boinet: An R Package for Bayesian Optimal Interval Design for Dose Finding Based on Both Efficacy and Toxicity Outcomes

by Yusuke Yamaguchi, Kentaro Takeda and Kazushi Maruo

Abstract Bayesian optimal interval based on both efficacy and toxicity outcomes (BOIN-ET) design is a model-assisted oncology phase I/II trial design, aiming to establish an optimal biological dose accounting for efficacy and toxicity in the framework of dose-finding. Some extensions of BOIN-ET design are also available to allow for time-to-event efficacy and toxicity outcomes based on cumulative and pending data (time-to-event BOIN-ET: TITE-BOIN-ET), ordinal graded efficacy and toxicity outcomes (generalized BOIN-ET: gBOIN-ET), and their combination (TITE-gBOIN-ET). boinet package is an R package to implement the BOIN-ET design family. The package supports the conduct of simulation studies to assess operating characteristics of BOIN-ET, TITE-BOIN-ET, gBOIN-ET, and TITE-gBOIN-ET designs, where users can choose design parameters in flexible and straightforward ways depending on their own application. We demonstrate its capability and effectiveness using several simulation studies.

1 Introduction

A paradigm shift in anti-cancer treatment with the emergence of molecular targeted agents and immunotherapies is calling for a reform of dose-finding strategy in phase I oncology clinical trials. Traditional dose-finding procedures designed to establish the maximum tolerated dose underlie an assumption of a monotonic dose-efficacy relationship; that is, the higher dose level is assumed to be more toxic and efficacious. However, the monotonic dose-efficacy relationship would be doubtful in many cases of the targeted agents and immunotherapies, where the efficacy may reach a plateau beyond a given dose level or even take an umbrella shape (Calabrese and Baldwin, 2001; LoRusso et al., 2010; Reynolds, 2010). Accordingly, a concept of optimal biological dose (OBD) has been introduced into the framework of dose-finding (Corbaux et al., 2019; Fraisse et al., 2021). The OBD, which is generally defined as a tolerable dose having adequate efficacy under an assumption of non-monotonic dose-efficacy relationships, is established with a dose-escalation rule based jointly on toxicity and efficacy. Numerous dose-finding designs to incorporate the trade-off between toxicity and efficacy have been proposed, broadly classified as algorithm-based designs (Lin and Ji, 2020), model-based designs (Thall and Cook, 2004; Yuan et al., 2016; Riviere et al., 2018), and model-assisted designs (Takeda et al., 2018; Yuan et al., 2019; Li et al., 2020; Lin et al., 2020; Yin and Yuan, 2020; Lin and Ji, 2021). Particularly, the model-assisted designs are known to yield superior performance compared with more complicated model-based designs and to be implemented in as simple a way as the algorithm-based designs (Yuan et al., 2019).

Bayesian optimal interval based on efficacy and toxicity outcomes (BOIN-ET) design proposed by Takeda et al. (2018) is one of the model-assisted designs and has been widely used to establish the OBD, where the dose-escalation decision is determined by minimizing a posterior probability of incorrect dose assignments in terms of both efficacy and toxicity. Furthermore, the BOIN-ET design has been extended in two ways to accommodate important clinical features unique to molecular targeted agents and immunotherapies. Firstly, the occurrence of efficacy responses and/or adverse events may be delayed due to the mechanism of the agents. Such late-onset efficacy and toxicity outcomes usually require a longer assessment window than the typical first one or two cycles for the traditional cytotoxic agents (Postel-Vinay et al., 2009; Weber et al., 2015). Ignoring the feature of late-onset toxicity

may cause potential dose reductions or interruptions in later phase trials. To allow for sequential enrollment even before patients have completed the required efficacy and toxicity assessment window, Takeda et al. (2020) proposed time-to-event BOIN-ET (TITE-BOIN-ET) design, which extends the BOIN-ET design to use cumulative and pending data of both efficacy and toxicity for the dose-escalation decision (Lin and Yuan, 2020). Secondly, the efficacy and toxicity outcomes are often summarized by binary outcomes, scored as an overall response (ORR) or not an ORR for efficacy, and scored as a dose-limiting toxicity (DLT) or not a DLT for toxicity; however, the molecular targeted agents and immunotherapies are more likely to be featured by multiple low or moderate grades of toxicities than DLTs (Penel et al., 2011). Besides, the efficacy response may prefer to be evaluated by the combination of ORR and long-term stable disease (SD) in solid tumors and by accounting for the difference between complete remission (CR) and partial remission (PR) in lymphoma. To allow for the ordinal graded efficacy and toxicity assessment, Takeda et al. (2022a) proposed generalized BOIN-ET (gBOIN-ET), which extends the BOIN-ET design to use ordinal graded efficacy and toxicity data for the dose-escalation decision. Takeda et al. (2023) proposed time-to-event generalized BOIN-ET (TITE-gBOIN-ET) to address simultaneously the aforementioned challenges of the late-onset ordinal graded efficacy and toxicity outcomes. **Keyboard** package (Yuan et al., 2025) and **escalation** package (Brock et al., 2024) are existing R packages that implement other types of the OBD-finding designs, while there are no R packages that provide functions to implement the BOIN-ET, TITE-BOIN-ET, gBOIN-ET, and TITE-gBOIN-ET design.

boinet package (Yamaguchi and Takeda, 2025) is an R package to implement the BOIN-ET design family. The package supports the conduct of simulation studies to assess operating characteristics of BOIN-ET, TITE-BOIN-ET, gBOIN-ET, and TITE-gBOIN-ET, where users can choose design parameters in flexible and straightforward ways depending on their own application. It should be noted that the selection of OBD in the BOIN-ET design family is carried out at the end of the trial after completing the dose-escalation procedure: (i) the toxicity and efficacy probabilities at each dose level are estimated using all data obtained from the entire cohort, and then (ii) given the estimated toxicity and efficacy probabilities, the utility to measure the toxicity-efficacy trade-off is computed and serves as a basis for the OBD selection. The **boinet** package offers several options regarding methods to estimate the efficacy probability and measures to select the OBD. The operating characteristics of the design are summarized by the percentage of times that each dose level is selected as OBD and the average number of patients who are treated at each dose level. The percentage of times that the study is terminated and the expected study duration are also provided. We demonstrate the capability and effectiveness of the **boinet** package.

This article is organized as follows. We first describe the BOIN-ET, TITE-BOIN-ET, gBOIN-ET, and TITE-gBOIN-ET designs implemented by the package. We then provide details on how to use the **boinet** package through simulation studies. The article is concluded with a brief discussion.

2 Case example

The BOIN-ET design family is a relatively new methodological development, and while they are being implemented in ongoing trials, published case results using these specific designs are still forthcoming. We here provide one case study that has actually implemented the TITE-BOIN-ET design.

A two-part phase I/II study evaluating a combination of luspatercept and erythropoiesis-stimulating agent (ESA) in patients with low-risk myelodysplastic syndromes without ring sideroblasts having failed to achieve a response after ESA without disease progression is being conducted (Ades et al., 2024). The primary objective of this dose-finding part of the trial was to determine the optimal dose in terms of both toxicity and efficacy for luspatercept + ESA, for selection in the randomized part B of the trial. Based on a TITE-BOIN-ET design, four dose levels were tested: Dose level 1 (n=3); Dose level 2 (n=3); Dose Level 3 (n=3); Dose

Level 4 (n=15). The toxicity was measured by a dose-limiting toxicity at Day 21 of cycle 1 for non-hematological toxicity, up to Day 42 for hematological toxicity. The efficacy was measured by a response rate (complete response + partial response + stable disease with hematological improvement), and the assessment window for efficacy was 21 days.

3 Methodology

Consider a dose-finding trial aiming to determine the OBD of a test drug, and suppose that J dose levels are investigated by sequentially assigning small cohorts of patients to the next dose levels until reaching a maximum number of patients enrolled. Suppose that patients in the current cohort have been treated at dose level j, and let n_j denote the cumulative number of patients treated at dose level j for $j = 1, \ldots, J$. Let τ_T and τ_E be the assessment window for toxicity and efficacy respectively. We below describe the dose-escalation procedure in the BOIN-ET, TITE-BOIN-ET, gBOIN-ET, and TITE-gBOIN-ET designs, followed by an introduction to the OBD selection. Further details are found in the original articles (Takeda et al., 2018, 2020, 2022b,a, 2023).

3.1 Dose-escalation procedure in BOIN-ET and TITE-BOIN-ET design

Let p_{Tj} and p_{Ej} be a true toxicity and efficacy probability at dose level j for $j = 1, \ldots, J$, respectively. Let ϕ^p and δ^p be a target toxicity and efficacy probability respectively, which are pre-specified by trial physicians according to the disease and the definition of toxicity and efficacy. Lower and upper toxicity boundaries are pre-specified as λ_1^p and λ_2^p respectively, which satisfy a condition of $0 \leq \lambda_1^p \leq \phi^p \leq \lambda_2^p < 1$. A lower efficacy boundary is also pre-specified as η_1^p with a condition of $0 \leq \eta_1^p < \delta^p < 1$. Let $\hat{p}_{Tj}$ and $\hat{p}_{Ej}$ be an observed toxicity and efficacy probability at dose level j respectively, which are calculated using the cumulative toxicity and efficacy data in different ways for the BOIN-ET and TITE-BOIN-ET design (see subsections below). Then, the dose level used in the next cohort is determined by the following rule:

1. If $\hat{p}_{Tj} \leq \lambda_1^p$ and $\hat{p}_{Ej} \leq \eta_1^p$, escalate to dose level $j + 1$.

2. If $\hat{p}_{Tj} < \lambda_2^p$ and $\hat{p}_{Ej} > \eta_1^p$, stay at the same dose level j.

3. If $\hat{p}_{Tj} \geq \lambda_2^p$, de-escalate to dose level $j - 1$.

4. If $\lambda_1^p < \hat{p}_{Tj} < \lambda_2^p$ and $\hat{p}_{Ej} \leq \eta_1^p$, consider the following additional rules:

   4-1 If dose level $j + 1$ has never been used until the current cohort, escalate to dose level $j + 1$.

   4-2 Otherwise, choose $\arg\max_{j' \in \{j-1, j, j+1\}} \hat{p}_{Ej'}$ as the dose level for the next cohort; if more than two dose levels have the same highest observed efficacy probability, randomly select one of them.

Optimal values of $(\lambda_1^p, \lambda_2^p, \eta_1^p)$ are determined as values that minimize a posterior probability of incorrect decisions under the following six hypotheses:

$$H_{1j} : (p_{Tj} = \phi_1^p, p_{Ej} = \delta_1^p), \quad H_{2j} : (p_{Tj} = \phi_1^p, p_{Ej} = \delta^p), \quad H_{3j} : (p_{Tj} = \phi^p, p_{Ej} = \delta_1^p),$$
$$H_{4j} : (p_{Tj} = \phi^p, p_{Ej} = \delta^p), \quad H_{5j} : (p_{Tj} = \phi_2^p, p_{Ej} = \delta_1^p), \quad H_{6j} : (p_{Tj} = \phi_2^p, p_{Ej} = \delta^p)$$

where ϕ_1^p is the highest toxicity probability that is deemed sub-therapeutic such that dose-escalation should be pursued, and ϕ_2^p is the lowest toxicity probability that is deemed overly toxic such that dose de-escalation is needed. δ_1^p is the minimum probability deemed efficacious such that the dose levels with less than δ_1^p are considered sub-therapeutic. Here, ϕ_1^p, ϕ_2^p, and δ_1^p are design parameters to be pre-specified along with the target probabilities of ϕ^p and δ^p. Given non-informative prior probabilities of the six hypotheses; i.e., $Pr(H_{1j}) =$

$Pr(H_{2j}) = Pr(H_{3j}) = Pr(H_{4j}) = Pr(H_{5j}) = Pr(H_{6j}) = 1/6$, a grid search approach can be used to find the optimal values of $(\lambda_1^p, \lambda_2^p, \eta_1^p)$ conditioned on $\phi_1^p < \lambda_1^p < \phi^p < \lambda_2^p < \phi_2^p$ and $\delta_2^p < \eta_1^p < \delta^p$. For example, considering $\phi^p = 0.3$, $\phi_1^p = 0.1\phi^p$, $\phi_2^p = 1.4\phi^p$, $\delta^p = 0.6$, and $\delta_1^p = 0.6\delta^p$, the optimal values of $(\lambda_1^p, \lambda_2^p, \eta_1^p)$ are given by $\lambda_1^p = 0.14$, $\lambda_2^p = 0.35$, and $\eta_1^p = 0.48$.

A key difference between the BOIN-ET and TITE-BOIN-ET design is the timing to determine the dose level used in the next cohort and to start the enrollment for the next cohort. The BOIN-ET design requires all the patients enrolled in the current cohort to complete the toxicity and efficacy assessment windows; in other words, the decision time is always the time when the last patient enrolled in the cohort completes the toxicity and efficacy assessment windows. In contrast, the TITE-BOIN-ET design allows for flexibility in the decision time, where the dose escalation decision is based on the cumulative data at a certain decision time. Differences in the decision time between the BOIN-ET and TITE-BOIN-ET designs lead to different calculations of the observed toxicity and efficacy probabilities, $\hat{p}_{Tj}$ and $\hat{p}_{Ej}$ for $i = 1, \ldots, n_j$, which are described below.

BOIN-ET (Takeda et al., 2018): Let y_{Tij} and y_{Eij} be a binary toxicity and efficacy outcome observed within the assessment windows of τ_T and τ_E for the i-th patient at dose level j, respectively. Given the cumulative toxicity and efficacy data for dose level j; i.e., $\{y_{Tij} : i = 1, \ldots, n_j\}$ and $\{y_{Eij} : i = 1, \ldots, n_j\}$, we have $\hat{p}_{Tj} = \sum_{i=1}^{n_j} y_{Tij}/n_j$ and $\hat{p}_{Ej} = \sum_{i=1}^{n_j} y_{Eij}/n_j$.

TITE-BOIN-ET (Takeda et al., 2020; Lin and Yuan, 2020): Let $\tilde{y}_{Tij}$ be a binary toxicity outcome observed at a certain decision time for the i-th patient at dose level j, where $\tilde{y}_{Tij} = 1$ if the patient has experienced the toxicity and $\tilde{y}_{Tij} = 0$ if the patient has not yet experienced the toxicity. Let γ_{Tij} denote that $\tilde{y}_{Tij}$ is ascertained ($\gamma_{Tij} = 1$) or is still pending ($\gamma_{Tij} = 0$) at the decision time, and u_{Tij} denote the time to the toxicity outcome if $\tilde{y}_{Tij} = 1$ and to censoring if $\tilde{y}_{Tij} = 0$. Note that $(\tilde{y}_{Tij}, \gamma_{Tij}) = (0, 1)$ indicates the censoring at $u_{Tij} = \tau_T$, and $(\tilde{y}_{Tij}, \gamma_{Tij}) = (0, 0)$ indicates the censoring at $u_{Tij} < \tau_T$. In the same way, $(\tilde{y}_{Eij}, \gamma_{Eij}, u_{Eij})$ is defined for efficacy data of the i-th patient at dose level j. Given the cumulative toxicity and efficacy data for dose level j at the decision time; i.e., $\{(\tilde{y}_{Tij}, \gamma_{Tij}, u_{Tij}) : i = 1, \ldots, n_j\}$ and $\{(\tilde{y}_{Eij}, \gamma_{Eij}, u_{Eij}) : i = 1, \ldots, n_j\}$, we have $\hat{p}_{Tj} = \sum_{i=1}^{n_j} \tilde{y}_{Tij}/\tilde{n}_{Tj}$ and $\hat{p}_{Ej} = \sum_{i=1}^{n_j} \tilde{y}_{Eij}/\tilde{n}_{Ej}$, where $\tilde{n}_{Tj} = \sum_{i=1}^{n_j}(\gamma_{Tij} + (1 - \gamma_{Tij})u_{Tij}/\tau_T)$ and $\tilde{n}_{Ej} = \sum_{i=1}^{n_j}(\gamma_{Eij} + (1 - \gamma_{Eij})u_{Eij}/\tau_E)$.

3.2 Dose-escalation procedure in gBOIN-ET and TITE-gBOIN-ET design

Yuan et al. (2007) proposed an equivalent toxicity score (ETS) to measure the relative severity of different toxicity grades in the dose allocation procedure and established the following severity weights related to DLT: (1) grade 0 and 1 toxicities are not of concern (no DLT); (2) two grade 2 toxicities are equivalent to a grade 3 toxicity (0.5 DLT); (3) a grade 3 toxicity is equivalent to a DLT (1 DLT); and (4) a grade 4 toxicity is equivalent to 1.5 grade 3 toxicities (1.5 DLT). A target ETS is obtained by a weighted sum of overall toxicity grades. Takeda et al. (2022a) extended this concept to an efficacy score named the equivalent efficacy score (EES) to measure the relative effectiveness of different efficacy outcomes and used the relative effectiveness of different efficacy outcomes: (1) progressive disease (PD) is not an effective outcome (no response); (2) an SD is equivalent to 0.25 PRs (0.25 responses); (3) a PR is equivalent to a response (1 response); and (4) a CR is equivalent to 3.0 PRs (3.0 responses). A target EES is obtained by a weighted sum of overall efficacy grades. In addition, the normalized ETS and EES are defined as $\text{ETS}^* = \text{ETS}/\text{ETS}_{max}$ and $\text{EES}^* = \text{EES}/\text{EES}_{max}$ respectively, where ETS_{max} is the ETS for the most severe toxicity grade and EES_{max} is the EES for the most desirable efficacy response. The normalized ETS and EES ranging from 0 to 1 are assumed to follow quasi-Bernoulli distributions.

Let μ_{Tj} and μ_{Ej} be a true quasi-Bernoulli toxicity and efficacy probability at dose level j for $j = 1, \ldots, J$, respectively. Let ϕ^μ and δ^μ be a target quasi-Bernoulli toxicity and efficacy

probability respectively. Lower and upper quasi-Bernoulli toxicity boundaries are pre-specified as λ_1^μ and λ_2^μ respectively, which satisfy a condition of $0 \leq \lambda_1^\mu \leq \phi^\mu \leq \lambda_2^\mu < 1$. A lower quasi-Bernoulli efficacy boundary is also pre-specified as η_1^μ with a condition of $0 \leq \eta_1^\mu < \delta^\mu < 1$. Let $\hat{\mu}_{Tj}$ and $\hat{\mu}_{Ej}$ be an observed quasi-Bernoulli toxicity and efficacy probability at dose level j respectively, which are calculated using the cumulative toxicity and efficacy data in different ways for the gBOIN-ET and TITE-gBOIN-ET design (see subsections below). Then, the dose level used in the next cohort is determined by the following rule:

1. If $\hat{\mu}_{Tj} \leq \lambda_1^\mu$ and $\hat{\mu}_{Ej} \leq \eta_1^\mu$, escalate to dose level $j + 1$.

2. If $\hat{\mu}_{Tj} < \lambda_2^\mu$ and $\hat{\mu}_{Ej} > \eta_1^\mu$, stay at the same dose level j.

3. If $\hat{\mu}_{Tj} \geq \lambda_2^\mu$, de-escalate to dose level $j - 1$.

4. If $\lambda_1^\mu < \hat{\mu}_{Tj} < \lambda_2^\mu$ and $\hat{\mu}_{Ej} \leq \eta_1^\mu$, consider the following additional rules:

    4-1 If dose level $j + 1$ has never been used until the current cohort, escalate to dose level $j + 1$.

    4-2 Otherwise, choose $\arg\max_{j' \in \{j-1, j, j+1\}} \hat{\mu}_{Ej'}$ as the dose level for the next cohort; if more than two dose levels have the same highest observed quasi-Bernoulli efficacy probability, randomly select one of them.

Optimal values of $(\lambda_1^\mu, \lambda_2^\mu, \eta_1^\mu)$ are determined as values that minimize a posterior probability of incorrect decisions under the following six hypotheses:

$$H_{1j} : (\mu_{Tj} = \phi_1^\mu, \mu_{Ej} = \delta_1^\mu), \quad H_{2j} : (\mu_{Tj} = \phi_1^\mu, \mu_{Ej} = \delta^\mu), \quad H_{3j} : (\mu_{Tj} = \phi^\mu, \mu_{Ej} = \delta_1^\mu),$$
$$H_{4j} : (\mu_{Tj} = \phi^\mu, \mu_{Ej} = \delta^\mu), \quad H_{5j} : (\mu_{Tj} = \phi_2^\mu, \mu_{Ej} = \delta_1^\mu), \quad H_{6j} : (\mu_{Tj} = \phi_2^\mu, \mu_{Ej} = \delta^\mu)$$

where ϕ_1^μ is the highest quasi-Bernoulli toxicity probability that is deemed sub-therapeutic such that dose-escalation should be pursued, and ϕ_2^μ is the lowest quasi-Bernoulli toxicity probability that is deemed overly toxic such that dose de-escalation is needed. δ_1^μ is the minimum quasi-Bernoulli probability deemed efficacious such that the dose levels with less than δ_1^μ are considered sub-therapeutic. Given non-informative prior probabilities of the six hypotheses; i.e., $Pr(H_{1j}) = Pr(H_{2j}) = Pr(H_{3j}) = Pr(H_{4j}) = Pr(H_{5j}) = Pr(H_{6j}) = 1/6$, a grid search approach can be used to find the optimal values of $(\lambda_1^\mu, \lambda_2^\mu, \eta_1^\mu)$ conditioned on $\phi_1^\mu < \lambda_1^\mu < \phi^\mu < \lambda_2^\mu < \phi_2^\mu$ and $\delta_2^\mu < \eta_2^\mu < \delta^\mu$. For example, considering $\phi^\mu = 0.313$ (which is the case of target ETS = 0.47 and $ETS_{max} = 1.5$), $\phi_1^\mu = 0.1\phi^\mu$, $\phi_2^\mu = 1.4\phi^\mu$, $\delta^\mu = 0.583$ (which is the case of target EES = 1.75 and $EES_{max} = 3.0$), and $\delta_1^\mu = 0.6\delta^\mu$, the optimal values of $(\lambda_1^\mu, \lambda_2^\mu, \eta_1^\mu)$ are given by $\lambda_1^\mu = 0.14$, $\lambda_2^\mu = 0.37$, and $\eta_1^\mu = 0.46$.

The relationship between the gBOIN-ET and TITE-gBOIN-ET designs is the same as that between the BOIN-ET and TITE-BOIN-ET designs. The gBOIN-ET design requires all the patients enrolled in the current cohort to complete the toxicity and efficacy assessment windows. In contrast, the TITE-gBOIN-ET design allows for flexibility in the decision time, where the dose escalation decision is based on the cumulative data at a certain decision time. The observed quasi-Bernoulli toxicity and efficacy probabilities, $\hat{\mu}_{Tj}$ and $\hat{\mu}_{Ej}$ for $i = 1, \ldots, n_j$, in the gBOIN-ET and TITE-gBOIN-ET design are calculated as follows.

gBOIN-ET (Takeda et al., 2022a): Let x_{Tij} and x_{Eij} be a quasi-Bernoulli toxicity and efficacy outcome (i.e., normalized ETS and EES) observed within the assessment windows of τ_T and τ_E for the i-th patient at dose level j, respectively. Given the cumulative toxicity and efficacy data for dose level j; i.e., $\{x_{Tij} : i = 1, \ldots, n_j\}$ and $\{x_{Eij} : i = 1, \ldots, n_j\}$, we have $\hat{\mu}_{Tj} = \sum_{i=1}^{n_j} x_{Tij}/n_j$ and $\hat{\mu}_{Ej} = \sum_{i=1}^{n_j} x_{Eij}/n_j$.

TITE-gBOIN-ET (Takeda et al., 2023): Let $\tilde{x}_{Tij}$ be a quasi-Bernoulli toxicity outcome (i.e., normalized ETS) observed at a certain decision time for the i-th patient at dose level j, where $\tilde{x}_{Tij} = 1$ if the patient has experienced the toxicity and $\tilde{x}_{Tij} = 0$ if the patient has not yet experienced the toxicity. Let κ_{Tij} denote that $\tilde{x}_{Tij}$ is ascertained ($\kappa_{Tij} = 1$) or is still pending ($\kappa_{Tij} = 0$) at the decision time, and v_{Tij} denote the time to the toxicity outcome if $\tilde{x}_{Tij} = 1$ and to censoring if $\tilde{x}_{Tij} = 0$. Note that $(\tilde{x}_{Tij}, \kappa_{Tij}) = (0, 1)$ indicates the censoring at $v_{Tij} = \tau_T$, and $(\tilde{x}_{Tij}, \kappa_{Tij}) = (0, 0)$ indicates the censoring at $v_{Tij} < \tau_T$. In the same way, $(\tilde{x}_{Eij}, \kappa_{Eij}, v_{Eij})$ is defined for efficacy data of the i-th patient at dose level j. Given the cumulative toxicity and efficacy data for dose level j at the decision time; i.e., $\{(\tilde{x}_{Tij}, \kappa_{Tij}, v_{Tij}) : i = 1, \ldots, n_j\}$ and $\{(\tilde{x}_{Eij}, \kappa_{Eij}, v_{Eij}) : i = 1, \ldots, n_j\}$, we have $\hat{\mu}_{Tj} = \sum_{i=1}^{n_j} \tilde{x}_{Tij} / \dot{n}_{Tj}$ and $\hat{\mu}_{Ej} = \sum_{i=1}^{n_j} \tilde{x}_{Eij} / \dot{n}_{Ej}$, where $\dot{n}_{Tj} = \sum_{i=1}^{n_j} (\kappa_{Tij} + (1 - \kappa_{Tij}) v_{Tij} / \tau_T)$ and $\dot{n}_{Ej} = \sum_{i=1}^{n_j} (\kappa_{Eij} + (1 - \kappa_{Eij}) v_{Eij} / \tau_E)$.

3.3 Optimal biological dose selection

In the BOIN-ET design family, the selection of OBD is carried out at the end of the trial after completing the dose-escalation procedure. Firstly, the (quasi-Bernoulli) toxicity and efficacy probabilities at each dose level are estimated using all data obtained from the entire cohort. Then, given the estimated (quasi-Bernoulli) toxicity and efficacy probabilities, the utility to measure the toxicity-efficacy trade-off is computed and serves as a basis for the OBD selection. We below describe the estimation method of the toxicity and efficacy probabilities and the OBD selection measures.

Estimation of (quasi-Bernoulli) toxicity probability To ensure the monotonically increasing dose-toxicity relationship, isotonic regression (Bril et al., 1984) is usually used to estimate the (quasi-Bernoulli) toxicity probability. Specifically, the pool-adjacent-violators algorithm (Barlow et al., 1972) is performed on the observed (quasi-Bernoulli) toxicity probability to obtain the estimated (quasi-Bernoulli) toxicity probability.

Estimation of (quasi-Bernoulli) efficacy probability The dose-efficacy relationship is typically unknown at the design stage and may show nonmonotonic patterns. There are three methods available to estimate the (quasi-Bernoulli) efficacy probability:

- Simple use of the observed (quasi-Bernoulli) efficacy probability

- Model averaging of multiple unimodal isotonic regressions (Lin and Yin, 2017)

- Fractional polynomial logistic regression (Takeda et al., 2018)

Measure to select the OBD Given the estimated (quasi-Bernoulli) toxicity and efficacy probabilities, there are four different measures to select the OBD. Three of them are to select a dose that maximizes a utility quantifying the toxicity-efficacy trade-off, and the other is to select an admissible dose that offers the highest estimated efficacy probability. Specifically, the OBD is selected as a dose level of:

- Maximizing utility defined by a weighted function (Lin and Yin, 2017; Zhou et al., 2019)

- Maximizing utility defined by truncated linear functions (Li et al., 2020; Lin and Ji, 2021)

- Maximizing utility defined by scoring (Lin et al., 2020)

- Having the highest (quasi-Bernoulli) efficacy probability among admissible doses (Takeda et al., 2018)

Yamaguchi et al. (2024) reported the comparative performance of several OBD selection approaches, in which users can find guidance on the choice of (i) the method to estimate the efficacy probability, and (ii) the measure to select the OBD. In general, the most straightforward measure, selecting an admissible dose with the highest efficacy, is recommended for selecting the OBD due to its simplicity, and the efficacy probability modeling would lead to limited gains in the OBD selection (Yamaguchi et al., 2024). Default options used in many functions of **boinet** package are (i) simple use of the observed (quasi-Bernoulli) efficacy probability, and (ii) having the highest (quasi-Bernoulli) efficacy probability among admissible doses.

3.4 Additional dose-escalation rules

Some additional dose-escalation rules implemented in the **boinet** package are described below. During the dose-escalation, dose skipping is not allowed from a safety viewpoint. In addition, a dose elimination rule is introduced to avoid patient assignment to inefficient or severely toxic dose levels. A dose level is eliminated from the investigation if a posterior probability that the toxicity probability is greater than the target toxicity probability is larger than a certain threshold, which is set to 0.95 as the default. Apart from this, a dose level is eliminated from the investigation if a posterior probability that the efficacy probability is less than the minimum efficacy probability is larger than a certain threshold, which is set to 0.99 as the default. When dose levels at which the next cohort patients would be assigned have been eliminated, the trial is terminated, and the patient enrollment is stopped. At the end of the trial, if all dose levels are eliminated due to these criteria, the OBD is determined as not applicable.

Other cases in which early trial termination is implemented include (i) when the decision is de-escalation from the lowest dose, and (ii) when the number of patients treated at the current dose level reaches a certain threshold, which is set to the maximum sample size as default. Moreover, when the decision is escalation from the highest dose, the stay decision is made alternatively. In addition, if no patients are allocated to some dose levels (for example, a case where the dose-escalation procedure does not reach the highest dose level), these dose levels are not included in the estimation of toxicity and efficacy probabilities.

4 Implementation

We below illustrate the four main functions from the **boinet** package: `boinet()`, `tite.boinet()`, `gboinet()`, and `tite.gboinet()` which conduct simulation studies with a particular scenario and provide operating characteristics of the BOIN-ET, TITE-BOIN-ET, gBOIN-ET, and TITE-gBOIN-ET designs, respectively. The **boinet** package is available from the Comprehensive R Archive Network (CRAN) at https://CRAN.R-project.org/package=boinet. Further to the conduct of simulation studies, the package offers a function, `obd.select()`, which can be used for selecting the OBD at the end of an actual trial. The illustration of this function is omitted here but it asks users to enter the estimated toxicity and efficacy probabilities for each dose level, the design parameters (e.g., target toxicity and efficacy probability), and the measure to select the OBD, returning an OBD.

As simulation study settings used commonly for the BOIN-ET, TITE-BOIN-ET, gBOIN-ET, and TITE-gBOIN-ET design, consider a dose-finding trial with six dose levels and suppose that three patients were enrolled per cohort. In particular, the trial treated three patients at the lowest dose level in the first cohort and then sequentially assigned three patients to the next dose level until reaching the maximum sample size of 36 patients (i.e., 12 cohorts at the maximum). The target (quasi-Bernoulli) toxicity and efficacy probability were set to 0.33 and 0.60 respectively. The toxicity and efficacy assessment windows were 30 and 45 days respectively, and the accrual rate (average number of days necessary to enroll 1 patient) was 10 days. These simulation study settings were given by:
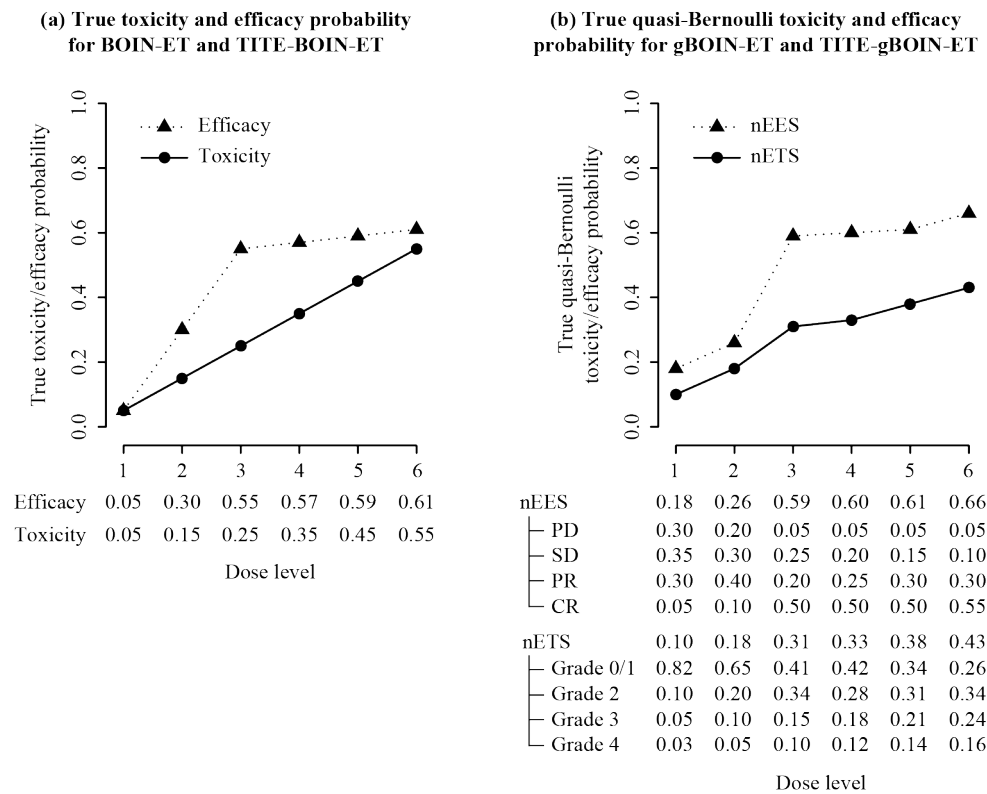
Figure 1: Dose-toxicity and dose-efficacy relationship considered in simulation study.

```
> n.dose      <- 6   # Six dose levels
> start.dose  <- 1   # Starting with the lowest dose level
> size.cohort <- 3   # Three patients assigned to each cohort
> n.cohort    <- 12  # Twelve cohorts at the maximum

> # Target (quasi-Bernoulli) toxicity and efficacy probability
> phi   <- 0.33
> delta <- 0.60
>
> tau.T   <- 30  # Thirty days of toxicity assessment windows
> tau.E   <- 45  # Forty-five days of efficacy assessment windows
> accrual <- 10  # Ten days of accrual rate
```

4.1 BOIN-ET

The `boinet()` function supports the conduct of simulation studies for the BOIN-ET design and returns a list of the summary results. The usage of the `boinet()` function is as follows:

```
boinet(
  n.dose, start.dose, size.cohort, n.cohort,
  toxprob, effprob,
  phi=0.3, phi1=phi*0.1, phi2=phi*1.4, delta=0.6, delta1=delta*0.6,
  alpha.T1=0.5, alpha.E1=0.5, tau.T, tau.E,
  te.corr=0.2, gen.event.time="weibull",
  accrual, gen.enroll.time="uniform",
  stopping.npts=size.cohort*n.cohort,
  stopping.prob.T=0.95, stopping.prob.E=0.99,
  estpt.method="obs.prob", obd.method="max.effprob",
  w1=0.33, w2=1.09,
```

```
  plow.ast=phi1, pupp.ast=phi2, qlow.ast=delta1/2, qupp.ast=delta,
  psi00=40, psi11=60,
  n.sim=1000, seed.sim=100)
```

Some of the key arguments are detailed in Table 1, 2, 3, and 4 in Appendix. Here, toxprob and effprob are vectors of true toxicity and efficacy probabilities with a length equal to the number of dose levels investigated.

We considered a scenario of the true toxicity and efficacy probabilities as given in Figure 1(a):

```
> toxprob <- c(0.05,0.15,0.25,0.35,0.45,0.55)
> effprob <- c(0.05,0.30,0.55,0.57,0.59,0.61)
```

and used (i) the observed efficacy probability for the efficacy probability estimation and (ii) the highest estimated efficacy probability as the measure to select the OBD:

```
> estpt.method <- "obs.prob"
> obd.method   <- "max.effprob"
```

Then, the boinet() is implemented with the previously specified arguments:

```
> boinet(
+   n.dose=n.dose, start.dose=start.dose,
+   size.cohort=size.cohort, n.cohort=n.cohort,
+   toxprob=toxprob, effprob=effprob,
+   phi=phi, delta=delta,
+   tau.T=tau.T, tau.E=tau.E, accrual=accrual,
+   estpt.method=estpt.method, obd.method=obd.method)
```

and returns the following output:

```
Simulation results:

                 Dose1  Dose2  Dose3  Dose4  Dose5  Dose6
Toxicity prob.    0.05   0.15   0.25   0.35   0.45   0.55
Efficacy prob.    0.05   0.30   0.55   0.57   0.59   0.61
No. Pts treated   3.40   7.00  15.90   7.00   2.00   0.60
Select %          1.70  12.00  54.40  25.50   5.30   0.80

No OBD %                 0.3
Trial duration (days)  778.9


Trial design settings:

Target toxicity prob.    0.330
Target efficacy prob.    0.600
Lower toxicity boundary  0.153
Upper toxicity boundary  0.390
Lower efficacy boundary  0.480


Tox. assessment window (days)  30
Eff. assessment window (days)  45
Accrual rate (days)            10


Efficacy prob. estimation: obs.prob


OBD selection: max.effprob
```

The simulation results include the number or the percentage of times that each dose level was selected as OBD (`Select %`), the average number of patients who were treated at each dose level (`No. Pts treated`), the percentage of times that the study was terminated (`Stop %`), and the expected study duration (`Trial duration (days)`). The trial design settings, such as the lower and upper toxicity/efficacy boundaries used for the dose-escalation, are also displayed. The average number of patients treated at Dose 3 was 15.90, and Dose 3 was selected as an OBD in 54.40% of the simulated trials, on average. The OBD was not selected in 0.3% of the simulated trials, on average. The trial duration was an average of 778.9 days.

4.2 TITE-BOIN-ET

The `tite.boinet()` function supports the conduct of simulation studies for the TITE-BOIN-ET design and returns a list of the summary results. The usage of the `tite.boinet()` function is as follows:

```
tite.boinet(
  n.dose, start.dose, size.cohort, n.cohort,
  toxprob, effprob,
  phi=0.3, phi1=phi*0.1, phi2=phi*1.4, delta=0.6, delta1=delta*0.6,
  alpha.T1=0.5, alpha.E1=0.5, tau.T, tau.E,
  te.corr=0.2, gen.event.time="weibull",
  accrual, gen.enroll.time="uniform",
  stopping.npts=size.cohort*n.cohort,
  stopping.prob.T=0.95, stopping.prob.E=0.99,
  estpt.method="obs.prob", obd.method="max.effprob",
  w1= 0.33, w2=1.09,
  plow.ast=phi1, pupp.ast=phi2, qlow.ast=delta1/2, qupp.ast=delta,
  psi00=40, psi11=60,
  n.sim=1000, seed.sim=100)
```

Some of the key arguments are detailed in Table 1, 2, 3, and 4 in Appendix. It should be noted that the TITE-BOIN-ET design has a suspension rule which holds off the decision on dose allocation (i.e., patient enrollment to the next cohort) until adequate information is collected, where the dose allocation is allowed only if more than 50% of patients have finished the assessment at the current dose level.

We considered a scenario of the true toxicity and efficacy probabilities as given in Figure 1(a):

```
> toxprob <- c(0.05,0.15,0.25,0.35,0.45,0.55)
> effprob <- c(0.05,0.30,0.55,0.57,0.59,0.61)
```

and used (i) the observed efficacy probability for the efficacy probability estimation and (ii) the highest estimated efficacy probability as the measure to select the OBD:

```
> estpt.method <- "obs.prob"
> obd.method   <- "max.effprob"
```

Then, the `tite.boinet()` is implemented with the previously specified arguments:

```
> tite.boinet(
+   n.dose=n.dose, start.dose=start.dose,
+   size.cohort=size.cohort, n.cohort=n.cohort,
+   toxprob=toxprob, effprob=effprob,
+   phi=phi, delta=delta,
+   tau.T=tau.T, tau.E=tau.E, accrual=accrual,
+   estpt.method=estpt.method, obd.method=obd.method)
```

and returns the following output:

```
Simulation results:

                 Dose1  Dose2  Dose3  Dose4  Dose5  Dose6
Toxicity prob.    0.05   0.15   0.25   0.35   0.45   0.55
Efficacy prob.    0.05   0.30   0.55   0.57   0.59   0.61
No. Pts treated   3.60   6.40  14.60   7.40   3.00   0.90
Select %          1.80  12.90  55.10  22.50   6.20   1.00

No OBD %                        0.5
Trial duration (days)  476.5

Trial design settings:

Target toxicity prob.    0.330
Target efficacy prob.    0.600
Lower toxicity boundary  0.153
Upper toxicity boundary  0.390
Lower efficacy boundary  0.480

Tox. assessment window (days)   30
Eff. assessment window (days)   45
Accrual rate (days)             10

Efficacy prob. estimation: obs.prob

OBD selection: max.effprob
```

The average number of patients treated at Dose 3 was 14.60, and Dose 3 was selected as an OBD in 55.10% of the simulated trials, on average. The OBD was not selected in 0.5% of the simulated trials, on average. The trial duration was an average of 476.5 days.

4.3 gBOIN-ET

The `gboinet()` function supports the conduct of simulation studies for the gBOIN-ET design and returns a list of the summary results. The usage of the `gboinet()` function is as follows:

```
gboinet(
  n.dose, start.dose, size.cohort, n.cohort,
  toxprob, effprob, sev.weight, res.weight,
  phi, phi1=phi*0.1, phi2=phi*1.4, delta, delta1=delta*0.6,
  alpha.T1=0.5, alpha.E1=0.5, tau.T, tau.E,
  te.corr=0.2, gen.event.time="weibull",
  accrual, gen.enroll.time="uniform",
  stopping.npts=size.cohort*n.cohort,
  stopping.prob.T=0.95, stopping.prob.E=0.99,
  estpt.method="obs.prob", obd.method="max.effprob",
  w1=0.33, w2=1.09,
  plow.ast=phi1, pupp.ast=phi2, qlow.ast=delta1/2, qupp.ast=delta,
  psi00=40, psi11=60,
  n.sim=1000, seed.sim=100)
```

Some of the key arguments are detailed in Table 1, 2, 3, and 4 in Appendix. Here, `toxprob` and `effprob` are matrices of true quasi-Bernoulli toxicity and efficacy probabilities with a row length equal to the number of outcome categories and a column length equal to the number of dose levels investigated. `sev.weight` and `res.weight` are weights assigned to toxicity and efficacy outcome categories respectively. For example, considering the ETS and EES, we can use sev.weight=c(0.00,0.50,1.00,1.50) for grade 0/1, 2, 3, and 4

toxicity and `res.weight=c(0.00,0.25,1.00,3.00)` for PD, SD, PR, and CR efficacy response, respectively

We considered a scenario of the true quasi-Bernoulli toxicity and efficacy probabilities as given in Figure 1(b):

```
> toxprob <- rbind(c(0.82,0.65,0.41,0.42,0.34,0.26),
+                  c(0.10,0.20,0.34,0.28,0.31,0.34),
+                  c(0.05,0.10,0.15,0.18,0.21,0.24),
+                  c(0.03,0.05,0.10,0.12,0.14,0.16))
>
> effprob <- rbind(c(0.30,0.20,0.05,0.05,0.05,0.05),
+                  c(0.35,0.30,0.25,0.20,0.15,0.10),
+                  c(0.30,0.40,0.20,0.25,0.30,0.30),
+                  c(0.05,0.10,0.50,0.50,0.50,0.55))
```

The weights assigned to toxicity and efficacy outcome categories were given by:

```
> sev.weight <- c(0.00,0.50,1.00,1.50)
> res.weight <- c(0.00,0.25,1.00,3.00)
```

and we used (i) the observed efficacy probability for the efficacy probability estimation and (ii) the highest estimated efficacy probability as the measure to select the OBD:

```
> estpt.method <- "obs.prob"
> obd.method   <- "max.effprob"
```

Then, the `gboinet()` is implemented with the previously specified arguments:

```
> gboinet(
+   n.dose=n.dose, start.dose=start.dose,
+   size.cohort=size.cohort, n.cohort=n.cohort,
+   toxprob=toxprob, effprob=effprob,
+   sev.weight=sev.weight, res.weight=res.weight,
+   phi=phi, delta=delta,
+   tau.T=tau.T, tau.E=tau.E, accrual=accrual,
+   estpt.method=estpt.method, obd.method=obd.method)
```

and returns the following output:

```
Simulation results:

                 Dose1  Dose2  Dose3  Dose4  Dose5  Dose6
Tox.cat1          0.82   0.65   0.41   0.42   0.34   0.26
Tox.cat2          0.10   0.20   0.34   0.28   0.31   0.34
Tox.cat3          0.05   0.10   0.15   0.18   0.21   0.24
Tox.cat4          0.03   0.05   0.10   0.12   0.14   0.16
nETS              0.10   0.18   0.31   0.33   0.38   0.43
Eff.cat1          0.30   0.20   0.05   0.05   0.05   0.05
Eff.cat2          0.35   0.30   0.25   0.20   0.15   0.10
Eff.cat3          0.30   0.40   0.20   0.25   0.30   0.30
Eff.cat4          0.05   0.10   0.50   0.50   0.50   0.55
nEES              0.18   0.26   0.59   0.60   0.61   0.66
No. Pts treated   3.80   7.20  17.60   5.40   1.70   0.30
Select %          2.90   8.20  63.10  19.50   5.50   0.70

No OBD %                 0.1
Trial duration (days)  780.1
```

```
Trial design settings:

Target toxicity prob.     0.330
Target efficacy prob.     0.600
Lower toxicity boundary   0.153
Upper toxicity boundary   0.390
Lower efficacy boundary   0.480

Tox. assessment window (days)   30
Eff. assessment window (days)   45
Accrual rate (days)             10

Efficacy prob. estimation: obs.prob

OBD selection: max.effprob
```

The average number of patients treated at Dose 3 was 17.60, and Dose 3 was selected as an OBD in 63.10% of the simulated trials, on average. The OBD was not selected in 0.1% of the simulated trials, on average. The trial duration was an average of 780.1 days.

4.4 TITE-gBOIN-ET

The `tite.gboinet()` function supports the conduct of simulation studies for the gBOIN-ET design and returns a list of the summary results. The usage of the `tite.gboinet()` function is as follows:

```
tite.gboinet(
  n.dose, start.dose, size.cohort, n.cohort,
  toxprob, effprob, sev.weight, res.weight,
  phi, phi1=phi*0.1, phi2=phi*1.4, delta, delta1=delta*0.6,
  alpha.T1=0.5, alpha.E1=0.5, tau.T, tau.E,
  te.corr=0.2, gen.event.time="weibull",
  accrual, gen.enroll.time="uniform",
  stopping.npts=size.cohort*n.cohort,
  stopping.prob.T=0.95, stopping.prob.E=0.99,
  estpt.method="obs.prob", obd.method="max.effprob",
  w1=0.33, w2=1.09,
  plow.ast=phi1, pupp.ast=phi2, qlow.ast=delta1/2, qupp.ast=delta,
  psi00=40, psi11=60,
  n.sim=1000, seed.sim=100)
```

Some of the key arguments are detailed in Table 1, 2, 3, and 4 in Appendix. It should be noted that the TITE-gBOIN-ET design has a suspension rule which holds off the decision on dose allocation (i.e., patient enrollment to the next cohort) until adequate information is collected, where the dose allocation is allowed only if more than 50% of patients have finished the assessment at the current dose level.

We considered a scenario of the true quasi-Bernoulli toxicity and efficacy probabilities as given in Figure 1(b):

```
> toxprob <- rbind(c(0.82,0.65,0.41,0.42,0.34,0.26),
+                  c(0.10,0.20,0.34,0.28,0.31,0.34),
+                  c(0.05,0.10,0.15,0.18,0.21,0.24),
+                  c(0.03,0.05,0.10,0.12,0.14,0.16))
>
> effprob <- rbind(c(0.30,0.20,0.05,0.05,0.05,0.05),
+                  c(0.35,0.30,0.25,0.20,0.15,0.10),
```

```
+                      c(0.30,0.40,0.20,0.25,0.30,0.30),
+                      c(0.05,0.10,0.50,0.50,0.50,0.55))
```

The weights assigned to toxicity and efficacy outcome categories were given by:

```
> sev.weight <- c(0.00,0.50,1.00,1.50)
> res.weight <- c(0.00,0.25,1.00,3.00)
```

and we used (i) the observed efficacy probability for the efficacy probability estimation and (ii) the highest estimated efficacy probability as the measure to select the OBD:

```
> estpt.method <- "obs.prob"
> obd.method   <- "max.effprob"
```

Then, the `tite.gboinet()` is implemented with the previously specified arguments:

```
> tite.gboinet(
+   n.dose=n.dose, start.dose=start.dose,
+   size.cohort=size.cohort, n.cohort=n.cohort,
+   toxprob=toxprob, effprob=effprob,
+   sev.weight=sev.weight, res.weight=res.weight,
+   phi=phi, delta=delta,
+   tau.T=tau.T, tau.E=tau.E, accrual=accrual,
+   estpt.method=estpt.method, obd.method=obd.method)
```

and returns the following output:

```
Simulation results:

                   Dose1  Dose2  Dose3  Dose4  Dose5  Dose6
Tox.cat1           0.82   0.65   0.41   0.42   0.34   0.26
Tox.cat2           0.10   0.20   0.34   0.28   0.31   0.34
Tox.cat3           0.05   0.10   0.15   0.18   0.21   0.24
Tox.cat4           0.03   0.05   0.10   0.12   0.14   0.16
nETS               0.10   0.18   0.31   0.33   0.38   0.43
Eff.cat1           0.30   0.20   0.05   0.05   0.05   0.05
Eff.cat2           0.35   0.30   0.25   0.20   0.15   0.10
Eff.cat3           0.30   0.40   0.20   0.25   0.30   0.30
Eff.cat4           0.05   0.10   0.50   0.50   0.50   0.55
nEES               0.18   0.26   0.59   0.60   0.61   0.66
No. Pts treated    3.90   7.20  15.70   6.20   2.40   0.60
Select %           3.00   8.10  61.00  20.70   5.90   1.20

No OBD %                  0.1
Trial duration (days)   441.0

Trial design settings:

Target toxicity prob.      0.330
Target efficacy prob.      0.600
Lower toxicity boundary    0.153
Upper toxicity boundary    0.390
Lower efficacy boundary    0.480

Tox. assessment window (days)   30
Eff. assessment window (days)   45
Accrual rate (days)             10
```

```
Efficacy prob. estimation: obs.prob

OBD selection: max.effprob
```

The average number of patients treated at Dose 3 was 15.70, and Dose 3 was selected as an OBD in 61.00% of the simulated trials, on average. The OBD was not selected in 0.1% of the simulated trials, on average. The trial duration was an average of 441.0 days.

5 Summary

The concept of OBD is now widely accepted as an alternative to MTD in oncology dose-finding trials, and the BOIN-ET design family is simple and flexible enough to establish the OBD while accounting for toxicity and efficacy in the framework of dose-finding. The **boinet** package has been developed to support the conduct of simulation studies to assess operating characteristics of BOIN-ET, TITE-BOIN-ET, gBOIN-ET, and TITE-gBOIN-ET designs. Users can choose design parameters in flexible and straightforward ways depending on their own application, and several options regarding methods to estimate the efficacy probability and measures to select the OBD are applicable. The package would assist practitioners in efficiently assessing the design features of their studies.

References

L. Ades, T. Cluzeau, T. Comont, L. Aguinaga, A. Stamatoullas, M. Meunier, E. Gyan, A. Garnier, M. D'Aveni, S. Thepot, M. Sebert, M. Moldovan, A. W. Petrich, K. Lemarie, F. Chermat, M. Fontenay, S. Chevret, and P. Fenaux. Combining ESA and Luspatercept in non-RS MDS patients having failed ESA - results of the phase 1-2 part a of the GFM combola study. *blood*, 144:351–352, 2024. URL https://doi.org/10.1182/blood-2024-204791. [p16]

R. E. Barlow, D. J. Bartholomew, J. M. Bremner, and H. D. Brunk. *Statistical Inference under Order Restrictions: The Theory and Application of Isotonic Regression*. Wiley, London, 1972. [p20]

G. Bril, R. Dykstra, C. Pillers, and T. Robertson. Isotonic regression in two independent variables. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 33(3):352–357, 1984. URL https://doi.org/10.2307/2347723. [p20]

K. Brock, D. Slade, and M. Sweeting. *escalation: A Modular Approach to Dose-Finding Clinical Trials*, 2024. URL https://CRAN.R-project.org/package=escalation. R package version 0.1.10. [p16]

E. J. Calabrese and L. A. Baldwin. U-shaped dose-responses in biology, toxicology, and public health. *Annual Review of Public Health*, 22:15–33, 2001. URL https://doi.org/10.1146/annurev.publhealth.22.1.15. [p15]

P. Corbaux, M. El-Madani, M. Tod, J. Peron, D. Maillet, J. Lopez, G. Freyer, and B. You. Clinical efficacy of the optimal biological dose in earlyphase trials of anti-cancer targeted therapies. *European Journal of Cancer*, 120:40–46, 2019. URL https://doi.org/10.1016/j.ejca.2019.08.002. [p15]

J. Fraisse, D. Dinart, D. Tosi, C. Bellera, and C. Mollevi. Optimal biological dose: a systematic review in cancer phase i clinical trials. *BMC Cancer*, 21(60), 2021. URL https://doi.org/10.1186/s12885-021-07782-z. [p15]

P. Li, R. Liu, J. Lin, and Y. Ji. TEPI-2 and UBI: designs for optimal immuno-oncology and cell therapy dose finding with toxicity and efficacy. *Journal of Biopharmaceutical Statistics*, 30(6):979–992, 2020. URL https://doi.org/10.1080/10543406.2020.1814802. [p15, 20]

R. Lin and G. Yin. STEIN: A simple toxicity and efficacy interval design for seamless phase i/ii clinical trials. *Statistics in Medicine*, 36(26):4106–4120, 2017. URL https://doi.org/10.1002/sim.7428. [p20]

R. Lin and Y. Yuan. Time-to-event model-assisted designs for dose-finding trials with delayed toxicity. *Biostatistics*, 21(4):807–824, 2020. URL https://doi.org/10.1093/biostatistics/kxz007. [p16, 18]

R. Lin, Y. Zhou, F. Yan, D. Li, and Y. Yuan. BOIN12: Bayesian optimal interval phase i/ii trial design for utility-based dose finding in immunotherapy and targeted therapies. *JCO Precision Oncology*, 4:PO.20.00257, 2020. URL https://doi.org/10.1200/PO.20.00257. [p15, 20]

X. Lin and Y. Ji. The joint i3+3 (ji3+3) design for phase i/ii adoptive cell therapy clinical trials. *Journal of Biopharmaceutical Statistics*, 30(6):993–1005, 2020. URL https://doi.org/10.1080/10543406.2020.1818250. [p15]

X. Lin and Y. Ji. Probability intervals of toxicity and efficacy design for dose-finding clinical trials in oncology. *Statistical Methods in Medical*, 30(3):843–856, 2021. URL https://doi.org/10.1177/0962280220977009. [p15, 20]

P. M. LoRusso, S. A. Boerner, and L. Seymour. An overview of the optimal planning, design, and conduct of phase i studies of new therapeutics. *Clinical Cancer Research*, 16(6):1710–1718, 2010. URL https://doi.org/10.1158/1078-0432.CCR-09-1993. [p15]

N. Penel, A. Adenis, S. Clisant, and J. Bonneterre. Nature and subjectivity of dose-limiting toxicities in contemporary phase 1 trials: comparison of cytotoxic versus non-cytotoxic drugs. *Investigational New Drugs*, 29(6):1414–1419, 2011. URL https://doi.org/10.1007/s10637-010-9490-7. [p16]

S. Postel-Vinay, H.-T. Arkenau, D. Olmos, J. E. Ang, J. Barriuso, S. Ashley, U. Banerji, J. De-Bono, I. Judson, and S. Kaye. Clinical benefit in phase-i trials of novel molecularly targeted agents: does dose matter? *British Journal of Cancer*, 100(9):1373–1378, 2009. URL https://doi.org/10.1038/sj.bjc.6605030. [p15]

A. R. Reynolds. Potential relevance of bell-shaped and u-shaped dose-responses for the therapeutic targeting of angiogenesis in cancer. *Dose Response*, 8(3):253–284, 2010. URL https://doi.org/10.2203/dose-response.09-049.Reynolds. [p15]

M.-K. Riviere, Y. Yuan, J.-H. Jourdan, F. Dubois, and S. Zohar. Phase i/ii dose-finding design for molecularly targeted agent: Plateau determination using adaptive randomization. *Statistical Methods in Medical Research*, 27(2):466–479, 2018. URL https://doi.org/10.1177/0962280216631763. [p15]

K. Takeda, M. Taguri, and S. Morita. BOIN-ET: Bayesian optimal interval design for dose finding based on both efficacy and toxicity outcomes. *Pharmaceutical Statistics*, 17(4):383–395, 2018. URL https://doi.org/10.1002/pst.1864. [p15, 17, 18, 20]

K. Takeda, S. Morita, and M. Taguri. TITE-BOIN-ET: Time-to-event Bayesian optimal interval design to accelerate dose-finding based on both efficacy and toxicity outcomes. *Pharmaceutical Statistics*, 19(3):335–349, 2020. URL https://doi.org/10.1002/pst.1995. [p16, 17, 18]

K. Takeda, S. Morita, and M. Taguri. gBOIN-ET: The generalized Bayesian optimal interval design for optimal dose-finding accounting for ordinal graded efficacy and toxicity in early clinical trials. *Biometrical Journal*, 64(7):1178–1191, 2022a. URL https://doi.org/10.1002/bimj.202100263. [p16, 17, 18, 19]

K. Takeda, Q. Xia, S. Liu, and A. Rong. TITE-gBOIN: Time-to-event Bayesian optimal interval design to accelerate dose-finding accounting for toxicity grades. *Pharmaceutical Statistics*, 21(2):496–506, 2022b. URL https://doi.org/10.1002/pst.2182. [p17]

K. Takeda, Y. Yamaguchi, M. Taguri, and S. Morita. TITE-gBOIN-ET: Time-to-event generalized Bayesian optimal interval design to accelerate dose-finding accounting for ordinal graded efficacy and toxicity outcomes. *Biometrical Journal*, 64(7):e2200265, 2023. URL https://doi.org/10.1002/bimj.202200265. [p16, 17, 20]

P. F. Thall and J. D. Cook. Dose-finding based on efficacy-toxicity trade-offs. *Biometrics*, 60 (3):684–693, 2004. URL https://doi.org/10.1111/j.0006-341X.2004.00218.x. [p15]

J. S. Weber, J. C. Yang, M. B. Atkins, and M. L. Disis. Toxicities of immunotherapy for the practitioner. *Journal of Clinical Oncology*, 33(18):2092–2099, 2015. URL https://doi.org/10.1200/JCO.2014.60.0379. [p15]

Y. Yamaguchi and K. Takeda. *boinet: Conduct Simulation Study of Bayesian Optimal Interval Design with BOIN-ET Family*, 2025. URL https://CRAN.R-project.org/package=boinet. R package version 1.3.0. [p16]

Y. Yamaguchi, K. Takeda, S. Yoshida, and K. Maruo. Optimal biological dose selection in dose-finding trials with model-assisted designs based on efficacy and toxicity: a simulation study. *Journal of Biopharmaceutical Statistics*, 34(3):379–393, 2024. URL https://doi.org/10.1080/10543406.2023.2202259. [p21]

J. Yin and Y. Yuan. Checkerboard: a Bayesian efficacy and toxicity interval design for phase i/ii dose-finding trials. *Journal of Biopharmaceutical Statistics*, 30(6):1006–1025, 2020. URL https://doi.org/10.1080/10543406.2020.1815033. [p15]

X. Yuan, C. Li, H. Sun, L. Tang, and H. Pan. *Keyboard: Bayesian Designs for Early Phase Clinical Trials*, 2025. URL https://CRAN.R-project.org/package=Keyboard. R package version 0.1.3. [p16]

Y. Yuan, R. Chappell, and H. Bailey. The continual reassessment method for multiple toxicity grades: a Bayesian quasi-likelihood approach. *Biometrics*, 63(1):173–179, 2007. URL https://doi.org/10.1111/j.1541-0420.2006.00666.x. [p18]

Y. Yuan, H. Q. Nguyen, and P. F. Thall. *Bayesian Designs for Phase I-II Clinical Trials*. Chapman and Hall/CRC, New York, 2016. [p15]

Y. Yuan, J. J. Lee, and S. G. Hilsenbeck. Model-assisted designs for early-phase clinical trials: simplicity meets superiority. *JCO Precision Oncology*, 3:PO.19.00032, 2019. URL https://ascopubs.org/doi/10.1200/PO.19.00032. [p15]

Y. Zhou, J. J. Lee, and Y. Yuan. A utility-based Bayesian optimal interval (U-BOIN) phase i/ii design to identify the optimal biological dose for targeted and immune therapies. *Statistics in Medicine*, 38(28):5299–5316, 2019. URL https://doi.org/10.1002/sim.8361. [p20]

Yusuke Yamaguchi
Quantitative Sciences and Evidence Generation, Astellas Pharma Global Development, Inc.
2375 Waterview Drive, Northbrook, Illinois 60062
United States
yusuke-yamaguchi@astellas.com

Kentaro Takeda
Quantitative Sciences and Evidence Generation, Astellas Pharma Global Development, Inc.
2375 Waterview Drive, Northbrook, Illinois 60062
United States
kentaro.takeda@astellas.com

Kazushi Maruo
Department of Biostatistics, Faculty of Medicine, University of Tsukuba

1-1-1, Tennodai, Tsukuba, Ibaraki 305-8577
Japan
kazushi.maruo@gmail.com

6 Appendix: Key arguments used in `boinet()`, `tite.boinet()`, `gboinet()`, and `tite.gboinet()` function

Argument	Description
phi	Numeric value between 0 and 1 specifying the target (quasi-Bernoulli) toxicity probability. Represents the maximum acceptable toxicity rate. Default is 0.3 (30%).
phi1	Numeric value specifying the highest (quasi-Bernoulli) toxicity probability that is deemed sub-therapeutic such that dose-escalation should be pursued. Default is phi*0.1.
phi2	Numeric value specifying the lowest (quasi-Bernoulli) toxicity probability that is deemed overly toxic such that dose de-escalation is needed. Default is phi*1.4.
delta	Numeric value between 0 and 1 specifying the target (quasi-Bernoulli) efficacy probability. Represents the desired minimum efficacy rate. Default is 0.6 (60%).
delta1	Numeric value specifying the minimum probability deemed efficacious such that the dose levels with efficacy < delta1 are considered sub-therapeutic. Default is delta*0.6.

Table 1: Arguments regarding target (quasi-Bernoulli) toxicity and efficacy probability.

Argument	Description
alpha.T1	Numeric value specifying the probability that a (quasi-Bernoulli) toxicity outcome occurs in the late half of the toxicity assessment window. Used for event time generation. Default is 0.5.
alpha.E1	Numeric value specifying the probability that a (quasi-Bernoulli) efficacy outcome occurs in the late half of the efficacy assessment window. Used for event time generation. Default is 0.5.
te.corr	Numeric value between -1 and 1 specifying the correlation between toxicity and efficacy, specified as Gaussian copula parameter. Default is 0.2 (weak positive correlation).
gen.event.time	Character string specifying the distribution for generating event times. Options are "weibull" (default) or "uniform". A bivariate Gaussian copula model is used to jointly generate the time to first toxicity and efficacy outcome, where the marginal distributions are set to Weibull distribution when gen.event.time="weibull", and uniform distribution when gen.event.time="uniform".
gen.enroll.time	Character string specifying the distribution for enrollment times. Options are "uniform" (default) or "exponential". Uniform distribution is used when gen.enroll.time="uniform", and exponential distribution is used when gen.enroll.time="exponential".

Table 2: Arguments regarding data generation.

Argument	Description
stopping.npts	Integer specifying the maximum number of patients per dose for early study termination. If the number of patients at the current dose reaches this criterion, the study stops the enrollment and is terminated. Default is size.cohort*n.cohort.
stopping.prob.T	Numeric value between 0 and 1 specifying the early study termination threshold for toxicity. If P(toxicity > phi) > stopping.prob.T, the dose levels are eliminated from the investigation. Default is 0.95.
stopping.prob.E	Numeric value between 0 and 1 specifying the early study termination threshold for efficacy. If P(efficacy < delta1) > stopping.prob.E, the dose levels are eliminated from the investigation. Default is 0.99.

Table 3: Arguments regarding early study termination criteria.

Argument	Description
estpt.method	Character string specifying the method for estimating efficacy probabilities. Options: "obs.prob" (observed efficacy probabilities/rates), "fp.logistic" (fractional polynomial), or "multi.iso" (model averaging of multiple unimodal isotonic regression). Default is "obs.prob".
obd.method	Character string specifying the method for OBD selection. Options: "utility.weighted", "utility.truncated.linear", "utility.scoring", or "max.effprob" (default).
w1	Numeric value specifying the weight for toxicity-efficacy trade-off in "utility.weighted" method. Default is 0.33.
w2	Numeric value specifying the penalty weight for toxic doses in "utility.weighted" method. Default is 1.09.
plow.ast	Numeric value specifying the lower toxicity threshold for "utility.truncated.linear" method. Default is phi1.
pupp.ast	Numeric value specifying the upper toxicity threshold for "utility.truncated.linear" method. Default is phi2.
qlow.ast	Numeric value specifying the lower efficacy threshold for "utility.truncated.linear" method. Default is delta1/2.
qupp.ast	Numeric value specifying the upper efficacy threshold for "utility.truncated.linear" method. Default is delta.
psi00	Numeric value specifying the utility score for (toxicity=no, efficacy=no) in "utility.scoring" method. Default is 40.
psi11	Numeric value specifying the utility score for (toxicity=yes, efficacy=yes) in "utility.scoring" method. Default is 60.

Table 4: Arguments regarding optimal biological dose (OBD) selection.