 0.10, target = 0.33, nu = 5, nlevel=6)
> p3 <- getprior(halfwidth = 0.10, target = 0.33, nu = 6, nlevel=6)
>
> p1;p2;p3
[1] 0.00286723 0.03466833 0.14506007 0.33000000 0.52905862 0.69377785
[1] 3.736508e-05 2.867230e-03 3.466833e-02 1.450601e-01 3.300000e-01 5.290586e-01
[1] 1.949679e-08 3.736508e-05 2.867230e-03 3.466833e-02 1.450601e-01 3.300000e-01
```

Scores of 0, 0.5, 1, and 1.5 corresponding to grade 0/1, 2, 3, and 4 were assigned by the investigator.

```
score <- c(0, 0.5, 1, 1.5)
```

The trial enrolled 3 patients for the first cohort at dose 1. After the first treatment course (4 weeks), the 1st and 2nd patients experienced grade 0 toxicity, and the 3rd patient experienced grade 1 toxicity. Thus, the computed normalized score value for the dose was (0 + 0 + 0)/1.5 = 0 since either toxicity grade 0 or 1 has the pre-assigned score value of 0. Therefore, the observed data at dose 1 can be input like below:

```
n <- c(3, 0, 0, 0, 0, 0)
y <- c(0, 0, 0, 0, 0, 0)
```

Then, we can use the below code to find the dose transition decision.

```
> next_RQ_CRM(target = target, n = n, y = y, dose.curr = 1,
+             score = score, skeleton = skeletons)
[1] 2
```

The decision is to escalate and we enrolled three patients for the 2nd cohort at dose 2. Among these patients, the 4th and 6th patients had experienced grade 0 toxicity, and the 5th patient had experienced grade 1 toxicity. Thus, the normalized score was (0 + 0 + 0)/1.5 = 0. The codes for the dose transition decision-making are as follows:

```
> n <- c(3, 3, 0, 0, 0, 0)
> y <- c(0, 0, 0, 0, 0, 0)
> next_RQ_CRM(target = target, n = n, y = y, dose.curr = 2,
+             score = score, skeleton = skeletons)
[1] 3
```

From the above output, the study escalated to dose 3 and among the three patients, the 7th patient had experienced grade 1 toxicity and the 8th and 9th patients had experienced grade 2 toxicity. Thus, the computed normalized score was (0 + 0.5 + 0.5)/1.5 = 0.6666667. The following codes show the dose transition decision making:

```
> n <- c(3, 3, 3, 0, 0, 0)
> y <- c(0, 0, 0.6666667, 0, 0, 0)
> next_RQ_CRM(target = target, n = n, y = y, dose.curr = 3,
+             score = score, skeleton = skeletons)
[1] 4
```

The decision is to escalate to dose 4. For the fourth cohort of the three patients, the 10th patient had experienced grade 2 toxicity, the 11th experienced grade 1 toxicity and the 12th experienced grade 3 toxicity. The computed normalized score was (0.5 + 0 + 1)/1.5 = 1. The following codes show the dose transition decision making.

```
> n <- c(3, 3, 3, 3, 0, 0)
> y <- c(0, 0, 0.6666667, 1, 0, 0)
> next_RQ_CRM(target = target, n = n, y = y, dose.curr = 4,
+             score = score, skeleton = skeletons)
[1] 4
```

The study continued to enroll the fifth cohort and treat patients at dose 4. For the newly enrolled three patients, the 13th patient had experienced grade 2 toxicity, the 14th experienced grade 3 toxicity and the 15th experienced grade 1 toxicity. The computed normalized score was (0.5 + 0 + 1 + 0.5 + 1 + 0)/1.5 = 2.

```
> n <- c(3, 3, 3, 6, 0, 0)
> y <- c(0, 0, 0.6666667, 2, 0, 0)
> next_RQ_CRM(target = target, n = n, y = y, dose.curr = 4,
+             score = score, skeleton = skeletons)
[1] 4
```

This study continued to treat patients at dose 4 and finally stopped the study when 15 patients were enrolled. To be specific, for the sixth cohort, the 16th patient had experienced grade 1 toxicity, the 17th experienced grade 1 toxicity and the 18th experienced grade 2 toxicity. The computed normalized score was (0.5 + 0 + 1 + 0.5 + 1 + 0 + 0 + 0 + 0.5)/1.5 = 2.3333333. For the seventh cohort, the 19th patient experienced grade 1 toxicity, the 20th patient experienced grade 3 toxicity and the 21st patient experienced grade 0 toxicity. The computed normalized score was (0.5 + 0 + 1 + 0.5 + 1 + 0 + 0 + 0 + 0.5 + 0 + 1 + 0)/1.5 = 3. For the eighth cohort, the 22nd patient experienced grade 3 toxicity, the 23rd patient experienced grade 3 toxicity and the 24th experienced grade 4 toxicity. The computed normalized score was (0.5 + 0 + 1 + 0.5 + 1 + 0 + 0 + 0 + 0.5 + 0 + 1 + 0 + 1 + 1 + 1.5)/1.5= 5.3333333. The following codes show the inputs and outputs based on this information.It must be noted that $5.3333333 \times 1.5/15 = 0.533$ is close to our target ET score of 0.535.

```
> #for the sixth cohort
> n <- c(3, 3, 3, 9, 0, 0)
> y <- c(0, 0, 0.6666667, 2.3333333, 0, 0)
> next_RQ_CRM(target = target, n = n, y = y, dose.curr = 4,
+             score = score, skeleton = skeletons)
[1] 4
>
> #for the seventh cohort
> n <- c(3, 3, 3, 12, 0, 0)
> y <- c(0, 0, 0.6666667, 3, 0, 0)
> next_RQ_CRM(target = target, n = n, y = y, dose.curr = 4,
+             score = score, skeleton = skeletons)
[1] 4
>
> #for the eighth cohort
> n <- c(3, 3, 3, 15, 0, 0)
> y <- c(0, 0, 0.6666667, 5.3333333, 0, 0)
> next_RQ_CRM(target = target, n = n, y = y, dose.curr = 4,
+             score = score, skeleton = skeletons)
[1] 4
```

5. Conclusion

The `UnifiedDoseFinding` package implements novel methods for phase I dose-finding clinical trials with non-binary endpoints. In this package, three methods, the Quasi-CRM (Yuan *et al.*, 2007, Pan *et al.*, 2014), the gBOIN design (Mu *et al.*, 2019), and the Ivanova design (Ivanova *et al.*, 2009) were implemented for the first time in a publicly available platform. For the quasi-binary outcomes that converted from the ordinal grade toxicity such as the equivalent toxicity score, we can use either the Quasi-CRM method or the gBOIN method. When the outcomes are continuous endpoints such as the total toxicity burden, both the gBOIN method and the Ivanova method can be applied. The functions starting with `next_` and `select_` for each method can be used during a prospective adaptive trial, where the dose for the next cohorts of patients depends on the outcomes of the previous cohorts, so that the dose transition for the next cohort of patients or the MTD identification for further clinical studies can be spontaneously carried out. The functions starting with `get_oc_(...)` can be used to perform simulations in the planning stage of a trial to study the robustness of the methods by using different parameters setting choices. We also provided three real examples with details for each of the methods to help readers adapt a more practical perspective. Overall, the `UnifiedDoseFinding` package is user-friendly and provides a useful tool for the dose-finding Phase I trials with non-binary outcomes.

Acknowledgement

The authors thank the associate editor and two reviewers for very insightful and constructive comments that substantially improved the article. The authors thank Dr. Vani Shanker and this Journal for editorial help on the manuscript. Pan and Hsu's research is partially supported by American Lebanese Syrian Associated Charities (ALSAC). Mu's research is partially supported by the National Natural Science Foundation of China (grant 11901519, grant 11801359) and the China Postdoctoral Science Foundation (grant 2019M661416).

References

Bekele BN and Thall PF (2004). Dose-finding based on multiple toxicities in a soft tissue sarcoma trial, *Journal of the American Statistical Association*, **99**, 26–35.

Brahmer JR, Charles GD, Ira W, *et al.* (2010). Phase I study of single-agent anti–programmed death-1 (MDX-1106) in refractory solid tumors: safety, clinical activity, pharmacodynamics, and immunologic correlates, *Journal of Clinical Oncology*, **8**, 3167.

Dent SF and Eisenhauer EA (1996). Phase I trial design: are new methodologies being put into practice?, *Annals of Oncology*, **7**, 561–566.

Ezzalfani M, Zohar S, Qin R, Mandrekar SJ, and Deley MCL (2013). Dose-finding designs using a novel quasi-continuous endpoint for multiple toxicities, *Statistics in Medicine*, **32**, 2728–2746.

Fiteni F, Le Ray I, Ousmen A, Isambert N, Anota A, and Bonnetain F (2019). Health-related quality of life as an endpoint in oncology phase I trials: a systematic review, *BMC Cancer*, **19**, 1–8.

Frankel PH, Chung V, Xing Y, Longmate J, Groshen S, and Newman EM (2020). Untenable dosing: A common pitfall of modern DLT-targeting Phase I designs in oncology, *Current Problems in Cancer*, **44**, 100583.

Friedman HS, Kokkinakis DM, Pluda J, *et al.* (1998). Phase I trial of O6-benzylguanine for patients undergoing surgery for malignant glioma, *Journal of Clinical Oncology*, **16**, 3570–3575.

Ivanova A and Kim SH (2009) Dose finding for continuous and ordinal outcomes with a monotone

objective function: a unified approach, *Biometrics*, **65**, 307–315.

Le Tourneau C, Lee JJ, and Siu LL (2009). Dose escalation methods in phase I cancer clinical trials, *JNCI: Journal of the National Cancer Institute*, **101**, 708–720.

Lee SM, Hershman DL, Martin P, Leonard JP, and Cheung YK (2012). Toxicity burden score: a novel approach to summarize multiple toxic effects, *Annals of Oncology*, **3**, 537–541.

Liu S and Yuan Y (2015). Bayesian optimal interval designs for phase I clinical trials, *Journal of the Royal Statistical Society: Series C (Applied Statistics)*,**64**, 507–523.

Meille C, Guerreiro N, Jullion A, *et al.* (2017). Abstract CT154: Optimization of the dose and schedule of an HDM2 inhibitor NVP-HDM201 in a first-in-human Phase I study using a mechanism-based PK/PD model, *Cancer Research*, **77**, CT154.

Meter EMV, Elizabeth GM, and Dipankar B (2011). Proportional odds model for dose -finding clinical trial designs with ordinal toxicity grading, *Statistics in Medicine*, **30**, 2070–2080.

Mu R, Yuan Y, Xu J, Mandrekar SJ, and Yin J (2019). gBOIN: a unified model-assisted phase I trial design accounting for toxicity grades, and binary or continuous end points, *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, **68**, 289–308.

Muenz D (2017). New statistical methods for phase I clinical trials of a single agent (Doctoral dissertation), The University of Michigan, Michigan.

National Cancer Institute: CTEP. Common Terminology Criteria for Adverse Events (CTCAE) Version 3.0. Bethesda, MD: National Institutes of Health 2003.

O'Quigley J, Pepe M, and Fisher L (1990). Continual reassessment method: A practical design for Phase I clinical trials in cancer, *Biometrics*, **46**, 33–48.

Pan H, Zhu C, Zhang F, *et al.* (2014). The continual reassessment method for multiple toxicity grades: a Bayesian model selection approach, *PloS One*, **9**, e98147.

Paoletti X, Le Tourneau C, Verweij J, *et al.* (2014). Defining dose-limiting toxicity for phase 1 trials of molecularly targeted agents: results of a DLT-TARGETT international survey, *European Journal of Cancer*, **50**, 2050–2056.

Papke L and Wooldrige J (1996). Econometric methods for fractional response variables with an application to 401(k) plan participation rates, *Journal of Applied Econometrics*, **11**, 619–632.

Plummer R, Jones C, Middleton M, *et al.* (2008). Phase I study of the poly (ADP-ribose) polymerase inhibitor, AG014699, in combination with temozolomide in patients with advanced solid tumors, *Clinical Cancer Research*, **14**, 7917–7923.

Storer BE (1989). Design and analysis of Phase I clinical trials, *Biometrics*, **45**, 925–937.

Yin G and Yuan Y (2009). Bayesian model averaging continual reassessment method in Phase I clinical trials, *Journal of the American Statistical Assoication*, **104**, 954–968.

Yin J, Du Y, Qin R, Shen S, and Mandrekar S (2021). phase1RMD: An R package for repeated measures dose-finding designs with novel toxicity and efficacy endpoints, *Plos One*, **16**, e0256391.

Yuan Z, Chappell R, and Bailey H (2007). The continual reassessment method for multiple toxicity grades: a bayesian Quasi-likelihood approach, *Biometrics*, **63**, 173–179.

Zhang L, Yu Q, Wu XC, Hsieh MC, Loch M, Chen VW, Fontham E, and Ferguson T (2018). Impact of chemotherapy relative dose intensity on cause-specific and overall survival for stage I–III breast cancer: ER+/PR+, HER2-vs. triple-negative, *Breast Cancer Research and Treatment*, **169**, 175–187.

Received November 22, 2021; Revised February 9, 2022; Accepted March 8, 2022