NOVA SCHOOL OF SCIENCE & TECHNOLOGY

DEPARTAMENT OF MATHEMATICS

DIOGO ALEXANDRE SILVÉRIO FERREIRA

BSc in Pharmacy

MODEL-BASED AND MODEL-ASSISTED DESIGNS FOR DOSE-ESCALATION IN CANCER SINGLE-AGENT PHASE I CLINICAL TRIALS: A SIMULATION STUDY

MASTER IN STATISTICS FOR HEALTH

NOVA University Lisbon

September, 2023

NOVA SCHOOL OF SCIENCE & TECHNOLOGY

DEPARTAMENT OF MATHEMATICS

MODEL-BASED AND MODEL-ASSISTED DESIGNS FOR DOSE-ESCALATION IN CANCER SINGLE-AGENT PHASE I CLINICAL TRIALS: A SIMULATION STUDY

DIOGO ALEXANDRE SILVÉRIO FERREIRA

BSc in Pharmacy

Adviser: Isabel Natário
Associate Professor, NOVA University Lisbon, NOVA School of Science and Technology

Examination Committee:

Chair: Maria Isabel Azevedo Rodrigues Gomes
Associate Professor, NOVA University Lisbon, NOVA School of Science and Technology

Rapporteurs: Frederico Almeida Gião Gonçalves Caeiro
Associate Professor, NOVA University Lisbon, NOVA School of Science and Technology

Adviser: Isabel Cristina Maciel Natário
Associate Professor, NOVA University Lisbon, NOVA School of Science and Technology

MASTER IN STATISTICS FOR HEALTH

NOVA University Lisbon
September, 2023

MODEL-BASED AND MODEL-ASSISTED DESIGNS FOR DOSE-ESCALATION IN CANCER SINGLE-AGENT PHASE I CLINICAL TRIALS: A SIMULATION STUDY

To all clinical trial participants, that make clinical research possible.

ACKNOWLEDGMENTS

I want to express my deepest gratitude to Professor Isabel Natário, for all the guidance, patience, and support throughout this journey. Her expertise and mentorship have been invaluable.

I am deeply thankful to my loving wife, Rosana, for her endless encouragement, understanding, and sacrifices.

I also want to thank my parents, José and Elisabete, for their endless love and sacrifices that have shaped me into the person I am today.

ABSTRACT

First-in-human clinical trials in oncology aim to determine the maximum tolerated dose (MTD) by quantifying the probability of participants experiencing a set of specified toxicities called dose-limiting toxicities.

These clinical trials are designed as dose-escalation studies, where the dose is escalated and de-escalated as the toxicity outcomes are observed. The rule or algorithm-based designs class has dominated the cancer clinical trial setting. Yet, over the last decades, two novel classes of dose-escalation methods emerged, model-based designs and model-assisted designs.

This thesis aims to describe and compare three model-based designs, including the Continual Reassessment Method (CRM), Escalation With Overdose Control (EWOC), and Bayesian Logistic Regression Method (BLRM), and two model-assisted designs, including the Modified Toxicity Probability Interval (mTPI) and the Bayesian Optimal Interval (BOIN). To compare the five methods, a numerical simulation study was performed, considering ten possible dose-toxicity scenarios, that evaluated the accuracy in identifying the MTD, the dose allocation and the toxicity exhibited. Dose allocation was also assessed by replicating a clinical trial using actual clinical trial data.

Our analysis suggested that the CRM outperforms other methods, being the most accurate method in identifying the MTD and offering a greater exposure of patients to therapeutic doses. However, it has also been shown to be more aggressive, exposing more patients to overly toxic doses. The EWOC, mTPI and BOIN showed comparable accuracy in determining the MTD, yet less accuracy than CRM. They were also shown to be more conservative, exposing less patients to higher doses. Nevertheless, the EWOC exhibited excessive aggressiveness in dose allocation when the MTD is at a lower dose, and proved to be less accurate and to expose too many patients to sub-therapeutic doses when the MTD is at higher doses. The BLRM imposes an overdose control that showed to be extremely conservative, causing the early termination of a significant number of the clinical trials simulated before the maximum

sample size was reached, suggesting that a more appropriate overdose control might be needed.

Our findings highlighted the trade-offs between the methods and can support research teams in their decision-making process on selecting the most appropriate dose-escalation method, given the specific characteristics of their clinical trials.

Keywords: model-based designs; model-assisted designs; maximum tolerated dose; Continual Reassessment Method; Escalation With Overdose Control; Bayesian Logistic Regression Method; Modified Toxicity Probability Interval; Bayesian Optimal Interval

RESUMO

Os ensaios clínicos *first-in-human* em oncologia têm como objetivo determinar a dose máxima tolerada (DMT), quantificando a probabilidade de os participantes manifestarem um conjunto de toxicidades específicas chamadas de toxicidades limitantes da dose.

Estes ensaios clínicos tomam a forma de estudos de escalonamento de dose, onde a dose é incrementada e reduzida à medida que os resultados de toxicidade são conhecidos. A classe de métodos baseados em regras ou algoritmos dominou os ensaios clínicos de oncologia. No entanto, nas últimas décadas, surgiram duas novas classes de métodos de escalonamento de dose, métodos baseados em modelos e métodos assistidos por modelos.

Esta tese tem como objetivo descrever e comparar três métodos baseados em modelos, incluindo o *Continual Reassessment Method* (CRM), *Escalation With Overdose Control* (EWOC) e *Bayesian Logistic Regression Method* (BLRM), e dois métodos assistidos por modelos, incluindo o *Modified Toxicity Probability Interval* (mTPI) e o *Bayesian Optimal Interval* (BOIN). Para comparar os cinco métodos, um estudo de simulação numérica foi realizado, que considerou dez possíveis cenários de dose-toxicidade, que avaliou a sua exatidão para identificar a DMT, a sua alocação de dose e a toxicidade exibida. A alocação de dose foi, também, avaliada através da replicação de um ensaio clínico utilizando dados reais.

A nossa análise sugeriu que o CRM supera outros métodos, sendo o método mais exato para identificar a DMT e expondo mais doentes a doses potencialmente terapêuticas. No entanto, também se mostrou mais agressivo, expondo mais doentes a doses excessivamente tóxicas. O EWOC, mTPI e o BOIN mostraram exatidão comparável na determinação da DMT, mas menor exatidão do que o CRM. Eles também se mostraram mais conservadores, expondo menos doentes a doses mais altas. No entanto, o EWOC exibiu excessiva agressividade na alocação de dose quando a DMT está em doses mais baixas, e mostrou ser menos exato, e expor muitos doentes a doses subterapêuticas quando a DMT está em doses mais altas. O BLRM impõe um controle de overdose que se mostrou extremamente conservador, causando a interrupção de um número significativo dos ensaios clínicos simulados, antes que o tamanho

máximo da amostra fosse atingido, sugerindo que é necessário um controle de overdose mais adequado.

Os nossos resultados identificaram as diferenças entre os métodos e podem apoiar as equipas de investigação no processo de tomada de decisão para a seleção do método de escalonamento de dose mais apropriado, considerando as características específicas dos seus ensaios clínicos.

Palavras chave: métodos baseados em modelos; métodos assistidos por modelos; dose máxima tolerada; *Continual Reassessment Method*; *Escalation With Overdose Control*; *Bayesian Logistic Regression Method*; *Modified Toxicity Probability Interval*; *Bayesian Optimal Interval*

CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACRONYMS

5-FU	5-fluorouracil
BLRM	Bayesian Logistic Regression Method
BOIN	Bayesian Optimal Interval
c.d.f.	Cumulative Distribution Function
CRM	Continual Reassessment Method
DLT	Dose-Limiting Toxicity
EI	Equivalence Interval
EWOC	Escalation With Overdose Control
FDA	Food and Drug Administration
FIH	First-in-Human
HNSTD	Highest Non-Severely Toxic Dose
MTD	Maximum Tolerated Dose
mTPI	Modified Toxicity Probability Interval
OI	Overdosing Interval
PAVA	Pool Adjacent Violators Algorithm
p.d.f.	Probability Density Function
p.m.f.	Probability Mass Function
UI	Underdosing Interval
UPM	Unit Probability Mass

1

INTRODUCTION

First-in-human (FIH) clinical trials are the foundation of the clinical development process and represent the first application of a new drug or drug combination in human beings [1]. Such clinical trials take place after extensive preclinical testing.

Early clinical trials for a new therapeutic agent typically aim to describe its safety and tolerability, determine a dose range likely to provide a clinical effect, and characterize its pharmacokinetics and pharmacodynamics [2]–[4]. Not all of these goals are usually completely met in any phase I trial though [5].

Anti-cancer agents are generally known for their inherent toxicity, reason why, unlike most FIH clinical trials, cancer clinical trials enroll patients instead of healthy volunteers.

In oncology, for cytotoxic therapies, it is assumed that higher doses will be more toxic, as well as, more effective [1], [3], [4]. Under this dual assumption, the toxicity is used as a surrogate for effectiveness [6].

Hence, phase I oncology trials aim to determine the maximum tolerated dose (MTD). The MTD will be the highest possible, but tolerable, dose, concerning some specific toxicities, called dose-limiting toxicities (DLTs) [7].

The DLTs, a prespecified set of unacceptable or unmanageable toxicities, constitute the variable of interest [4], [8]. The toxicities' severity, frequency and number is usually reduced to a single binary variable [9], i.e., a patient treated in a cancer FIH clinical trial experiences DLT(s) or not.

In practice, the MTD will correspond to the dose that yields a specified DLT proportion or DLT probability, defined by the research team, depending on the method being utilized. Some methods treat the MTD as a simple statistical quantity calculated directly from the observed data, usually in terms of DLT proportions, while others adopt a more complex approach and treat the MTD as a toxicity probability quantile estimated from a monotonic dose-toxicity curve [10], [11].

It is essential to highlight that the notion that both effectiveness and toxicity increase with dose, applies to cytotoxic therapies. For non-cytotoxic therapies, such as molecular targeted therapies and immunotherapy, the relationship between the dose, toxicity and effectiveness is not always direct [12]. These therapies are less toxic than standard cytotoxic agents, and therapeutical effects can occur at doses lower than the MTD. Consequently, additional non-toxicity primary endpoints, such as efficacy endpoints, must be considered for clinical trials involving such therapies [1].

Cancer FIH clinical trials are designed as dose-escalation studies, in which the patients are sequentially recruited and treated in small cohorts, with doses that are gradually increased, to describe the drug safety and tolerability and to establish the MTD.

The patient population for FIH clinical trials is relatively small, usually composed of patients with no alternative standard therapies. It can also be very heterogeneous, as the same trial can enroll patients with different types of malignant tumors at various disease stages [7], [13]. Consequently, cancer FIH clinical trials "are typically small , single-arm and open-label " [1], that is, small studies, without a comparator or placebo arm, and unblinded.

Dose-escalation studies start by treating the first cohort of patients with a dose deemed to be safe, the starting dose. The following cohorts are treated at a higher dose level — escalation — lower dose level — de-escalation — or at the same dose level, depending on the toxicity responses observed.

Thus, dose-escalation clinical trials result from a sequence of dose assignment decisions [14]. The cohorts of patients are successively treated until sufficient information is collected to estimate the MTD. Based on the observed toxicity responses, before treating a new cohort of patients, the clinical research team takes one of three actions: escalate, de-escalate, or maintain the current dose [14].

These trials pose some ethical concerns. They should be carefully designed and conducted to minimize the patients exposed to sub-therapeutic doses, maximize the chance of therapeutic effects, and avoid exposure to overly toxic doses [3], [15].

As Liu and Yuan [14] expose in their 2015 paper, under the assumption that toxicity and efficacy monotonically increase with dose, an ideal trial design would:

- escalate the dose if the current dose is below the MTD, to reduce exposure to sub-therapeutic doses;
- maintain the dose if the current dose corresponds to the MTD, to maximize the therapeutic effect;
- de-escalate the dose when the current dose is above the MTD, to avoid exposure to overly toxic doses.

Such an ideal trial is not available because it is impossible to know whether the current dose is below, above or equal (or close) to the MTD [14]. Thus, the dose-escalation methods must provide a theoretical framework to achieve good dose-escalation decisions and a reasonable estimate of the MTD, through algorithms or statistical inference, based on the data collected in the trial.

Many different methods have been proposed, looking to optimize, from a practical and ethical viewpoint, the dose assignment decisions and the determination of the MTD. The published literature is vast, but few dose-escalation methods were broadly put into practice.

The dose-escalation methods can be classified into three broad classes: algorithm-based designs, model-based designs, and model-assisted designs.

In the algorithm-based (or rule-based) designs, the dose is progressively escalated throughout the trial from the lowest dose level until a specific number of DTLs is observed, based exclusively on rules [15]. Such designs are easy to comprehend and implement because the decision rules are intuitive and do not involve complex calculations [3], [8].

The model-based and the model-assisted designs represent *novel* classes of dose-escalation methods. The first rely on parametric dose-toxicity models that use the cumulative toxicity data to drive the dose escalation and estimate the MTD [1], [16]. The latter combine the advantages of the algorithm-based designs and model-based designs [17], they use a set of pre-tabulated dose-escalation and de-escalation rules formulated in accordance with statistical models [12].

This thesis aims to describe and compare the statistical methods underlying some of the most relevant and utilized model-based and model-assisted designs used in cancer FIH clinical trials.

Three model-based designs, the Continual Reassessment Method (CRM), the Escalation With Overdose Control (EWOC) and the Bayesian Logistic Regression Method (BLRM), and two model-assisted designs, the Modified Toxicity Probability Interval (mTPI) and the Bayesian Optimal Interval (BOIN), were described and compared.

To compare the five methods, a numerical simulation study was performed, considering ten possible dose-toxicity scenarios, that evaluated the accuracy in identifying the MTD, the dose allocation and the toxicity exhibited. Dose allocation was also assessed by replicating a clinical trial using actual clinical trial data.

The present thesis is organized into five chapters. The opening chapter introduces the topic of dose-escalation studies and establishes some definitions necessary to understand the planning and conduction of cancer early development clinical trials.

Chapter 2 introduces the three classes of dose-escalation methods and describes in detail the five methods identified previously.

Chapter 3 presents the simulation study, which explores ten hypothetical scenarios, and compares the dose-escalation methods, examining their performance using commonly employed metrics.

In Chapter 4, we replicate a clinical trial, using actual clinical trial data, to understand possible differences in the dose-escalation actions of the different dose-escalation methods.

Finally, in Chapter 5, we summarize and discuss our findings and present the limitations of this study.

1.1 Basic Concepts

A dose-escalation study has many elements that must be defined, fundamental for the clinical trial design and its successful conduction. Below are described some key concepts essential to understanding the methods described ahead.

1.1.1 Dose

The dose can be defined as the amount of an active substance given in a single administration or repeated over a given period, according to a certain administration schedule of equal or unequal single doses at equal or unequal intervals [2], [5], [15].

In oncology, therapies are frequently delivered in cycles. In the dose-escalation trials, the patients are often treated with a single cycle [3], although more prolonged exposures can occur.

1.1.2 Starting dose

The starting dose is the dose deemed to be safe, assigned to the first patient or cohort of patients enrolled in a clinical trial. In FIH clinical trials, the starting dose is determined based on pre-clinical studies data and is, sometimes, set by the regulatory authorities [7].

The International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use (ICH) S9 [18] suggests as an appropriate starting dose for small molecule

drugs, like cytotoxic agents and some molecular targeted agents, 1/10 of the Severely Toxic Dose in 10% of the animals (STD 10) in rodents or 1/6 of the Highest Non-Severely Toxic Dose (HNSTD) in non-rodents. HNSTD corresponds to the highest dose that does not cause signs of lethality, life-threatening or irreversible toxicities.

For immunotherapy, the minimal anticipated biologic effect level (MABEL) is typically used to calculate the starting dose [18], [19].

1.1.3 Dose levels

In a dose-escalation clinical trial, we are interested in testing several doses to understand their safety. The doses within the dose range available in a clinical trial, are called dose levels. The dose levels can be selected by the investigational team [1], [3], based on the existing pre-clinical data, characteristics of the population who will receive the drug [3], type of drug and data available from drugs of the same class.

The size of the dose increments between dose levels should take into consideration the uncertainties associated with the estimation of the dose-toxicity relationship [20].

Common practices are setting dose increments on a logarithmic scale (e.g., 25 mg, 50 mg, 100 mg, 200 mg), the most common approach, or on a linear scale (e.g., 25 mg, 50 mg, 75 mg, 100 mg) [21], [22]. The first approach is sometimes adjusted to give the so called modified Fibonacci sequence, in which the dose increases through dose increments of 100%, 65%, 52%, 40%, 33%, 33%, and so on, [1], [3], [4], ensuring that dose increments are initially large but smaller at higher dose levels [16].

Presently, it's common to find FIH clinical trials where only the starting dose and the maximum fold increase are defined in the clinical trial protocol, being the dose levels chosen as the trial progresses, based on the toxicity data observed [20].

1.1.4 Cohort

The cohorts are groups of patients treated sequentially at a designated dose level. Typically, cohorts of 3 to 6 patients are utilized in the clinical trials [23].

1.1.5 Dose-limiting toxicity

DLTs are clinically significant adverse events or abnormal laboratory values, usually serious or life-threatening, ideally reversible, that a clinical trial participant can experience when exposed to an investigational anti-cancer agent. The DLTs vary by disease and type of drug under analysis [4], [5].

The DLTs are defined in a clinical trial protocol and usually are established by some criteria. The National Cancer Institute Common Terminology Criteria for Adverse Events (CTCAE) is widely used to define the DLTs [24].

DLTs are typically handled as a single binary variable. A patient treated with an investigational anti-cancer agent, at a certain dose level, experiences one or more DLTs, or not.

1.1.6 Maximum tolerated dose

MTD was defined by Storer [25] as the highest dose of an experimental agent that produces an acceptable risk of toxicity.

From a statistical viewpoint, it can be defined as the dose that, at the end of the trial, has an estimated DLT proportion, or estimated probability of causing DLTs, as close as possible to some predetermined target toxicity level [26]. Some methods use the DLT proportion of the current dose level to declare the MTD; others use statistical probability models to estimate the MTD.

1.1.7 Target toxicity level

The specification of the target toxicity level is required by some dose-escalation methods, such as the model-based and model-assisted methods. It can be defined as the maximum probability of DLT that is considered acceptable in the trial [1]. The target toxicity probability in cancer FIH trials is typically between 20% and 33% [1], [27], but it can be an higher or lower value.

The value chosen for the target toxicity level depends on the nature of the DLTs expected to be observed. It is set relatively high when the DLTs are transient and correctable, and low when potentially lethal or life-threatening [28].

In the following chapters, the terms target toxicity level and target toxicity probability will be used interchangeably.

2

DOSE-ESCALATION METHODS

2.1 Algorithm-Based Designs

The algorithm-based or rule-based designs use a set of simple pre-tabulated rules to determine dose escalation and de-escalation actions based on the observed DLT proportion. These designs are simple and easy to implement, but their decision rules lack statistical justification [29].

The standard 3+3 design is the most well-known dose-escalation design. It is an algorithmic design where a set of rules are followed for dose escalation. At the end of the trial, the MTD is also acknowledged based on a pre-specified rule, without any relevant statistical analysis of the data [30].

To describe the standard 3+3 design without de-escalation, we consider j dose levels in, $J = \{d_1 <, \dots, < d_j\}$, available in a clinical trial. The subjects are sequentially assigned to the dose levels J, in cohorts of 3, following the below rules:

- The first cohort of patients is treated at the starting dose d_1;
- If 0 out of 3 patients experience a DLT, the dose is escalated to dose level d_{j+1};
- If 1 out of 3 patients experiences a DLT, additional 3 patients are treated at the current dose level, d_j;
- If 1 out of 6 patients experiences a DLT, the dose is escalated to next dose level d_{j+1};
- If 2 or more out of 6 patients experience a DLT, the MTD is considered to have been exceeded and further dose escalation is not pursued [3].

Dose escalation ends when the DLT rate is at least one-third at a given dose level, in which six patients are treated. The MTD is declared to be the next lower dose level. Thus, the MTD is defined as the highest dose, with a DLT rate of $0/6$ or $1/6$ [30].

If escalation occurs at the last dose level, then the MTD is at or above the last dose level. If the trial stops at the first dose, then the MTD is below the first dose level. In either case, the dose levels under study need to be revised to be possible to determine the MTD [31].

The standard 3+3 design with de-escalation allows more patients to be treated at a lower dose level when an excessive DLT rate is reached, i.e., DLT rate $\geq 2/6$. Thus, if only three patients were treated at the next lower dose level d_{j-1}, additional three patients can be treated. Hence, the dose may be de-escalated until reaching a dose level at which six patients are treated, and at most, one DLT is observed [7], [31].

Due to their simplicity, the 3+3 design and its variations (e.g., 3+3 with intrapatient dose escalation and accelerated titration design) have been dominant in phase I clinical trials for decades [32].

Some authors reviewed the literature published for the dose-escalation methods utilized in phase I clinical trials to highlight such dominance. For example, a review of 1235 phase I dose-escalation studies for new anti-cancer agents published between 1991 and 2006 found that 98,4% of these studies utilized the traditional 3+3 design or some variant [33]. Another review of published studies between 2007 and 2008, performed by Le Tourneau *et al* [1], showed that of 181 published studies, 96,7% used the traditional 3+3 design or some variant. More recently, Yuan *et al.* [34] reviewed the phase I trials published in the Clinical Cancer Research journal in 2015 and noticed that 32 out of 34 used the 3+3 design or some variant.

The 3+3 and its variations will not be described in detail, as the present section aims to provide context to the importance of the *novel* dose-escalation methods, the model-based and model-assisted designs, which is related to the several drawbacks associated with the 3+3 design.

One of the main criticisms of the 3+3 design is that the MTD is observed from the data, from an algorithm-based perspective, and is not interpreted as an estimate of the dose that yields a target toxicity probability[31].

Another drawback is that the 3+3 design only considers the data of the current dose level to determine the MTD. It treats very few patients at each dose level (3 or 6 patients), so it lacks precision [3], [17], leading to higher uncertainty about the MTD chosen at the end of the trial.

Furthermore, it was shown that it has low accuracy in determining the MTD, compared to other designs [17], [35].

Finally, an also frequent criticism, is that they tend to expose too many patients to subtherapeutic doses [12], [31] and treat few patients at *optimal* dose levels, with potential therapeutic effect [36].

All the downsides outlined above are widely acknowledged and contributed to the development of novel techniques.

2.2 Model-Based and Model-Assisted Designs

2.2.1 Bayesian inference

The frequentist statistical approach, largely developed during the first half of the XX century, has had a central role in scientific research. It was soon adopted in clinical trials, and its use has been dominant ever since [37], [38].

Despite its simple probability calculation, rigor, and focus on the experiment, the frequentist framework suffers from some limitations [37], [38]. It requires large samples, resulting in recruitment challenges and elevated drug development costs. Additionally, frequentist statistics are often inflexible, limiting innovation in the design and analysis of clinical trials [37].

The interest in Bayesian statistics has been increasing in the last decades, and that has been observed in many scientific fields. This expansion was possible due to the advances in computational techniques.

The pharmaceutical and medical device industries were no exception [37], [39], and even the regulatory authorities have encouraged the use of Bayesian designs. In 2004, the US Food and Drug Administration (FDA) co-sponsored a workshop to discuss the role of Bayesian methods in drug and medical device development [37]. Since then, the FDA and the European Medicines Agency have issued guidance on innovative trial designs on several occasions.

The frequentist and the Bayesian approaches differ and represent the two main schools of statistical inference.

The frequentist approach is based on a definition of probability built on the long-run frequency of repeated events. It considers parameters as fixed unknown constants and infers from random data, collected from the repetition of an experiment a great number of times, usually in terms of frequency or proportion [40], [41].

The Bayesian approach considers probability as an expression of the degree of belief in one event, based on prior knowledge about that event (expressed as a prior distribution) and the observed data. It assumes that both data and model parameter(s) are random variables, since we are uncertain about the true value of the parameter(s).

Bayesian inference applies the Bayes' Theorem, formulated by Thomas Bayes, and posthumously published in 1763, to determine the updated probability of a hypothesis given new information.

The Bayes theorem is a restatement of the conditional probability, that can be used to determine the probability of an event, based on the prior knowledge of the conditions associated with that event. It is often expressed by

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}, \qquad P(B) \neq 0, \tag{1}$$

where A and B are two events.

The applications of the Bayes theorem are diverse and not limited to simple conditioned events. In Bayesian inference, we are often interested in parameter(s) estimation. Hence, assuming θ as an unknown parameter or vector of parameters for the observed data Y, $Y = (y_1, \dots, y_n)$, the Bayes theorem can be expressed as,

$$P(\theta|Y) = \frac{P(\theta)P(Y|\theta)}{P(Y)} \tag{2}$$

where:

- $P(\theta)$ is the prior distribution of the unknown θ, representing our prior degree of belief about θ before the data is observed. Its probability distribution gives the weight attributed to each value θ in our prior belief. The prior belief, although subjective, should be carefully selected, based on the existing knowledge about θ.
- $P(Y|\theta) \equiv l(\theta|y)$, gives the data likelihood and represents the plausibility of each value of θ, given the observed values of Y. All evidence obtained from an experiment about an unknown quantity θ is included in the likelihood function of θ for the data [42].
- $P(\theta|Y)$ represents the posterior probability of θ given the data. It combines our prior beliefs with the evidence provided by the data.
- $P(Y)$ is the marginal probability of the data. By the law of total probability, the overall probability of observing the data Y, across all possible values of θ, is given by $P(Y) = \sum_j P(\theta_j)P(Y|\theta_j)$, when θ is discrete, i.e., the sum of the information about all possible values of θ. When θ is continuous, $P(Y) = \int P(\theta)P(Y|\theta)d\theta$, i.e., the integral of the prior probability multiplied by the likelihood function over the sample space for θ [43]. In the Bayes theorem $P(Y)$ works as a normalizing constant to make the posterior probability to integrate to one, taking a proper value in the interval [0; 1].

Because the denominator $P(Y)$ is a constant, and does not depend on the parameter(s) for which we want to make inference, the Bayes rule is often re-written, the denominator is removed from the equation, and the posterior probability of the parameters is obtained proportionally to the product of the prior distribution of the parameter(s) and the likelihood,

$$P(\theta|Y) \propto P(\theta)P(Y|\theta), \tag{3}$$

The above is frequently expressed in words as

$$Posterior \propto Prior \times Likelihood.$$

All Bayesian inference is made through the posterior distribution of the parameter(s), calculated proportionally to the product of the prior probability of the model parameter(s) and the data likelihood, to express how a degree of belief or prior knowledge should rationally change to account for the new evidence [44].

Thus, the Bayesian approach uses a cohesive framework to combine, through statistical modelling, the observed data with external knowledge, expert opinion, or both [39].

The Bayesian statistical framework has been utilized for decades in cancer FIH clinical trials, which represent the most significant share of Bayesian clinical trial designs employed in oncology [45].

The Bayesian approach has some key characteristics appealing to the dose-escalation studies that have contributed to its application in many dose-escalation methods. Characteristics like:

- Capacity to handle small samples:

  The dose-escalation studies have small samples, as the number of patients available is usually very small, and the Bayesian approach is better equipped to model data from small sample sizes [46].

- Incorporation of prior information:

  The Bayes theorem provides a mathematical framework to include the prior belief about the shape of the dose-toxicity curve, along with any surrounding uncertainty [9]. In the early development stage, the uncertainty about the safety and tolerability of an anti-cancer agent is high and the knowledge about the dose-toxicity relationship is scarce. By incorporating information from pre-clinical data and clinical opinion, the dose-toxicity models can be strengthened. Mostly because the number of patients willing to participate in early development cancer clinical trials is typically small, and the number of patients dosed at each dose level is reduced. Therefore, the inclusion of prior knowledge or expert opinion can be beneficial to modelling the dose-toxicity relationship.

  Nonetheless, it should be noted that the model must be carefully calibrated so that the prior knowledge does not dominate over the data collected during the trial [9]. The regulatory authorities recommend that as many sources of good prior information as possible should be identified and utilized [47].

- Bayesian sequential analysis:

  Another important characteristic is the Bayesian sequential analysis. In a Bayesian sequential analysis, the posterior model is successively updated as more data come in [48], providing support for interim analysis. As the data from each patient or cohort of patients is observed, the dose-toxicity model can be updated, and the posterior toxicity

probabilities for each dose calculated. The Bayesian sequential analysis is utilized on several dose-escalation methods to provide a statistical foundation for the dose assignment decisions.

- Flexibility:

  Under the Bayesian approach, a study design can be adaptively changed during its conduction, without undermining its validity and integrity [47]. Pre-planned modifications to one or more aspects can be made, based on the accumulated data (interim data) [49]. In FIH clinical trials, examples of such modifications are modifications to the sample size, changes in the cohort size, adjustments to the dose levels under experiment, or changes to the stopping rules.

The dose-escalation methods described in the following sections use Bayesian methods, although to different extents. But first, we will continue exploring the Bayesian statistics.

2.2.1.1 Likelihood function

The likelihood, earlier identified as $l(\theta|y)$, is one of the fundamental concepts in the Bayes theorem. The likelihood function is expressed as the joint probability mass function (p.m.f.) or probability density function (p.d.f.), $h(Y|\theta)$, of a random variable realization, regarded as a function of θ [50].

In Bayesian inference, the observed data is held constant, and the parameter values vary. Hence, the likelihood function does not behave as a valid probability distribution, as it doesn't integrate to 1. Instead, it provides a framework to compare the compatibility of the observed data [48], indicating which parameter values are more plausible. Consequently, it is often notated as $l(\theta|Y)$, instead of as $h(Y|\theta)$.

We have seen previously that, in the dose-escalation studies, the toxicity observed is considered a binary variable, in the sense that, the patients treated at a certain dose level experience a DLT, or not.

Thus, we could consider a Bernoulli trial or a binomial model as good choices to model the dependence of the response variable DLT, Y, on θ.

Taking the binomial model as an example, we will have Y given θ following a binomial distribution with parameters n and θ,

$$Y|\theta \sim \mathrm{Bin}(n, \theta). \tag{4}$$

Consequently, the likelihood function for the binomial distribution will have the same form as the binomial p.m.f., given by

$$l(\theta|Y = y) = h(Y|\theta) = \binom{n}{y}\theta^{y}(1-\theta)^{n-y}. \quad (5)$$

When the data is observed, $\binom{n}{y}$ in the equation above becomes a constant, thus, by dropping the terms that don't depend upon θ, we can simplify as

$$l(\theta|Y = y) \propto \theta^{y}(1-\theta)^{n-y}. \quad (6)$$

2.2.1.2 Priors

The prior probability distribution, identified in equation (2) as $P(\theta)$, is the other quantity fed in the Bayes theorem to calculate the posterior probability distribution.

In Bayesian inference, priors are chosen to define the subjective knowledge about the data before the data is observed [50]. They are modelled as probability distributions that summarize the existing information about the random variable(s), obtained from previous studies or expert opinion. The prior distributions depend on the chosen prior model parameters, designated hyperparameters.

Depending on the amount of information about the random variable(s) that is expressed by the priors, before the data is observed, they can be classified as noninformative or informative.

Noninformative priors, also called vague or diffuse, add little or no information about the random variable(s) [48], minimizing the impact of any pre-existing beliefs [51]. They are usually characterized by having a large variability, with little influence on the posterior. They can also take the form of improper probability distributions.

A flat prior is a special case, where equal prior plausibility is assigned to all possible values of the parameter(s) [48].

Conversely, informative priors reflect specific information about the random variable(s) with some certainty, i.e., lower variability [48]. The more informative the prior, the greater the influence on the posterior distribution.

2.2.1.3 Conjugate priors

In Bayesian statistics, when the prior and the posterior distributions belong to the same probability distribution family, the prior distribution is called a conjugate prior for the likelihood function.

In other words, if a prior distribution for the parameter θ with p.d.f. $h(\theta)$, and the resulting posterior p.d.f., $h(\theta|Y)$, calculated proportionally to the product of the prior with the likelihood function, $h(\theta|Y) \propto h(\theta)l(\theta|Y)$, are of the same family, the prior and posterior are called conjugate distributions, and, consequently, the prior distribution is designated a conjugate prior [48].

The advantage of utilizing conjugate priors, and the reason of their importance in Bayesian statistics, is that no numeric integration is required to find the posterior. The data is used to update the parameters of the conjugate prior to find the conjugate posterior, without requiring any complex computation [41].

An example of a conjugate prior is the beta distribution for the binomial likelihood function. The beta-binomial Bayesian model is an important tool to model the number of successes in n trials, when the probability of success θ is not fixed, but rather randomly drawn from a beta distribution with hyperparameters α and β.

In building the beta-binomial model we start with the beta prior, adjusted to reflect the knowledge about θ before data is observed,

$$\theta \sim \text{Beta}(\alpha, \beta) \tag{7}$$

with p.d.f. $h(\theta)$ given by

$$h(\theta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\theta^{\alpha-1}(1-\theta)^{\beta-1} \tag{8}$$

where $\theta \in [0;1], \alpha > 0, \beta > 0$ and $\Gamma(.)$ is the gamma function.

Regarding the likelihood function, we have seen earlier in section 2.2.1.1, that the binomial distribution was a good choice to model the dependence of a binary response variable Y, on θ, like the DLTs in the dose-escalation studies. So, we assume

$$Y|\theta \sim \text{Bin}(n, \theta). \tag{9}$$

We have also seen that the likelihood function for the binomial distribution was given by (5).

Combining the beta prior and the likelihood function, through the Bayes theorem, we obtain the posterior,

$$h(\theta|Y) \propto h(\theta)l(\theta|Y = y) =$$
$$h(\theta|Y) \propto \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\theta^{\alpha-1}(1-\theta)^{\beta-1}\binom{n}{y}\theta^{y}(1-\theta)^{n-y}. \tag{10}$$

Combining all the terms that don't depend upon θ in the constant C, $C = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\binom{n}{y}$, the above equation can be simplified as

$$h(\theta|Y) \propto C\theta^{\alpha+y-1}(1-\theta)^{\beta+n-y-1}. \quad (11)$$

Examining equation (11), we observe that the binomial likelihood has the same form as the beta prior distribution, that is, a product of θ to a power, by $1-\theta$ to another power.

The product of the prior and respective likelihood with the same exponential form produces a posterior that keeps that form. As a result, we obtain a posterior with the functional form of a beta p.d.f. with parameters $\alpha + y$ and $\beta + n - y$,

$$\text{Beta}(\alpha + y, \beta + n - y). \quad (12)$$

Thus, whenever we have a conjugate beta prior for a binomial likelihood, the posterior distribution obtained will also be a beta distribution, with parameter $\alpha' = \alpha + y$ and $\beta' = \beta + n - y$.

In summary, as we have just seen for the beta conjugate prior for the binomial likelihood, the computation of the posterior distribution, when one uses conjugate distributions, is simple and can be found analytically. Many other conjugate priors exist for their respective likelihood functions, nevertheless, conjugate priors are limited to likelihood functions that belong to the exponential family.

2.2.1.4 Markov chain Monte Carlo

All Bayesian inference is made through the calculation of the posterior distribution, nevertheless, finding the posterior through the Bayes theorem is not always easy. The posterior cannot always be calculated through closed-form expressions, as it happens for the conjugate priors.

We have seen previously, from the Bayesian theorem, that the posterior is calculated proportionally to the product of the prior and the likelihood. This equation, however, does not give a proper density, as it does not integrate to 1.

The marginal probability of the data, $P(Y)$, i.e., the denominator of the Bayes theorem, needs to be found to derive a proper posterior, to perform inference. This can be achieved by dividing the proportional posterior by the integral parameter values, as earlier mentioned in section 2.2.1. When closed forms for the integral can't be found, it needs to be evaluated numerically, however, this is often difficult or even impossible.

With the advances in computational techniques, alternative methods to estimate the posterior became available, expanding the range of models available to be used in Bayesian statistics.

The Markov chain Monte Carlo (MCMC) methods use algorithms for sampling parameter values of a probability distribution, without knowing its closed analytic form. The Metropolis-Hastings algorithm and the Gibbs sampling algorithm are the two main MCMC methods [41].

They draw samples from the posterior $h(\theta|Y)$, despite only the proportional form of the posterior $h(\theta|Y) \propto h(\theta)l(Y|\theta)$ is known, simulating values from the proper posterior density [41].

To do that, they generate Markov chain(s) composed of a large number of parameter values that are sequentially sampled, where each subsequent sample value is dependent of the previous one, that is, the parameter value θ^{i+1} depends exclusively on θ^{i} and the data Y [48]. These chains converge to the distribution of interest, the posterior distribution.

When the Markov chain(s) convergence to the target distribution is achieved, the values sampled belong to the posterior distribution and reflect the properties of the posterior [48], making statistical inference possible.

Having presented the fundamentals of the Bayesian analysis, essential to understand the following sections, we will next start describing the model-based and model-assisted methods.

2.2.2 Model-based designs

The model-based designs use statistical models to estimate the dose-toxicity relationship, under the assumption that the toxicity increases monotonically with dose.

They utilize Bayesian methods to model the dose-toxicity curve as the toxicity data accrues. One starts with a statistical model and the prior parameter(s) distribution, which characterize the initial beliefs about the dose–toxicity relationship between the dose levels. The dose-toxicity curve is then updated as the DLT outcomes from the sequentially enrolled cohorts are observed.

The dose allocation decisions are based on the predicted probabilities of DLT for each dose, estimated from the updated posterior distribution. The MTD is equally estimated from the posterior distribution by seeking the dose that yields a prespecified target toxicity probability, i.e., the defined target toxicity level [1].

Simple one or two-parameter models are usually utilized to describe the dose-toxicity relationship [39].

To ensure patient safety, safeguards are also usually set in most model-based designs, to restrict dose increments and avoid overshooting the MTD, exposing patients to excessively toxic doses [1], [16].

The model-based designs use the data of all patients treated in a clinical trial, in contrast to other methods. They have demonstrated higher accuracy in estimating the MTD [12], [17], [52] and to expose fewer patients to sub-therapeutic doses, in comparison to rule-based designs [12]. They are very flexible, the clinical research team can choose the desired target toxicity level, and the sample and cohort sizes are not restricted and can be adjusted as needed.

Despite their solid statistical foundation [29], they too have some drawbacks.

They require statistical expertise, adequate software [16], [53] and repeated model fitting and estimation, which is perceived as conceptually and computationally complex by clinicians [1], [54]. A frequent criticism made by clinicians is that the dose allocation decisions seem to be coming from a *black box* [34], [52], [54].

Another drawback is that the dose assignment decisions are based on the ongoing cumulative toxicity data, consequently the clinical research team requires access to real-time data, making the trial operationally more demanding. In addition, close collaboration between investigators, statisticians, and computer programmers is required [12].

The perception that the model-based designs are more complex has prevented the research teams from embracing these designs in practice [12], [17], [54], nevertheless, their use has been increasing.

The CRM, the EWOC and the BLRM, are examples of model-based designs and will be described ahead.

2.2.2.1 Notation

Before presenting the model-based designs, we start by establishing the notation.

Let J notate the ordered dose levels of a new drug under investigation in a FIH clinical trial, such that $J = \{d_1 <, d_2, \dots, < d_j\}$. Let $x_k \in J$ be the dose level that patient k receives and y_k denote the DLT outcome, a Bernoulli random variable, $Y = \{0, 1\}$, where $y_k = 1$ when the DLT is observed and $y_k = 0$ otherwise, for the kth patient enrolled in the trial. The data after observation of k patients is given by $D_k = \{(x_k, y_k), k = 1, \dots, n\}$, where n represents the total number of patients treated in a FIH clinical trial.

Let p_j denote the probability of DLT, $p_j = P(Y = 1|X = x_j)$, we aim to determine the MTD, the dose level with the probability of DLT equal (or closest) to the targeted toxicity level ϕ, $P(Y = 1|X = MTD) \approx \phi$.

2.2.2.2 Continual Reassessment Method

The CRM was presented by O'Quigley *et al.* [35] in 1990 and is among the first model-based designs in which toxicity is the primary endpoint [55].

This method uses a one-parameter parametric model to estimate the dose-toxicity relationship, assuming a monotonically non-decreasing toxicity function[9], [56]. As the name of the present method suggests, it continually updates the dose-toxicity curve, based on the posterior distribution of the model parameter, after each patient's response is observed [56], and reassesses the dose allocation for the next patient.

The CRM, as proposed by O'Quigley *et al.* [35], treats one patient at a time and all patients are treated with the dose thought to have toxicity probability closest to the target toxicity level ϕ. In other words, the first patient receives the dose deemed to be closest to the MTD, that is, the dose with prior probability of DLT closest to target toxicity level. Subsequently, the dose with estimated mean probability of DLT closest to the target toxicity level is designated for all next patients, based on the adjusted posterior toxicity probabilities [53], [56].

To describe the CRM, let's start by denoting $f(x_k, \theta)$ the dose function describing the probability of DLT with parameter θ in a one-parameter model, such that $f(x_k, \theta) = E(Y|X, \theta) = p_j$.

O'Quigley *et al.* [35], in their original paper, utilized the hyperbolic tangent model and the one-parameter logistic model, which can be defined, respectively, by

$$f(x_k, \theta) = \left[\frac{\tanh x_k + 1}{2}\right]^{\theta} \tag{13}$$

and

$$f(x_k, \theta) = \frac{\exp(\theta_0 + \theta x_k)}{1 + \exp(\theta_0 + \theta x_k)} \tag{14}$$

where θ_0 represents the fixed intercept.

However, the power and the one-parameter logistic models are the most commonly utilized working models [9], [55].

The power model can be given by

$$f(x_k, \theta) = {x_k}^{\theta} \tag{15}$$

or through reparameterization of θ, replacing θ by $\exp(\theta)$, by

$$f(x_k, \theta) = x_k^{\exp(\theta)} \tag{16}$$

[9], [55], [57].

The one-parameter logistic model can be given by the expression in (14), or through the above reparameterization by

$$f(x_k, \theta) = \frac{\exp(\theta_0 + \exp(\theta) x_k)}{1 + \exp(\theta_0 + \exp(\theta) x_k)} \tag{17}$$

[9], [55], [57].

The parameter θ is restricted to take only positive values in (14) and (15) to ensure the non-decreasing monotonicity. Under the reparameterization in (16) and (17), with the parameter θ exponentialized, θ is free to take on any real value [55].

A key point about the CRM is that the numerical dose labels do not need to correspond to the actual dose levels under experimentation [55].

Usually, the initial toxicity probability guesses, p_0, are available, for each one of the J dose levels, $p_0 = \{p_{01} <, \dots, < p_{0j}\}$. These initial dose-toxicity probability guesses are known as the *skeleton* [9], [58] and may be elicited from clinical opinion, pre-clinical data and data from previous studies, when available [9], [53].

Based on the initial toxicity probability guesses, p_0, a transformed set of the dose levels $J, J^* = \{d_1^*, \dots, d_j^*\}$, are calculated to adjust the prior dose-toxicity curve [9]. The use of the dose labels J^* ensures that the model provides an exact fit over the initial guesses of the toxicity probabilities, before the trial start [9], [55], [59].

The dose labels J^* can be calculated by backward substitution, based on the initial dose-toxicity probability guesses, p_0. For example, consider a scenario where d_1 has prior probability of DLT equal to $p_{01} = 0{,}05$, in a clinical trial using the one-parameter logistic model with fixed intercept $\theta_0 = 3$ and prior slope $\theta = 0$. The dose label d_1^* can be calculated by the expression below

$$0{,}05 = \frac{\exp(3 + \exp(0) d_1^*)}{1 + \exp(3 + \exp(0) d_1^*)}. \tag{18}$$

The same process would be used to calculate the remaining dose labels.

The only constraint is that the J^* dose labels are ordered in a strictly increasing sequence in terms of toxic probabilities [57], as in Table I. Table II summarizes the common dose-toxicity models and the resulting dose labels.

TABLE I
CRM initial dose-toxicity relationship

Dose labels	d_1^*	d_2^*	d_3^*	…	d_j^*
Prior probability of toxicity	$p_{01} <$	$p_{02} <$	$p_{03} <$	…	$< p_{0j}$

Adapted from [57]

TABLE II
CRM common dose-toxicity models and the resulting dose labels

Model name	Model $f(x_k, \theta)$	General form of dose labels (d_j^*)	Prior θ	Dose labels (d_j^*) given the prior θ
Power model	${x_k}^{\exp(\theta)}$	$\frac{1}{p_{0j}^{\exp(\theta)}}$	$\theta = 0$	p_{0j}
One-parameter logistic model	$\frac{\exp(3 + \exp(\theta)x_k)}{1 + \exp(3 + \exp(\theta)x_k)}$	$\frac{\ln\left(\frac{p_{0j}}{1-p_{0j}}\right) - 3}{\exp(\theta)}$	$\theta = 0$	$\ln\left(\frac{p_{0j}}{1-p_{0j}}\right) - 3$

Adapted from [59]

Under the CRM, one assumes that the MTD is within the dose range J, and its probability is equal to the specified target toxicity level, ϕ. After the first patient being treated with the dose level with prior toxicity probability closest to the target toxicity level, the CRM sequentially assigns the upcoming patients to one of the possible J dose levels, based on the updated dose-toxicity curve, derived from the Bayes theorem.

The posterior distribution of θ, $h(\theta|\Omega_k)$, is obtained through the Bayes theorem, given the set of the trial data, D_k, by:

$$h(\theta|D_k) \propto h(\theta) l(\theta|D_k) \tag{19}$$

where $l(\theta|D_k)$ represents the Bernoulli likelihood function, given by

$$l(\theta|D_k) \equiv f(x_k|\theta) = \prod_{k=1}^{k} p_k^{y_k}(1 - p_k)^{1-y_k} \quad \text{, where } p_k = f(x_k, \theta) \tag{20}$$

and $h(\theta)$ denotes the prior distribution of θ [56].

The dose-toxicity curve $f(x_k, \theta)$ is updated using one of the below estimators:

$$\hat{p}_j = \int_\theta f(d_j, \theta) h(\theta|D_k) \mathrm{d}\theta, \tag{21}$$

or

$$\hat{p}_j = f(d_j, \hat{\theta}_k). \tag{22}$$

where $\hat{\theta}_k = \int \theta h(\theta|D_k)\, \mathrm{d}\theta$ [56].

The dose level for the next patient $k + 1$ is selected using the updated dose-toxicity curve and will correspond to the dose level that has an estimated probability of DLT as close to the target toxicity level ϕ as possible [9]. This can be iteratively performed after each observation, minimizing a quadratic loss, through

$$x_{k+1} = arg \min_{d_j \in J} \left(\int_\theta f(d_j, \theta) h(\theta|D_k) d\theta - \phi \right)^2 \tag{23}$$

or through

$$x_{k+1} = arg\min_{d_j \in J} \left(f\left(d_j, \hat{\theta}_k\right) - \phi\right)^2. \quad (24)$$

The approach that uses the posterior mean of θ, given by $\hat{\theta}_k$, as a plug-in to determine the dose for patient $k+1$ is the convention in the literature, possibly, because it is computationally less demanding and easier to implement [9], [55].

Once patient $k+1$ is treated, the above computation is repeated based on the trial data set D_{k+1} [9]. As the trial progresses and the data from the patients treated accrues, the CRM is expected to converge to the MTD [55].

The trial ends when the maximum sample size n is reached or when an MTD estimate with a sufficient level of certainty is obtained [9].

At the end of the trial, the MTD can be derived by (23) or (24), by determining the dose that would have been given to patient $n+1$ [55].

The one-parameter models utilized by the CRM might not be flexible enough to provide an accurate estimation of the dose-toxicity relationship for the entire dose range, yet, as the CRM tends to gather information at the dose levels around the target toxicity level, it is thought to obtain reasonable estimates of the MTD [56].

Under the CRM, as presented originally by O'Quigley *et al.* [35], each cohort is composed of a single patient and all patients are treated at the dose level thought to be closest to the target toxicity level. This approach wasn't well accepted due to the risk of overshooting the MTD and exposing patients to unacceptable high doses [1].

Due to the single-patient cohorts used, the CRM is not shielded against patient variability. Heterogeneity among patients is common and can make the dose escalation too aggressive. In addition, Møller [60] demonstrated, using a simulation scenario, that if the CRM is not well calibrated, the second patient can be treated at the highest dose level when the first patient is treated at the lowest dose level and no DLT is observed.

Such criticisms led to several proposed and utilized safety modifications. Examples of such modifications are starting at the lowest dose [61], not skipping dose levels [62] and using cohorts of three patients, instead of a single-patient cohort [63].

2.2.2.3 Escalation With Overdose Control

Babb *et al.* [28] proposed the EWOC, a Bayesian adaptive design that is a derivation of the method presented previously. It has several differences, nonetheless. The EWOC uses a two-parameter model, a more flexible model relating the dose and toxicity; this two-parameter model is reparametrized using parameters that have clinical interpretation; and, the most

important difference, it has in place an overdose control, a safety measure to avoid exposing patients to potentially toxic dose-levels.

The EWOC method was designed to approach the MTD subject to the ethical constraint that an anticipated proportion of patients who experience toxicity do not exceed a chosen percentile [28]. That is, the EWOC selects the dose levels for the next cohorts so that the predicted probability of the dose administered to future patients exceeding the MTD is less or equal to a chosen percentile [28].

This percentile, denoted as α, is known as the *feasibility bound* [28] and is recommended to be a value $\leq 50\%$ [39]. The value chosen for the *feasibility bound* reflects how conservative escalating the dose between patients will be [9], corresponding lower values of α to more conservative approaches. Babb *et al.* [28] suggested using $\alpha = 25$.

The EWOC assumes that the MTD is a dose within a minimum and a maximum dose range available in a clinical trial, in such a way that $X_{min} \leq MTD \leq X_{max}$. X_{min} is a dose assumed to be safe and can be the dose assigned to the first patient or cohort of patients [28].

The relationship between dose levels and the toxicity is modelled through a two-parameter model, $f(x_k, \theta_0, \theta_1) = P(Y = 1|X = x) = p_j$, where the parameters θ_0 and θ_1 are unknown [64]. Babb *et al.* [28] utilized a two-parameter logistic function, which is more flexible than the functions described in the CRM,

$$f(x_k, \theta_0, \theta_1) = \frac{\exp(\theta_0 + \theta_1 x_k)}{1 + \exp(\theta_0 + \theta_1 x_k)} \tag{25}$$

It is assumed that $\theta_1 > 0$ so that the probability of a DLT is a monotonic increasing function of dose.

Babb *et al.* [28] propose deriving the cumulative distribution function (c.d.f.) of the posterior distribution of the MTD after the application of the Bayes theorem. To do that, the authors propose additional reparameterization in terms of two clinically interpretable measures: the MTD, denoted as γ, where $P(Y = 1|X = \gamma) = \phi$; and the prior probability of DLT at the lowest dose, denoted as ρ_0, $\rho_0 = P(Y = 1|X = X_{min}), \rho_0 \in (0,1)$. ρ_0 may come from data from preclinical studies, and γ is the parameter of interest [28], [64].

Given the above, the logistic model can be rewritten in terms of ρ_0 and γ to give

$$\text{logit}(\rho_0) = \theta_0 + \theta_1 X_{min} \tag{26}$$

and

$$\text{logit}(\phi) = \theta_0 + \theta_1 \gamma \tag{27}$$

and the parameters of interest derived by

$$\gamma = \frac{\text{logit}(\phi) - \theta_0}{\theta_1} \tag{28}$$

and

$$\rho_0 = \left(\frac{\exp(\theta_0 + \theta_1 X_{min})}{1 + \exp(\theta_0 + \theta_1 X_{min})}\right). \quad (29)$$

[9].

This reparameterization is used in the Bayesian updating procedure by placing prior distributions upon γ and ρ_0 [9]. Babb *et al.* [28] suggested uniform distributions for γ and ρ_0, over $[X_{min}, X_{max}]$ and $[0, \phi]$, respectively.

Given the joint prior distribution of the parameters γ and ρ_0, $h(\gamma, \rho_0)$, and the available trial data, the joint posterior density distribution, $h(\gamma, \rho_0|D_k)$, can be calculated. As the MTD is the parameter of interest, the marginal posterior distribution of γ, given D_k, is obtained integrating over ρ_0,

$$g(\gamma|D_k) = \int_0^{\rho_0} h\,(\gamma, \rho_0|D_k) d\rho_0 \quad (30)$$

and the c.d.f. of the marginal posterior determined by

$$G_k(x) = \int_{\mathrm{X}_{min}}^{x} h\,(\gamma|D_k) d\gamma, \qquad x \in [\mathrm{X}_{min}, \mathrm{X}_{max}]. \quad (31)$$

[28], [56].

The dose allocation for each subsequent patient, or cohort of patients, enrolled in the trial under the EWOC is determined by selecting from the c.d.f. of the marginal posterior distribution of γ, the dose corresponding to the *feasibility bound* α. Hence, the cohort $k+1$ receives the dose x_{k+1}, where

$$x_{k+1} = G_k^{-1}(\alpha), \qquad k = 2, \ldots n. \quad (32)$$

This means that the overdose control applied in the EWOC method is $P(x_{k+1} > \gamma|D_k) \leq \alpha$, and therefore, the dose given to patient or cohort of patients $k+1$ is restricted to be lower than the α^{th} quantile of the c.d.f. of the posterior distribution of γ [56].

The above equation assumes that the dose is a continuous variable and any dose between $[X_{min}, X_{max}]$ can be selected to treat patients, nevertheless, due to financial and logistic constraints, that doesn't usually happen. The dose is, most of the times, a discrete variable as a limited number of dose levels is available to be used in a clinical trial.

Thus, the dose for the next patient, or cohort of patients, $k+1$ needs to be calculated using an alternative method. It can be determined using the c.d.f. of the marginal posterior, identifying the dose that has marginal posterior probability closer to the *feasibility bound* α. This can be achieved through a quadratic loss function,

$$x_{k+1} = \arg\min_{d_j \in J} (G_k(d_j) - \alpha)^2, \qquad k = 2, \ldots n. \quad (33)$$

[9].

This process is repeated until a sufficiently precise estimate of the MTD is achieved or the maximum sample size n is reached.

At the end of the trial, all information accrued is used to estimate the MTD. Babb *et al.* [28] suggested that the MTD could be estimated by one of the measures of central tendency (mean, median and mode) of the c.d.f. of the marginal posterior of the MTD, γ, by minimizing an asymmetric loss function.

$$\widehat{\text{MTD}} = \underset{x \in [X_{Min}, X_{Max}]}{\arg\inf} \int_{X_{Min}}^{X_{Max}} L_\alpha(x,\gamma) h(\gamma | D_k) d\gamma \tag{34}$$

where

$$L_\alpha(x,\gamma) \begin{cases} \alpha(\gamma - x) & \text{if } x \leq \gamma, \text{ i.e., if } x \text{ is an underdose} \\ (1-\alpha)(x-\gamma) & \text{if } x > \gamma, \text{ i.e., if } x \text{ is an overdose.} \end{cases} \tag{35}$$

This asymmetric loss function penalizes overdosing over underdosing [9] and can be interpreted as the loss caused by treating a patient at doses above the MTD by δ units, $\delta > 0$, is $(1 - \alpha)/\alpha$ times greater than the loss caused by treating the patient below the MTD by dose δ units [28].

In clinical practice, the MTD selection often follows a simpler approach, it is chosen as the dose that would have been given to the patient $n + 1$ [65].

2.2.2.4 Bayesian Logistic Regression Method

Neuenschwander *et al.* [27] presented the BLRM in 2008, another modification of the CRM. The BLRM is currently a relevant method, widely used by some clinical trial sponsors [66]. It uses a Bayesian logistic regression model to drive the dose allocation so that it maximizes the posterior probability that the toxicity falls within a target toxicity range [67].

The authors assume that it is not possible to define a single precise target toxicity level ϕ. This way, the BLRM was designed to extend the target toxicity level to an ordinal scale with four intervals of toxicity probabilities, as defined below [39].

Underdosing	$p_j \in (0;\ 0{,}20] \mid D_k$
Target Toxicity	$p_j \in (0{,}20; 0{,}35] \mid D_k$
Excessive Toxicity	$p_j \in (0{,}35; 0{,}60] \mid D_k$
Unacceptable Toxicity	$p_j \in (0{,}60; 1] \mid D_k$

Although the original paper described four toxicity intervals, it is common to find in the literature the use of only three intervals, the underdosing interval (UI), the target toxicity interval, and the overdosing interval (OI). Considering $[\delta_1; \delta_2]$ as the target toxicity interval, the three intervals can be adapted and written in terms of δ_1 and δ_2,

Underdosing	$p_j \in [0;\ \delta_1[\,\|\, D_k$
Target Toxicity	$p_j \in [(\delta_1; \delta_2]\|\| D_k$
Overdosing	$p_j \in]\delta_2; 1]\|\| D_k.$

Like the previous methods, the DLT is assumed to be a Bernoulli trial

$$Y_j|\theta_0, \theta_1 \sim \text{Bern}(p_j) \tag{36}$$

with probability of a DLT at dose d_j denoted by p_j.

The DLT probability is described by a reparametrized two-parameter logistic model,

$$\text{logit}(p_j) = \log(\theta_0) + \theta_1 \log\left(\frac{d_j}{d^*}\right), \qquad \theta_0, \theta_1 > 0, \tag{37}$$

where θ_0 and θ_1 are unknown parameters and d^* is a reference dose used for the standardization of dose [27].

The probability of DLT, under the reparametrized model, is given by

$$p_j = \frac{\exp\left(\log(\theta_0) + \theta_1 \log\left(\frac{d_j}{d^*}\right)\right)}{1 + \exp\left(\log(\theta_0) + \theta_1 \log\left(\frac{d_j}{d^*}\right)\right)}. \tag{38}$$

Under the above reparameterization, the BLRM estimates the probability of DLT through the logarithm of a standardized dose [27], [68], and the model parameters are interpreted as:

- θ_0 represents the odds of DLT at the reference dose d^* ($\log(\theta_0)$ is the log-odds of DLT at reference d^*)
- θ_1 represents the increase log-odds of DLT by unit increase of log-dose.

According to Neuenschwander *et al.* [27], the two-parameter model utilized is more flexible and gives a more realistic representation of the dose-toxicity curve than the one-parameter power models used frequently in the CRM.

With the BLRM, the first cohort of patients is treated at the lowest dose level. The following cohorts follow a similar process to the other methods, the subsequent doses are selected based on the updated dose-toxicity curve, that reflects the accumulated toxicity data across all dose levels.

The posterior distribution of the model parameters is determined through the Bayes theorem, to give the updated dose-toxicity curve, by

$$h(\theta_0, \theta_1|D_k) \propto h(\theta_0, \theta_1) l(\theta_0, \theta_1|D_k) \tag{39}$$

where $h(\theta_0, \theta_1|D_k)$ corresponds to the posterior, $h(\theta_0, \theta_1)$ the join prior and $l(\theta_0, \theta_1|D_k)$ represents the Bernoulli data model,

$$l(\theta_0, \theta_1|D_k) \equiv f(x_k|\theta_0, \theta_1) = \prod_{k=1}^{k} p_k^{y_k}(1 - p_k)^{1-y_k} \tag{40}$$

and D_k represents the data after observation of k patients, $x_k \in J$ the dose level that patient k receives and y_k denotes the DLT outcome.

The authors [27] recommended using a weekly informative prior, like a bivariate normal prior,

$$(\log(\theta_0), \log(\theta_1) \sim \text{BVN}\left(\begin{pmatrix}\mu_0\\ \mu_1\end{pmatrix}, \Sigma\right), \qquad \Sigma = \begin{pmatrix}\sigma_0^2 & \rho\sigma_0\sigma_1\\ \rho\sigma_0\sigma_1 & \sigma_1^2\end{pmatrix} \tag{41}$$

They also recommended using informative priors whenever available.

Dose recommendation for the following cohorts will be based on the estimated posterior probabilities of DLT, $\hat{p}_j$, and on an overdose control. Like the EWOC, the BLRM can impose an overdose control. This overdose control criterion is satisfied for the dose levels J with posterior probability of exceeding δ_2 less than α,

$$P(\hat{p}_j > \delta_2 | D_k) < \alpha. \tag{42}$$

The authors [27] utilized $\alpha = 0{,}25$ in their original paper.

For each candidate dose level, the estimated posterior probabilities of DLT are summarized by the toxicity probability intervals and utilized in the dose recommendation process [27]. The dose chosen for the next cohort will be, among the doses that fulfil the overdose control criterion, the dose that has a maximal posterior toxicity probability for the target toxicity interval, i.e., the dose for cohort $k + 1$ will be given by

$$x_{k+1} = \max\{P(\hat{p}_j \in (\delta_1, \delta_2) | D_k)\} \tag{43}$$

[27], [68].

The trial ends when the maximum sample size is reached or when no further escalation is possible due to no higher dose levels satisfying the overdose control criterion [32], [68]. The highest dose that falls in the target toxicity interval and satisfies the overdose control criterion is acknowledged as the MTD [68].

2.2.3 Model-assisted designs

The model-assisted designs are a class of designs that emerged recently. They are hybrid methods that combine model-based designs' superior performance with the algorithm-based designs' simplicity [69].

Like the model-based designs, the model-assisted designs utilize a statistical model for efficient decision-making, but their dose-escalation and de-escalation rules can be pre-tabulated before the trials start [69], thus, their implementation is simple and clear, as the algorithm-based designs [12], [17].

Unlike the model-based designs, where the dose-toxicity relationship is estimated across all doses, the model-assisted designs model only local data, i.e., model the data observed for each dose level independently, typically, using a beta-binomial model [54], [67].

Several simulation studies were performed and determined that the model-assisted designs yield greater performance than the algorithm-based designs and are comparable to the more complicated model-based designs [17].

Examples of model-assisted designs include the mTPI and the BOIN and will be described below.

2.2.3.1 Notation

Before describing the model-assisted designs, we start by establishing the notation. We will assume J pre-specified dose levels being utilized in a clinical trial, such that $J = \{d_1 < d_2 < \cdots < d_j\}$. Let p_j denote the probability of DLT corresponding to the dose level d_j and ϕ the chosen target toxicity level. Assuming a total of n patients participating in the trial, n_j will denote the number of patients treated at dose level d_j until a particular point of the trial and y_j the respective number of patients that experienced DLT at the same dose level.

Under the model-assisted designs, described in the following sections, the escalation decisions are based exclusively on the DLT probabilities calculated for the current dose-level, therefore, when appropriate, we will denote n_{cur}, y_{cur} and p_{cur}, respectively as the number of patients treated at the current dose level, the number of DLTs observed at the current dose and the probability of DLT at the current dose.

2.2.3.2 Modified Toxicity Probability Interval

The mTPI was proposed by Ji *et al.* [53] based on the toxicity probability interval method published by Ji *et al.* [70], which was the founding method that established the utilization of toxicity probability intervals.

The mTPI is a model-assisted design in which subsequent cohorts are assigned to dose levels according to the toxicity outcome at the current dose, by calculating the posterior probabilities of DLT belonging to one of three intervals [1], [9].

One advantage of the mTPI design is that the dose decision rules can be simply understood and implemented. Like the 3+3 design, the escalation decision-making is algorithmic, and all possible dose assignment actions are available in advance. Nevertheless, this design is assisted by a beta-binomial model that independently models the accumulated data available for each dose level [67].

Under the mTPI, we assume a Bayesian beta-binomial model. We assume Y as binomially distributed,

$$y_{cur}|p_{cur} \sim \text{Bin}(n_{cur}, p_{cur}) \tag{44}$$

and assume that the probability of toxicity at the current dose d_{cur}, p_{cur}, has a beta conjugate prior with hyperparameters α and β,

$$p_{cur} \sim \text{Beta}(\alpha, \beta). \tag{45}$$

The mTPI obtained its best performance under the prior $\text{Beta}(1,1) \equiv \text{Unif}(0,1)$, since this method was calibrated for the uniform prior. For that reason, the authors recommended using the prior $\text{Beta}(1,1)$, yet others can be used [53].

Consequently, the posterior toxicity probabilities for each dose level, p_{cur}, are derived by the following beta posterior

$$p_{cur}|y_{cur} \sim \text{Beta}(1 + y_{cur}, 1 + n_{cur} - y_{cur}). \tag{46}$$

To establish the decision framework for the dose allocation in this design, the authors proposed breaking the probability unit interval [0,1] in three subintervals, using two constants that can be elicited from clinicians, ε_1 and ε_2. The constants ε_1 and ε_2 are used to define the equivalence interval (EI), $[\phi - \varepsilon_1, \phi + \varepsilon_2]$, $\varepsilon_1, \varepsilon_2 \geq 0$, which corresponds to the interval for the potential candidates for the true MTD [53]. For example, if the target toxicity level is $\phi = 0{,}3$, then a clinician may agree to select any dose with toxicity probability between $[0{,}25;\ 0{,}35]$. In this example $\varepsilon_1 = \varepsilon_2 = 0{,}05$.

The two other intervals are given by $[0, \phi - \varepsilon_1[$ and $]\phi + \varepsilon_2, 1]$. Each interval corresponds to:

- UI, $\text{UI} = [0, \phi - \varepsilon_1[$, the current dose is deemed below MTD;
- EI or proper dosing, $\text{EI} = [\phi - \varepsilon_1, \phi + \varepsilon_2]$, the current dose is deemed close to the MTD;
- OI, $\text{OI} =]\phi + \varepsilon_2, 1]$, the dose is deemed higher than the MTD [53].

Given the above intervals, the decision rules of the mTPI are driven by the Unit Probability Mass (UPM), which is defined as the quotient of the posterior probability of the interval and the length of the interval [53].

Assuming the above definition and the posterior DLT probability for the current dose, it is possible to calculate the UPM for each one of the three intervals,

$$\text{UPM}_{\text{UI}} = \frac{P(\hat{p}_{cur} \in \text{UI})}{(\phi - \varepsilon_1)} \tag{47}$$

$$\mathrm{UPM}_{\mathrm{EI}} = \frac{P(\hat{p}_{cur} \in \mathrm{EI})}{(\varepsilon_1 + \varepsilon_2)} \tag{48}$$

$$\mathrm{UPM}_{\mathrm{OI}} = \frac{P(\hat{p}_{cur} \in \mathrm{OI})}{(1 - \phi + \varepsilon_2)} \tag{49}$$

[71].

The UPMs give the posterior probability of the toxicity probability falling in each of the three intervals divided by the width of the interval and corresponds to the average probability density of the interval [67], [72]. The dose-escalation decisions follow the largest UPM.

Given what was exposed above, the dose-finding algorithm of the mTPI design works as follows. The first cohort of patients is treated with the lowest dose level. Given the observed toxicity outcomes, the UPMs are calculated. If the $\mathrm{UPM}_{\mathrm{UI}}$ is maximized, the dose is escalated to dose level d_{cur+1}; if the $\mathrm{UPM}_{\mathrm{EI}}$ is maximized, the dose is retained at the current dose level d_{cur}; if the $\mathrm{UPM}_{\mathrm{OI}}$ is maximized, the dose is de-escalated to dose level d_{cur-1}.

The following cohorts are treated, and the previous step is repeated until a stopping rule is satisfied (e.g., the maximum sample size). Fig. 1. illustrates the mTPI flowchart.

As mentioned earlier, once all the variables are defined (ϕ, ε_1, ε_2 and samples size), the mTPI can give in advance all possible dose assignment actions. In appendix A1. we find all possible dose allocation actions to be implemented for a clinical trial with the following parameters: $\phi = 0{,}25$, $\varepsilon_1 = \varepsilon_2 = 0{,}05$ and $n = 30$.

Like other methods, the mTPI has safety rules built into its design. Namely, an early termination rule and a dose exclusion rule.

The early termination rule establishes that if the first cohort of patients was treated with the lowest dose level and $P(\hat{p}_1 > \phi | y_1, n_1) > \xi$ for ξ close to 1 (e.g., ξ=0,95), the trial should be terminated due to excessive toxicity [53].

The dose exclusion rule determines that, given the observed data, if the decision is to escalate to dose d_{cur+1} and the $P(\hat{p}_{cur+1} > \phi | y_j, n_j) > \xi$, for ξ close to 1 (e.g., ξ=0,95), then the next cohort of patients should be treated at the same dose level d_{cur} and dose levels d_{cur+1} and higher should be excluded from the trial [53].

At the end of the trial, after all the toxicity outcomes are observed for all patients, the MTD is selected through an independent method [71]. The MTD is selected using the Pool Adjacent Violators Algorithm (PAVA), proposed by Barlow *et al.* A method that borrows strength across doses [73].

First, for each dose level, the estimated posterior toxicity probability means, $\hat{p}_j$, are computed. Then, the PAVA is applied on the $\hat{p}_j$ values so that the resulting transformed values $\hat{p}_j^*$ increase with dose, ensuring the desired non-decreasing monotonicity.

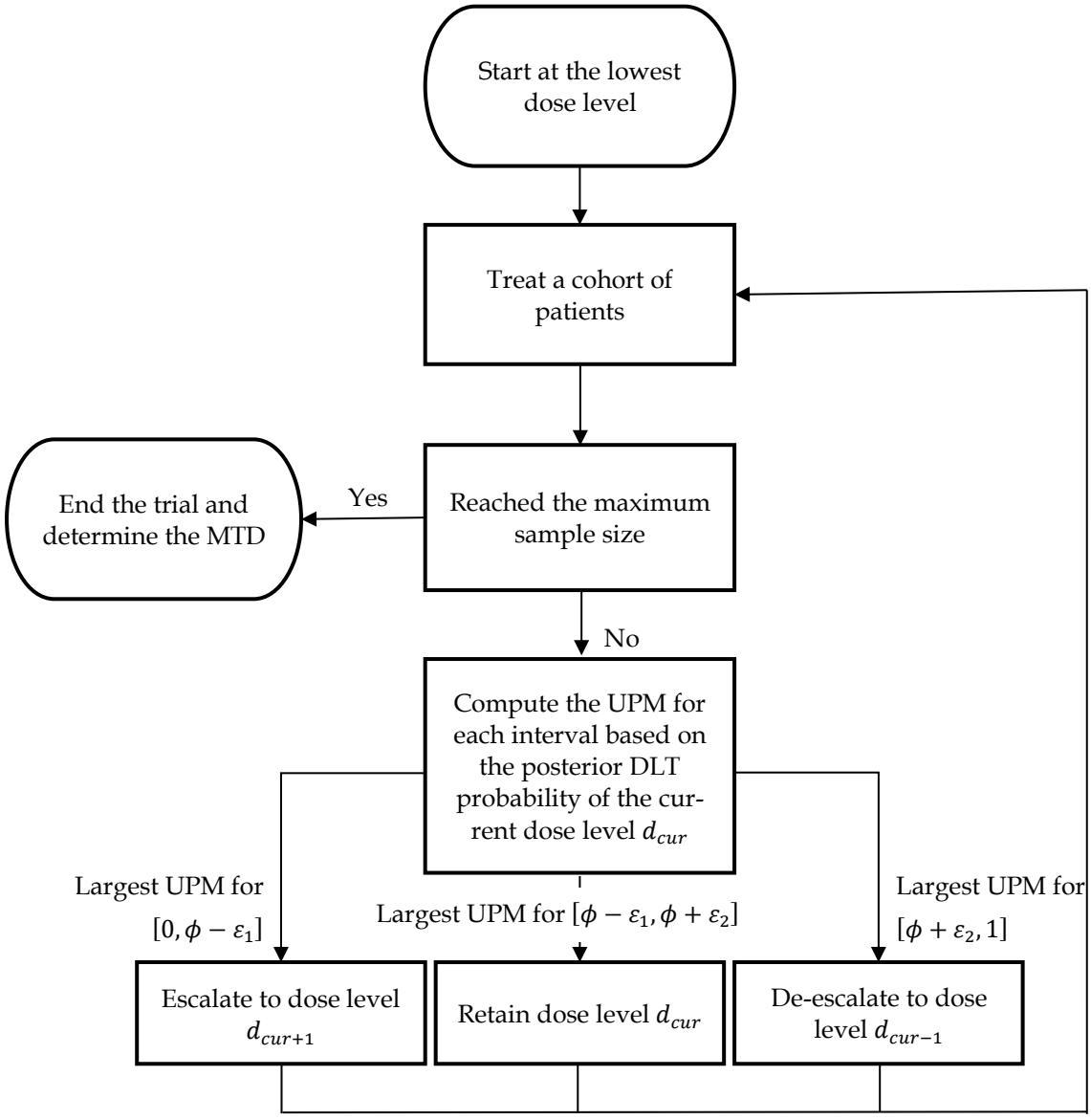


Fig. 1. Dose escalation flowchart for the mTPI.

Operatively, the PAVA replaces any adjacent $\hat{p}_j$ that violates the non-decreasing monotonicity by their weighted average [14]. That is, for the ordered dose levels $J = \{d_1 <, \dots, < d_j\}$, if the estimated posterior toxicity probability means increase with dose for all doses, such that $\{\hat{p}_1 \leq, \dots, \leq \hat{p}_j\}$, the observed toxicity probabilities are monotonic and $\hat{p}_j = \hat{p}_j^*$ [14]. If $\hat{p}_j > \hat{p}_{j+1}$, $\hat{p}_j$ is considered a monotonicity violator and the algorithm will replace both estimates by their weighted average, such that, $\hat{p}_j^* = \hat{p}_{j+1}^* = (\hat{p}_j n_j + \hat{p}_{j+1} n_{j+1})/(n_j + n_{j+1})$ [74]. This process is iteratively repeated until we obtain a non-decreasing set of transformed values $\hat{p}_j^*$, $\{\hat{p}_1^* \leq, \dots, \leq \hat{p}_j^*\}$.

The dose level that corresponds to the MTD will be the dose level with the minimum absolute difference between the transformed posterior mean and the target toxicity ϕ, i.e.,

$$\widehat{\text{MTD}} = \arg\min_j \left|\hat{p}_j^* - \phi\right| \tag{50}$$

[53].

The authors proposed the following rules to select the MTD in the case of ties. The tied doses can only be $(\phi + q)$ or $(\phi - q)$ for some $q \in (0, \phi)$.

- If there is at least one dose where $\hat{p}_j^* \in (\phi - q)$, choose the highest dose among those at which $\hat{p}_j^* \in (\phi - q)$
- Otherwise, choose the lowest dose among those at which $\hat{p}_j^* \in (\phi + q)$ [53].

2.2.3.3 Bayesian Optimal Interval

Liu and Yuan [14] presented the local and general BOIN designs. The first showed better performance and, for that reason, was the method endorsed by the authors and is the method described below [14], [67].

The BOIN design is easy to implement and doesn't require complex statistical knowledge, similar to traditional 3+3 design. Yet, it is more flexible, as it allows choosing the target toxicity level and allows changes in the cohort size during the trial. It also yields substantially better performance, comparable to the model-based designs [34].

The BOIN recommends dose escalation, de-escalation, and dose maintenance by comparing the estimated DLT rate for the current dose level with a pair of fixed dose-escalation and de-escalation boundaries [67].

Assuming a total of J prespecified doses under investigation and the target toxicity level ϕ, we assume that patients are treated in cohorts whose size may vary [14]. Let $\hat{p}_{cur}$ denotes the observed DLT rate for the current dose level d_{cur}, y_{cur} the respective total number of DLTs observed and n_{cur} the total number of total patients treated at the same dose level, such that

$$\hat{p}_{cur} = \frac{y_{cur}}{n_{cur}} \tag{51}$$

[14], [34].

Let $\lambda_{1j}(n_j, \phi)$ and $\lambda_{2j}(n_j, \phi)$ denote the lower and upper boundaries such that $0 < \lambda_{1j}(n_j, \phi) < \lambda_{2j}(n_j, \phi) < 1$.

Given the above boundaries, the dose assignment actions are:

if $\hat{p}_{cur} \leq \lambda_{1j}(n_j, \phi)$, the dose is escalated to dose level d_{cur+1}

if $\lambda_{1j}(n_j, \phi) < \hat{p}_{cur} < \lambda_{2j}(n_j, \phi)$, the dose is retained at level d_{cur}

if $\hat{p}_{cur} \geq \lambda_{2j}(n_j, \phi)$, the dose is de-escalated to dose level d_{cur-1}.

Thus, the logic behind this method is simple, one just needs to establish the dose-escalation boundaries. To do so, Liu and Yuan [14] started by defining the correct and incorrect dose assignment decisions, in order to optimize the decision boundaries and minimize the chance of making incorrect dose assignments.

Assuming p_j as the true toxicity probability for dose level d_j, for $J = \{d_1 <, \dots, < d_j\}$, the authors formulated three hypotheses:

H_{0j}: $p_j = \phi$, the current dose d_j is the MTD (retain dose-level)

H_{1j}: $p_j = \phi_1$, the current dose d_j is below the MTD (escalate dose)

H_{2j}: $p_j = \phi_2$, the current dose d_j is overly toxic (de-escalate dose)

where ϕ_1 and ϕ_2 denotes, respectively, the highest toxicity probability considered subtherapeutic, and the lowest toxicity probability deemed toxic [14].

Correct decisions under hypotheses H_{0j}, H_{1j}, and H_{2j} are retaining dose-level (R), escalating dose (E), and de-escalating dose (D), respectively. On the other hand, incorrect decisions under the three hypotheses are given by the complements $\bar{R}$, $\bar{E}$ and $\bar{D}$ [14].

Under the Bayesian paradigm, each hypothesis was assigned a prior probability of being true: $\pi_{kj} = P(H_{kj}), k = 0, 1, 2$. It was assumed that the three hypotheses have an equal probability of being true, $P(H_{0j}) = P(H_{1j}) = P(H_{2j}) = 1/3$ (non-informative prior) [14].

Considering the dose assignment actions defined above, the probability of making an incorrect decision (the decision error rate) at each one of the dose assignments can be given by $\alpha(\lambda_{1j}, \lambda_{2j})$,

$$\begin{aligned}\alpha(\lambda_{1j}, \lambda_{2j}) &= P(H_{0j})P(\bar{R}|H_{0j}) + P(H_{1j})P(\bar{E}|H_{1j}) + P(H_{2j})P(\bar{D}|H_{2j}) = \\ &= P(H_{0j})P(y_j \leq n_j\lambda_{1j} \; or \; y_j \geq n_j\lambda_{2j}|H_{0j}) + P(H_{1j})P(y_j \\ &> n_j\lambda_{1j}|H_{1j}) + P(H_{2j})P(y_j < n_j\lambda_{2j}|H_{2j}) = \\ &= \pi_{0j}\{\text{Binomial}(n_j\lambda_{1j}; n_j\phi) + 1 - \text{Binomial}(n_j\lambda_{2j} - 1; n_j\phi)\} \\ &+ \pi_{1j}\{1 - \text{Binomial}(n_j\lambda_{1j}; n_j\phi_1)\} + \pi_{2j}\text{Binomial}(n_j\lambda_{2j} - 1; n_j\phi_2)\end{aligned} \quad (52)$$

where the binomial c.d.f. is represented by $\text{Binomial}(b; n; \phi)$, with parameter ϕ and size n at the value b [14].

Liu and Yuan [14] showed that the decision error rate is minimized when the boundaries λ_{1j} and λ_{2j} are given by:

$$\lambda_{1j} = \frac{\log\left(\frac{1-\phi_1}{1-\phi}\right) + n_j^{-1}\log\left(\frac{\pi_{1j}}{\pi_{0j}}\right)}{\log\left(\frac{\phi(1-\phi_1)}{\phi_1(1-\phi)}\right)} \quad (53)$$

$$\lambda_{2j} = \frac{\log\left(\frac{1-\phi}{1-\phi_2}\right) + n_j^{-1}\log\left(\frac{\pi_{0j}}{\pi_{2j}}\right)}{\log\left(\frac{\phi_2(1-\phi)}{\phi(1-\phi_2)}\right)}. \quad (54)$$

Equations (53) and (54) can be further simplified, as the values of π_{0j}, π_{1j} and π_{2j} are known, giving origin to:

$$\lambda_{1j} = \frac{\log\left(\frac{1-\phi_1}{1-\phi}\right)}{\log\left(\frac{\phi(1-\phi_1)}{\phi_1(1-\phi)}\right)} \quad (55)$$

$$\lambda_{2j} = \frac{\log\left(\frac{1-\phi}{1-\phi_2}\right)}{\log\left(\frac{\phi_2(1-\phi)}{\phi(1-\phi_2)}\right)}. \quad (56)$$

Thus, as we can observe in (55) and (56), the determination of the boundaries depends only on the values of ϕ, ϕ_1 and ϕ_2. These boundaries are fixed for all dose levels and it's common to find them in the literature denoted as λ_e and λ_d.

The values for $\phi_1$and ϕ_2 need to be specified. Liu and Yuan [14] proposed as default values $\phi_1 = 0.6\phi$ and $\phi_2 = 1.4\phi$ (i.e., 40% deviation from the target) for general use. Different values for ϕ_1 and ϕ_2 can be used, yet they shouldn't be too close to ϕ, because the method won't have enough power to distinguish when to escalate, deescalate or stay in the same dose [29]. Setting ϕ_1 and ϕ_2 close to ϕ prevents this method from being able to discriminate the target toxicity rate from the rates close to it, due to the small sample sizes that are utilized in phase I trials [14].

As the target toxicity rate ϕ is specified and known, the escalation and de-escalation boundaries, λ_{1j} and λ_{2j} can be easily determined using the default values for ϕ_1 and ϕ_2 suggested by Liu and Yuan [14].

Table III below gives the values for the boundaries, assuming $\phi_1 = 0{,}6\phi$ and $\phi_2 = 1{,}4\phi$, for some of the common target toxicity levels, ϕ, utilized in cancer FIH clinical trials.

TABLE III
BOIN interval boundaries for different target toxicity levels with $\phi_1 = 0{,}6\phi$ and $\phi_2 = 1{,}4\phi$

Interval boundaries	Target toxicity level ϕ					
	0,15	0,20	0,25	0,30	0,35	0,40
λ_{1j} (λ_e)	0,118	0,157	0,197	0,236	0,276	0,316
λ_{2j} (λ_d)	0,179	0,238	0,298	0,358	0,419	0,479

Adapted from [14]

Once the boundaries are defined, the BOIN design works as follows. The first cohort of patients is treated with the dose corresponding to the lowest dose. The escalation decisions are determined by comparing the estimated toxicity rate $\hat{p}_{cur}$ to the escalation boundaries. If $\hat{p}_{cur} \leq \lambda_{1j}$ the dose is escalated to level d_{cur+1}; if $\hat{p}_{cur} \geq \lambda_{2j}$ the dose is de-escalated to level

d_{cur-1}; if $\lambda_{1j} \leq \hat{p}_{cur} \leq \lambda_{2j}$ the dose is maintained at the same dose level. The dose also remains at the same dose level when the DLT rate for the lowest dose level suggests de-escalation, $\hat{p}_1 \geq \lambda_{2j}$, and when the DLT rate obtained for a cohort treated at the highest dose level suggests escalation, $\hat{p} \leq \lambda_{1j}$ [75].

The process continues until the maximum sample is reached or the trial is terminated due to observed excessive toxicity. Fig. 2. illustrates the BOIN design through a flowchart.

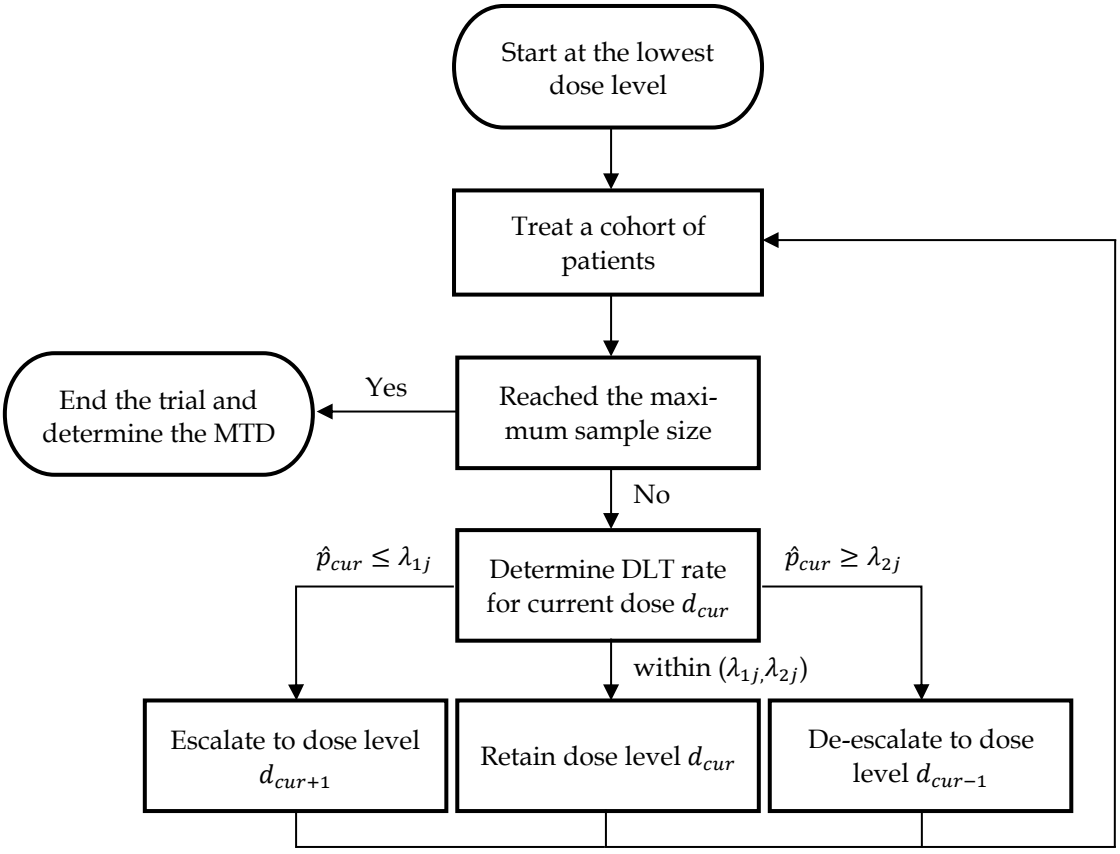


Fig. 2. Dose escalation flowchart for the BOIN.
Source: Adapted from [76]

The BOIN design has also in place a safety rule to avoid exposing patients to overly toxic doses. The current dose d_{cur} cannot exceed the target toxicity rate ϕ on 95%, $P(p_{cur} > \phi | y_{cur}, n_{cur}) > 95\%, n_{cur} \geq 3$. If this happens, and we have at least three patients treated at that dose level, dose levels d_{cur} and above are eliminated from the trial. The trial is terminated if the first dose level is eliminated [14].

The above posterior probability, $P(p_{cur} > \phi | y_{cur}, n_{cur})$, is calculated through a beta-binomial model, where $y_{cur}|p_{cur} \sim \text{Bin}(n_{cur}, p_{cu})$ and p_{cur} follows a vague prior, $p_{cur} \sim \text{Beta}(1,1)$. Under this beta-binomial model the posterior distribution of $\hat{p}_{cur}$ is given by

$$\hat{p}_{cur}|y_{cur} \sim Beta(1 + y_{cur}, 1 + \ n_{cur} - y_{cur}). \quad (57)$$

At the end of the trial, the MTD is selected using the same approach utilized in the mTPI method. The PAVA is applied to pool information across doses and obtain an efficient estimate of the MTD [34]. The PAVA identifies the dose levels that violate the monotonicity assumption and adjusts their toxicity rates to maintain monocity [75].

Thus, considering $\hat{p}_j^*$ as the transformed values of the respective estimated toxicity rates $\hat{p}_j$, the MTD will be selected as the dose d_j for which the monotonic estimate of the toxicity rate $\hat{p}_j^*$ is closest to ϕ,

$$\widehat{\text{MTD}} = \arg\min_j |\hat{p}_j^* - \phi|. \quad (58)$$

In the presence of ties for $\hat{p}_j^*$, the MTD is selected as the highest dose level when $\hat{p}_j^* < \phi$ or the lowest dose level when $\hat{p}_j^* > \phi$ [14].

3

SIMULATION STUDY - SIMULATED CLINICAL TRIALS

A simulation study was performed with ten true dose-toxicity scenarios to compare the performance of the five dose-escalation methods described in the previous chapters.

Based on the ten true dose-toxicity scenarios, and respective true DLT probabilities, DLT outcomes were randomly generated to simulate a clinical trial and compare the methods. One thousand trials were simulated for each method and each scenario.

A maximum sample size of $n = 30$ was assumed, and fixed cohorts of 3 patients were utilized. All simulations used the first dose-level as the starting dose and dose-skipping was not allowed, meaning that dose escalation occurred one level at a time. The target toxicity level was assumed to be $\phi = 0{,}33$.

3.1 Simulation Scenarios

Ten hypothetical true dose-toxicity scenarios were used to examine the performance of the dose-escalation methods. These scenarios were adopted from the paper by Wheeler *et al.* [65], which examined the EWOC toxicity-dependent *feasibility bound* approach, an approach where the *feasibility bound* α increases as the clinical trial unfolds.

The true dose-toxicity scenarios consider six discrete dose-levels with increments of 50 mg/m^2 between then, $J = \{150, 200, 250, 300, 350, 400\}$ mg/m^2. The true dose-toxicity scenarios were generated from a logistic function, to give steep, shallow and plateauing dose-toxicity curves. Table IV presents the true DLT probabilities for each dose level and each scenario. The dose-levels that represent the true MTD for each scenario are highlighted in bold.

TABLE IV
True dose-toxicity scenarios and respective DLT probabilities per dose-level

Scenario	Dose (mg/m²)					
	150	200	250	300	350	400
1	**0,28**	0,47	0,66	0,82	0,91	0,96
2	**0,31**	0,36	0,41	0,46	0,52	0,57
3	0,05	**0,33**	0,84	0,98	1,00	1,00
4	0,06	0,15	**0,33**	0,58	0,79	0,92
5	0,00	0,01	0,07	**0,33**	0,78	0,96
6	0,02	0,06	0,16	**0,33**	0,58	0,79
7	0,01	0,03	0,07	0,16	**0,33**	0,56
8	0,06	0,09	0,15	0,23	**0,33**	0,46
9	0,00	0,00	0,01	0,04	0,13	**0,33**
10	0,03	0,06	0,09	0,15	0,23	**0,33**

DLT probabilities corresponding to the true MTD identified in bold for each scenario

Scenarios 1 and 2 represent two extreme scenarios where only the first dose-level has a true probability of DLT below (and close to) the target toxicity level, $\phi = 0{,}33$. Both scenarios have an almost linear increase. Scenario 1 was generated to reflect a steep increase in the DLT probability and scenario 2 a shallower increase.

We must note that the first dose-level, which we utilized as the starting dose, should be a safe dose, but in scenarios 1 and 2 it has an elevated true DLT probability (0,28 and 0,31, respectively), which will translate into an elevated number of DLTs. In a real clinical trial, the research teams would, most likely, interrupt the trial if such a high number of DLTs were observed.

In scenario number 3 the true DLT probability increases steeply for the first three dose-levels and plateaus for the last three dose-levels. The second dose level, 200 mg/m², corresponds to the true MTD. Scenario number 4, on the other hand, represents a steep increase in the true toxicity probability across all dose levels. In this scenario, the true MTD corresponds to the third dose level, 250 mg/m².

Scenarios 5 and 6 have the dose-level number four, 300 mg/m², as the true MTD. Scenario number 5 reflects a shallow DLT probability increase for the first three dose levels and a steeper increase for the last three. In scenario number 6 the DLT probability increase is more pronounced across the dose-levels.

In scenarios number 7 and 8 the true MTD corresponds to the fifth dose-level, 350 mg/m². Scenario 7 represents a curve where the probability of DLT increase is low on the first three doses, followed by larger increases. In scenario number 8 the toxicity probability increases in an almost linear fashion.

Scenarios 9 and 10 have all the dose-levels with true DLT probability below the target toxicity level, except the last one, 400 mg/m², which has a true DLT probability equal to 0,33, corresponding to the true MTD.

Fig. 3. illustrates the true dose-toxicity scenarios utilized in the simulation study. The dashed horizontal line corresponds to the target toxicity level, $\phi = 0{,}33$.

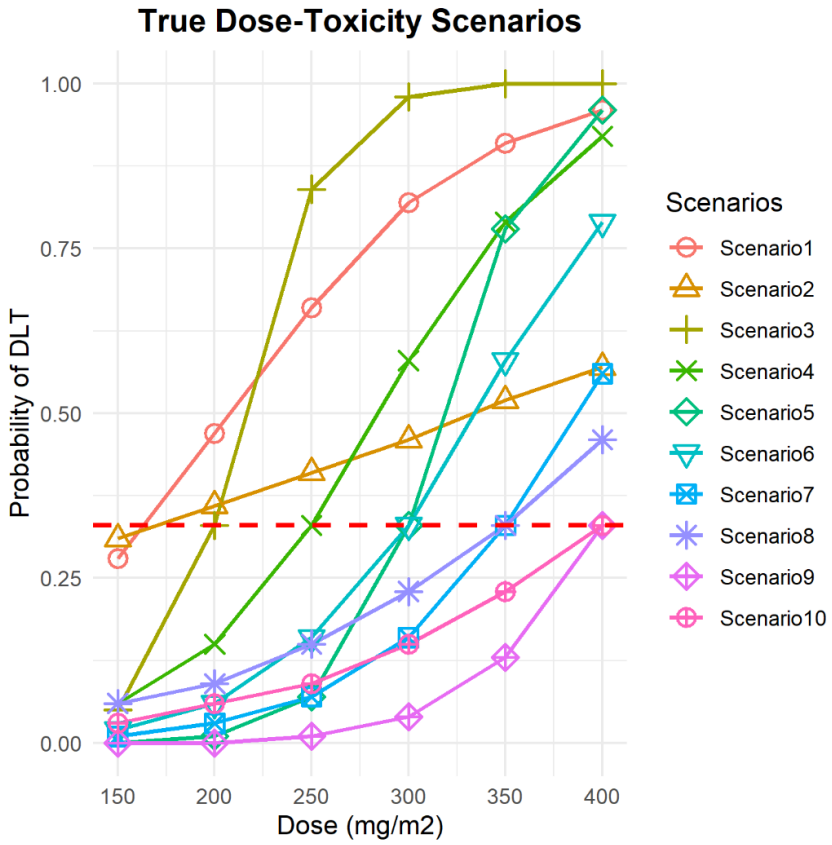


Fig. 3. Illustration of the ten true dose-toxicity scenarios relating the dose-levels with the probability of DLT.

3.2 Performance Metrics

Two general aspects were considered to assess the performance of the dose-escalation designs, the accuracy of the dose-escalation methods to find the true MTD and the ethical concern of optimizing the dose allocation. Consequently, four performance metrics were used to evaluate the dose-escalation methods, organized under three topics, accuracy, dose allocation and toxicity:

- **Accuracy:**
    - Percentage of correct selection (PCS) of the true MTD, defined as the percentage of the 1000 simulated trials in which the correct dose that corresponds to the true MTD was selected as the MTD. The average PCS across the ten scenarios was determined to summarize the accuracy of each method.
- **Dose allocation:**
    - Mean number (and percentage) of patients treated at the dose level corresponding to the true MTD across the 1000 simulated trials, calculated as the mean number of patients treated at the dose level corresponding to the true MTD in the 1000 simulated trials divided by the mean number of patients dosed in the 1000 simulated trials. Higher values of this metric indicate less exposure to subtherapeutic or potentially overly toxic doses. The average number of patients treated at the dose level corresponding to the true MTD across the ten scenarios was determined to summarize the methods that expose more patients to potentially therapeutic doses.
    - Mean number (and percentage) of patients treated above the true MTD in the 1000 simulated trials, defined as the quotient of the mean number of patients treated at dose levels above the MTD in the 1000 simulated trials by the mean number of patients treated in the 1000 simulated trials. This metric evaluates which dose-escalation methods are more aggressive and expose patients to higher dose levels, increasing the risk of exposing patients to overly toxic doses. The average number of patients treated above the true MTD across the ten scenarios was determined to summarize the methods that are more aggressive and expose more patients to potentially overly toxic dose levels.
- **Toxicity:**
    - Mean number of DLTs observed (and percentage) for each dose-escalation method across the 1000 simulated trials, defined as the mean number of DLTs observed divided by the mean number of patients dosed. Along with the previous metric, this measure allows evaluating which method subjects patients to higher doses, as patients dosed at higher dose levels tend to manifest more DLTs. The average number of DLTs observed across the ten scenarios was determined to summarize the dose-escalation methods that exhibit a more aggressive approach.

Note that to facilitate the interpretation of the results, the word *mean* was used when referring to the simulations' results for each scenario individually, and the word *average* was used when referring to the results of the ten aggregated scenarios.

3.3 Dose-Escalation Methods Configuration

3.3.1 Continual Reassessment Method

A modified single-parameter CRM was employed in the simulation study, utilizing the reparametrized power model, $f(x_k, \theta) = p_j = x^{\exp(\theta)}$. The CRM was set up to start at the lowest dose, to avoid overly aggressive escalation, and dose escalation was restricted to one level at the time.

A non-informative normal prior was selected, with zero mean and variance large enough to be considered non-informative, as suggested by O'Quigley [77]. $\log(\theta) \sim N(0; 1{,}34^2)$ was chosen by convenience, as it is a prior commonly found in the literature.

Given that we did not assume any prior knowledge regarding the dose-toxicity relationship and the CRM requires the *skeleton* to get the transformed dose toxicity labels, the initial toxicity guesses were obtained through the algorithm proposed by Lee and Cheung [78], the indifference interval technique.

Under this approach, one specifies the target toxicity level, ϕ, the prior dose level that is deemed to correspond to the MTD, denoted by ν, $\nu = \{1, \dots, j\}$, and an acceptable halfwidth indifference interval, δ. Once all the values are selected, the initial toxicity probability guesses, $p_0 = \{p_{01}, \dots, p_{0j}\}$, are computed recursively from the equations below, for the power model:

$$\begin{aligned} p_{0j-1} &= \exp\left(\frac{\log(\phi - \delta)\log(p_j)}{\log(\phi + \delta)}\right), \qquad j = 2, \dots, \nu \\ p_{0\nu} &= \phi \\ p_{0j+1} &= \exp\left(\frac{\log(\phi + \delta)\log(p_j)}{\log(\phi - \delta)}\right), \qquad j = \nu, \dots, J-1. \end{aligned} \tag{58}$$

Once the initial toxicity probability guesses are calculated, the dose labels for the six dose levels, $J^* = \{d_1^*, \dots, d_6^*\}$, can be determined.

Lee and Cheung [78] suggested performing simulations with a range of indifference intervals to identify the best *skeleton*, which can be determined as the *skeleton* that yields the highest PCS.

The third dose level was assumed as being the prior MTD, $\nu = 3$. The half-width indifference interval $\delta = 0{,}06$ was chosen, as it was the *optimal* half-width indifference interval in the simulations performed by Lee and Cheung [78], for 6 dose levels, $\phi = 0{,}33$, $\nu = 3$ and $n = 30$.

Table V presents the *skeleton* obtained from the above algorithm and the respective dose labels for the power model $f(x_k, \theta) = p_j = x^{\exp(\theta)}$. Fig. 4. shows the initial dose-toxicity curve before observing any data for CRM, EWOC and BLRM.

TABLE V
CRM initial dose-toxicity guesses

Skeleton obtained from Lee and Cheung's algorithm (p_0)	0,117	0,214	0,330	0,451	0,564	0,662
Dose labels (d_j^*)	0,117	0,214	0,330	0,451	0,564	0,662

3.3.2 Escalation With Overdose Control

A *feasibility bound* of $\alpha = 0{,}25$ and the non-informative priors proposed by Babb *et al.* [28], $\gamma \sim \text{Unif}(X_{min}, X_{max})$ and $\rho_0 \sim \text{Unif}(0, \phi)$, were used in the 1000 clinical trials simulated for the EWOC. $Xmin$ and $Xmax$ were defined as $Xmin = 100$ and $Xmax = 450$, since the authors suggest defining $Xmin$ as a dose lower or equal to d_1 and $Xmax$ as a dose greater than the highest allowable dose in the clinical trial. $Xmin$ was defined as being a dose below the lowest dose available, 150 mg/m², because $Xmin$ should be a dose without any expected DLTs, and that is not true for some of the dose-toxicity scenarios.

3.3.3 Bayesian Logistic Regression Method

The vague default prior proposed by Neuenschwander *et al.* [27] was utilized as the prior distribution of the parameters in the BLRM, the bivariate normal distribution for $\log(\theta_0), \log(\theta_1)$,

$$(\log(\theta_0), \log(\theta_1) \sim \text{BVN}\left(\begin{pmatrix} -0{,}847 \\ 0{,}381 \end{pmatrix}, \Sigma\right), \qquad \Sigma = \begin{pmatrix} 2{,}015^2 & 0 \\ 0 & 1{,}027^2 \end{pmatrix}.$$

Considering the target toxicity level $\phi = 0{,}33$, the target toxicity interval [0,28; 0,38] was utilized, i.e., we considered $\delta_1 = \delta_2 = 0{,}05$ to define the target toxicity interval. Consequently, the UI and OI were defined as [0; 0,28[and]0,38; 1], respectively.

The BLRM was implemented with the sixth dose-level, 400 mg/m², as the reference dose-level and the overdose control was defined considering $\alpha = 0{,}25$.

The last dose level was selected as the reference dose, d^*, because it was the approach utilized by other authors [27], [32], [79] and because it was the approach that yielded better results in the initial simulations (results not shown).

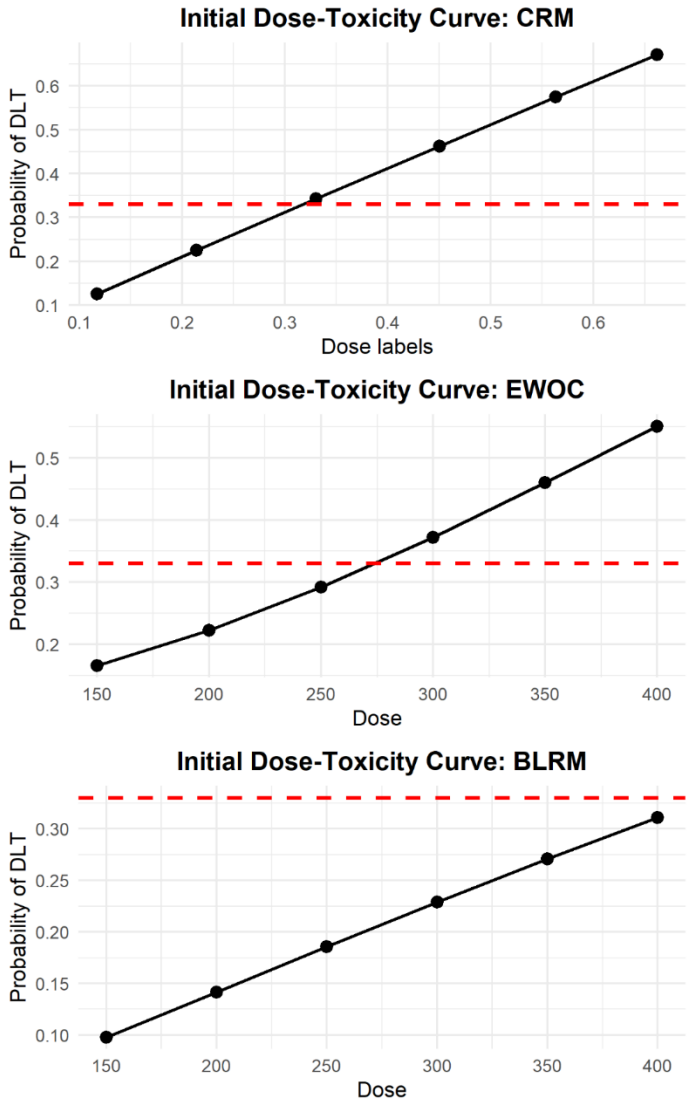


Fig. 4. Initial dose-toxicity curve before observing any data, for the CRM, EWOC and BLRM.

3.3.4 Modified Toxicity Probability Interval

p_{cur}~Beta(1,1) was used as the prior distribution in the mTPI, since it was the prior recommended by the authors [53]. The EI was defined to match the target toxicity interval used in BLRM. EI = [0,28; 0,38], considering $\phi = 0{,}33$ and $\varepsilon_1 = \varepsilon_2 = 0{,}05$.

3.3.5 Bayesian Optimal Interval

For the BOIN, the default vague prior was utilized, p_{cur}~Beta(1,1). The default dose-escalation boundaries were also employed, resulting in $\lambda_e = 0{,}260$ and $\lambda_d = 0{,}395$ as the escalation and de-escalation boundaries, respectively, considering $\phi = 0{,}33$, $\phi_1 = 0.6\phi$ and $\phi_2 = 1.4\phi$.

3.4 Software

Statistical analysis was performed with software R version 4.2.2 [80] and R-studio version 2023.03.1 [81].

CRM was implemented with the R package "bcrm" version 0.5.4 [82]. The *skeleton* for the CRM was obtained using the function getprior() on the "dfcrm" R package, version 0.2-2.1 [83], that implements the algorithm presented in 3.3.1.

BLRM and EWOC simulations were performed using the R package "crmPack" version 1.0.3 [84].

The mPTI was implemented using the R code available in the MD Anderson Cancer Center website, at [85], which was adapted to run the simulations with the specified configurations.

Simulations for the BOIN were executed with the R package "BOIN" version 2.7.2 [76].

It's important to highlight that the software available is extensive and that a significant portion of time was spent learning how to use software to simulate dose-escalation studies. Simulations were performed with different R packages, Excel™ macros, standalone applications, and online software. It is worth noting that some of this software is not user-friendly, making its utilization challenging.

The choice of the software for conducting the final simulations was driven by its compatibility with the defined simulation setup and the dose-escalation methods configuration. Additionally, the obtention of the simulation results in R facilitated subsequent statistical analysis.

3.5 Results and Discussion

Our simulation study examined ten true dose-toxicity scenarios, and 1000 clinical trials were simulated for each method and scenario. The number of simulated trials was chosen to be 1000, as we considered this number sufficiently large to allow the comparison between the methods, while ensuring that the simulations could be executed within a reasonable timeframe, given their computationally intensive nature.

In each simulation, the DLT outcomes were randomly drawn from the true dose-toxicity curves, and based on the simulated clinical trial data, the dose-escalation methods were run to identify the MTD.

It's important to note that the different DLT outcomes in each simulation can originate different dose escalation and de-escalation actions and, therefore, the identification of different dose levels as the MTD. It's also important to note that the general stopping rule implemented in the simulations was the treatment of 30 patients on each one of the clinical trials simulated.

The results of the 1000 simulated clinical trials for each method and each scenario are presented in Tables VI and VII, and in Fig. 5., which summarizes the simulation results through boxplots.

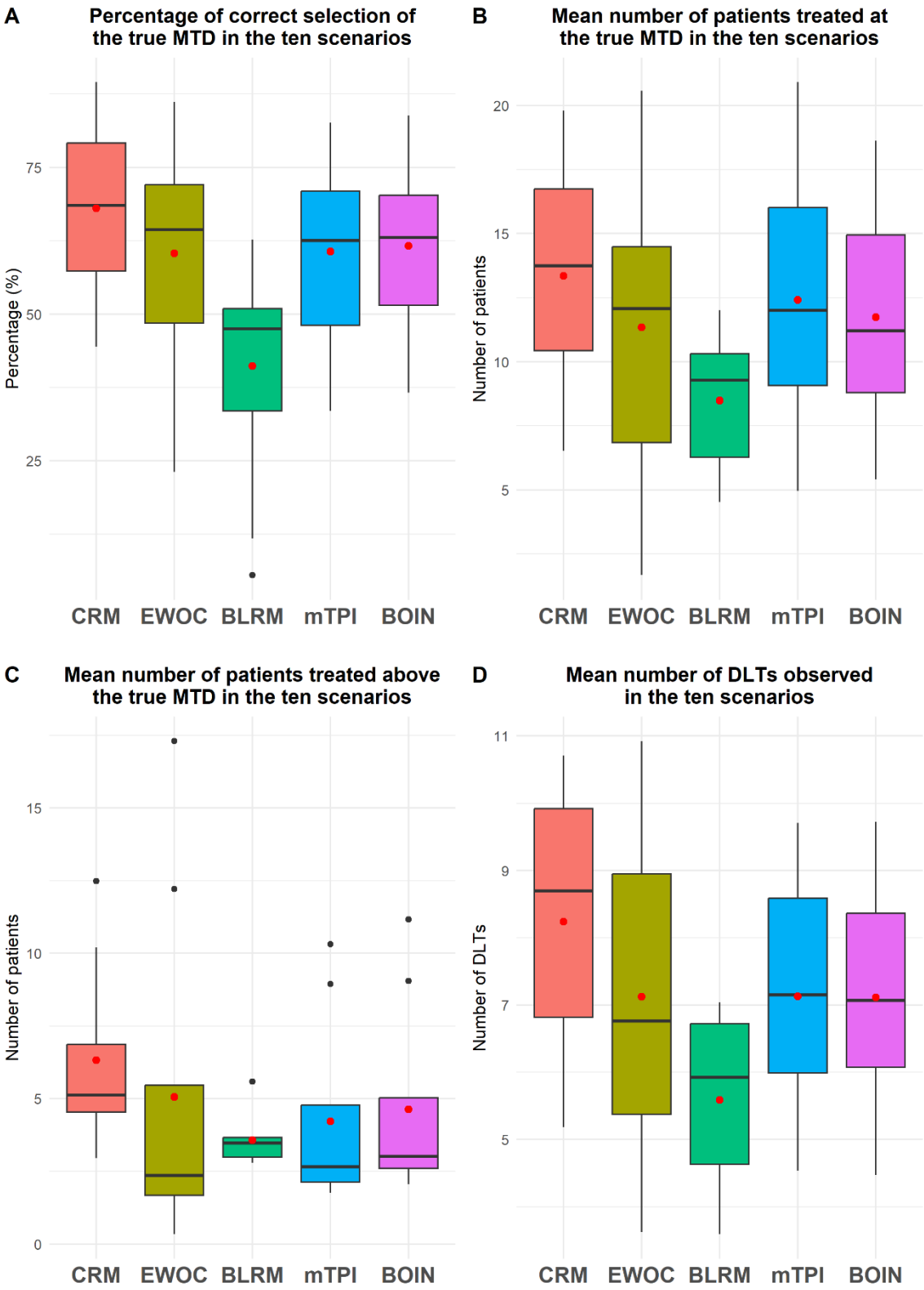


Fig. 5. Boxplots for the PCS, mean number of patients treated at the MTD, mean number of patients treated above the MTD and mean number of DLTs, for the ten scenarios. The red dots indicate the average value calculated across the ten scenarios.

3.5.1 Accuracy

We will start by analyzing the PCS, and, specifically, the simulated clinical trials where the MTD could not be determined due to the clinical trials being stopped before treating 30 patients (identified as No MTD in Table VI).

As described in the previous chapters, some dose-escalation methods implement safety rules to avoid exposing patients to excessively toxic doses. It's the case of the model-assisted designs, mTPI and BOIN, where a clinical trial is terminated if the estimated posterior probability of DLT at the lowest dose level exceeds a certain value (e.g., 95%), $P(\hat{p}_1 > \phi | y_1) > 0{,}95$, which happened in scenarios 1 and 2.

Scenarios 1 and 2 are examples of misspecification of the dose levels under study, where the first dose level corresponds to the true MTD (with true probability of DLT of 0,28 and 0,31, respectively), meaning that all the dose levels above are potentially toxic. It's important to note that the first dose level should correspond to a safe dose, where the probability of DLT is assumed to be low, and consequently, no DLTs are usually expected.

The percentage of clinical trials stopped with mTPI and BOIN on those two scenarios, due to the above safety rule, were 10,8% and 16,3%, and 11,3% and 16,2%, respectively.

The other method where stopped clinical trials were observed was the BLRM, on a much larger scale. The overdose control criterion in BLRM leads to a natural stopping rule, which is, if the lowest dose is an overdose, $P(\hat{p}_1 > 0{,}38 | D_k) \geq 0{,}25$, no further escalation can be performed, and the trial is stopped.

A significant number of the simulated trials were stopped due to the above stopping rule. That is, after observing the toxicity data and estimating the posterior probability of DLT, no dose levels satisfied the overdose criterion, and the clinical trials were stopped.

In the simulations performed, 83,6% of the simulated trials in scenario 1 were stopped; 79,1% in scenario 2; 39,0% in scenario 3; 19,8% in scenario 4; 6,3% in scenario 6; 3,3% in scenario 7; 18,9% in scenario 8; and 9,1% in scenario 10. Only scenarios 5 and 9 simulations had no trials stopped. A significative number of simulated trials stopped were also observed in Yao's [79] and Zhou *et al.*'s [32] research.

The findings from the stopped clinical trials indicate that the BLRM displays high sensitivity towards toxicity outcomes, particularly to the occurrence of DTL outcomes early in the trial. While the considerable amount of stopped clinical trials complicates the assessment of BLRM's performance, it's essential to acknowledge that this heightened sensitivity to DLTs prioritizes patient safety and welfare, reason that may contribute to the extensive utilization of this method presently.

TABLE VI
Percentage of correct selection of the true MTD (%)

Dose levels (mg/m²)		150	200	250	300	350	400	No MTD	Dose levels (mg/m²)		150	200	250	300	350	400	No MTD	Average PCS across the ten scenarios	
Scenario 1	CRM	**67,7**	31,4	0,9	-	-	-	N/A-	**Scenario 6**	CRM	-	0,2	15,3	**69,2**	15,3	-	N/A		
	EWOC	**74,0**	25,2	0,8	-	-	-	N/A		EWOC	-	0,4	25,8	**63,2**	10,6	-	N/A		
	BLRM	**11,8**	4,3	0,3	-	-	-	83,6		BLRM	-	1,4	34,1	**52,0**	5,6	0,6	6,3	CRM	68,0
	mTPI	**64,0**	25,0	0,2	-	-	-	10,8		mTPI	-	1,1	27,4	**61,0**	10,5	-	N/A		
	BOIN	**63,0**	25,5	0,2	-	-	-	11,3		BOIN	-	1,1	24,6	**63,0**	11,2	0,1	N/A		
Scenario 2	CRM	**51,3**	29,9	13,7	4,5	0,6	-	N/A	**Scenario 7**	CRM	-	-	0,2	16,3	**67,8**	15,7	N/A		
	EWOC	**43,5**	37,8	16,3	2,4	-	-	N/A		EWOC	-	-	0,6	28,6	**64,9**	5,9	N/A		
	BLRM	**5,5**	9,3	4,2	1,5	0,3	0,1	79,1		BLRM	-	-	2,0	39,1	**47,5**	8,1	3,3	EWOC	60,3
	mTPI	**44,9**	24,4	11,6	2,4	0,4	-	16,3		mTPI	-	-	2,2	28,5	**57,6**	11,7	N/A		
	BOIN	**41,2**	26,5	11,6	3,8	0,7		16,2		BOIN	-	-	1,3	27,1	**58,8**	12,8	N/A		
Scenario 3	CRM	9,9	**83,1**	7,0	-	-	-	N/A	**Scenario 8**	CRM	-	0,3	6,5	33,3	**44,4**	15,5	N/A		
	EWOC	12,2	**86,1**	1,7	-	-	-	N/A		EWOC	-	0,40	10,50	43,80	**41,20**	4,10	N/A		
	BLRM	16,0	**44,5**	0,5	-	-	-	39,0		BLRM	-	0,5	8,3	30,7	**29,8**	11,8	18,9	BLRM	41,1
	mTPI	17,1	**82,6**	0,3	-	-	-	N/A		mTPI	0,4	2,2	14,9	36,9	**33,5**	12,1	N/A		
	BOIN	16,8	**83,1**	0,1		-	-	N/A		BOIN	0,3	1,5	12,0	34,1	**36,6**	15,5	N/A		
Scenario 4	CRM	-	12,6	**71,4**	16,0	-	-	N/A	**Scenario 9**	CRM	-	-	-	-	10,5	**89,5**	N/A		
	EWOC	0,1	23,2	**66,2**	10,4	0,1	-	N/A		EWOC	-	-	-	0,4	35,8	**63,8**	N/A		
	BLRM	0,7	27,9	**47,8**	3,4	0,3	0,1	19,8		BLRM	-	-	-	1,6	35,9	**62,5**	N/A	mTPI	60,7
	mTPI	1,5	24,2	**65,7**	8,5	0,1	-	N/A		mTPI	-	-	-	0,8	26,5	**72,7**	N/A		
	BOIN	1,5	23,8	**65,8**	8,9	-	-	N/A		BOIN	-	-	-	0,8	27,5	**71,7**	N/A		
Scenario 5	CRM	-	-	11,1	**81,7**	7,2	-	N/A	**Scenario 10**	CRM	-	-	0,9	8,7	36,5	**53,9**	N/A		
	EWOC	-	-	18,6	**77,3**	4,1	-	N/A		EWOC	-	0,1	0,8	17,7	58,3	**23,1**	N/A		
	BLRM	-	-	35,6	**62,7**	1,7	-	N/A		BLRM	-	-	1,2	11,8	30,5	**47,4**	9,1	BOIN	61,6
	mTPI	-	-	17,4	**81,4**	1,2	-	N/A		mTPI	-	-	4,2	16,4	36,1	**43,3**	N/A		
	BOIN	-	-	15,2	**83,8**	1,0	-	N/A		BOIN	-	-	2,3	14,3	34,3	**49,1**	N/A		

PCS for the dose levels corresponding to the true MTD identified in bold for each scenario

In their studies, Yao [73] and Zhou *et al.* [27] performed simulations with the BLRM with no overdose control criterion. Their simulations with the BLRM with no overdose control obtained no stopped trials [73] or a low number of stopped trials [27] (Yao [73] didn't use any stopping rules, while Zhou *et al.* [27] used a stopping rule, $P(\hat{p}_1 > \phi | D_k) > 0{,}95$). Regarding PCS, the BLRM without overdose control enhanced precision in determining the true MTD, achieving a level of accuracy similar or even higher than the CRM. In the present study, simulations with the BLRM without overdose control were not performed, as the BLRM with no overdose control is not generally used in early development clinical research.

Continuing the examination of the PCS of the true MTD, the simulation results revealed that the CRM was the method with the highest accuracy to correctly identify the true MTD.

When considering the average PCS across the ten scenarios, the CRM obtained an average PCS of 68,0%, followed by the BOIN, 61,6%, mTPI, 60,7%, EWOC, 60,3%, and the BLRM, 41,1%. It's also clear that the BLRM was the method that performed worst, with an average PCS of 41,1%. Regarding the other three methods, the differences between the BOIN, mTPI and EWOC were minimal, and the methods can be considered comparable in terms of average PCS.

The CRM consistently emerged as the method with higher accuracy for determining the true MTD in the simulation study, it obtained the highest PCS values in most scenarios, in seven out of the ten scenarios: 2 (51,3%), 4 (71,4%), 6 (69,2%), 7 (67,8%), 8 (44,4%), 9 (89,5%), and 10 (53,9%).

Ulas and Karaman's [86] numerical study, comparing the CRM, mTPI and BOIN, among other designs, utilizing twelve different scenarios, showed that the CRM was the method that performed better, exhibiting the highest PCS. Conversely, other authors [32], [54] performed extensive simulations and found no significant differences between the CRM and other methods, such as mTPI and the BOIN design. Zhu *et al.* [75], in turn, did not demonstrate relevant differences between the CRM and BOIN, in their study comparing the 3+3 design, CRM, BOIN and Keyboard designs.

Zhou *et al.*'s [32] simulation study also revealed that the BLRM and EWOC had the weakest performance, however, the results here presented showed that the EWOC doesn't perform as poorly as the BLRM.

Analyzing Fig. 5., it's also possible to verify that the CRM, mTPI and the BOIN design were the most consistent methods, as they exhibited less variability in the PCS across the ten scenarios, comparing to the EWOC and the BLRM.

The EWOC exhibited the highest PCS in the first and third scenarios and had the second-highest PCS in scenarios 4, 6, 7 and 8. Examining those scenarios, we found no pattern in the dose-toxicity curves to justify the positive performance. However, it's noticeable that the

EWOC loses accuracy when the MTD is on the last dose level. The EWOC was the method with the second-lowest PCS in scenario 9 (63,8% compared to CRM's 89,5%) and the method with the lowest PCS in scenario 10 (23,1% compared to CRM's 53,9%).

The average PCS is negatively affected by the EWOC's aforementioned poor performance in scenarios 9 and 10. Excluding the scenarios where the true MTD corresponds to the highest dose level would enhance EWOC's average PCS, bringing it closer to CRM's.

As mentioned before, the simulation results revealed low accuracy for the BLRM. Zhou *et al.* [32] and Yao [79] made the same observation about the low accuracy.

As previously noted, the BLRM is very conservative, stopping a great number of the simulated clinical trials, particularly when DLT outcomes are observed in the first cohorts. However, watching BLRM's performance in scenarios 5 (no trials stopped), 7 (only 3,3% of the trials stopped) and 9 (no trials stopped), we see that the BLRM was the method with the lowest PCS, suggesting limited capability for accurately determining the true MTD in these scenarios too.

In regard to the mTPI, when examining the results of each scenario individually, we observe that this method demonstrated a reasonable performance. It exhibited the second-highest PCS in scenarios 2 and 9, but in most scenarios, it exhibited the third and fourth-highest PCS.

As we have noted earlier, Zhou *et al.* [32] and Zhou *et al.* [54] concluded, in their research, that the mTPI was, in general, comparable, in terms of PCS, to the BOIN. The results of the simulation study here presented reaffirm such findings.

It's important to emphasize nonetheless, that exists literature suggesting that the BOIN outperforms the mTPI for lower target toxicity levels. Yuan *et al.* [34] performed a simulation study to compare the BOIN design, the mTPI and the 3+3 design, considering sixteen different scenarios and four target toxicity levels, ϕ, of 0,15, 0,20, 0,25 and 0,30. Their results showed that the BOIN design outperformed the mTPI (and largely the 3+3), particularly when the target toxicity level was low, such as 0,15 or 0,20, where the BOIN design frequently showed PCS 6 to 10% higher than the mTPI. Similar results regarding mTPI's underperformance for lower target toxicity levels are also observable in [32] (for $\phi = 0{,}20$) and [54] (for $\phi = 0{,}20$ with cohorts of one patient).

Regarding the BOIN, Zhu *et al.* [75] claim, in their 2019 paper, that the BOIN has performance comparable to the CRM. Zhou *et al.*'s [32] and Zhou *et al.*'s [54] numerical studies also showed such evidence. Nevertheless, the simulation study here presented didn't corroborate that statement. The BOIN exhibited PCS comparable to the mTPI and EWOC, and 6,39%, on average, lower than the CRM. Across the ten scenarios, the BOIN only had the highest PCS value in one scenario, scenario number 5. Additionally, it consistently exhibited the third-

highest accuracy in determining the true MTD in five scenarios, specifically, scenarios 4, 6, 7, 8 and 9.

3.5.2 Dose allocation

Concerning dose allocation, Table VII presents the simulation results, outlining the mean number of patients treated at each dose level (including the dose that corresponds to the true MTD, in bold) and the mean number of patients treated below and above the true MTD, (as well as the mean number of DLTs observed), across the 1000 clinical trials simulated. Table VII also presents the average of the above metrics, calculated across the ten scenarios.

The CRM has shown to be the method that maximizes the patients treated at the true MTD, with an average of 13,352 patients, followed by mTPI, 12,411 patients, BOIN, 11,736 patients, EWOC, 11,339 patients and, finally, BLRM, with 8,487 patients.

Given the above results, it's possible to observe that the CRM and mTPI tend to allocate more patients to the dose level that corresponds to the true MTD. Similar observation was made by Zhou *et al.* [32].

Looking at each scenario individually, the CRM treated the highest mean number of patients at the true MTD in six scenarios, 1, 2, 6, 7, 8 and 9; and was the second method treating more patients at the true MTD in scenarios 4 and 10. On the other hand, it was the method treating more patients above the MTD in six out of eight scenarios, scenarios 3 to 8.

Thus, in the conducted simulations, the CRM method exhibited the highest average PCS and highest number of patients treated at the true MTD. However, this positive performance came at the cost of a higher risk of overdosing patients.

Examining the metrics obtained for the EWOC, we can see it had, in general, a conservative approach. The average number of patients treated at the true MTD was the third-lowest (11,339), and that is due to EWOC's disposition to treat patients at dose levels below the true MTD (18,266 against CRM's 14,485). In addition, excluding the scenarios where the true MTD is at the lowest dose level (scenarios 1 and 2), where the EWOC showed an aggressive performance, the EWOC was the method that exposed fewer patients to dose levels above the MTD on five out of six scenarios (scenarios 4 to 8), which reflects this method conservativeness.

The simulation results also showed a curious behavior in the extreme scenarios.

In scenarios 1 and 2, the EWOC was the most aggressive method, treating a mean number of patients above the MTD of 12,210 and 17,298, respectively.

In scenarios 9 and 10, we have seen earlier that the EWOC exhibited low PCSs, which might be justified by overly conservative dose escalation to higher dose levels. We can see that the mean number of patients treated below the MTD was very high, 23,928 (against CRM's 16,446; BLRM's 19,644; mTPI's 19,809; and BOIN's 20,067) in scenario 9 and 28,335 (against CRM's 22,683; BLRM's 20,160; mTPI's 24,888; and BOIN's 24,450) in scenario 10, mainly because this method dosed a very high number of patients in dose level number five (29,56% and 36,87%, respectively), suggesting that the escalation to the highest dose level was difficult to achieve, due to the overdose control criterion.

Fig. 5. also highlights EWOC's behavior mentioned above, of being extremely aggressive when the true MTD is the first dose level, and being extremely conservative when the true MTD is at the last dose level, illustrated by the largest variability in the mean number of patients treated at the true MTD.

Regarding the BLRM, as previously observed, the frequency of stopped clinical trials was notably elevated. Consequently, drawing a meaningful comparison to the other methods becomes challenging. Unsurprisingly, this method showed the lowest average number of patients dosed at the dose corresponding to the true MTD, and the lowest average number of patients dosed above the true MTD. Similar results are also evident in Zhou *et al.*'s [32] paper, where we can also observe similar results for the EWOC in regards with treating patients at the dose that corresponds to the true MTD (second worst method).

The mTPI displayed the second-highest average number of patients treated at the true MTD (12,411) and the second-lowest average number of patients treated above the true MTD (4,222). Concerning the mean number of patients treated above the true MTD, it's important to highlight that the mTPI exhibited always the lowest or second-lowest values, suggesting lower tendency to treat patients above the MTD, when compared to the other methods.

The last method tested, the BOIN, also showed a conservative approach compared to the CRM. The simulation results revealed that, across the ten scenarios, the BOIN dosed fewer patients, on average, with the dose corresponding to the true MTD (11,736) than the CRM (13,352) or the mTPI (12,411), yet higher than the BLRM (8,487) and the EWOC (11,339). Furthermore, it also dosed, on average, fewer patients above the true MTD, with results only higher than those of the mTPI and BLRM (4,637 patients compared to 4,222 and 3,574 patients, respectively).

Yuan *et al.* [34] and Zhou *et al.* [54], showed through their simulation studies that the BOIN design outperforms the mTPI in assigning patients to the dose level that corresponds to the true MTD, particularly when the target toxicity level was low (0,15 or 0,20), and that both methods were comparable when the targeted toxicity level is higher (0,25 or 0,30). The simulation study performed didn't corroborate Yuan *et al.*'s [34] and Zhou *et al.*'s [54] results; the

mTPI was slightly better, on average, treating patients at the true MTD, as evidenced when observing the results for each scenario individually. The mTPI outperformed the BOIN in terms of mean number of patients treated at the dose level that corresponds to the true MTD on eight of the ten scenarios.

TABLE VII

Mean numbers of patients treated at each dose-level, below the MTD, above the MTD, and mean number of DLTs in the 1000 clinical trials simulated

Dose levels (mg/m²)		150	200	250	300	350	400	Mean number of patients treated below the true MTD	Mean number of patients treated above the true MTD	Means number of DLTs
		$\bar{n}_{d=150}$	$\bar{n}_{d=200}$	$\bar{n}_{d=250}$	$\bar{n}_{d=300}$	$\bar{n}_{d=350}$	$\bar{n}_{d=400}$	$\bar{n}_{d<MTD}$	$\bar{n}_{d>MTD}$	$\overline{DLT}$
Scenario 1	CRM	**19,806 (66,02%)**	8,688 (28,96%)	1,422 (4,74%)	0,084 (0,28%)	-	-	N/A	10,194 (33,98%)	10,709 (35,70%)
	EWOC	**17,790 (59,30%)**	11,016 (36,72%)	1,173 (3,91%)	0,021 (0,07%)	-	-	N/A	12,21 (40,70%)	10,923 (36,41%)
	BLRM	**5,901 (61,49%)**	2,661 (27,73%)	0,942 (9,82%)	0,090 (0,94%)	0,003 (0,03%)	-	N/A	3,696 (38,51%)	3,595 (37,46%)
	mTPI	**18,810 (67,77%)**	8,184 (29,49%)	0,744 (2,68%)	0,018 (0,06%)	-	-	N/A	8,946 (32,23%	9,706 (34,97%)
	BOIN	**18,621 (67,31%)**	8,046 (29,08%)	0,969 (3,50%)	0,030 (0,11%)	-	-	N/A	9,045 (32,69%)	9,723 (35,14%)
Scenario 2	CRM	**17,511 (58,37%)**	7,506 (25,02%)	3,675 (12,25%)	1,101 (3,67%)	0,180 (0,60%)	0,027 (0,09%)	N/A	12.489 (41,63%)	10,280 (34,27%)
	EWOC	**12,702 (42,34%)**	11,961 (39,87%)	4,551 (15,17%)	0,750 (2,50%)	0,036 (0,12%)	-	N/A	17,298 (57,66%)	10,401 (34,67%)
	BLRM	**4,524 (44,75%)**	2,688 (26,59%)	1,998 (19,76%)	0,654 (6,47%)	0,201 (1,99%)	0,045 (0,45%)	N/A	5,586 (55,25%)	3,657 (36,17%)
	mTPI	**16,299 (61,24%)**	7,059 (26,52%)	2,541 (9,55%)	0,618 (2,32%)	0,084 (0,32%	0,012 (0,05%)	N/A	10,314 (38,76%)	9,013 (33,87%)
	BOIN	**15,453 (58,07%)**	7,230 (27,17%)	2,964 (11,14%)	0,825 (3,10%)	0,129 (0,48%)	0,012 (0,05%)	N/A	11,16 (41,93%)	9,087 (34,14%)
Scenario 3	CRM	7,245 (24,15%)	**18,375 (61,25%)**	4,308 (14,36%)	0,072 (0,24%)	-	-	7,245 (24,15%)	4,380 (0,146%)	10,140 (33,8%)
	EWOC	6,819 (22,73%)	**20,583 (68,61%)**	2,592 (8,64%)	0,006 (0,02%)	-	-	6,819 (22,73%)	2,598 (8.66%)	9,244 (30,81%)
	BLRM	7,410 (35,44%)	**10,620 (50,80%)**	2,793 (13,36%)	0,084 (0,40%)	-	-	7,410 (35,44%)	2,877 (13,76%)	6,305 (30,16%)
	mTPI	7,251 (24,17%)	**20,910 (69,70%)**	1,827 (6,09%)	0,012 (0,04%)	-	-	7,251 (24,17%)	1,839 (6,13%)	8,797 (29,32%)
	BOIN	9,393 (31,31%)	**18,123 (60,41%)**	2,469 (8,23%)	0,015 (0,05%)	-	-	9,393 (31,31%)	2,484 (8,28%)	8,502 (28,34%)

TABLE VII (continued)
Mean numbers of patients treated at each dose-level, below the MTD, above the MTD, and mean number of DLTs in the 1000 trials simulated

	Dose levels (mg/m²)	150	200	250	300	350	400	Mean number of patients treated below the true MTD	Mean number of patients treated above the true MTD	Mean number of DLTs
		$\bar{n}_{d=150}$	$\bar{n}_{d=200}$	$\bar{n}_{d=250}$	$\bar{n}_{d=300}$	$\bar{n}_{d=350}$	$\bar{n}_{d=400}$	$\bar{n}_{d<MTD}$	$\bar{n}_{d>MTD}$	$\overline{DLT}$
Scenario 4	CRM	3,855 (12,85%)	6,465 (21,55%)	**13,923 (46,41%)**	5,292 (17,64)	0,453 (1,51%)	0,012 (0,04%)	10,320 (34,4%)	5,757 (19,19%)	9.249 (30,83%)
	EWOC	3,219 (10,73%)	9,288 (30,96%)	**14,280 (47,60%)**	3,159 (10,53%)	0,054 (0,18%)	-	12,507 (41,69%)	3,213 (10,71%)	8,059 (26,86%)
	BLRM	3,558 (14,39%)	7,317 (29,59%)	**10,203 (41,25%)**	3,297 (13,33%)	0,345 (1,39%)	0,012 (0,05%)	10,875 (43,97%)	3,654 (14,77%)	6,860 (27,74%)
	mTPI	4,416 (14,72%)	9,231 (30,77%)	**12,969 (43,23%)**	3,291 (10,97%)	0,093 (0,31%)	-	13,647 (45,49%)	3,384 (11,28%)	7,954 (26,51%)
	BOIN	4,470 (14,90%)	9,765 (32,55%)	**12,078 (40,26%)**	3,546 (11,82%)	0,138 (0,46%)	0,003 (0,01%)	14,235 (47,45%)	3,687 (12,29%)	7,949 (26,50%)
Scenario 5	CRM	3.000 (10,00%)	3,006 (10,02%)	4,653 (15,51%)	**14,460 (48,20%)**	4,779 (15,93%)	0,102 (0,34%)	10,659 (35,53%)	4,881 (16,27%)	8,950 (29,83%)
	EWOC	3,000 (10,00%)	3,150 (10,50%)	7,284 (24,28%)	**14,568 (48,56%)**	1,998 (6,66%)	-	13,434 (44,78%)	1,998 (6,66%)	6,845 (22,82%)
	BLRM	3,003 (10,01%)	3,132 (10,44%)	8,832 (29,44%)	**12,021 (40,07%)**	2,892 (9,64%)	0,120 (0,40%)	14,967 (49,89%)	3,012 (10,04%)	7,008 (23,36%)
	mTPI	3,003 (10,01%)	3,183 (10,61%)	6,435 (21,45%)	**15,162 (50,54%)**	2,190 (7,30%)	0,027 (0,09%)	12,621 (42,07%)	2,217 (7,39%)	7,244 (24,15%)
	BOIN	3,003 (10,01%)	3,195 (10,65%)	7,731 (25,77%)	**13,434 (44,78%)**	2,604 (8,68%)	0,033 (0,11%)	13,929 (46,43%)	2,637 (8,79%)	7,059 (23,53%)
Scenario 6	CRM	3.216 (10,72)	3,195 (10,65)	6,027 (20,09%)	**12,210 (40,70%)**	4,932 (16,44%)	0,420 (1,40%)	12,438 (41,46%)	5,352 (17,84%)	8,448 (28,16%)
	EWOC	3,006 (10,02%)	4,056 (13,52%)	9,387 (31,29%)	**11,433 (38,11%)**	2,118 (7,06%)	-	16,449 (54,83%)	2,118 (7,06%)	6,674 (22,25%)
	BLRM	3,066 (10,83%)	3,330 (11,76%)	8,142 (28,77%)	**10,119 (35,75%)**	3,231 (11,41%)	0,417 (1,47%)	14,538 (51,36%)	3,648 (12,89%)	7,039 (24,87%)
	mTPI	3,300 (11,00%)	4,251 (14,17%)	8,547 (28,49%)	**11,028 (36,76%)**	2,757 (9,19%)	0,117 (0,39%)	16,098 (53,66%)	2,874 (9,58%)	7,054 (23,51%)
	BOIN	3,282 (10,94%)	4,371 (14,57%)	8,724 (29,08%)	**10,347 (34,49%)**	3,126 (10,42%)	0,150 (0,50%)	16,377 (54,59%	3,276 (10,92%)	7,071 (23,57%)

TABLE VII (continued)
Mean number of patients treated at each dose-level, below the MTD and above the MTD, and mean number of DLTs, in the 1000 trials simulated

	Dose levels (mg/m²)	150	200	250	300	350	400	Mean number of patients treated below the true MTD	Mean number of patients treated above the true MTD	Mean number of DLTs
		$\bar{n}_{d=150}$	$\bar{n}_{d=200}$	$\bar{n}_{d=250}$	$\bar{n}_{d=300}$	$\bar{n}_{d=350}$	$\bar{n}_{d=400}$	$\bar{n}_{d<MTD}$	$\bar{n}_{d>MTD}$	$\overline{DLT}$
Scenario 7	CRM	3,093 10,31%)	3,015 (10,05%)	3,273 (10,91%)	6,171 (20,57%)	**9,849 (32,83%)**	4,599 (15,33%)	15,552 (51,84%)	4,599 (15,33%)	7,290 (24,30%)
	EWOC	3,003 (10,01%)	3,276 (10,92%)	4,434 (14,78%)	9,417 (31,39%)	**9,165 (30,55%)**	0,705 (2,35%)	20,130 (67,10%)	0,705 (2,35%)	5,265 (17,55%)
	BLRM	3,006 (10,33%)	2,952 (10,14%)	3,495 (12,01%)	7,950 (27,31%)	**8,424 (28,94%)**	3,282 (11,27%)	17,403 (59,79%)	3,282 (11,27%)	6,240 (21,44%)
	mTPI	3,102 (10,34%)	3,402 (11,34%)	4,395 (14,65%)	7,968 (26,56%)	**8,688 (28,96%)**	2,445 (8,15%)	18,867 (62,89%)	2,445 (8,15%)	6,018 (20,06%)
	BOIN	3,099 (10,33%)	3,390 (11,30%)	4,425 (14,75%)	7,920 (26,40%)	**8,418 (28,06%)**	2,748 (9,16%)	18,834 (62,78%)	2,748 (9,16%)	6,054 (20,18%)
Scenario 8	CRM	3,708 (12,36%)	3,528 (11,76%)	5,223 (17,41%)	8,076 (26,92%)	**6,516 (21,72%)**	2,949 (9,83%)	20,535 (68,45%)	2,949 (9,83)	6,655 (22,18)
	EWOC	3,099 (10,33%)	4,401 (14,67%)	7,161 (23,87%)	9,861 (32,87%)	**5,136 (17,12%)**	0,342 (1,14%)	24,522 (81,74%)	0,342 (1,14%)	5,701 (19,00%)
	BLRM	3,081 (12,37%)	2,754 (11,06%)	4,389 (17,62%)	6,525 (26,20%)	**5,232 (21,01%)**	2,922 (11,73%)	16,749 (67,26%)	2,922 (11,73%)	5,604 (22,50%)
	mTPI	4,053 (13,51%)	4,833 (16,11%)	6,597 (21,99%)	7,812 (26,04%)	**4,944 (16,48%)**	1,761 (5,87%)	23,295 (77,65%)	1,761 (5,87%)	5,979 (19,93%)
	BOIN	3,918 (13,06%)	4,713 (15,71%)	6,414 (21,38%)	7,491 (24,97%)	**5,403 (18,01%)**	2,061 (6,87%)	22,536 (75,12%)	2,061 (6,87%)	6,143 (20,48%)
Scenario 9	CRM	3,000 (10,00%)	3,000 (10,00%)	3,000 (10,00%)	3,051 (10,17%)	4,395 (14,65%)	**13,554 (45,18%)**	16,446 (54,82%)	N/A	5,185 (17,28%)
	EWOC	3,000 (10,00%)	3,003 (10,01%)	3,114 (10,38%)	3,750 (12,50%)	11,061 (36,87%)	**6,072 (20,24%)**	23,928 (79,76%)	N/A	3,624 (12,08%)
	BLRM	3,000 (10,00%)	3,000 (10,00%)	3,003 (10,01%)	3,333 (11,11%)	7,308 (24,36%)	**10,356 (34,52%)**	19,644 (65,48%)	N/A	4,485 (14,95%)
	mTPI	3,000 (10,00%)	3,000 (10,00%)	3,096 (10,32%)	3,777 (12,59%)	6,936 (23,12%)	**10,191 (33,97%)**	19,809 (66,03%)	N/A	4,541 (15,14%)
	BOIN	3,000 (10,00%)	3,000 (10,00%)	3,102 (10,34%)	3,822 (12,74%)	7,143 (23,81%)	**9,933 (33,11%)**	20,067 (66,89%)	N/A	4,473 (14,91%)

TABLE VII (continued)
Mean number of patients treated at each dose-level, below the MTD and above the MTD, and mean number of DLTs, in the 1000 trials simulated

	Dose levels (mg/m²)	150	200	250	300	350	400	Mean number of patients treated below the true MTD	Mean number of patients treated above the true MTD	Mean number of DLTs
		$\bar{n}_{d=150}$	$\bar{n}_{d=200}$	$\bar{n}_{d=250}$	$\bar{n}_{d=300}$	$\bar{n}_{d=350}$	$\bar{n}_{d=400}$	$\bar{n}_{d<MTD}$	$\bar{n}_{d>MTD}$	$\overline{DLT}$
Scenario 10	CRM	3,342 (11,14%)	3,102 (10,34%)	3,687 (12,29%)	5,472 (18,24%)	7,080 (23,60%)	**7,317 (24,39%)**	22,683 (75,61%)	N/A	5,488 (18,29%)
	EWOC	3,018 (10,06%)	3,699 (12,33%)	4,809 (16,03%)	7,941 (26,47%)	8,868 (29,56%)	**1,665 (5,55%)**	28,335 (94,45%)	N/A	4,500 (15,00%)
	BLRM	3,045 (11,06%)	2,823 (10,25%)	3,357 (12,19%)	4,824 (17,51%)	6,111 (22,19%)	**7,383 (26,81%)**	20,16 (73,19%)	N/A	5,072 (18,41%)
	mTPI	3,408 (11,36%)	3,945 (13,15%)	4,743 (15,81%)	6,435 (21,45%)	6,357 (21,19%)	**5,112 (17,04%)**	24,888 (82,96%)	N/A	4,970 (16,57%)
	BOIN	3,378 (11,26%)	3,828 (12,76%)	4,593 (15,31%)	6,276 (20,92)	6,375 (21,25%)	**5,550 (18,50%)**	24,450 (81,50%)	N/A	5,071 (16,90%)

		Average number of patients treated at the true MTD	Average number of patients treated below the true MTD	Average number of patients treated above the true MTD	Average number of DLTs
Average across the ten scenarios	CRM	13,352	14,485	6,325	8,239
	EWOC	11,339	18,266	5,060	7,124
	BLRM	8,478	15,218	3,585	5,586
	mTPI	12,411	17,060	4,222	7,128
	BOIN	11,736	17,478	4,637	7,113

Mean number of patients treated at the dose levels corresponding to the MTD in bold

3.5.3 Toxicity

Regarding toxicity, the number of DLTs is expected to be larger for more aggressive dose-escalation methods. We observed in Table VII that methods that exposed more patients to higher dose levels, such as CRM, yielded a higher number of DLTs.

The CRM exhibited the highest average number of DLTs in the simulation study, 8,239, across the ten scenarios. It displayed the highest mean number of DLTs in eight out of ten scenarios, scenarios 3 to 10.

The EWOC, mTPI, and BOIN didn't show significant differences in terms of the number of DLTs observed. The average numbers of DLTs observed across the ten scenarios were 7,124, 7,128, and 7,113, respectively. This suggests that there are no relevant differences between the three methods in terms of aggressiveness in dose escalation, as more aggressive methods tend to reflect a higher number of DLTs.

Regarding the EWOC, it is important to note that its average number of DLTs was highly influenced by the aggressive performance in scenarios 1 and 2. Excluding these two scenarios, the average number of DLTs would have been only higher than of the BLRM, which was the most conservative method.

4

SIMULATION STUDY - REPLICATION OF A REAL CLINICAL TRIAL

4.1 Deflexifol™ Clinical Trial

In 2019, the results of a clinical trial comparing infusional and bolus schedules of Deflexifol™, a new formulation of 5-fluorouracil (5-FU) with leucovorin, were published [87]. In this study, the dose-escalation method utilized was the 3+3, with five dose levels.

Table VIII summarizes the DLT outcomes observed for the bolus schedule during the clinical trial, which we used to evaluate the dose allocation of the model-based and model-assisted designs. The criteria utilized for assessing the DLTs are described in [87]. The MTD was declared the next lower dose level to the one where two or more DLTs were observed out of six patients [87].

We will focus on the results presented below to assess how the CRM, EWOC, BLRM, mTPI and BOIN would have performed.

TABLE VIII
Deflexifol™ clinical trial dose escalation summary for the bolus schedule

Dose level	Deflexifol™ dose (mg/m² 5-FU equivalents)	Number of patients treated (n_j)	Number of DLTs observed (y_j)
1	375	3	0
2	425	3	0
3	475	3	0
4	525	6	0
5	575	4	2
	Total	19	2

Adapted from [87]

In the published article [87], two DLTs were reported at dose level number five (575 mg/m²), where four patients were treated. As a result, the MTD declared was 525 mg/m².

The 3+3 design does not require the definition of a target toxicity level, therefore, a target toxicity level of $\phi = 0{,}25$ will be considered, given that it was established that the target toxicity level of such design is around 0,25 [88]. To facilitate comparison between methods, the lowest dose level was considered as starting dose and dose escalation was restricted one level at the time.

4.1.1 Dose-escalation methods configuration

4.1.1.1 Continual Reassessment Method

The same configuration described in 3.3.1 was utilized.

Lee and Cheung's [78] algorithm was used to derive the *skeleton*. To run this algorithm, the third dose level was assumed as being the prior MTD, $\nu = 3$, and the half-width indifference interval $\delta = 0{,}06$ was utilized, since the authors shown that it is the *optimal* half-width indifference interval for 5 dose levels, $\phi = 0{,}25$, $\nu = 3$ and $n = 30$.

The dose labels derived for the dose levels used in this clinical trial, were $J^* = \{0{,}062;\ 0{,}140;\ 0{,}250;\ 0{,}376;\ 0{,}502\}$.

4.1.1.2 Escalation With Overdose Control

The same *feasibility bound* and prior distributions described in 3.3.2, were used in the reanalysis of this clinical trial. $Xmin$ and $Xmax$ were defined as $Xmin = 325$ and $Xmax = 625$.

4.1.1.3 Bayesian Logistic Regression Method

The same prior distribution described in 3.3.3 was utilized. $\delta_1 = \delta_2 = 0{,}05$ was used to define the target interval, $[0{,}20; 0{,}30]$, considering the target toxicity level $\phi = 0{,}25$.

The highest dose level was defined as being the reference dose-level, 575 mg/m², and the overdose control was defined considering $\alpha = 0{,}25$.

4.1.1.4 Modified Toxicity Probability Interval

The standard prior distribution for p_{cur} was considered, p_{cur}~Beta(1,1). The EI was defined to match the target toxicity interval used in BLRM. EI $= [0{,}20; 0{,}30]$.

4.1.1.5 Bayesian Optimal Interval

The default vague prior for p_{cur}, p_{cur}~Beta(1,1), was used in the BOIN's safety rule. The default dose-escalation boundaries for $\phi = 0{,}25$ were also utilized, $\lambda_e = 0{,}197$and $\lambda_d = 0{,}298$.

4.1.2 Software

Statistical analysis was performed with software R version 4.2.2 [80], R-studio version 2023.03.1 [81], and WinBUGS 14 [89].

The *skeleton* for the CRM was obtained using the function getprior() on the "dfcrm" R package, version 0.2-2.1 [83].

The Excel™ macro tool spreadsheet obtained from [85] and the R package "BOIN" version 2.7.2 [76] were used to get the pre-tabulated dose-escalation actions presented in Appendix A.

4.1.3 Results

4.1.3.1 Continual Reassessment Method

The CRM chooses the dose for the next patient or cohort of patients based on the updated posterior probability of DLT, by selecting the dose with posterior probability of DLT closer to

the target toxicity probability ϕ, $x_{k+1} = arg\ \underset{d_j \in J}{min}\left(f\left(d_j, \hat{\theta}_k\right) - \phi\right)^2$. For the Deflexifol™ clinical trial, we assumed $\phi = 0{,}25$.

The results of the re-analysis of the Deflexifol™ clinical trial using the CRM are presented in Table IX.

TABLE IX
CRM statistics summary for the Deflexifol™ clinical trial (bolus schedule)

Cohort	Dose level (mg/m²)	n_j	y_j	Estimated mean posterior probability of DLT $\hat{p}_j(\theta \mid D_k)$ [95% credible intervals] (dose allocation decision)	Gelman-Rubin statistic for the parameters estimated
1	1 (375)	3	0	$\hat{p}_{375}(\theta \mid D_k) = 0{,}060\ [0{,}000; 0{,}397]$ $\hat{p}_{425}(\theta \mid D_k) = 0{,}101\ [0{,}000; 0{,}521]$ $\hat{p}_{475}(\theta \mid D_k) = 0{,}159\ [0{,}000; 0{,}631]$ $\hat{p}_{525}(\theta \mid D_k) = 0{,}230\ [0{,}000; 0{,}723]$ $\hat{p}_{575}(\theta \mid D_k) = 0{,}313\ [0{,}000; 0{,}795\ ]$ $(x_{k+1} = 425)$	1,001-1,001
2	2 (425)	3	0	$\hat{p}_{375}(\theta \mid D_k) = 0{,}025\ [0{,}000; 0{,}198\ ]$ $\hat{p}_{425}(\theta \mid D_k) = 0{,}051\ [0{,}000; 0{,}319\ ]$ $\hat{p}_{475}(\theta \mid D_k) = 0{,}094\ [0{,}000; 0{,}446\ ]$ $\hat{p}_{525}(\theta \mid D_k) = 0{,}154\ [0{,}000; 0{,}566\ ]$ $\hat{p}_{575}(\theta \mid D_k) = 0{,}231\ [0{,}000; 0{,}670\ ]$ $(x_{k+1} = 475)$	1,001-1,001
3	3 (475)	3	0	$\hat{p}_{375}(\theta \mid D_k) = 0{,}010\ [0{,}000; 0{,}091]$ $\hat{p}_{425}(\theta \mid D_k) = 0{,}026\ [0{,}000; 0{,}184\ ]$ $\hat{p}_{475}(\theta \mid D_k) = 0{,}055\ [0{,}000; 0{,}303\ ]$ $\hat{p}_{525}(\theta \mid D_k) = 0{,}103\ [0{,}000; 0{,}430]$ $\hat{p}_{575}(\theta \mid D_k) = 0{,}170\ [0{,}000; 0{,}552\ ]$ $(x_{k+1} = 525)$	1,001-1,001
4	4 (525)	6	0	$\hat{p}_{375}(\theta \mid D_k) = 0{,}002\ [0{,}000; 0{,}016]$ $\hat{p}_{425}(\theta \mid D_k) = 0{,}006\ [0{,}000; 0{,}053]$ $\hat{p}_{475}(\theta \mid D_k) = 0{,}018\ [0{,}000; 0{,}127]$ $\hat{p}_{525}(\theta \mid D_k) = 0{,}042\ [0{,}000; 0{,}233]$ $\hat{p}_{575}(\theta \mid D_k) = 0{,}086\ [0{,}000; 0{,}358]$ $(x_{k+1} = 575)$	1,001-1,001
5	5 (575)	4	2	$\hat{p}_{375}(\theta \mid D_k) = 0{,}008\ [0{,}000; 0{,}048]$ $\hat{p}_{425}(\theta \mid D_k) = 0{,}026\ [0{,}000; 0{,}116]$ $\hat{p}_{475}(\theta \mid D_k) = 0{,}067\ [0{,}000; 0{,}219]$ $\hat{p}_{525}(\theta \mid D_k) = 0{,}138\ [0{,}018; 0{,}343]$ $\hat{p}_{575}(\theta \mid D_k) = 0{,}237\ [0{,}059; 0{,}470]$ $(x_{k+1} = 575)$	1,001-1,001

We started by assigning the lowest dose to the first cohort of patients, 375 mg/m², which manifested no DLTs. After updating the power model implemented in the CRM, we observed that the fourth dose level was the dose recommended by the CRM, since it is the dose with posterior probability of DLT closer to the target toxicity level.

It's important to note the aggressiveness of this method. After treating a single cohort of three patients it is already recommending escalation to the fourth dose level. As we restricted dose escalation to one level at a time, we will escalate the dose to the next higher dose level, 425 mg/m².

For the second cohort, treated with 425 mg/m², we had three patients dosed without DLTs. After obtaining the posterior probabilities of DLT for each dose level, we observed that the dose with posterior probability of DLT closer to the target probability was the fifth dose level, 575 mg/m². We would then escalate one dose level.

The same happened for the following two cohorts, cohorts three and four, and therefore, the dose would be escalated to the next higher dose level, dose levels four and five, respectively.

Four patients were treated in cohort number five, and two manifested DLTs. After updating the posterior probabilities of DLT for each dose, we observed that it was the fifth dose level that had posterior probability of DLT closer to the target toxicity level, which means that the dose would remain at the same dose level, instead of de-escalating to the next lower dose.

In the Deflexifol™ clinical trial, the MTD was declared, as the 3+3 method assumes that the MTD has been exceeded when two or more DLTs are observed out of six patients. Despite only four patients being treated at this dose level, two DLTs were observed, and the clinical trial was stopped.

Other methods would not stop the trial but would recommend dose de-escalation. The CRM, on the other hand, recommends staying at the same dose level, which reflects the aggressiveness of this method regarding dose allocation. Aggressiveness that we had previously noted when analyzing the simulation study results with the ten true dose-toxicity scenarios.

With the CRM, the clinical trial could have continued, for example, until the maximum sample size was exhausted.

The Gelman-Rubin convergence diagnostic was assessed for the CRM, EWOC and BLRM, and was provided in Tables IX, X and XI. All values obtained suggest that the chains have converged to the target distribution.

4.1.3.2 Escalation With Overdose Control

The EWOC selects the dose for the next patient or cohort of patients based on the c.d.f. of the marginal posterior distribution of the parameter γ, by selecting the dose corresponding to the *feasibility bound* α. Since we are assuming five discrete dose levels, the dose for the next cohort of patients would be chosen by determining from the marginal posterior c.d.f. of the parameter γ, the dose that has marginal posterior probability closer to the percentile α ($\alpha = 0{,}25$), through the quadratic loss in equation (33), $x_{k+1} = \arg\min_{d_j \in J}(G_k(d_j) - \alpha)^2\,,\ k = 2, \dots n$.

Hence, we start with the first three patients, treated at the first dose level, where no DLTs were observed in the clinical trial. After obtaining the c.d.f. of the posterior distribution of γ, we found that the α^{th} percentile is equal to 433 mg/m^2. This is the dose that would have been given to the next cohort of patients, if we were assuming that the explanatory variable dose was a continuous variable and that any dose between $Xmax$ and $Xmin$ could be administered to the clinical trial participants.

As this isn't the case, we need to choose from one of the possible five dose levels, $J = \{375, 425, 475, 525, 575\}$. To do so, we determined which dose had marginal posterior cumulative probability closer to the percentile α, which was dose level two, 425 mg/m^2. Results are available in Table X. The dose would then be escalated to the second dose level.

As observed in Table VIII, three additional patients were treated at the second dose level, and no DLTs were observed. Updating the c.d.f. of the marginal posterior distribution of γ, and calculating the respective probabilities for each one of the dose levels, we determined that dose level number three, 475 mg/m^2, was the dose level with marginal posterior cumulative probability closer to α. The dose would be then escalated to the third dose level.

After treating the third cohort, composed of three patients, at the third dose level, 475 mg/m^2, and without DLTs, the marginal posterior distribution for γ was obtained. The dose that minimized the difference to α was the third dose level (475 mg/m^2), which means that following the EWOC, the dose is not escalated to the next higher dose level. The differences between the Deflexifol™ clinical trial (using the 3+3 method) and the replication of this clinical trial using the EWOC start here. In the Deflexifol™ clinical trial, the dose was escalated to the fourth dose level (525 mg/m^2), and with the EWOC, we wouldn't escalate the dose.

We will assume that three additional patients were treated with 475 mg/m^2 and no DLTs were observed (the same outcomes as the previous cohort). After updating the marginal posterior cumulative distribution of γ, we verified that the dose that had marginal posterior cumulative probability closer to α was the dose level number four, 525 mg/m^2, which means that the dose can be escalated this time. Cohort 5 would be treated with a dose of 525 mg/m^2.

In the Deflexifol™ clinical trial, six patients were treated with 525 mg/m^2 (dose level number four) and no DLTs were reported in the article. We will assume that these six patients were treated in two cohorts of three patients (cohorts five and six) to assess the capacity of the EWOC method to escalate the dose.

The updated marginal posterior cumulative distribution of γ showed that, both for cohorts five and six, the dose should not be escalated to the next higher dose level. We will then assume that an additional cohort of three patients (cohort number seven) was treated at dose level four, 525 mg/m^2, and no DLTs were observed (the same outcomes as the previous two cohorts).

TABLE X
EWOC statistics summary for the Deflexifol™ clinical trial (bolus schedule)

Cohort	Dose level (mg/m²)	n_j	y_j	γ_α	$P(\gamma \leq d_j \mid D_k)$ (dose allocation decision)	Gelman-Rubin statistic for the parameters estimated
1	1 (375)	3	0	$\gamma_\alpha = 433{,}00$	$P(\gamma \leq 375 \mid D_k) = 0{,}063$ $P(\gamma \leq 425 \mid D_k) = 0{,}223$ $P(\gamma \leq 475 \mid D_k) = 0{,}408$ $P(\gamma \leq 525 \mid D_k) = 0{,}601$ $P(\gamma \leq 575 \mid D_k) = 0{,}797$ $(x_{k+1} = 425)$	1,001-1,001
2	2 (425)	3	0	$\gamma_\alpha = 463{,}00$	$P(\gamma \leq 375 \mid D_k) = 0{,}020$ $P(\gamma \leq 425 \mid D_k) = 0{,}114$ $P(\gamma \leq 475 \mid D_k) = 0{,}298$ $P(\gamma \leq 525 \mid D_k) = 0{,}513$ $P(\gamma \leq 575 \mid D_k) = 0{,}753$ $(x_{k+1} = 475)$	1,001-1,001
3	3 (475)	3	0	$\gamma_\alpha = 494{,}28$	$P(\gamma \leq 375 \mid D_k) = 0{,}009$ $P(\gamma \leq 425 \mid D_k) = 0{,}051$ $P(\gamma \leq 475 \mid D_k) = 0{,}175$ $P(\gamma \leq 525 \mid D_k) = 0{,}397$ $P(\gamma \leq 575 \mid D_k) = 0{,}681$ $(x_{k+1} = 475)$	1,001-1,001
4	3 (475)	3	0	$\gamma_\alpha = 511{,}50$	$P(\gamma \leq 375 \mid D_k) = 0{,}005$ $P(\gamma \leq 425 \mid D_k) = 0{,}028$ $P(\gamma \leq 475 \mid D_k) = 0{,}111$ $P(\gamma \leq 525 \mid D_k) = 0{,}316$ $P(\gamma \leq 575 \mid D_k) = 0{,}622$ $(x_{k+1} = 525)$	1,001-1,001
5	4 (525)	3	0	$\gamma_\alpha = 534{,}40$	$P(\gamma \leq 375 \mid D_k) = 0{,}002$ $P(\gamma \leq 425 \mid D_k) = 0{,}014$ $P(\gamma \leq 475 \mid D_k) = 0{,}057$ $P(\gamma \leq 525 \mid D_k) = 0{,}203$ $P(\gamma \leq 575 \mid D_k) = 0{,}526$ $(x_{k+1} = 525)$	1,001-1,001
6	4 (525)	3	0	$\gamma_\alpha = 547{,}30$	$P(\gamma \leq 375 \mid D_k) = 0{,}002$ $P(\gamma \leq 425 \mid D_k) = 0{,}009$ $P(\gamma \leq 475 \mid D_k) = 0{,}033$ $P(\gamma \leq 525 \mid D_k) = 0{,}141$ $P(\gamma \leq 575 \mid D_k) = 0{,}451$ $(x_{k+1} = 525)$	1,001-1,002
7	4 (525)	3	0	$\gamma_\alpha = 558{,}30$	$P(\gamma \leq 375 \mid D_k) = 0{,}001$ $P(\gamma \leq 425 \mid D_k) = 0{,}006$ $P(\gamma \leq 475 \mid D_k) = 0{,}023$ $P(\gamma \leq 525 \mid D_k) = 0{,}093$ $P(\gamma \leq 575 \mid D_k) = 0{,}379$ $(x_{k+1} = 575)$	1,001-1,001
8	5 (575)	4	2	$\gamma_\alpha = 544{,}20$	$P(\gamma \leq 375 \mid D_k) = 0{,}001$ $P(\gamma \leq 425 \mid D_k) = 0{,}008$ $P(\gamma \leq 475 \mid D_k) = 0{,}030$ $P(\gamma \leq 525 \mid D_k) = 0{,}139$ $P(\gamma \leq 575 \mid D_k) = 0{,}511$ $(x_{k+1} = 525)$	1,001-1,001

It's important to note that with the 3+3 design the dose would have been escalated, as in the Deflexifol™ clinical trial.

For cohort number seven, the dose for which the marginal posterior cumulative probability is closer to the *feasibility bound* was dose level five, 575 mg/m². This means the dose can be escalated, despite nine patients being treated and no DLTs observed.

For dose level number five, we will assume the outcomes of the Deflexifol™ clinical trial, a cohort of four patients treated (cohort number eight) with two DLTs observed.

For cohort number eight, the updated c.d.f. of the marginal posterior distribution of the parameter showed that the dose with marginal posterior probability closer to the percentile α, was dose level four, 525 mg/m², which means that the dose should be de-escalated to the next lower dose level.

The Deflexifol™ clinical trial using 3+3 ends here, since excessive toxicity was observed, according to the 3+3 method criteria. With the EWOC, the clinical trial could have continued until a maximum number of patients was treated or the MTD was estimated with a certain precision.

It's important to note that it took the EWOC seven cohorts of three patients (21 patients) to reach the highest dose level, while the 3+3 method only required five cohorts (15 patients) in the Deflexifol™ clinical trial. This performance highlights the conservativeness of the EWOC method.

As a workaround, the *feasibility bound* α could have been increased as the trial progressed, increasing α in z units (for example, $z = 0{,}05$) towards 0,50 (median) for each cohort treated with no DLTs observed [65].

4.1.3.3 Bayesian Logistic Regression Method

In BLRM, the dose allocation is determined by the candidate dose level that maximizes the posterior probability of belonging to the target interval, defined as $[0{,}20; 0{,}30]$. Yet, dose escalation is also restricted by an overdose control criterion, $P\left(\hat{p}_j > 0{,}30 \middle| D_k\right) < 0{,}25$.

Thus, after treating the first cohort of patients at the lower dose level and estimating the updated dose-toxicity curve, the dose chosen for the next cohort would be, among the doses that fulfill the overdose control criterion, the dose that maximizes the posterior probability that the toxicity falls within the range of the target interval.

As we didn't allow dose skipping, we will escalate one level when the dose that maximizes the posterior probability of the target interval and fulfils the overdose control criterion is above the current dose level. We will de-escalate when the dose that maximizes the target

interval's posterior probability and satisfies the overdose control criterion is below the current dose level.

In the Deflexifol™ clinical trial, as we have seen previously, the first cohort of three patients was treated at the lowest dose, 375 mg/m^2. No DLTs were observed.

After updating the BLRM probability model, we observed that dose levels one to four satisfied the overdose control criterion. Among those four dose levels, dose level number four was the dose that maximized the posterior probability for the target interval. If we were not restricting the dose escalation one level at a time, dose level four, 525 mg/m^2, would be the dose administered to the second cohort of patients. As we restricted dose escalation, we will assume that the next cohort was treated with a 425 mg/m^2 dose.

Before continuing the reanalysis of this clinical trial, just a note to highlight that the BLRM is suggesting a considerable dose jump, and therefore an aggressive approach to dose escalation. As on the current setting we are not allowing dose skipping, that would not be a problem, yet this is a remark that needs to be emphasized.

For the second cohort, we have seen in Table VIII that three patients were treated at 425 mg/m^2 and no DLTs were observed. Analyzing the data in Table XI, we can see that dose levels one to five satisfied the overdose control. Among all doses, it's dose level five, 575 mg/m^2, that maximized the posterior probability of the target interval.

Thus, after treating just six patients, with no DLTs observed, the BLRM is already recommending escalation to the highest dose level, the same dose-escalation decisions provided by the CRM. We emphasize once again the aggressiveness of this method. This may be the reason why the authors of the BLRM do not recommend dose skipping [27]. Since we are restricting dose escalation, the dose would be escalated just one level, to 475 mg/m^2.

Cohort number three received 475 mg/m^2 without manifesting DLTs. After updating the dose-toxicity curve and obtaining the BLRM statistics, we saw that all dose levels fulfilled the overdose control criterion and that the highest dose level maximized the target interval's posterior probability. The dose would be escalated one level to dose level four, 525 mg/m^2. The same was observed for the fourth cohort, and the dose would be escalated one level, to dose level number five, 575 mg/m^2.

At dose level five, four patients were treated and two DLTs were observed. After updating the BLRM statistical model, we saw that dose level five no longer fulfils the overdose control criterion. Among doses one to four, it was dose level four that maximized the probability of the target interval, which means that the dose would have been de-escalated one level, to dose level four, 525 mg/m^2.

Like the other methods, dose escalation could have been continued with the BLRM, for example, until the maximum sample size was reached.

When analyzing the results of the simulation using the ten true dose-toxicity scenarios, we noted a significant number of clinical trials stopped for the BLRM. We also noted that most clinical trials were stopped due to toxicities observed in the initial cohorts. To demonstrate that, we ran a simulation of one cohort, composed of three patients, treated at the lowest dose level and where one DLT was manifested. We observed that none of the doses satisfied the overdose control criterion, $P(\hat{p}_{375} > 0{,}30|D_k) = 0{,}516$, $P(\hat{p}_{425} > 0{,}30|D_k) = 0{,}591$, $P(\hat{p}_{475} > 0{,}30|D_k) = 0{,}649$, $P(\hat{p}_{525} > 0{,}30|D_k) = 0{,}693$, $P(\hat{p}_{575} > 0{,}30|D_k) = 0{,}729$. Then, the clinical trial would be interrupted.

TABLE XI
BLRM statistics summary for the Deflexifol™ clinical trial (bolus schedule)

Cohort	Dose level (mg/m²)	n_j	y_j	$\hat{p}_j \in (0{,}20; 0{,}30]\mid D_k$	Overdose control $P(\hat{p}_j > 0{,}30\mid D_k) < 0{,}25$ (dose allocation decision)	Gelman-Rubin statistic for the parameters estimated
1	1 (375)	3	0	$\hat{p}_{375} \in (0{,}20; 0{,}30]\mid D_k = 0{,}070$ $\hat{p}_{425} \in (0{,}20; 0{,}30]\mid D_k = 0{,}084$ $\hat{p}_{475} \in (0{,}20; 0{,}30]\mid D_k = 0{,}099$ $\hat{p}_{525} \in (0{,}20; 0{,}30]\mid D_k = 0{,}109$ $\hat{p}_{575} \in (0{,}20; 0{,}30]\mid D_k = 0{,}121$	$P(\hat{p}_{375} > 0{,}30\mid D_k) = 0{,}098$ $P(\hat{p}_{425} > 0{,}30\mid D_k) = 0{,}134$ $P(\hat{p}_{475} > 0{,}30\mid D_k) = 0{,}180$ $P(\hat{p}_{525} > 0{,}30\mid D_k) = 0{,}234$ $P(\hat{p}_{575} > 0{,}30\mid D_k) = 0{,}294$ $(x_{k+1} = 425)$	1,001-1,001
2	2 (425)	3	0	$\hat{p}_{375} \in (0{,}20; 0{,}30]\mid D_k = 0{,}037$ $\hat{p}_{425} \in (0{,}20; 0{,}30]\mid D_k = 0{,}054$ $\hat{p}_{475} \in (0{,}20; 0{,}30]\mid D_k = 0{,}074$ $\hat{p}_{525} \in (0{,}20; 0{,}30]\mid D_k = 0{,}094$ $\hat{p}_{575} \in (0{,}20; 0{,}30]\mid D_k = 0{,}106$	$P(\hat{p}_{375} > 0{,}30\mid D_k) = 0{,}028$ $P(\hat{p}_{425} > 0{,}30\mid D_k) = 0{,}045$ $P(\hat{p}_{475} > 0{,}30\mid D_k) = 0{,}078$ $P(\hat{p}_{525} > 0{,}30\mid D_k) = 0{,}124$ $P(\hat{p}_{575} > 0{,}30\mid D_k) = 0{,}192$ $(x_{k+1} = 475)$	1,001-1,001
3	3 (475)	3	0	$\hat{p}_{375} \in (0{,}20; 0{,}30]\mid D_k = 0{,}015$ $\hat{p}_{425} \in (0{,}20; 0{,}30]\mid D_k = 0{,}023$ $\hat{p}_{475} \in (0{,}20; 0{,}30]\mid D_k = 0{,}041$ $\hat{p}_{525} \in (0{,}20; 0{,}30]\mid D_k = 0{,}067$ $\hat{p}_{575} \in (0{,}20; 0{,}30]\mid D_k = 0{,}092$	$P(\hat{p}_{375} > 0{,}30\mid D_k) = 0{,}008$ $P(\hat{p}_{425} > 0{,}30\mid D_k) = 0{,}013$ $P(\hat{p}_{475} > 0{,}30\mid D_k) = 0{,}026$ $P(\hat{p}_{525} > 0{,}30\mid D_k) = 0{,}061$ $P(\hat{p}_{575} > 0{,}30\mid D_k) = 0{,}119$ $(x_{k+1} = 525)$	1,001-1,001
4	4 (525)	6	0	$\hat{p}_{375} \in (0{,}20; 0{,}30]\mid D_k = 0{,}003$ $\hat{p}_{425} \in (0{,}20; 0{,}30]\mid D_k = 0{,}005$ $\hat{p}_{475} \in (0{,}20; 0{,}30]\mid D_k = 0{,}010$ $\hat{p}_{525} \in (0{,}20; 0{,}30]\mid D_k = 0{,}019$ $\hat{p}_{575} \in (0{,}20; 0{,}30]\mid D_k = 0{,}041$	$P(\hat{p}_{375} > 0{,}30\mid D_k) = 0{,}001$ $P(\hat{p}_{425} > 0{,}30\mid D_k) = 0{,}001$ $P(\hat{p}_{475} > 0{,}30\mid D_k) = 0{,}002$ $P(\hat{p}_{525} > 0{,}30\mid D_k) = 0{,}009$ $P(\hat{p}_{575} > 0{,}30\mid D_k) = 0{,}036$ $(x_{k+1} = 575)$	1,001-1,001
5	5 (575)	4	2	$\hat{p}_{375} \in (0{,}20; 0{,}30]\mid D_k = 0{,}019$ $\hat{p}_{425} \in (0{,}20; 0{,}30]\mid D_k = 0{,}032$ $\hat{p}_{475} \in (0{,}20; 0{,}30]\mid D_k = 0{,}059$ $\hat{p}_{525} \in (0{,}20; 0{,}30]\mid D_k = 0{,}132$ $\hat{p}_{575} \in (0{,}20; 0{,}30]\mid D_k = 0{,}216$	$P(\hat{p}_{375} > 0{,}30\mid D_k) = 0{,}007$ $P(\hat{p}_{425} > 0{,}30\mid D_k) = 0{,}013$ $P(\hat{p}_{475} > 0{,}30\mid D_k) = 0{,}026$ $P(\hat{p}_{525} > 0{,}30\mid D_k) = 0{,}079$ $P(\hat{p}_{575} > 0{,}30\mid D_k) = 0{,}300$ $(x_{k+1} = 525)$	1,001-1,001

4.1.3.4 Modified Toxicity Probability Interval

With the mTPI, all possible dose-escalation actions can be pre-tabulated in advance, before the trial starts. Appendix A.1 illustrates all possible dose-escalation actions in this trial.

The mTPI models the data locally, in each dose level individually, using the beta-binomial model with prior distribution Beta(1,1) and a posterior probability distribution given by (46).

With the mTPI, the dose allocation recommendation after treating the first cohort of three patients with no DLTs observed is to escalate one level (the mTPI escalates and de-escalates one level at a time). The same for the subsequent two cohorts, treated with dose levels two and three, respectively.

In Table XII, we can observe that the estimated probability of DTL, $\hat{p}_{cur}$, for the first three dose levels is obtained from the posterior distribution Beta(1,4). The estimated posterior mean probability of DLT is 0,20 with 95% credible intervals of [0,006; 0,602]. The estimated posterior probability of DLT belonging to the UI was, $P(\hat{p}_{cur} \in \text{UI}) = 0{,}5904$. Consequently, for the first three cohorts, the largest UPM obtained was the UPM for the UI, $\text{UI} = [0, \phi - \varepsilon_1[$. Thus, the mTPI recommends escalating the dose one level.

In dose level number four, six patients were treated and no DLTs were observed. The estimated probability of DLT for this dose level is given by a posterior beta distribution $\text{Beta}(1, 7)$, with estimated mean probability of DLT of $\hat{p}_4 = 0{,}125$, and, once again, the largest UPM was obtained for the UI interval. Then, with the mTPI, the dose would be escalated to the next higher level.

Up to this point, the mTPI is aligned with the 3+3 in regards with dose-escalation actions.

The only difference between the Deflexifol™ clinical trial and the replication of this clinical trial using the mTPI is in dose level number five. Four patients were treated and two DLTs were observed. The mTPI, in contrast to the 3+3, recommends dose de-escalation to the next lower dose and continuing with the trial. The 3+3, on the other hand, stops the trial and considers the MTD as the next lower dose level.

At dose level number five, the estimated posterior probability distribution is obtained from the posterior distribution $Beta(3;\ 3)$, with estimated mean probability of DLT of $\hat{p}_5 = 0{,}50$ with 95% credible intervals of [0,144; 0,853]. $\text{UPM}_{\text{OI}} = 1{,}196$ was the largest UPM, recommending dose de-escalation, as mentioned earlier. Thus, further cohorts could be treated with the mTPI.

Dose level number five also doesn't have a probability of exceeding the target toxicity level higher than 95%, $P(\hat{p}_5 > \phi | y_j) = 0{,}896$, therefore, this dose would not yet be excluded from the trial.

TABLE XII
mTPI statistics summary for the Deflexifol™ clinical trial (bolus schedule)

Cohort	Dose level (mg/m^2)	n_j	y_j	Posterior distribution	Mean probability of DLT [95% credible intervals]	UPM calculation	Safety rule $P(\hat{p} > \phi \mid y_j)$ below 95%
1	1 (375)	3	0	$Beta(1;4)$	$\hat{p}_1 = 0{,}20$ $[0{,}006; 0{,}602]$	$\mathrm{UPM}_{\mathrm{UI}} = \frac{P(\hat{p}_{cur} \in \mathrm{UI})}{(\phi - \varepsilon_1)} = \frac{0{,}59}{(0{,}25 - 0{,}05)} = 2{,}952$	$P(\hat{p}_1 > \phi \mid y_j) = 0{,}316$
2	2 (425)	3	0	$Beta(1;4)$	$\hat{p}_2 = 0{,}20$ $[0{,}006; 0{,}602]$	$\mathrm{UPM}_{\mathrm{EI}} = \frac{P(\hat{p}_{cur} \in \mathrm{EI})}{(\varepsilon_1 + \varepsilon_2)} = \frac{0{,}17}{(0{,}05 + 0{,}05)} = 1{,}695$	$P(\hat{p}_2 > \phi \mid y_j) = 0{,}316$
3	3 (475)	3	0	$Beta(1;4)$	$\hat{p}_3 = 0{,}20$ $[0{,}006; 0{,}602]$	$\mathrm{UPM}_{\mathrm{OI}} = \frac{P(\hat{p}_{cur} \in \mathrm{OI})}{(1 - \phi + \varepsilon_2)} = \frac{0{,}24}{(1 - 0{,}25 + 0{,}05)} = 0{,}343$	$P(\hat{p}_3 > \phi \mid y_j) = 0{,}316$
4	4 (525)	6	0	$Beta(1;7)$	$\hat{p}_4 = 0{,}125$ $[0{,}004; 0{,}410]$	$\mathrm{UPM}_{\mathrm{UI}} = \cdots = 3{,}951$ $\mathrm{UPM}_{\mathrm{EI}} = \cdots = 1{,}273$ $\mathrm{UPM}_{\mathrm{OI}} = \cdots = 0{,}118$	$P(\hat{p}_4 > \phi \mid y_j) = 0{,}133$
5	5 (575)	4	2	$Beta(3;3)$	$\hat{p}_5 = 0{,}50$ $[0{,}147; 0{,}853]$	$\mathrm{UPM}_{\mathrm{UI}} = \cdots = 0{,}29$ $\mathrm{UPM}_{\mathrm{EI}} = \cdots = 1{,}052$ $\mathrm{UPM}_{\mathrm{OI}} = \cdots = 1{,}196$	$P(\hat{p}_5 > \phi \mid y_j) = 0{,}896$

Examining Table XII, we can verify that the estimated mean probability of DLT for the five dose levels is not monotonically increasing. Thus, if a similar situation was obtained at the end of the trial, the PAVA would be used to ensure non-decreasing monotonicity.

4.1.3.5 Bayesian Optimal Interval

The BOIN design shares some similarities with the mTPI. Nevertheless, it bases the dose allocation decisions on frequentist statistics. The Bayesian approach is used only in the safety rule, where a beta-binomial model like the one employed in mTPI is utilized. This model assesses whether the probability of any dose level exceeding the predefined target toxicity level is greater than a specified threshold (95% was considered).

Looking at Table XIII it's evident that, with the BOIN design, the first four dose levels have an estimated probability of DLT equal to zero and, consequently, lower than the escalation boundary, $\hat{p}_{cur} < \lambda_e$. As a result, the BOIN design recommends escalating the dose one level for the first four cohorts.

TABLE XIII
BOIN statistics summary for the Deflexifol™ clinical trial (bolus schedule)

Cohort	Dose level (mg/m²)	n_{cur}	y_{cur}	Probability of DLT $\hat{p}_{cur} = \frac{y_{cur}}{n_{cur}}$ [95% confidence intervals]	Interval boundaries $\lambda_e = 0{,}197$ $\lambda_d = 0{,}298$	Posterior distribution for the safety rule	Safety rule $P(\hat{p} > \phi \mid y_j, n_j)$ below 95%
1	1 (375)	3	0	$\hat{p}_1 = \frac{0}{3} = 0$	$\hat{p}_1 < \lambda_e$	$Beta(1;4)$	$P(\hat{p}_1 > \phi \mid y_1, n_1) = 0{,}316$
2	2 (425)	3	0	$\hat{p}_2 = \frac{0}{3} = 0$	$\hat{p}_2 < \lambda_e$	$Beta(1;4)$	$P(\hat{p}_2 > \phi \mid y_2, n_2) = 0{,}316$
3	3 (475)	3	0	$\hat{p}_3 = \frac{0}{3} = 0$	$\hat{p}_3 < \lambda_e$	$Beta(1;4)$	$P(\hat{p}_3 > \phi \mid y_3, n_3) = 0{,}316$
4	4 (525)	6	0	$\hat{p}_4 = \frac{0}{6} = 0$	$\hat{p}_4 < \lambda_e$	$Beta(1;7)$	$P(\hat{p}_4 > \phi \mid y_4, n_4) = 0{,}133$
5	5 (575)	4	2	$\hat{p}_5 = \frac{2}{4} = 0{,}5$ $[0{,}010; 0{,}990]$	$\hat{p}_5 > \lambda_d$	$Beta(3;3)$	$P(\hat{p}_5 > \phi \mid y_5, n_5) = 0{,}896$

On dose level five, where four patients were treated and two DLTs were observed, the estimated probability of DLT is 0,5 and, consequently, higher than the de-escalation boundary λ_d. Thus, the BOIN recommends dose de-escalation, the same dose-escalation action recommended by the other methods, except the CRM.

Like the mTPI, all possible dose-escalation actions can be pre-tabulated in advance before the trial starts. Appendix A.2 depicts all possible dose-escalation actions for this trial.

In summary, all methods allow additional cohorts to be treated, in contrast to the 3+3 method, that stops the trial. Dose escalation and de-escalation would continue until a stopping rule is observed (e.g., maximum sample size is reached, an MTD estimate with a sufficient level of certainty, etc.).

5

CONCLUSION

The dose-escalation methods provide a theoretical and statistical framework that serves as the foundation of the dose-escalation decisions and estimation of the MTD in FIH oncology clinical trials. However, oncology clinical trials are complex and include many additional endpoints, other than toxicity, to help determine therapies' efficacy and safety. Thus, in FIH clinical research, the dose-escalation and de-escalation decisions are not exclusively based on the DLTs observed, but also on other variables, like pharmacokinetic data or efficacy data.

In this thesis, we described the three classes of dose-escalation designs, the algorithm-based designs, the model-based designs, and the model-assisted designs, and presented in detail five of the most important *novel* single-agent dose-escalation methods, the CRM, EWOC, BLRM, mTPI and BOIN.

A simulation study compared the five dose-escalation methods concerning accuracy, dose allocation and toxicity. In addition, a replication of a clinical trial, using real clinical trial data, was conducted to assess potential dose allocation differences in the methods.

The simulation study results suggest that the CRM is more likely to correctly select the MTD and treat more patients with the dose corresponding to the MTD. Nevertheless, the CRM showed to be more aggressive than the others, exposing more patients to doses above the MTD, translating into a higher number of DLTs.

The CRM's aggressive nature was also observable in the replication of the Deflexifol™ clinical trial, where we noted that the CRM maintained the same dose level upon the occurrence of two DLTs, in a total of four patients, while the other dose-escalation methods de-escalated to the next lower dose level.

The EWOC, mTPI and BOIN showed comparable accuracy in determining the true MTD, with PCS that ranged between 60,33% and 61,61%. Yet some issues were identified in the EWOC. It exhibited overly aggressive behavior when the MTD was at the lowest dose level, and low accuracy when the MTD was at the highest dose level. Thus, given the

simulation results, the mTPI and the BOIN can be a better choice than the EWOC, when a more conservative method is desired. Furthermore, we noted that, both in the simulation study and the replication of the Deflexifol™ clinical trial, the EWOC tends to treat a significant number of patients below the MTD, a characteristic that does not work in its favor.

The BLRM was shown to be overly conservative, exhibiting a very high number of trials stopped due to the overdose control criterion, low accuracy, and poor dose allocation. The overdose control criterion employed in the BLRM appears to be excessively conservative, suggesting that a more suitable overdose control criterion may be required to improve the BLRM's performance.

The simulations conducted utilized the default configuration for each one of the dose-escalation methods, as we assumed no prior knowledge about the drug under study in each one of the scenarios, which usually doesn't happen in clinical research, given that the FIH clinical trials take place after extensive pre-clinical research.

The model-based designs require the specification of the prior distribution of the parameters, and the use of non-informative default priors can potentially impact the accuracy and dose allocation of these methods. Zhou *et al.* [54] and Zhu *et al.* [75], for example, showed, respectively, that the CRM's performance is affected by the specification of the *skeleton*, and the specification of the prior.

The model-assisted designs, on the other hand, don't have the risk of biased results caused by inaccurate initial dose-toxicity curves, which can be seen as an advantage.

Despite the evaluation conducted through the simulated clinical trials and the established metrics, it's important to note that the choice of the different dose-escalation methods in clinical research can be based on each method's characteristics and suitability to specific clinical trials and specific anti-cancer agents. For example, more aggressive dose-escalation methods can be used for anti-cancer agents that cause DLTs transient and correctable, and more conservative methods to anti-cancer agents that cause potentially lethal or life-threatening DLTs. Additionally, the choice of the method can be influenced by the investigation teams' preferences, for example, the model-assisted designs are statistically and computationally less demanding and easier to implement.

It's also important to emphasize that all the dose-escalation methods operate under the assumption that the toxicity monotonically increases with the dose. Many of these methods were initially designed for application in clinical trials involving cytotoxic anti-cancer agents, generally known for their potential toxicity.

Most novel anti-cancer treatments comprise targeted therapies or immunotherapy, which generally exhibit lower toxicity compared to cytotoxic therapies. Additionally, their therapeutic effects can occur at doses below the MTD. These characteristics may have

contributed to the extensive use of conservative methods, such as the BLRM, despite its lower accuracy in identifying the correct MTD in several published simulation studies.

The literature on dose-escalation methods is vast, and what was presented in this thesis is just the tip of the iceberg. Many authors proposed solutions to enhance the dose-escalation methods performance, which range from prior specification to actual changes in the methods, proposed modifications and new approaches to deal with late-onset toxicity (time-to-event (TITE) CRM, data augmentation CRM, TITE-EWOC, TITE-BOIN, to name a few) and proposed modifications and new methods for drug combination dose escalation.

Despite that, we believe that the results from our simulations have helped identify the differences between the methods, and can, together with the published literature's findings, assist in choosing the most suitable method for each clinical trial's specific characteristics.

FIH oncology clinical trials are the cornerstone of the development of new anti-cancer agents. Despite all the research on improving dose allocation and estimation of the MTD, the uncertainty surrounding the candidate anti-cancer agents, regarding safety and tolerability, is always high, and some patients inevitably suffer toxicities. Hence, it's essential to acknowledge that patients who participate in early development clinical research are the real heroes, without which science would not move forward.

5.1 Limitations

The simulation study included ten scenarios covering the MTD at all dose levels and several types of dose-toxicity curves. Nevertheless, we recognize that the ten scenarios do not comprehensively represent all theoretically possible dose-toxicity curves. A more thorough simulation could have been performed. Some authors, for example, used algorithms to generate a great number of scenarios (e.g., one thousand scenarios) and completed in-depth simulation studies.

The many elements inherent to the FIH clinical trials also limited the simulation study. For example, the simulation study considered only six dose levels. Still, it could have compared the methods with a smaller number of dose levels (e.g., four) and a larger number of dose levels (e.g., eight). Moreover, in the simulation study, we focused solely on a single target toxicity level, 0,33, and additional target toxicity levels could have been included for a more comprehensive analysis. Furthermore, fixed cohorts of three patients were utilized in the simulation study and it could have included cohorts of one patient, for example. Another

consideration is that our dose escalation was limited to increments of one dose level at a time, and the model-based designs could have been allowed to perform dose jumps.

In summary, the possible configurations for the simulation study are vast. Still, the setup chosen was based on practical clinical research experience (e.g., the use of one-patient cohorts during the entire clinical trial is not likely), comparability of the methods (e.g., dose escalation was limited to one level at the time to facilitate the comparison of model-based and model-assisted designs, that do not allow dose jumps), and time constraints, due to the computationally demanding nature of the simulations. Nevertheless, we acknowledge that including additional target toxicity levels would have improved the simulation study performed.

Another limitation pertains to the use of R packages to simulate the clinical trials, which restricted the obtention of raw data for conducting a more complete statistical analysis. For example, the "BOIN" R package only gives the simulation output with the main statistics. The computation of the whole simulation study could have been an improvement.

REFERENCES

[1] C. Le Tourneau, J. J. Lee, and L. L. Siu, "Dose escalation methods in phase i cancer clinical trials," *J. Natl. Cancer Inst.*, vol. 101, no. 10, pp. 708–720, 2009.

[2] T. A. Durham and J. R. Turner, "Early phase clinical trials," in *Introduction to Statistics in Pharmaceutical Clinical Trials*, London: Pharmaceutical Press, 2008, pp. 85–95.

[3] A. Ivanova, "Phase I Trials in Oncology," in *Methods and Applications of Statistics in Clinical Trials: Volume 1 Concepts Principles, Trials and Designs*, N. Balakrishnan, Ed. New Jersey: John Wiley & Sons, Inc., 2014, pp. 719–724.

[4] S. Green, A. Smith, J. Benedetti, and J. Crowley, "Phase I and Phase I/II Trials," in *Clinical Trials in Oncology*, 3rd ed., Boca Raton: CRC Press, 2012, pp. 57–71.

[5] S. Piantadosi, "Dose-Finding Designs," in *Clinical Trials: A Methodologic Perspective*, 2nd ed., New Jersey: John Wiley & Sons, Inc., 2005, pp. 223–250.

[6] J. O'Quigley and A. Iasonos, "Dose-Finding Designs Based on the Continual Reassessment Method," in *Handbook of Statistics in Clinical Oncology*, 3rd ed., J. Crowley and A. Hoering, Eds. Boca Raton: CRC Press, 2012, pp. 21–51.

[7] S.-C. Chow and J.-P. Liu, "Designs For Cancer Clinical Trials," in *Design and Analysis of Clinical Trials: Concepts and Methodologies*, 2nd ed., New Jersey: John Wiley & Sons, Inc., 2004, pp. 215–238.

[8] A. Ivanova and N. Flournoy, "Phase I Clinical Trials," in *Statistical Advances in the Biomedical Sciences: Clinical Trials, Epidemiology, Survival Analysis, and Bioinformatics*, A. Biswas, S. Datta, J. P. Fine, and M. R. Segal, Eds. New Jersey: John Wiley & Sons, Inc., 2008, pp. 3–14.

[9] G. M. Wheeler, "Overview of Phase I Designs," in *Handbook of Methods for Designing, Monitoring, and Analyzing Dose-Finding Trials*, J. O'Quigley, A. Iasonos, and B. Bornkamp, Eds. Boca Raton: CRC Press, 2017, pp. 3–26.

[10] W. F. Rosenberger and L. M. Haines, "Competing designs for phase I clinical trials: a review," *Stat. Med.*, vol. 21, no. 18, pp. 2757–2770, 2002.

[11] O. Sverdlov, W. K. Wong, and Y. Ryeznik, "Adaptive clinical trial designs for phase I cancer studies," *Stat. Surv.*, vol. 8, pp. 2–44, 2014.

[12] D. V Araujo, M. Oliva, K. Li, R. Fazelzad, Z. Amy, and L. L. Siu, "Contemporary dose-

escalation methods for early phase studies in the immunotherapeutics era," *Eur. J. Cancer*, vol. 158, pp. 85–98, 2021.

[13] Cytel, "Practical Model-based Approaches for Phase 1 Oncology Trials," 2020. [Online]. Available: https://www.youtube.com/watch?v=vAMhQ0JDQBM. [Accessed: 05-May-2022].

[14] S. Liu and Y. Yuan, "Bayesian optimal interval designs for phase I clinical trials," *J. R. Stat. Soc. Ser. C Appl. Stat.*, vol. 64, no. 3, pp. 507–523, 2015.

[15] S. Chevret, "Basic concepts in dose-finding," in *Statistical Methods for Dose-Finding Experiments*, S. Chevret, Ed. Chichester: John Wiley & Sons Ltd, 2006, pp. 5–18.

[16] A. R. Hansen, D. M. Graham, G. R. Pond, and L. L. Siu, "Phase 1 Trial Design: Is 3 + 3 the Best?," *Cancer Control*, vol. 21, no. 3, pp. 200–208, 2014.

[17] Y. Yuan, J. J. Lee, and S. G. Hilsenbeck, "Model-Assisted Designs for Early-Phase Clinical Trials: Simplicity Meets Superiority," *JCO Precis. Oncol.*, vol. 3, no. 3, pp. 1–12, 2019.

[18] The International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use, "ICH HARMONISED TRIPARTITE GUIDELINE: NONCLINICAL EVALUATION FOR ANTICANCER PHARMACEUTICALS S9," 2009. [Online]. Available: https://database.ich.org/sites/default/files/S9_Guideline.pdf.

[19] R. K. Mittapalli, D. Yin, D. Beaupre, and R. Palaparthy, "Starting dose selection and dose escalation for oncology small molecule first-in-patient trials: learnings from a survey of FDA-approved drugs," *Cancer Chemother. Pharmacol.*, vol. 87, pp. 23–30, 2021.

[20] European Medicines Agency, "Guideline on strategies to identify and mitigate risks for first-in-human and early clinical trials with investigational medicinal products," 2017. [Online]. Available: https://www.ema.europa.eu/en/documents/scientific-guideline/guideline-strategies-identify-mitigate-risks-first-human-early-clinical-trials-investigational_en.pdf.

[21] Food and Drug Administration, "Good Review Practice: Clinical Review of Investigational New Drug Applications," 2013. [Online]. Available: https://www.fda.gov/media/87621/download. [Accessed: 05-May-2022].

[22] B. De Heras *et al.*, "Healthy volunteers in first-in-human oncology drug development for small molecules," vol. 88, pp. 1773–1784, 2022.

[23] K. Panageas, S. and A. Knezevic, "Selection of Endpoints," in *Textbook of Clinical Trials in Oncology: A Statistical Perspective*, S. Halabi and S. Michiels, Eds. Boca Raton: CRC Press, 2019, pp. 9–26.

[24] National Cancer Institute, "Common Terminology Criteria for Adverse Events (CTCAE) Vrsion 5.0," 2017. [Online]. Available: https://ctep.cancer.gov/protocoldevelopment/electronic_applications/docs/CTCAE_v5_Quick_Reference_5x7.pdf. [Accessed: 05-May-2022].

[25] B. E. Storer, "Phase I Trials," in *Biostatistics in Clinical Trials*, C. K. Redmond and T. Colton, Eds. Chichester: John Wiley & Sons Ltd, 2001, pp. 337–340.

[26] J. S. Babb and A. Rogatko, "Bayesian Methods for Cancer Phase I Clinical Trials," in *Advances in Clinical Trial Biostatistics*, N. L. Geller, Ed. New York: Marcel Dekker, Inc.,

2004, pp. 1–40.

[27] B. Neuenschwander, M. Branson, and T. Gsponer, "Critical aspects of the Bayesian approach to phase I cancer trials," *Stat. Med.*, vol. 27, no. 13, pp. 2420–2439, 2008.

[28] J. Babb, A. Rogatko, and S. Zacks, "Cancer phase I clinical trials: efficient dose escalation with overdose control," *Stat. Med.*, vol. 17, no. 10, pp. 1103–1120, 1998.

[29] Y. Zhou, R. Li, F. Yan, J. J. Lee, and Y. Yuan, "A comparative study of Bayesian optimal interval (BOIN) design with interval 3+3 (i3+3) design for phase I oncology dose-finding trials," *Stat. Biopharm. Res.*, vol. 13, no. 2, pp. 147–155, 2021.

[30] E. Garrett-Mayer, "Application of New Designs in Phase I Trials," in *Methods and Applications of Statistics in Clinical Trials: Volume 1 Concepts Principles, Trials and Designs*, N. Balakrishnan, Ed. New Jersey: John Wiley & Sons, Inc., 2014, pp. 65–73.

[31] W. He and J. Liu, "Algorithm-Based Designs," in *Methods and Applications of Statistics in Clinical Trials: Volume 1 Concepts Principles, Trials and Designs*, N. Balakrishnan, Ed. New Jersey: John Wiley & Sons, Inc., 2014, pp. 40–52.

[32] H. Zhou, Y. Yuan, and L. Nie, "Accuracy, Safety, and Reliability of Novel Phase I Trial Designs," *Clin. Cancer Res.*, vol. 24, no. 18, pp. 4357–4364, 2018.

[33] A. Rogatko, D. Schoeneck, W. Jonas, M. Tighiouart, F. R. Khuri, and A. Porter, "Translation of innovative designs into phase I trials," *J. Clin. Oncol.*, vol. 25, no. 31, pp. 4982–4986, 2007.

[34] Y. Yuan, K. R. Hess, S. G. Hilsenbeck, and M. R. Gilbert, "Bayesian Optimal Interval Design: A Simple and Well-Performing Design for Phase I Oncology Trials," *Clin. Cancer Res.*, vol. 22, no. 17, pp. 4291–4301, 2016.

[35] J. O'Quigley, M. Pepe, and L. Fisher, "Continual Reassessment Method: A Practical Design for Phase 1 Clinical Trials in Cancer," *Biometrics*, vol. 46, no. 1, pp. 33–48, 1990.

[36] Adaptive Designs Working Group of the MRC Network of Hubs for Trials Methodology Research, "A quick guide why not to use A + B designs." [Online]. Available: https://www.methodologyhubs.mrc.ac.uk/files/6814/6253/2385/A_quick_guide_why_not_to_use_AB_designs.pdf.

[37] D. A. Berry, "Bayesian clinical trials," *Nat. Rev. Drug Discov.*, vol. 5, no. 1, pp. 27–36, 2006.

[38] J. J. Lee and C. T. Chu, "Bayesian clinical trials in action," *Stat. Med.*, vol. 31, no. 25, pp. 2955–2972, 2012.

[39] S. M. Berry, B. P. Carlin, J. J. Lee, and P. Müller, *Bayesian Adaptive Methods for Clinical Trials*. Boca Raton: CRC Press, 2010.

[40] G. A. Young and R. L. Smith, "Bayesian methods," in *Essentials of Statistical Inference*, Cambridge: Cambridge University Press, 2005, pp. 22–64.

[41] J. M. Curran and W. M. Bolstad, *Introduction to Bayesian Statistics*, 3rd ed. New Jersey: John Wiley & Sons, Inc., 2017.

[42] S. K. Gupta, "Use of Bayesian statistics in drug development: Advantages and challenges," *Int. J. Appl. basic Med. Res.*, vol. 2, no. 1, pp. 3–6, 2012.

[43] S. M. Lynch, "Basics of Bayesian Statistics," in *Introduction to Applied Bayesian Statistics and Estimation for Social Scientists*, New York: Springer Science+Media, LLC, 2007, pp. 47–75.

[44] T. Donovan and R. Mickey, *Bayesian Statistics for Beginners: A Step-by-Step Approach*. Oxford: Oxford University Press, 2019.

[45] M. Fors and P. González, "Current status of Bayesian clinical trials for oncology, 2020," *Contemp. Clin. trials Commun.*, vol. 20, pp. 1–20, 2020.

[46] D. McNeish, "On using Bayesian methods to address small sample problems," *Struct. Equ. Model.*, vol. 23, no. 5, pp. 750–773, 2016.

[47] U.S. Department of Health and Human Services Food and Drug Administration, "Guidance for the Use of Bayesian Statistics in Medical Device Clinical Trials," 2010. [Online]. Available: https://www.fda.gov/media/71512/download.

[48] A. A. Johnson, M. Q. Ott, and M. Dogucu, *Bayes Rules! An Introduction to Applied Bayesian Modeling*. Boca Raton: CRC Press, 2022.

[49] U.S. Department of Health and Human Services Food and Drug Administration, "Adaptive Designs for Clinical Trials of Drugs and Biologics Guidance for Industry," 2019. [Online]. Available: https://www.fda.gov/media/78495/download.

[50] J. K. Ghosh, M. Delampady, and T. Samanta, *An Introduction to Bayesian Analysis: Theory and Methods*. New York: Springer Science+Media, LLC, 2006.

[51] B. Lambert, *A Student's Guide to Bayesian Statistics*. London: SAGE Publications Ltd, 2018.

[52] S. B. Love *et al.*, "Embracing model-based designs for dose-finding trials," *Br. J. Cancer*, vol. 117, no. 3, pp. 332–339, 2017.

[53] Y. Ji, P. Liu, Y. Li, and B. N. Bekele, "A modified toxicity probability interval method for dose-finding trials," *Clin. Trials*, vol. 7, no. 6, pp. 653–663, 2010.

[54] H. Zhou, T. A. Murray, H. Pan, and Y. Yuan, "Comparative review of novel model-assisted designs for phase I clinical trials," *Stat. Med.*, vol. 37, no. 14, pp. 2208–2222, 2018.

[55] Y. K. Cheung, *Dose Finding by the Continual Reassessment Method*. Boca Raton: CRC Press, 2011.

[56] H. Yang and S. J. Novick, *Bayesian Analysis with R for Drug Development: Concepts, Algorithms, and Case Studies*. Boca Raton: CRC Press, 2019.

[57] S. Chevret and S. Zohar, "The continual reassessment method," in *Statistical Methods for Dose-Finding Experiments*, S. Chevret, Ed. Chichester: John Wiley & Sons Ltd, 2006, pp. 133–148.

[58] G. D. James, S. Symeonides, J. Marshall, J. Young, and G. Clack, "Assessment of various continual reassessment method models for dose-escalation phase 1 oncology clinical trials: using real clinical data and simulation studies," *BMC Cancer*, vol. 21, no. 1, pp. 1–10, 2021.

[59] G. M. Wheeler *et al.*, "How to design a dose-finding study using the continual reassessment method," *BMC Med. Res. Methodol.*, vol. 19, no. 1, pp. 1–15, 2019.

[60] S. Møller, "An extension of the continual reassessment methods using a preliminary

up-and-down design in a dose finding study in cancer patients, in order to investigate a greater range of doses," *Stat. Med.*, vol. 14, no. 9–10, pp. 911–922, 1995.

[61] S. N. Goodman, M. L. Zahurak, and S. Piantadosi, "Some practical improvements in the continual reassessment method for phase I studies," *Stat. Med.*, vol. 14, no. 11, pp. 1149–1161, 1995.

[62] C. Ahn, "An evaluation of phase I cancer clinical trial designs," *Stat. Med.*, vol. 17, no. 14, pp. 1537–1549, 1998.

[63] J. O'Quigley and L. Z. Shen, "Continual reassessment method: a likelihood approach," *Biometrics*, vol. 52, no. 2, pp. 673–684, 1996.

[64] M. Tighiouart and A. Rogatko, "Dose-escalation with overdose control," in *Statistical Methods for Dose-Finding Experiments*, S. Chevret, Ed. Chichester: John Wiley & Sons Ltd, 2006, pp. 173–188.

[65] G. M. Wheeler, A. P. Mander, and M. J. Sweeting, "Toxicity-dependent feasibility bounds for the escalation with overdose control approach in phase I cancer trials," *Stat. Med.*, vol. 36, no. 16, pp. 2499–2513, 2017.

[66] S. Y. Yum, "What is the optimal fit-for-purpose design in an oncology first-in-human trial?: 3+3 vs. Bayesian logistic regression model," *Transl. Clin. Pharmacol.*, vol. 28, no. 1, pp. 1–6, 2020.

[67] T. Daimon, A. Hirakawa, and S. Matsui, *Dose-Finding Designs for Early- Phase Cancer Clinical Trials: A Brief Guidebook to Theory and Practice*. Tokyo: Springer Japan KK, 2019.

[68] B. K. Günhan, S. Weber, A. Seroutou, and T. Friede, "Phase I dose-escalation oncology trials with sequential multiple schedules," *BMC Med. Res. Methodol.*, vol. 21, pp. 1–14, 2021.

[69] F. Yan, S. J. Mandrekar, and Y. Yuan, "Keyboard: A Novel Bayesian Toxicity Probability Interval Design for Phase I Clinical Trials," *Clin. Cancer Res.*, vol. 23, no. 15, pp. 3994–4003, 2017.

[70] Y. Ji, Y. Li, and B. N. Bekele, "Dose-finding in phase I clinical trials based on toxicity probability intervals," *Clin. Trials*, vol. 4, no. 3, pp. 235–244, 2007.

[71] S. Yang, S. Wang, and Y. Ji, "An Integrated Dose-Finding Tool for Phase I Trials in Oncology," *Contemp. Clin. Trials*, vol. 45, pp. 426–434, 2015.

[72] CU Cancer Center, "Modified Toxicity Probability Interval and Keyboard Design Fundamentals," 2022. [Online]. Available: https://www.youtube.com/watch?v=GzsB1RHfVho.

[73] 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*. London, New York: John Wiley & Sons, 1972.

[74] Y. Wang, "Dose-finding Designs for Phase I Clinical Trials in Oncology and Use of Selective Phenotypyng to Increase Power of Genetic Association Studies," University of North Carolina at Chapel Hill, 2014.

[75] Y. Zhu, W. Hwang, and Y. Li, "Evaluating the effects of design parameters on the performances of phase I trial designs," *Contemp. Clin. trials Commun.*, vol. 15, pp. 1–10, 2019.

[76] F. Yan, L. Zhang, Y. Zhou, H. Pan, S. Liu, and Y. Yuan, "Boin: An R package for designing single-agent and drug-combination dose-finding trials using bayesian optimal interval designs," *J. Stat. Softw.*, vol. 94, no. 13, pp. 1–32, 2020.

[77] J. O'Quigley, "Estimating the Probability of Toxicity at the Recommended Dose Following a Phase I Clinical Trial in Cancer," *Biometrics*, vol. 48, no. 3, pp. 853–862, 1992.

[78] S. M. Lee and Y. K. Cheung, "Model calibration in the continual reassessment method," *Clin. Trials*, vol. 6, no. 3, pp. 227–238, 2009.

[79] Y. Yao, "Comparison of Novel Phase I Clinical Trial Designs Based on Toxicity Probability Intervals," University of Pittsburgh, 2017.

[80] R Core Team, "R: A language and environment for statistical computing." R Foundation for Statistical Computing, Vienna, 2022.

[81] RStudio Team, "RStudio: Integrated Development Environment for R." RStudio, PBC., Boston, 2023.

[82] M. Sweeting and G. Wheeler, "bcrm: Bayesian Continual Reassessment Method for Phase I Dose-Escalation Trials." 2019.

[83] K. Cheung, "dfcrm: Dose-Finding by the Continual Reassessment Method." 2019.

[84] D. S. Bové, W. Y. Yeung, G. Palermo, and T. Jaki, "Model-Based Dose Escalation Designs in R with crmPack," *J. Stat. Softw.*, vol. 89, no. 10, pp. 1–22, 2019.

[85] Y. Ji and R. Herrick, "Software Download Kiosk: Toxicity Probability Intervals," 2012. [Online]. Available: https://biostatistics.mdanderson.org/SoftwareDownload/SingleSoftware/Index/72. [Accessed: 20-May-2023].

[86] E. Ulas and F. Karaman, "A comprehensive comparison of phase I dose escalation methods," *Kuwait J. Sci.*, vol. 45, no. 4, pp. 26–36, 2018.

[87] P. R. Clingan *et al.*, "First-in-human phase I study of infusional and bolus schedules of Deflexifol , a novel 5-fluorouracil and leucovorin formulation , after failure of standard treatment," *Asia Pac J Clin Oncol.*, vol. 15, no. 3, pp. 151–157, 2019.

[88] S. Liu, C. Cai, and J. Ning, "Up-and-down designs for phase I clinical trials," *Contemp Clin Trials.*, vol. 36, no. 1, pp. 218–227, 2013.

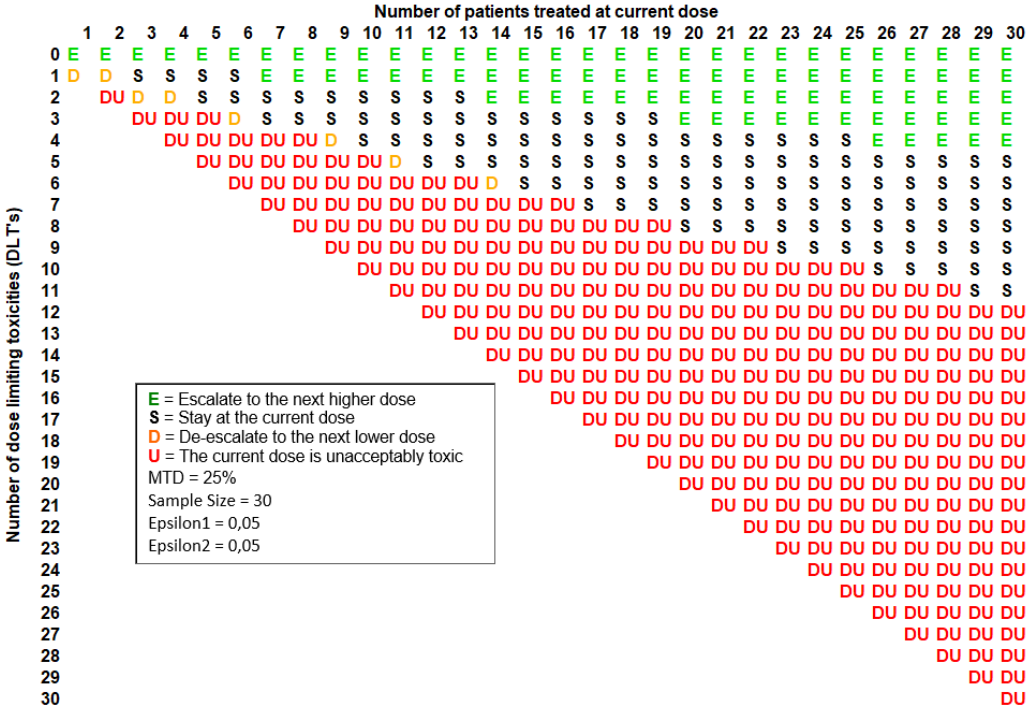
[89] D. J. Lunn, A. Thomas, N. Best, and D. Spiegelhalter, "WinBUGS — a Bayesian modelling framework: concepts, structure, and extensibility," *Stat. Comput.*, vol. 10, pp. 325–337, 2000.

A

APPENDIX A – MODEL-ASSISTED DESIGNS DOSE-ESCALATION PRETABULATED ACTIONS

A.1 mTPI Dose-Escalation Pretabulated Actions

mTPI spreadsheet output depicting all possible dose-escalation and de-escalation actions for a target toxicity level of $\phi = 0{,}25$, $\varepsilon_1 = \varepsilon_2 = 0{,}05$ and $n = 30$. mTPI Excel™ macro tool spreadsheet output obtained from [85].



A.2 BOIN Dose-Escalation Pretabulated Actions

R output depicting all possible dose-escalation and de-escalation actions for a target toxicity level of $\phi = 0{,}25$ and $n = 30$. R output obtained from the R package "BOIN" [76].

```
> library(BOIN)
> get.boundary(0.25,30,1)
$lambda_e
[1] 0.1968009

$lambda_d
[1] 0.2983922

$boundary_tab

Number of patients treated  1  2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 1
9 20 21 22 23 24 25 26 27 28
Escalate if # of DLT <=     0  0 0 0 0 1 1 1 1  1  2  2  2  2  2  3  3  3
3  3  4  4  4  4  4  5  5  5
Deescalate if # of DLT >=   1  1 1 2 2 2 3 3 3  3  4  4  4  5  5  5  6  6
6  6  7  7  7  8  8  8  9  9
Eliminate if # of DLT >=   NA NA 3 3 3 4 4 4 5  5  6  6  6  7  7  7  8  8
8  9  9  9 10 10 10 11 11 11

Number of patients treated 29 30
Escalate if # of DLT <=     5  5
Deescalate if # of DLT >=   9  9
Eliminate if # of DLT >=   12 12

attr(,"class")
[1] "boin"
```

.

B

ANNEX B – R CODE

B.1 R Code to Run the Simulation with the Ten Hypothetic True Dose-Toxicity Scenarios

```
####################
####################
### Simulation using ten hypothetical dose-toxicity scenarios
####################
####################
# Data txt file - name given to file:Dados em txt.txt
#Dose,Scenario1,Scenario2,Scenario3,Scenario4,Scenario5,Scenario6,Sce-
nario7,Scenario8,Scenario9,Scenario10
#150,0.28,0.31,0.05,0.06,0,0.02,0.01,0.06,0,0.03
#200,0.47,0.36,0.33,0.15,0.01,0.06,0.03,0.09,0,0.06
#250,0.66,0.41,0.84,0.33,0.07,0.16,0.07,0.15,0.01,0.09
#300,0.82,0.46,0.98,0.58,0.33,0.33,0.16,0.23,0.04,0.15
#350,0.91,0.52,1,0.79,0.78,0.58,0.33,0.33,0.13,0.23
#400,0.96,0.57,1,0.92,0.96,0.79,0.56,0.46,0.33,0.33
##############################################################################
#Loading true dose-toxicity scenarios
Dados<-read.csv("Dados em txt.txt", header = TRUE, sep=",")
Dados
Dados$Dose
Dados$Scenario1

###### NOTE ######
R PACKAGES SHOULD BE LOADED IN THE ORDER DESCRIBED BELOW, OTHERWISE, THERE
IS A RISK OF THE CODE NOT RUNNING DUE TO SOME R PACKAGES HAVING FUNCTIONS
WITH THE SAME NAME
###### NOTE ######

##############################################################################
##############################################################################
##############################################################################
```

```
###### CRM
################################################################################
################################################################
################################################################################
### bcrm R package ###
install.packages("dfcrm")
library(dfcrm)
################################################################################
#Obtaining the Skeleton through the method described by Lee & Cheung
(2009)
#(getprior function in the dfcrm package)
#halfwidth=0.06 was chosen, following Lee & Cheung (2009) simulations, and
nu was selected as
#corresponding to the 3rd dose level
crm.skeleton <- getprior(halfwidth=0.06, target=0.33, nu=3, nlevel=6,
model="empiric")
#Skeleton used on the initial fit of the dose-toxicity curve was the same
for all scenarios, which
#means that it was not a good fit, except for Scenario4
#0.1172232 0.2140326 0.3300000 0.4505456 0.5636193 0.6620963
################################################################################
#Installing and loading the bcrm package to run simulations for CRM
install.packages("bcrm")
library(bcrm)
################################################################################
##### CRM SIMULATION: SCENARIOS 1-10 #####
#Target toxicity level phi=0.33
phi=0.33
#Dose-skipping not allowed
#Non-informative prior was utilized (common prior in the literature):
log(theta)~N(0,1.34^2)
################################################################################
#SCENARIO 1
CRM.Scenario1<-bcrm(stop=list(nmax=30),
                    p.tox0=crm.skeleton, dose=Dados$Dose, ff="power",
                    prior.alpha=list(3,0,1.34^2), cohort=3, tar-
get.tox=phi, constrain=TRUE,
                    pointest="plugin", start=1, simulate=TRUE, nsims=1000,
truep=Dados$Scenario1,
                    method="exact", set.seed(1313))

###################################
### (Repeat for scenarios 2 to 10)
###################################
################################################################################
##### Performance Analysis ########
##### SCENARIO 1 #####
#1. Percentage of correct selection (PCS) of the true MTD
print.CRM.Scenario1 <- print(CRM.Scenario1)
print.CRM.Scenario1
CRM.Recommended.MTDs.Scenario1  <-table(print.CRM.Scenario1$rec)/1000*100
CRM.Recommended.MTDs.Scenario1
```

```
#2. Mean number (and percentage) of patients treated at the dose level
corresponding to the true MTD
#across the 1000 simulated trials
CRM.Scenario1.Matrices <- print(CRM.Scenario1, trajectories = TRUE)
CRM.Scenario1.Matrices$doses
#n
#Function to change matrix to give the number of patients dosed at each
dose level.
#Each column is a dose level and each row is a simulated simulated clini-
cal trial
count.x.row <- function(row){
  x.counts <- table(row)
  counts <- integer(6)
  for (i in 1:6) {
    if (i %in% names(x.counts)) {
      counts[i] <- x.counts[[as.character(i)]]
    } else {
      counts[i] <- 0
    }
  }
  return(counts)
}
# Apply the function to each row of the CRM.ScenarioX.Matrices$doses ma-
trix
Dose.counts.Scenario1.Matrix <- t(apply(CRM.Scenario1.Matrices$doses, 1,
count.x.row))
CRM.Scenario1.x.mean <-colMeans(Dose.counts.Scenario1.Matrix)
CRM.Scenario1.x.mean
#Percentage (%)
CRM.Scenario1.x.mean.percentage <-colMeans(Dose.counts.Scenario1.Ma-
trix)/mean(rowSums(Dose.counts.Scenario1.Matrix))*100
CRM.Scenario1.x.mean.percentage

#3.A Average percentage of patients treated below the MTD
#N/A

#3.B Mean number (and percentage) of patients treated above the true MTD
in the 1000 simulated trials
#n
CRM.Scenario1.x.mean.above.true.MTD <-mean(rowSums(Dose.counts.Sce-
nario1.Matrix[, 2:6]))
CRM.Scenario1.x.mean.above.true.MTD
#percentage
CRM.Scenario1.x.mean.above.true.MTD.percentage <-mean(row-
Sums(Dose.counts.Scenario1.Matrix[, 2:6]))/mean(rowSums(Dose.counts.Sce-
nario1.Matrix))*100
CRM.Scenario1.x.mean.above.true.MTD.percentage

#4. Mean number of DLTs observed (and percentage) for each dose-escalation
method and each scenario,
```

```
#across the 1000 simulated trials
CRM.Scenario1.Matrices$outcomes
#n
CRM.Scenario1.mean.number.DLTs <-mean(rowSums(CRM.Scenario1.Matrices$out-
comes))
CRM.Scenario1.mean.number.DLTs
#Percentage
CRM.Scenario1.mean.number.DLTs.percentage <-mean(rowSums(CRM.Scenario1.Ma-
trices$outcomes))/30
CRM.Scenario1.mean.number.DLTs.percentage
##################################
### (Repeat for scenarios 2 to 10)
##################################

#############################################################################
#############################################################################
###### EWOC
#############################################################################
#############################################################
#############################################################################
#Installing and loading the crmPack package
install.packages("crmPack")
library(crmPack)
#############################################################################
##### EWOC SIMULATION: SCENARIOS 1-10 #####
# Dose range
emptydata <- Data(doseGrid=seq(from=150, to=400, by=50))
# Initialize the EWOC model with Unif priors on gamma and raw0
phi=0.33
xmin=100
xmax=450
#Defininng the EWOC model
modelEWOC <- LogisticKadane(theta = phi, xmin=xmin, xmax=xmax)
#Unif prior in rhozero - Unif(0,phi)
#Unif prior in gama - Unif(Xmin,Xmax)

# Rule for selecting the next dose - escalation with overdose criterion,
P(x(k+1)>gamma) closer to alpha
#alpha=0.25
NextBestEWOC <- NextBestMTD(target=0.33, derive=function(mtdSamples){quan-
tile(mtdSamples, probs=0.25)})
# Cohort-size
CohortSizeEWOC <- CohortSizeConst(size=3)
# Stopping rule - minimum of 30 patients treated
StoppingRulesEWOC <- StoppingMinPatients(nPatients=30)
# Dose-skipping no allowed
DoseIncrementsEWOC <-  IncrementsNumDoseLevels(maxLevels=1)
# Initialize the design
designEWOC <- Design(model=modelEWOC,
                     nextBest=NextBestEWOC,
                     stopping=StoppingRulesEWOC,
```

```
                     increments=DoseIncrementsEWOC,
                     cohortSize=CohortSizeEWOC,
                     data=emptydata,
                     startingDose=150)
# MCMC methods - burn-in of 10000 iterations and then take every other it-
eration up to a collection of 50000 samples (JAGS)
mcmc.definitions <- McmcOptions(burnin=10000, step=2, samples=50000)
################################
#SCENARIO 1
## True DLT probability function
Scenario1TrueDLTProbabilitiesMatrix <- cbind(c(Dados$Dose),
                                             c(Dados$Scenario1))
Scenario1TrueDLTProbabilitiesFunction <- function(dose){Scenario1TrueDLT-
ProbabilitiesMatrix[match(dose,
                                Scenario1TrueDLTProbabilitiesMatrix[, 1]),
2]}
#Run
time <- system.time(EWOCScenario1 <- simulate(designEWOC, nsim=1000,
                                              seed=1313, args=NULL,
                                              truth=Scenario1TrueDLTProba-
bilitiesFunction,
                                              mcmcOptions=mcmc.defini-
tions,
                                              parallel=FALSE))[3]
###################################
### (Repeat for scenarios 2 to 10)
###################################

##### Performance Analysis ########
##### SCENARIO 1 #####
#1. Percentage of correct selection (PCS) of the true MTD
print(EWOCScenario1)
#Code to analyse each trial individually
EWOCScenario1@data[[1]]
#Number of NAs - trials interrupted due to excessive toxicity
sum(is.na(EWOCScenario1@doses))
#Percentage of NAs - trials interrupted due to excessive toxicity
sum(is.na(EWOCScenario1@doses))/1000*100
#Number of doses selected as MTD in the simulated trials
table(EWOCScenario1@doses)
#Percentage of doses selected as the MTD
EWOC.Recommended.MTDs.Scenario1 <- table(EWOCScenario1@doses)/1000*100
EWOC.Recommended.MTDs.Scenario1

#2. Mean number (and percentage) of patients treated at the dose level
corresponding to the true MTD
#across the 1000 simulated trials
#Total number of patients dosed at each dose level
EWOC.Scenario1.x <- lapply(EWOCScenario1@data, function(data) data@x)
EWOC.Scenario1.x
#Creating a matrix combining the 1000 x objects
```

```
EWOC.Scenario1.x.matrix <- do.call(rbind, EWOC.Scenario1.x)
EWOC.Scenario1.x.matrix
#Replacing the doses by dose-levels 1 to 6
EWOC.Scenario1.x.matrix[EWOC.Scenario1.x.matrix==150] <- 1
EWOC.Scenario1.x.matrix[EWOC.Scenario1.x.matrix==200] <- 2
EWOC.Scenario1.x.matrix[EWOC.Scenario1.x.matrix==250] <- 3
EWOC.Scenario1.x.matrix[EWOC.Scenario1.x.matrix==300] <- 4
EWOC.Scenario1.x.matrix[EWOC.Scenario1.x.matrix==350] <- 5
EWOC.Scenario1.x.matrix[EWOC.Scenario1.x.matrix==400] <- 6
EWOC.Scenario1.x.matrix
#Function to change matrix to give the number of patients dosed at each
dose level.
#Each column is a dose level and each row is a simulated clinical trial
count.x.row <- function(row){
  x.counts <- table(row)
  counts <- integer(6)
  for (i in 1:6) {
    if (i %in% names(x.counts)) {
      counts[i] <- x.counts[[as.character(i)]]
    } else {
      counts[i] <- 0
    }
  }
  return(counts)
}
# Apply the function to each row of the EWOC.Scenario1.x.matrix
Dose.counts.EWOC.Scenario1.Matrix <- t(apply(EWOC.Scenario1.x.matrix, 1,
count.x.row))
Dose.counts.EWOC.Scenario1.Matrix
#n
EWOC.Scenario1.x.mean <-colMeans(Dose.counts.EWOC.Scenario1.Matrix)
EWOC.Scenario1.x.mean
#Percentage
EWOC.Scenario1.x.mean.percentage <-colMeans(Dose.counts.EWOC.Scenario1.Ma-
trix)/mean(rowSums(Dose.counts.EWOC.Scenario1.Matrix))*100
round(EWOC.Scenario1.x.mean.percentage, 2)

#3.A Average percentage of patients treated below the MTD
#N/A

#3.B    Mean number (and percentage) of patients treated above the true
MTD in the 1000 simulated trials
#n
EWOC.Scenario1.x.mean.above.true.MTD <-mean(rowSums(Dose.counts.EWOC.Sce-
nario1.Matrix[, 2:6]))
EWOC.Scenario1.x.mean.above.true.MTD
#percentage
EWOC.Scenario1.x.mean.above.true.MTD.percentage <-mean(row-
Sums(Dose.counts.EWOC.Scenario1.Matrix[, 2:6]))/mean(row-
Sums(Dose.counts.EWOC.Scenario1.Matrix))*100
round(EWOC.Scenario1.x.mean.above.true.MTD.percentage, 2)
```

```
#4. Mean number of DLTs observed (and percentage) for each dose-escalation
method and each scenario,
#across the 1000 simulated trials
EWOC.Scenario1.DLTs <- lapply(EWOCScenario1@data, function(data) data@y)
EWOC.Scenario1.DLTs
#Creating a matrix combining the 1000 y objects
EWOC.Scenario1.DLTs.matrix <- do.call(rbind, EWOC.Scenario1.DLTs)
EWOC.Scenario1.DLTs.matrix
#n
EWOC.Scenario1.mean.number.DLTs <-mean(rowSums(EWOC.Scenario1.DLTs.ma-
trix))
EWOC.Scenario1.mean.number.DLTs
#Percentage
EWOC.Scenario1.mean.number.DLTs.percentage <-EWOC.Scenario1.mean.num-
ber.DLTs/mean(rowSums(Dose.counts.EWOC.Scenario1.Matrix))*100
round(EWOC.Scenario1.mean.number.DLTs.percentage, 2)
###################################
### (Repeat for scenarios 2 to 10)
###################################

#############################################################################
#############################################################################
###### BLRM  ################################################################
#############################################################################
#############################################################################
#Installing and loading the crmPack package
install.packages("crmPack")
library(crmPack)
#############################################################################
##### BLRM SIMULATION: SCENARIOS 1-10 #####
# Dose range (dose-grid)
emptydata <- Data(doseGrid=seq(from=150, to=400, by=50))
# Initialize the BLRM model with bivariate normal prior
modelBLRM <- LogisticLogNormal(mean=c(-0.847, 0.381), cov=ma-
trix(c(2.015^2, 0, 0, 1.027^2),nrow=2),
                               refDose=400)
# Rule for selecting the next dose - target toxicity interval [0.28-0.38]
(i.e. [delta1; delta2]),
#and overdose control criterion, satisfied for the dose levels J with pos-
terior probability of
#exceeding delta2 less than alpha. P(phat(j)>delta2|Dk)<alpha
NextBestBLRM <- NextBestNCRM(target=c(0.28, 0.38),
                             overdose=c(0.38, 1),
                             maxOverdoseProb=0.25)
# Cohort-size
CohortSizeBLRM <- CohortSizeConst(size=3)
# Stopping rule - minimum of 30 patients treated
StoppingRulesBLRM <- StoppingMinPatients(nPatients=30)
# Dose-skipping no allowed
DoseIncrementsBLRM <- IncrementsNumDoseLevels(maxLevels=1)
```

```
# Initialize the design
designBLRM <- Design(model=modelBLRM,
                     nextBest=NextBestBLRM,
                     stopping=StoppingRulesBLRM,
                     increments=DoseIncrementsBLRM,
                     cohortSize=CohortSizeBLRM,
                     data=emptydata,
                     startingDose=150)
# MCMC methods - burn-in of 10000 iterations and then take every other it-
eration up to a collection of 50000 samples (JAGS)
mcmc.definitions <- McmcOptions(burnin=10000, step=2, samples=50000)
################################
#SCENARIO 1
## True DLT probability function
Scenario1TrueDLTProbabilitiesMatrix <- cbind(c(Dados$Dose),
                                             c(Dados$Scenario1))
Scenario1TrueDLTProbabilitiesFunction <- function(dose){Scenario1TrueDLT-
ProbabilitiesMatrix[match(dose,
                                                    Scenario1TrueDLT-
ProbabilitiesMatrix[, 1]), 2]}
#Run
time <- system.time(BLRMScenario1 <- simulate(designBLRM, nsim=1000,
                                              seed=1313, args=NULL,
                                              truth=Scenario1TrueDLTProba-
bilitiesFunction,
                                              mcmcOptions=mcmc.defini-
tions,
                                              parallel=FALSE))[3]
###################################
### (Repeat for scenarios 2 to 10)
###################################
##### Performance Analysis ########
##### SCENARIO 1 #####
#1. Proportion of correct selection (PCS) of the true MTD
print(BLRMScenario1)
#Analyse each trial individually
BLRMScenario1@data[[1]]
#Number of NAs - trials interrupted due to excessive toxicity
sum(is.na(BLRMScenario1@doses))
#Proportion of NAs - trials interrupted due to excessive toxicity
sum(is.na(BLRMScenario1@doses))/1000*100
#Number of doses selected as MTD in the simulated trials
table(BLRMScenario1@doses)
#Proportion of doses selected as the MTD
BLRM.Recommended.MTDs.Scenario1 <- table(BLRMScenario1@doses)/1000*100
BLRM.Recommended.MTDs.Scenario1



#2. Mean number (and percentage) of patients treated at the dose level
corresponding to the true MTD
#across the 1000 simulated trials
```

```
#Total number of patients dosed at each dose level
BLRM.Scenario1.x <- sapply(BLRMScenario1@data, function(data) data@x)
BLRM.Scenario1.x
#Function to replace blank values by 0
Blanks.BLRM.Scenario1.x <- lapply(BLRM.Scenario1.x, function(vector) {
  if (length(vector) < 30) {
    c(vector, rep(0, 30 - length(vector)))
  } else {
    vector
  }
})
#Creating a matrix where the blank values are replaced by 0
BLRM.Scenario1.x.matrix <- do.call(rbind, Blanks.BLRM.Scenario1.x)
BLRM.Scenario1.x.matrix
#Replacing the doses by dose-levels 1 to 6
BLRM.Scenario1.x.matrix[BLRM.Scenario1.x.matrix==150] <- 1
BLRM.Scenario1.x.matrix[BLRM.Scenario1.x.matrix==200] <- 2
BLRM.Scenario1.x.matrix[BLRM.Scenario1.x.matrix==250] <- 3
BLRM.Scenario1.x.matrix[BLRM.Scenario1.x.matrix==300] <- 4
BLRM.Scenario1.x.matrix[BLRM.Scenario1.x.matrix==350] <- 5
BLRM.Scenario1.x.matrix[BLRM.Scenario1.x.matrix==400] <- 6
BLRM.Scenario1.x.matrix
#Function to change matrix to give the number of patients dosed at each
dose level.
#Each column is a dose level and each row is a simulated clinical trial
count.x.row <- function(row) {
  x.counts <- table(row)
  counts <- integer(6)
  for (i in 1:6) {
    if (i %in% names(x.counts)) {
      counts[i] <- x.counts[[as.character(i)]]
    } else {
      counts[i] <- 0
    }
  }
  return(counts)
}
# Apply the function to each row of the BLRM.Scenario1.x.matrix
Dose.counts.BLRM.Scenario1.Matrix <- t(apply(BLRM.Scenario1.x.matrix, 1,
count.x.row))
Dose.counts.BLRM.Scenario1.Matrix
#n
BLRM.Scenario1.x.mean <-colMeans(Dose.counts.BLRM.Scenario1.Matrix)
BLRM.Scenario1.x.mean
#Percentage
BLRM.Scenario1.x.mean.percentage <-colMeans(Dose.counts.BLRM.Scenario1.Ma-
trix)/mean(rowSums(Dose.counts.BLRM.Scenario1.Matrix))*100
round(BLRM.Scenario1.x.mean.percentage, 2)

#3.A Average percentage of patients treated below the MTD
#N/A
```

```
#3.B    Mean number (and percentage) of patients treated above the true
MTD in the 1000 simulated trials
#n
BLRM.Scenario1.x.mean.above.true.MTD <-mean(rowSums(Dose.counts.BLRM.Sce-
nario1.Matrix[, 2:6]))
BLRM.Scenario1.x.mean.above.true.MTD
#percentage
BLRM.Scenario1.x.mean.above.true.MTD.percentage <-mean(row-
Sums(Dose.counts.BLRM.Scenario1.Matrix[, 2:6]))/mean(row-
Sums(Dose.counts.BLRM.Scenario1.Matrix))*100
round(BLRM.Scenario1.x.mean.above.true.MTD.percentage, 2)

#4. Mean number of DLTs observed (and percentage) for each dose-escalation
method and each scenario,
#across the 1000 simulated trials
#Total number of DLTs observed
BLRM.Scenario1.DLTs <- sapply(BLRMScenario1@data, function(data) data@y)
BLRM.Scenario1.DLTs
#Function to replace blank values by 0
Blanks.BLRM.Scenario1.DLTs <- lapply(BLRM.Scenario1.DLTs, function(vector)
{
  if (length(vector) < 30) {
    c(vector, rep(0, 30 - length(vector)))
  } else {
    vector
  }})
BLRM.Scenario1.DLTs.matrix <- do.call(rbind, Blanks.BLRM.Scenario1.DLTs)
BLRM.Scenario1.DLTs.matrix
#n
BLRM.Scenario1.mean.number.DLTs <-mean(rowSums(BLRM.Scenario1.DLTs.ma-
trix))
BLRM.Scenario1.mean.number.DLTs
#Percentage
BLRM.Scenario1.mean.number.DLTs.percentage <-BLRM.Scenario1.mean.num-
ber.DLTs/mean(rowSums(Dose.counts.BLRM.Scenario1.Matrix))*100
round(BLRM.Scenario1.mean.number.DLTs.percentage, 2)
###################################
### (Repeat for scenarios 2 to 10)
###################################

#########################################################################
#########################################################################
###### mTPI  ############################################################
#########################################################################
#########################################################################
#Adapted from the MD Anderson R Code (see references)
#########################################################################
#Definition of the PAVA function
## PAVA is the pool adjacent violators algorithm to perform isotonic
#transformation for the posterior means
```

```
pava <- function (x, wt = rep(1, length(x)))
{
  n <- length(x)
  if (n <= 1)
    return(x)
  if (any(is.na(x)) || any(is.na(wt))) {
    stop("Missing values in 'x' or 'wt' not allowed")
  }
  lvlsets <- (1:n)
  repeat {
    viol <- (as.vector(diff(x)) < 0)
    if (!(any(viol)))
      break
    i <- min((1:(n - 1))[viol])
    lvl1 <- lvlsets[i]
    lvl2 <- lvlsets[i + 1]
    ilvl <- (lvlsets == lvl1 | lvlsets == lvl2)
    x[ilvl] <- sum(x[ilvl] * wt[ilvl])/sum(wt[ilvl])
    lvlsets[ilvl] <- lvl1
  }
  x
}
## betavar computes variances of beta distributions
betavar<-function(a,b){
  resp <- a*b/((a+b)^2*(a+b+1))
  return(resp)
}
#### mTPI SIMULATION: SCENARIOS 1-10 #####
################################################################
#simulation configuration
pT<-.33 #Target toxicity level
D<-6 #Six dose levels
sampsize <- 30 #Máximum sample size of 30 patients (10 cohorts)
csize <-3 #cohort size=3
startdose <- 1 #start dose is d1
simN<-1000  # number of simulations
#Equivalence interval de difed as [pT-eps1,pT+eps2]
#esp1=esp2=0.05
eps1<-0.05
eps2<-0.05
#Parameters of the beta prior, beta(1,1)
a<-1; b<-1;

#Dose Exclusion Rule termination rule
xi<-.95  ## the cutoff probability for the early termination and dose ex-
clusion rule

#Additional rule in the MD Anderson code
xi2 <- 1 ## if the last dose is much lower probability than the MTD, the
trial will be terminated
#and no dose selected - Disabled (xi2=1)
```

```
################################
#SCENARIO 1
set.seed(1313)
for(sc in c(1:1)){
  if (sc==1){
    p <- Dados$Scenario1
  }
}

datan1<-matrix(rep(0,simN*D),ncol=D)
datax1<-matrix(rep(0,simN*D),ncol=D)

rez<-rep(0,simN)

# Initializing simulation
for(sim in 1:simN){

  x<-rep(0,D); n<-rep(0,D)
  pa<-rep(0, D); pb<-rep(0,D)
  q <- rep(0,3)
  d<-startdose; st<-0; nodose<-0; maxdose<-1; toxdose<-D+1; seldose<-0;
mindose <- D

  while(st==0){  ## st = 1 indicates the trial must be terminated
    maxdose<-max(maxdose, d)
    mindose <- min(mindose, d)

    ### generate random toxicity response
    xx <- 0
    for(i in 1:csize){
      ttt <- runif(1)
      if(ttt < p[d]) xx <- xx + 1
    }

    x[d] <- x[d] + xx; n[d] <- n[d] + csize

    #### Update posterior beta distribution
    pa[d]<-x[d]+a; pb[d]<-n[d]-x[d]+b
    pa[d+1]<-x[d+1]+a; pb[d+1]<-n[d+1]-x[d+1]+b

    ###Compute the indicator T_{i+1} to see if the next dose is too toxic
    if(d<D){
      temp<-1-pbeta(pT, 1+x[d+1], 1+n[d+1]-x[d+1])
      if(temp>xi) {tt<-1; toxdose<-d+1} else {tt<-0}
    }



    ##Compute the UPM for three intervals defined by the equivalence in-
terval
    q[1] <- (1-pbeta(eps2+pT, pa[d], pb[d]))/(1-eps2-pT)
    q[2] <- (pbeta(eps2+pT, pa[d], pb[d]) - pbeta(pT-eps1, pa[d],
```

```
pb[d]))/(eps2+eps1)
    q[3] <- (pbeta(pT-eps1, pa[d], pb[d])/(pT-eps1))*(1-tt)


    ### implement the dose-assignment rules based on the UPM
    if(d==1){
      ## if the first dose is too toxic, the trial will be terminated
      if((1-pbeta(pT, 1+x[d], 1+n[d]-x[d]))>xi){st=1; nodose<-1;}
      else{
        #if((pbeta(pT, 1+x[d], 1+n[d]-x[d])>xi2)){maxdose<-d+1}
        #else{
        if((q[2]>q[1])&&(q[2]>q[3])){d<-d}
        if((q[3]>q[1])&&(q[3]>q[2])){d<-d+1}
        #}
      }
    }
    else{
      if(d==D){
        ## if the last dose is much lower than the MTD, the trial will be
terminated and no dose selected - disabled (xi2=1)
        if(pbeta(pT, 1+x[d], 1+n[d]-x[d])>xi2){st<-1; nodose<-1;}
        else{
          if((q[1]>q[2])&&(q[1]>q[3])){d<-d-1}
          if((q[2]>q[1])&&(q[2]>q[3])){d<-d}
        }
      }
      else{
        if((d>1)&&(d<D)){
          #if(pbeta(pT, pa[d], pb[d])>xi){maxdose<-d+1}
          #else{
          if((q[1]>q[2])&&(q[1]>q[3])){d<-d-1}
          if((q[2]>q[1])&&(q[2]>q[3])){d<-d}
          if((q[3]>q[1])&&(q[3]>q[2])){d<-d+1}
          #}
        }
      }
    }
    total<-sum(n)
    if(total >= sampsize){st<-1}
  }

  ### compute the posterior mean from the beta distribution
  if(nodose==0){
    tdose<-min(maxdose, toxdose-1)
    ##tdose <- toxdose - maxdose

    ##pp <- rep(-100, D)
    pp<-rep(-100,tdose)

    pp.var<-rep(0, tdose)
    ##pp.var <- rep(0, D)
```

```
    ##for(i in maxdose:(toxdose-1)){
    for(i in 1:tdose){
      pp[i] <- (x[i]+.005)/(n[i]+.01); pp.var[i] <- betavar(x[i]+0.005,
n[i]-x[i]+0.005) ### here adding 0.005 is to use beta(0.005, 0.005) for
estimating the MTD, which is different from the dose-finding.
    }

    pp<-pava(pp, wt=1/pp.var)
    ##pp[maxdose:(toxdose-1)]<-pava(pp[maxdose:(toxdose-1)],
wt=1/pp.var[maxdose:(toxdose-1)])  ## perform the isotonic transformation
using PAVA with weights being posterior variances

    ##for(i in maxdose:(toxdose-1)){
    for(i in 2:tdose){
      pp[i] <- pp[i] + i*1E-10 ## by adding an increasingly small number
to tox prob at higher doses, it will break the ties and make the lower
dose level the MTD if the ties were larger than pT or make the higher dose
level the MTD if the ties are smaller than pT
    }
    seldose<-c(mindose:tdose)[sort(abs(pp[mindose:tdose]-pT), index.re-
turn=T)$ix[1]]
    ##seldose is the final MTD that is selected based on the order-trans-
formed posterior means
  }
  rez[sim] <- seldose;
  for(i in 1:D){
    datan1[sim,i] <- n[i]
    datax1[sim,i] <- x[i]
  }
}
##rez[is.na(rez)]<-0
aaa1<-rep(0,D)

for(i in 1:D){
  aaa1[i] <- sum(rez==i)/simN ### aaa is the proportion of the dose se-
lected as the MTD
}
###################################
### (Repeat for scenarios 2 to 10)
###################################
##### Performance Analysis ########
##### SCENARIO 1 #####
#1. Proportion of correct selection (PCS) of the true MTD
mTPI.Recommended.MTDs.Scenario1 <-c(aaa1, 1-sum(aaa1))*100
mTPI.Recommended.MTDs.Scenario1
round(mTPI.Recommended.MTDs.Scenario1, 2)

#2. Mean number (and percentage) of patients treated at the dose level
corresponding to the true MTD
#across the 1000 simulated trials
```

```
#n
mTPI.Scenario1.x.mean <-colMeans(datan1)
mTPI.Scenario1.x.mean
#Percentage (%)
mTPI.Scenario1.x.mean.percentage <- round(colMeans(datan1)/mean(row-
Sums(datan1))*100, 2)
mTPI.Scenario1.x.mean.percentage
#Looking at the trials that were interrupted due to excess toxicity
Interrupted.Trials.Scenario1 <- which(rowSums(datan1) < 30)
Interrupted.Trials.Scenario1
datan1[3,]
datax1[3,]

#3.A Average percentage of patients treated below the MTD
#N/A

#3.B    Mean number (and percentage) of patients treated above the true
MTD in the 1000 simulated trials
#n
mTPI.Scenario1.x.mean.above.true.MTD <- sum(colMeans(datan1)[2:6])
mTPI.Scenario1.x.mean.above.true.MTD
#Percentage
mTPI.Scenario1.x.mean.above.true.MTD.percentage <-
round(sum(colMeans(datan1)[2:6])/mean(rowSums(datan1))*100,2)
mTPI.Scenario1.x.mean.above.true.MTD.percentage
#4. Mean number of DLTs observed (and percentage) for each dose-escalation
method and each scenario,
#across the 1000 simulated trials
#n
mTPI.Scenario1.mean.number.DLTs <-round(mean(rowSums(datax1)),3)
mTPI.Scenario1.mean.number.DLTs
#Percentage
mTPI.Scenario1.mean.number.DLTs.percentage <-round(mean(row-
Sums(datax1))/mean(rowSums(datan1))*100, 2)
mTPI.Scenario1.mean.number.DLTs.percentage
###################################
### (Repeat for scenarios 2 to 10)
###################################

#########################################################################
#########################################################################
###### BOIN
#########################################################################
#########################################################################
#########################################################################
#Installing and loading the BOIN R package
install.packages("BOIN")
library(BOIN)
#########################################################################
### mTPI SIMULATION: SCENARIOS 1-10 #####
#Definition of the target toxicity level phi=0.33
```

```
phi=0.33
#Safety measures implemented in BOIN:
#current dose dcur cannot exceed the target toxicity probability phi on
95%,
#P(pcur>phi|ycur,ncur)>95%,ncur≥3. If this happens, and we have at least
three patients treated
#at the current dose level, dose levels dcur and above are eliminated from
the trial.
#The trial is terminated if the first dose level is eliminated.
################################
#### BOIN SIMULATION: SCENARIOS 1-10 #####
#SCENARIO 1
BOIN.Scenario1 <- get.oc(target=phi, p.true=Dados$Scenario1, ncohort=10,
cohortsize=3,
                         n.earlystop=100, ntrial=1000, startdose = 1, ti-
tration=FALSE,
                         p.saf=0.6*phi, p.tox=1.4*phi, cutoff.eli=0.95,
seed = 1313)
###################################
### (Repeat for scenarios 2 to 10)
###################################
##### Performance Analysis ########
##### SCENARIO 1 #####
#1. Percentage of correct selection (PCS) of the true MTD
summary(BOIN.Scenario1)

#2. Mean number (and percentage) of patients treated at the dose level
corre-sponding to the true MTD
#across the 1000 simulated trials
#n
BOIN.Scenario1.x.mean <-BOIN.Scenario1$npatients
BOIN.Scenario1.x.mean
#Percentage
BOIN.Scenario1.x.mean.percentage <-BOIN.Scenario1$npatients/BOIN.Sce-
nario1$totaln*100
BOIN.Scenario1.x.mean.percentage
round(BOIN.Scenario1.x.mean.percentage, 2)

#3.A Average percentage of patients treated below the MTD
#N/A

#3.B    Mean number (and percentage) of patients treated above the true
MTD in the 1000 simulated trials
#n
BOIN.Scenario1$npatients

BOIN.Scenario1.x.mean.above.true.MTD <- sum(BOIN.Scenario1$npatients[2:6])
BOIN.Scenario1.x.mean.above.true.MTD
#Percentage
BOIN.Scenario1.x.mean.above.true.MTD.percentage <-sumfun(BOIN.Sce-
nario1$npatients, 2,6)/BOIN.Scenario1$totaln*100
```

```
BOIN.Scenario1.x.mean.above.true.MTD.percentage

#4. Mean number of DLTs observed (and percentage) for each dose-escalation
method and each scenario,
#across the 1000 simulated trials
#n
BOIN.Scenario1.mean.number.DLTs <-BOIN.Scenario1$totaltox
BOIN.Scenario1.mean.number.DLTs
#Percentage
BOIN.Scenario1.mean.number.DLTs.percentage <-BOIN.Scenario1$to-
taltox/BOIN.Scenario1$totaln*100
BOIN.Scenario1.mean.number.DLTs.percentage
###################################
### (Repeat for scenarios 2 to 10)
###################################
```

B.2 R Code to Run the Replication of the Deflexifol™ Clinical Trial Using the CRM, EWOC, BLRM, mTPI and BOIN Designs

```
### WinBUGS models ###
#CRM - CRMmodelFINAL
model {
  #Prior on log(theta)
  #Precision 1/(1.34^2)=0.5569169
  logtheta ~ dnorm(0, 0.5569169)

  # Calculate theta from log(theta)
  theta <- exp(logtheta)

  # Likelihood - Bernoulli trial
  for (i in 1:n) {
    y[i] ~ dbern(p[i])
    #Power model
    p[i]<-pow(x[i],theta)
  }

  # for each dose
  for (k in 1:N) {
    p.pred[k]<- pow(xk[k],theta)
  }
}
#EWOC - EWOCmodelFINAL
model
```

```
{
  #Doses available in the trial: 375, 425, 475, 525, 575. Xmin defined as
325 and Xmax defined as 625
  #phi=0.25
  theta0 <- ((gamma * logit(rho0) - 325 * logit(0.25))/(gamma - 325))
  theta1 <- ((logit(0.25)-logit(rho0))/(gamma-325))

  for (i in 1:n) {
    # Linear regression on logit
    logit(p[i]) <- theta0 + theta1*x[i]
    # Likelihood function for each data point
    y[i] ~ dbern(p[i])
  }

  # Prior distributions
  gamma ~ dunif(325, 625)
  rho0 ~ dunif(0, 0.25)
}
}
#BLRM
#Cohort1 - BLRMmodelFINALfirstcohort
model{
  # prior covariance matrix
  cova[1,1] <- Prior[3]*Prior[3]
  cova[2,2] <- Prior[4]*Prior[4]
  cova[1,2] <- Prior[3]*Prior[4]*Prior[5]
  cova[2,1] <- cova[1,2]
  prec[1:2,1:2] <- inverse(cova[,])
  log.alphabeta[1:2] ~ dmnorm(Prior[1:2],prec[1:2,1:2])
  alphabeta[1] <- exp(log.alphabeta[1])
  alphabeta[2] <- exp(log.alphabeta[2])

  # sampling model
  for (j in 1:Ncohorts){
    logit(Pr.Tox1[j]) <- log.alphabeta[1]+alphabeta[2]*log(DosesAdm/DoseRe
f)
    Ntox ~ dbin(Pr.Tox1[j],Npat)
  }

  # for each dose: probabilities of toxicity, category probabilities, risk
s
  for (i in 1:Ndoses) {
    lin[i]<-log.alphabeta[1]+alphabeta[2]*log(Doses[i]/DoseRef)
    logit(Pr.Tox[i]) <- lin[i]
    Pr.Cat[i,1] <- step(Pint[1]-Pr.Tox[i])
    Pr.Cat[i,2] <- step(Pint[2]-Pr.Tox[i])-step(Pint[1]-Pr.Tox[i])
    Pr.Cat[i,3] <- step(Pint[3]-Pr.Tox[i])-step(Pint[2]-Pr.Tox[i])
    Pr.Cat[i,4] <- step(1-Pr.Tox[i])-step(Pint[3]-Pr.Tox[i])
    Risk1[i] <- inprod(Pr.Cat[i,],LossFunction1[1:4])
    Risk2[i] <- inprod(Pr.Cat[i,],LossFunction2[1:4])
    Risk3[i] <- inprod(Pr.Cat[i,],LossFunction3[1:4])
  }
```

```
}
#Other cohorts - BLRMmodelFINAL
model{
  # prior covariance matrix
  cova[1,1] <- Prior[3]*Prior[3]
  cova[2,2] <- Prior[4]*Prior[4]
  cova[1,2] <- Prior[3]*Prior[4]*Prior[5]
  cova[2,1] <- cova[1,2]
  prec[1:2,1:2] <- inverse(cova[,])
  log.alphabeta[1:2] ~ dmnorm(Prior[1:2],prec[1:2,1:2])
  alphabeta[1] <- exp(log.alphabeta[1])
  alphabeta[2] <- exp(log.alphabeta[2])
  # sampling model
  for (j in 1:Ncohorts){
    logit(Pr.Tox1[j]) <- log.alphabeta[1]+alphabeta[2]*log(DosesAdm[j]/Dos
eRef)
    Ntox[j] ~ dbin(Pr.Tox1[j],Npat[j])
  }
  # for each dose: probabilities of toxicity, category probabilities, risk
s
  for (i in 1:Ndoses) {
    lin[i]<-log.alphabeta[1]+alphabeta[2]*log(Doses[i]/DoseRef)
    logit(Pr.Tox[i]) <- lin[i]
    Pr.Cat[i,1] <- step(Pint[1]-Pr.Tox[i])
    Pr.Cat[i,2] <- step(Pint[2]-Pr.Tox[i])-step(Pint[1]-Pr.Tox[i])
    Pr.Cat[i,3] <- step(Pint[3]-Pr.Tox[i])-step(Pint[2]-Pr.Tox[i])
    Pr.Cat[i,4] <- step(1-Pr.Tox[i])-step(Pint[3]-Pr.Tox[i])
    Risk1[i] <- inprod(Pr.Cat[i,],LossFunction1[1:4])
    Risk2[i] <- inprod(Pr.Cat[i,],LossFunction2[1:4])
    Risk3[i] <- inprod(Pr.Cat[i,],LossFunction3[1:4])
  }
}
##############################################################################
##############################################################################
##############################################################################
###################### CRM ############
##############################################################################
##############################################################################
library("R2WinBUGS")
dir.bugs <- "C:/Programas/WinBUGS14/"
#### Deflexifol clinical trial (bolus schedule)####
# Doses x: 375, 425, 475, 525, 575 mg/m2
library(dfcrm)
#Obtaining the Skeleton through the method described by Lee & Cheung (2009
)
#(getprior function in the dfcrm package)
#halfwidth=0.06 was chosen, following Lee & Cheung (2009) simulations, and
nu was selected as
#corresponding to the 3rd dose level
crm.skeleton.CT1 <- getprior(halfwidth=0.06, target=0.25, nu=3, nlevel=5,
model="empiric")
crm.skeleton.CT1
```

```
#Dose labels: 0.0615790 0.1400497 0.2500000 0.3761963 0.5018492
round(crm.skeleton.CT1,3)
#Dose labels: 0.062 0.140 0.250 0.376 0.502
#COHORT1 - Three patients treated at 375 and no DLTs
#data
dataCRMcohort1 <- list(y=c(0, 0, 0),
                       x=c(0.062,0.062,0.062),
                       n=3, N=5,
                       xk=c(0.062, 0.140, 0.250, 0.376, 0.502))

CRMpowermodelcohort1 <-bugs(dataCRMcohort1, inits=NULL, parameters.to.save
=c("theta","p.pred"),
                            "CRMmodelFINAL.txt", n.chains=3,
                            n.iter=20000, n.burnin=5000, n.thin=1, bugs.di
rectory=dir.bugs) #debug=T
print(CRMpowermodelcohort1, digits.summary = 3)
#Dose that has posterior probability of DLT closer to the target toxicity
level is the dose level 4 (525)
#This would be the dose level administered to the next cohort of patients
if we were not restricting dose
# escalation to one level at a time. Dose will be escalated to the next hi
gher level
##############################################################################
#COHORT2 - Three patients treated at 425 with no DLTs
#data
dataCRMcohort2 <- list(y=c(0,0,0,0,0,0),
                       x=c(0.062,0.062,0.062,0.140,0.140,0.140),
                       n=6, N=5,
                       xk=c(0.062, 0.140, 0.250, 0.376, 0.502))
CRMpowermodelcohort2 <-bugs(dataCRMcohort2, inits=NULL, parameters.to.save
=c("theta","p.pred"),
                            "CRMmodelFINAL.txt", n.chains=3,
                            n.iter=20000, n.burnin=5000, n.thin=1, bugs.di
rectory=dir.bugs) #debug=T
print(CRMpowermodelcohort2, digits.summary = 3)
#Dose that has posterior probability of DLT closer to the target toxicity
level is the dose level 5 (575)
#This would be the dose level administered to the next cohort of patients
if we were not restricting dose
# escalation to one level at a time. Dose will be escalated to the next hi
gher level
##############################################################################
#COHORT3 - Three patients treated at 475 with no DLTs
#data
dataCRMcohort3 <- list(y=c(0,0,0,0,0,0,0,0,0),
                       x=c(0.062,0.062,0.062,0.140,0.140,0.140,0.250,0.250
,0.250),
                       n=9, N=5,
                       xk=c(0.062, 0.140, 0.250, 0.376, 0.502))
CRMpowermodelcohort3 <-bugs(dataCRMcohort3, inits=NULL, parameters.to.save
=c("theta","p.pred"),
                            "CRMmodelFINAL.txt", n.chains=3,
```

```
                             n.iter=20000, n.burnin=5000, n.thin=1, bugs.di
rectory=dir.bugs) #debug=T
print(CRMpowermodelcohort3, digits.summary = 3)
#Dose that has posterior probability of DLT closer to the target toxicity
level is the dose level  5 (575)
#This would be the dose level administered to the next cohort of patients
if we were not restricting dose
# escalation to one level at a time. Dose will be escalated to the next hi
gher level
##############################################################################
#COHORT4 - Six patients treated at 525 and no DLTs
#data
dataCRMcohort4 <- list(y=c(0,0,0,0,0,0,0,0,0,0,0,0,0,0,0),
                       x=c(0.062,0.062,0.062,0.140,0.140,0.140,0.250,0.250
,0.250,0.376,0.376,0.376,0.376,0.376,0.376),
                       n=15, N=5,
                       xk=c(0.062, 0.140, 0.250, 0.376, 0.502))
CRMpowermodelcohort4 <-bugs(dataCRMcohort4, inits=NULL, parameters.to.save
=c("theta","p.pred"),
                            "CRMmodelFINAL.txt", n.chains=3,
                            n.iter=20000, n.burnin=5000, n.thin=1, bugs.di
rectory=dir.bugs) #debug=T
print(CRMpowermodelcohort4, digits.summary = 3)
#Dose that has posterior probability of DLT closer to the target toxicity
level is the dose level number 5 (575)
#This would be the dose level administered to the next cohort of patients
if we were not restricting dose
# escalation to one level at a time. Dose will be escalated to the next hi
gher level
##############################################################################
#COHORT5 - Four patients treated at 575 and two DLTs
#data
dataCRMcohort5 <- list(y=c(0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,0,1),
                       x=c(0.062,0.062,0.062,0.140,0.140,0.140,0.250,0.250
,0.250,0.376,0.376,0.376,0.376,
                           0.376,0.376,0.502,0.502,0.502,0.502),
                       n=19, N=5,
                       xk=c(0.062, 0.140, 0.250, 0.376, 0.502))
CRMpowermodelcohort5 <-bugs(dataCRMcohort5, inits=NULL, parameters.to.save
=c("theta","p.pred"),
                            "CRMmodelFINAL.txt", n.chains=3,
                            n.iter=20000, n.burnin=5000, n.thin=1, bugs.di
rectory=dir.bugs) #debug=T
print(CRMpowermodelcohort5, digits.summary = 3)
#Dose that has posterior probability of DLT closer to the target toxicity
level is the dose level number 5 (575)
#This would be the dose level administered to the next cohort of patients
if we were not restricting dose
# escalation to one level at a time. Dose will be escalated to the next lo
wer level
CRM.cohort5.vector.post.prob <-c(0.011, 0.029, 0.067, 0.131, 0.224)
which.min(c((CRM.cohort5.vector.post.prob-0.25)^2))
```

```
##############################################################################
##############################################################################
##############################################################################
##################### EWOC ############
##############################################################################
##############################################################################
#### Deflexifol clinical trial (bolus schedule)####
# Doses x: 375, 425, 475, 525, 575 mg/m2
# xmin was defined as xmin=325 and xmax=625

#COHORT1 - Three patients treated at 375 and no DLTs
#data
dataEWOCcohort1 <- list(y=c(0, 0, 0), x=c(375, 375, 375), n=3)

#inits
#To initiate we assumed rho0 with probability of DLT as 0.05 and
#that the MTD is at the middle dose level, 475 mg/mg2.
initsEWOCmodel <-function(){list(rho0=0.05, gamma=475)}

#WinBugs code to run after the FIRST cohort outcomes are known
EWOCmodelcohort1 <-bugs(dataEWOCcohort1, inits=initsEWOCmodel, parameters.
to.save =c("gamma","rho0"),
                        "EWOCmodelFINAL.txt", n.chains=3,
                        n.iter=10000, n.burnin=2000, n.thin=1, bugs.direct
ory=dir.bugs) #debug=T
print(EWOCmodelcohort1, digits.summary = 3)
#alpha quantile for the marginal posterior cumulative distribution of gamm
a:
#x=433
#Posterior probability of toxicity for each dose that minimizes the differ
ence to alpha (through a quadratic loss)
gamma.samples.EWOC.cohort1<-EWOCmodelcohort1$sims.list$gamma
gamma.samples.EWOC.cohort1.375<-sum(gamma.samples.EWOC.cohort1<=375)/lengt
h(gamma.samples.EWOC.cohort1)
gamma.samples.EWOC.cohort1.425<-sum(gamma.samples.EWOC.cohort1<=425)/lengt
h(gamma.samples.EWOC.cohort1)
gamma.samples.EWOC.cohort1.475<-sum(gamma.samples.EWOC.cohort1<=475)/lengt
h(gamma.samples.EWOC.cohort1)
gamma.samples.EWOC.cohort1.525<-sum(gamma.samples.EWOC.cohort1<=525)/lengt
h(gamma.samples.EWOC.cohort1)
gamma.samples.EWOC.cohort1.575<-sum(gamma.samples.EWOC.cohort1<=575)/lengt
h(gamma.samples.EWOC.cohort1)
vector.of.posterior.probabilities.for.each.dose.EWOC.cohort1 <-c(gamma.sam
ples.EWOC.cohort1.375,gamma.samples.EWOC.cohort1.425,
                                                                gamma.sam
ples.EWOC.cohort1.475,gamma.samples.EWOC.cohort1.525,
                                                                gamma.sam
ples.EWOC.cohort1.575)
round(vector.of.posterior.probabilities.for.each.dose.EWOC.cohort1,3)
#Quadractic loss
(vector.of.posterior.probabilities.for.each.dose.EWOC.cohort1-0.25)^2
#0.0351406267 0.0007177934 0.0250694444 0.1230547934 0.2991634184
```

```
which.min(c((vector.of.posterior.probabilities.for.each.dose.EWOC.cohort1-
0.25)^2))
#Dose level number 2 (425) is the dose that minimizes the difference. Dose
escalated to the next higher level
##############################################################################
#COHORT2 - Three patients treated at 425 with no DLTs
#data
dataEWOCcohort2 <- list(y=c(0,0,0,0,0,0), x=c(375, 375, 375, 425, 425, 425
), n=6)

#WinBugs code to run after the SECOND cohort outcomes are known
EWOCmodelcohort2 <-bugs(dataEWOCcohort2, inits=initsEWOCmodel, parameters.
to.save =c("gamma","rho0"),
                        "EWOCmodelFINAL.txt", n.chains=3,
                        n.iter=10000, n.burnin=2000, n.thin=1, bugs.direct
ory=dir.bugs) #debug=T
print(EWOCmodelcohort2, digits.summary = 3)
#alpha quantile for the marginal posterior cumulative distribution of gamm
a:
#x=463
#Posterior probability of toxicity for each dose that minimizes the differ
ence to alpha (through a quadratic loss)
gamma.samples.EWOC.cohort2<-EWOCmodelcohort2$sims.list$gamma
gamma.samples.EWOC.cohort2.375<-sum(gamma.samples.EWOC.cohort2<=375)/lengt
h(gamma.samples.EWOC.cohort2)
gamma.samples.EWOC.cohort2.425<-sum(gamma.samples.EWOC.cohort2<=425)/lengt
h(gamma.samples.EWOC.cohort2)
gamma.samples.EWOC.cohort2.475<-sum(gamma.samples.EWOC.cohort2<=475)/lengt
h(gamma.samples.EWOC.cohort2)
gamma.samples.EWOC.cohort2.525<-sum(gamma.samples.EWOC.cohort2<=525)/lengt
h(gamma.samples.EWOC.cohort2)
gamma.samples.EWOC.cohort2.575<-sum(gamma.samples.EWOC.cohort2<=575)/lengt
h(gamma.samples.EWOC.cohort2)
vector.of.posterior.probabilities.for.each.dose.EWOC.cohort2 <-c(
  gamma.samples.EWOC.cohort2.375,gamma.samples.EWOC.cohort2.425,
  gamma.samples.EWOC.cohort2.475,gamma.samples.EWOC.cohort2.525,
  gamma.samples.EWOC.cohort2.575)
round(vector.of.posterior.probabilities.for.each.dose.EWOC.cohort2,3)
#Quadractic loss
(vector.of.posterior.probabilities.for.each.dose.EWOC.cohort2-0.25)^2
#0.052957516 0.018609507 0.002272111 0.069432250 0.252799460
which.min(c((vector.of.posterior.probabilities.for.each.dose.EWOC.cohort2-
0.25)^2))
#Dose level number 3 (475) is the dose that minimizes the difference. Dose
escalated to the next higher level
##############################################################################
#COHORT3 - Three patients treated at 475 with no DLTs
#data
dataEWOCcohort3 <- list(y=c(0,0,0,0,0,0,0,0,0), x=c(375,375,375,425,425,42
5,475,475,475), n=9)

#WinBugs code to run after the THIRD cohort outcomes are known
```

```
EWOCmodelcohort3 <-bugs(dataEWOCcohort3, inits=initsEWOCmodel, parameters.
to.save =c("gamma","rho0"),
                        "EWOCmodelFINAL.txt", n.chains=3,
                        n.iter=10000, n.burnin=2000, n.thin=1, bugs.direct
ory=dir.bugs) #debug=T
print(EWOCmodelcohort3, digits.summary = 3)
#alpha quantile for the marginal posterior cumulative distribution of gamm
a:
#x=494.28
#Posterior probability of toxicity for each dose that minimizes the differ
ence to alpha (through a quadratic loss)
gamma.samples.EWOC.cohort3<-EWOCmodelcohort3$sims.list$gamma
gamma.samples.EWOC.cohort3.375<-sum(gamma.samples.EWOC.cohort3<=375)/lengt
h(gamma.samples.EWOC.cohort3)
gamma.samples.EWOC.cohort3.425<-sum(gamma.samples.EWOC.cohort3<=425)/lengt
h(gamma.samples.EWOC.cohort3)
gamma.samples.EWOC.cohort3.475<-sum(gamma.samples.EWOC.cohort3<=475)/lengt
h(gamma.samples.EWOC.cohort3)
gamma.samples.EWOC.cohort3.525<-sum(gamma.samples.EWOC.cohort3<=525)/lengt
h(gamma.samples.EWOC.cohort3)
gamma.samples.EWOC.cohort3.575<-sum(gamma.samples.EWOC.cohort3<=575)/lengt
h(gamma.samples.EWOC.cohort3)
vector.of.posterior.probabilities.for.each.dose.EWOC.cohort3 <-c(
  gamma.samples.EWOC.cohort3.375,gamma.samples.EWOC.cohort3.425,
  gamma.samples.EWOC.cohort3.475,gamma.samples.EWOC.cohort3.525,
  gamma.samples.EWOC.cohort3.575)
round(vector.of.posterior.probabilities.for.each.dose.EWOC.cohort3,3)
#Quadractic loss
(vector.of.posterior.probabilities.for.each.dose.EWOC.cohort3-0.25)^2
#0.05806092 0.03941879 0.00569396 0.02168256 0.18597656
which.min(c((vector.of.posterior.probabilities.for.each.dose.EWOC.cohort3-
0.25)^2))
#Dose level number 3 (475) is the dose that minimizes the difference. Stay
at the current dose level
# The trial diverged from the actual trial. As we don't have data we will
assume that additional three patients
#were treated at 475 and no DLTs were observed (same outcomes as before)
#############################################################################
#COHORT4 - Three additional patients at 475 with no DLTs
#data
dataEWOCcohort4 <- list(y=c(0,0,0,0,0,0,0,0,0,0,0,0),
                        x=c(375,375,375,425,425,425,475,475,475,475,475,47
5), n=12)

#WinBugs code to run after the FOURTH cohort outcomes are known
EWOCmodelcohort4 <-bugs(dataEWOCcohort4, inits=initsEWOCmodel, parameters.
to.save =c("gamma","rho0"),
                        "EWOCmodelFINAL.txt", n.chains=3,
                        n.iter=10000, n.burnin=2000, n.thin=1, bugs.direct
ory=dir.bugs) #debug=T
print(EWOCmodelcohort4, digits.summary = 3)
#alpha quantile for the marginal posterior cumulative distribution of gamm
```

```
a:
#x=511.50
#Posterior probability of toxicity for each dose that minimizes the differ
ence to alpha (through a quadratic loss)
gamma.samples.EWOC.cohort4<-EWOCmodelcohort4$sims.list$gamma
gamma.samples.EWOC.cohort4.375<-sum(gamma.samples.EWOC.cohort4<=375)/lengt
h(gamma.samples.EWOC.cohort4)
gamma.samples.EWOC.cohort4.425<-sum(gamma.samples.EWOC.cohort4<=425)/lengt
h(gamma.samples.EWOC.cohort4)
gamma.samples.EWOC.cohort4.475<-sum(gamma.samples.EWOC.cohort4<=475)/lengt
h(gamma.samples.EWOC.cohort4)
gamma.samples.EWOC.cohort4.525<-sum(gamma.samples.EWOC.cohort4<=525)/lengt
h(gamma.samples.EWOC.cohort4)
gamma.samples.EWOC.cohort4.575<-sum(gamma.samples.EWOC.cohort4<=575)/lengt
h(gamma.samples.EWOC.cohort4)
vector.of.posterior.probabilities.for.each.dose.EWOC.cohort4 <-c(
  gamma.samples.EWOC.cohort4.375,gamma.samples.EWOC.cohort4.425,
  gamma.samples.EWOC.cohort4.475,gamma.samples.EWOC.cohort4.525,
  gamma.samples.EWOC.cohort4.575)
round(vector.of.posterior.probabilities.for.each.dose.EWOC.cohort4,3)
#Quadractic loss
(vector.of.posterior.probabilities.for.each.dose.EWOC.cohort4-0.25)^2
#0.060208891 0.049247007 0.019402168 0.004389062 0.138601085
which.min(c((vector.of.posterior.probabilities.for.each.dose.EWOC.cohort4-
0.25)^2))
#Dose level number 4 (525) is the dose that minimizes the difference. Dose
escalated to the next higher level
##############################################################################
#COHORT5 - Three patients treated at 525 with no DLTs
#data
dataEWOCcohort5 <- list(y=c(0,0,0,0,0,0,0,0,0,0,0,0,0,0,0),
                        x=c(375,375,375,425,425,425,475,475,475,475,475,47
5,525,525,525), n=15)

#WinBugs code to run after the FIFTH cohort outcomes are known
EWOCmodelcohort5 <-bugs(dataEWOCcohort5, inits=initsEWOCmodel, parameters.
to.save =c("gamma","rho0"),
                        "EWOCmodelFINAL.txt", n.chains=3,
                        n.iter=10000, n.burnin=2000, n.thin=1, bugs.direct
ory=dir.bugs) #debug=T
print(EWOCmodelcohort5, digits.summary = 3)
#alpha quantile for the marginal posterior cumulative distribution of gamm
a:
#x=534.40
#Posterior probability of toxicity for each dose that minimizes the differ
ence to alpha (through a quadratic loss)
gamma.samples.EWOC.cohort5<-EWOCmodelcohort5$sims.list$gamma
gamma.samples.EWOC.cohort5.375<-sum(gamma.samples.EWOC.cohort5<=375)/lengt
h(gamma.samples.EWOC.cohort5)
gamma.samples.EWOC.cohort5.425<-sum(gamma.samples.EWOC.cohort5<=425)/lengt
h(gamma.samples.EWOC.cohort5)
gamma.samples.EWOC.cohort5.475<-sum(gamma.samples.EWOC.cohort5<=475)/lengt
```

```
h(gamma.samples.EWOC.cohort5)
gamma.samples.EWOC.cohort5.525<-sum(gamma.samples.EWOC.cohort5<=525)/lengt
h(gamma.samples.EWOC.cohort5)
gamma.samples.EWOC.cohort5.575<-sum(gamma.samples.EWOC.cohort5<=575)/lengt
h(gamma.samples.EWOC.cohort5)
vector.of.posterior.probabilities.for.each.dose.EWOC.cohort5 <-c(
  gamma.samples.EWOC.cohort5.375,gamma.samples.EWOC.cohort5.425,
  gamma.samples.EWOC.cohort5.475,gamma.samples.EWOC.cohort5.525,
  gamma.samples.EWOC.cohort5.575)
round(vector.of.posterior.probabilities.for.each.dose.EWOC.cohort5,3)
#Quadractic loss
(vector.of.posterior.probabilities.for.each.dose.EWOC.cohort5-0.25)^2
#0.061359418 0.055932250 0.037216840 0.002205085 0.076084028
which.min(c((vector.of.posterior.probabilities.for.each.dose.EWOC.cohort5-
0.25)^2))
#Dose level number 4 (525) is the dose that minimizes the difference. Stay
at the same dose level
#The trial diverged again. We will stay at the same dose level and we will
assume that three additional
#patients were treated and no DLTs were observed.
##############################################################################
#COHORT6 - Additional three patients treated at 525 with no DLTs
#data
dataEWOCcohort6 <- list(y=c(0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0),
                        x=c(375,375,375,425,425,425,475,475,475,475,475,47
5,525,525,525,525,525,525), n=18)

#WinBugs code to run after the SIXTH cohort outcomes are known
EWOCmodelcohort6 <-bugs(dataEWOCcohort6, inits=initsEWOCmodel, parameters.
to.save =c("gamma","rho0"),
                        "EWOCmodelFINAL.txt", n.chains=3,
                        n.iter=10000, n.burnin=2000, n.thin=1, bugs.direct
ory=dir.bugs) #debug=T
print(EWOCmodelcohort6, digits.summary = 3)
#alpha quantile for the marginal posterior cumulative distribution of gamm
a:
#x=547.30
#Posterior probability of toxicity for each dose that minimizes the differ
ence to alpha (through a quadratic loss)
gamma.samples.EWOC.cohort6<-EWOCmodelcohort6$sims.list$gamma
gamma.samples.EWOC.cohort6.375<-sum(gamma.samples.EWOC.cohort6<=375)/lengt
h(gamma.samples.EWOC.cohort6)
gamma.samples.EWOC.cohort6.425<-sum(gamma.samples.EWOC.cohort6<=425)/lengt
h(gamma.samples.EWOC.cohort6)
gamma.samples.EWOC.cohort6.475<-sum(gamma.samples.EWOC.cohort6<=475)/lengt
h(gamma.samples.EWOC.cohort6)
gamma.samples.EWOC.cohort6.525<-sum(gamma.samples.EWOC.cohort6<=525)/lengt
h(gamma.samples.EWOC.cohort6)
gamma.samples.EWOC.cohort6.575<-sum(gamma.samples.EWOC.cohort6<=575)/lengt
h(gamma.samples.EWOC.cohort6)
vector.of.posterior.probabilities.for.each.dose.EWOC.cohort6 <-c(
  gamma.samples.EWOC.cohort6.375,gamma.samples.EWOC.cohort6.425,
```

```
  gamma.samples.EWOC.cohort6.475,gamma.samples.EWOC.cohort6.525,
  gamma.samples.EWOC.cohort6.575)
round(vector.of.posterior.probabilities.for.each.dose.EWOC.cohort6,3)
#Quadractic loss
(vector.of.posterior.probabilities.for.each.dose.EWOC.cohort6-0.25)^2
#0.06166944 0.05802077 0.04690834 0.01196289 0.04051834
which.min(c((vector.of.posterior.probabilities.for.each.dose.EWOC.cohort6-
0.25)^2))
#Dose level number 4 (525) is again the dose that minimizes the difference
. Stay at the same dose level
#The trial diverged again. We will stay at the same dose level and we will
assume that three additional patients
#were treated and no DLTs were observed.
#########################################################################
#COHORT7 - Additional three patients treated at 525 with no DLTs
#data
dataEWOCcohort7 <- list(y=c(0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0),
                        x=c(375,375,375,425,425,425,475,475,475,475,475,47
5,525,525,525,525,525,525,
                            525,525,525), n=21)

#WinBugs code to run after the SEVENTH cohort outcomes are known
EWOCmodelcohort7 <-bugs(dataEWOCcohort7, inits=initsEWOCmodel, parameters.
to.save =c("gamma","rho0"),
                        "EWOCmodelFINAL.txt", n.chains=3,
                        n.iter=10000, n.burnin=2000, n.thin=1, bugs.direct
ory=dir.bugs) #debug=T
print(EWOCmodelcohort7, digits.summary = 3)
#alpha quantile for the marginal posterior cumulative distribution of gamm
a:
#x=558.30
#Posterior probability of toxicity for each dose that minimizes the differ
ence to alpha (through a quadratic loss)
gamma.samples.EWOC.cohort7<-EWOCmodelcohort7$sims.list$gamma
gamma.samples.EWOC.cohort7.375<-sum(gamma.samples.EWOC.cohort7<=375)/lengt
h(gamma.samples.EWOC.cohort7)
gamma.samples.EWOC.cohort7.425<-sum(gamma.samples.EWOC.cohort7<=425)/lengt
h(gamma.samples.EWOC.cohort7)
gamma.samples.EWOC.cohort7.475<-sum(gamma.samples.EWOC.cohort7<=475)/lengt
h(gamma.samples.EWOC.cohort7)
gamma.samples.EWOC.cohort7.525<-sum(gamma.samples.EWOC.cohort7<=525)/lengt
h(gamma.samples.EWOC.cohort7)
gamma.samples.EWOC.cohort7.575<-sum(gamma.samples.EWOC.cohort7<=575)/lengt
h(gamma.samples.EWOC.cohort7)
vector.of.posterior.probabilities.for.each.dose.EWOC.cohort7 <-c(
  gamma.samples.EWOC.cohort7.375,gamma.samples.EWOC.cohort7.425,
  gamma.samples.EWOC.cohort7.475,gamma.samples.EWOC.cohort7.525,
  gamma.samples.EWOC.cohort7.575)
round(vector.of.posterior.probabilities.for.each.dose.EWOC.cohort7,3)
#Quadractic loss
(vector.of.posterior.probabilities.for.each.dose.EWOC.cohort7-0.25)^2
#0.06202175 0.05967842 0.05173729 0.02462284 0.01670556
```

```
which.min(c((vector.of.posterior.probabilities.for.each.dose.EWOC.cohort7-
0.25)^2))
#Dose level number 5 (575) is the dose that minimizes the difference. Dose
escalated to the next higher level
#############################################################################
#COHORT8 - Now we will use the data from the trial, four patients treated
at 575 with 2 DLTs
#data
dataEWOCcohort8 <- list(y=c(0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,
0,1),
                        x=c(375,375,375,425,425,425,475,475,475,475,475,47
5,525,525,525,525,525,525,525,
                            525,525,525,575,575,575,575), n=25)

#Dose level number 5 (575) is the dose that minimizes the difference. Dose
escalated to the next higher level
#WinBugs code to run after the EIGHTH cohort outcomes are known
EWOCmodelcohort8 <-bugs(dataEWOCcohort8, inits=initsEWOCmodel, parameters.
to.save =c("gamma","rho0"),
                        "EWOCmodelFINAL.txt", n.chains=3,
                        n.iter=10000, n.burnin=2000, n.thin=1, bugs.direct
ory=dir.bugs) #debug=T
print(EWOCmodelcohort8, digits.summary = 3)
#alpha quantile for the marginal posterior cumulative distribution of gamm
a:
#x=544.20
#Posterior probability of toxicity for each dose that minimizes the differ
ence to alpha (through a quadratic loss)
gamma.samples.EWOC.cohort8<-EWOCmodelcohort8$sims.list$gamma
gamma.samples.EWOC.cohort8.375<-sum(gamma.samples.EWOC.cohort8<=375)/lengt
h(gamma.samples.EWOC.cohort8)
gamma.samples.EWOC.cohort8.425<-sum(gamma.samples.EWOC.cohort8<=425)/lengt
h(gamma.samples.EWOC.cohort8)
gamma.samples.EWOC.cohort8.475<-sum(gamma.samples.EWOC.cohort8<=475)/lengt
h(gamma.samples.EWOC.cohort8)
gamma.samples.EWOC.cohort8.525<-sum(gamma.samples.EWOC.cohort8<=525)/lengt
h(gamma.samples.EWOC.cohort8)
gamma.samples.EWOC.cohort8.575<-sum(gamma.samples.EWOC.cohort8<=575)/lengt
h(gamma.samples.EWOC.cohort8)
vector.of.posterior.probabilities.for.each.dose.EWOC.cohort8 <-c(
  gamma.samples.EWOC.cohort8.375,gamma.samples.EWOC.cohort8.425,
  gamma.samples.EWOC.cohort8.475,gamma.samples.EWOC.cohort8.525,
  gamma.samples.EWOC.cohort8.575)
round(vector.of.posterior.probabilities.for.each.dose.EWOC.cohort8,3)
#Quadractic loss
(vector.of.posterior.probabilities.for.each.dose.EWOC.cohort8-0.25)^2
#0.06179367 0.05874564 0.04858351 0.01224711 0.06838225
which.min(c((vector.of.posterior.probabilities.for.each.dose.EWOC.cohort8-
0.25)^2))
##Dose level number 4 (525) is the dose that minimizes the difference. Dos
e de-escalated to the next lower level
#########################################################
```

```
##############################################################################
##############################################################################
##############################################################################
##################### BLRM ############
##############################################################################
##############################################################################
#### Deflexifol clinical trial (bolus schedule)####
# Doses x: 375, 425, 475, 525, 575 mg/m2
#COHORT1 - Three patients treated at 375 and no DLTs
#data
dataBLRMmodelcohort1 <- list(Ncohorts=1, DoseRef=575,
                             DosesAdm=c(375),Npat=c(3), Ntox=c(0), Ndoses=
5,Doses=c(375, 425, 475, 525, 575),
                             Prior=c(-0.847, 0.381, 2.015, 1.027, 0),Pint=
c(0.2,0.30,0.6), #Interval definition
                             LossFunction1=c(1,0,1,1),LossFunction2=c(1,0,
1,2),LossFunction3=c(1,0,2,4)) #Ignore loss functions N/A

#WinBugs code to run after the first cohort
BLRMmodelcohort1 <-bugs(dataBLRMmodelcohort1, inits=NULL, parameters.to.sa
ve =c("Pr.Tox","Pr.Cat[,2]"),
                        "BLRMmodelFINALfirstcohort.txt", n.chains=3,n.iter
=20000, n.burnin=5000, n.thin=1,
                        bugs.directory=dir.bugs) #debug=T
print(BLRMmodelcohort1, digits.summary = 3)
#Calculation of the overdose control
#The dose chosen for the next cohort will be, among the doses that fulfill
the overdose control criterion,
#the dose that has a maximal posterior toxicity probability for the target
toxicity interval, i.e.,
Pr.Tox.samples.BLRM.cohort1.375 <- BLRMmodelcohort1$sims.list$Pr.Tox[,1]
Pr.Tox.samples.BLRM.cohort1.425 <- BLRMmodelcohort1$sims.list$Pr.Tox[,2]
Pr.Tox.samples.BLRM.cohort1.475 <- BLRMmodelcohort1$sims.list$Pr.Tox[,3]
Pr.Tox.samples.BLRM.cohort1.525 <- BLRMmodelcohort1$sims.list$Pr.Tox[,4]
Pr.Tox.samples.BLRM.cohort1.575 <- BLRMmodelcohort1$sims.list$Pr.Tox[,5]
Prob.Tox.BLRM.cohort1.375.larger0.25 <-sum(Pr.Tox.samples.BLRM.cohort1.375
>0.25)/length(Pr.Tox.samples.BLRM.cohort1.375)
Prob.Tox.BLRM.cohort1.425.larger0.25 <-sum(Pr.Tox.samples.BLRM.cohort1.425
>0.25)/length(Pr.Tox.samples.BLRM.cohort1.425)
Prob.Tox.BLRM.cohort1.475.larger0.25 <-sum(Pr.Tox.samples.BLRM.cohort1.475
>0.25)/length(Pr.Tox.samples.BLRM.cohort1.475)
Prob.Tox.BLRM.cohort1.525.larger0.25 <-sum(Pr.Tox.samples.BLRM.cohort1.525
>0.25)/length(Pr.Tox.samples.BLRM.cohort1.525)
Prob.Tox.BLRM.cohort1.575.larger0.25 <-sum(Pr.Tox.samples.BLRM.cohort1.575
>0.25)/length(Pr.Tox.samples.BLRM.cohort1.575)
round(Prob.Tox.BLRM.cohort1.375.larger0.25,3)
round(Prob.Tox.BLRM.cohort1.425.larger0.25,3)
round(Prob.Tox.BLRM.cohort1.475.larger0.25,3)
round(Prob.Tox.BLRM.cohort1.525.larger0.25,3)
round(Prob.Tox.BLRM.cohort1.575.larger0.25,3)
# Doses 1 to 4 fulfill the overdose control criterion. Among them dose lev
el number 4 (525) maximizes the posterior
```

```
#probability of the target interval. Dose 525 is the dose that would have
been given to the next cohort if
#we were not restricting the dose escalation to one level. Next cohort tre
ated with dose level 2 (425)
##############################################################################
#COHORT2 - Three patients treated at 425 and no DLTs
#data
dataBLRMmodelcohort2 <- list(Ncohorts=2, DoseRef=575,
                             DosesAdm=c(375,425),Npat=c(3,3), Ntox=c(0,0),
Ndoses=5,Doses=c(375, 425, 475, 525, 575),
                             Prior=c(-0.847, 0.381, 2.015, 1.027, 0),Pint=
c(0.2,0.30,0.6), #Interval definition
                             LossFunction1=c(1,0,1,1),LossFunction2=c(1,0,
1,2),LossFunction3=c(1,0,2,4)) #Ignore loss functions N/A

#WinBugs code to run after the first cohort
BLRMmodelcohort2 <-bugs(dataBLRMmodelcohort2, inits=NULL, parameters.to.sa
ve =c("Pr.Tox","Pr.Cat[,2]"),
                        "BLRMmodelFINAL.txt", n.chains=3,n.iter=20000, n.b
urnin=5000, n.thin=1,
                        bugs.directory=dir.bugs) #debug=T
print(BLRMmodelcohort2, digits.summary = 3)
#Calculation of the overdose control
#The dose chosen for the next cohort will be, among the doses that fulfill
the overdose control criterion,
#the dose that has a maximal posterior toxicity probability for the target
toxicity interval, i.e.,
Pr.Tox.samples.BLRM.cohort2.375 <- BLRMmodelcohort2$sims.list$Pr.Tox[,1]
Pr.Tox.samples.BLRM.cohort2.425 <- BLRMmodelcohort2$sims.list$Pr.Tox[,2]
Pr.Tox.samples.BLRM.cohort2.475 <- BLRMmodelcohort2$sims.list$Pr.Tox[,3]
Pr.Tox.samples.BLRM.cohort2.525 <- BLRMmodelcohort2$sims.list$Pr.Tox[,4]
Pr.Tox.samples.BLRM.cohort2.575 <- BLRMmodelcohort2$sims.list$Pr.Tox[,5]
Prob.Tox.BLRM.cohort2.375.larger0.25 <-sum(Pr.Tox.samples.BLRM.cohort2.375
>0.25)/length(Pr.Tox.samples.BLRM.cohort2.375)
Prob.Tox.BLRM.cohort2.425.larger0.25 <-sum(Pr.Tox.samples.BLRM.cohort2.425
>0.25)/length(Pr.Tox.samples.BLRM.cohort2.425)
Prob.Tox.BLRM.cohort2.475.larger0.25 <-sum(Pr.Tox.samples.BLRM.cohort2.475
>0.25)/length(Pr.Tox.samples.BLRM.cohort2.475)
Prob.Tox.BLRM.cohort2.525.larger0.25 <-sum(Pr.Tox.samples.BLRM.cohort2.525
>0.25)/length(Pr.Tox.samples.BLRM.cohort2.525)
Prob.Tox.BLRM.cohort2.575.larger0.25 <-sum(Pr.Tox.samples.BLRM.cohort2.575
>0.25)/length(Pr.Tox.samples.BLRM.cohort2.575)
round(Prob.Tox.BLRM.cohort2.375.larger0.25,3)
round(Prob.Tox.BLRM.cohort2.425.larger0.25,3)
round(Prob.Tox.BLRM.cohort2.475.larger0.25,3)
round(Prob.Tox.BLRM.cohort2.525.larger0.25,3)
round(Prob.Tox.BLRM.cohort2.575.larger0.25,3)
# Doses 1 to 5 fulfill the overdose control criterion. Among them dose lev
el number 5 (575) maximizes the posterior
#probability of the target interval. Dose 575 is the dose that would have
been given to the next cohort if
#we were not restricting the dose escalation to one level. Next cohort tre
```

```
ated with dose level 3 (475)
##############################################################################
#COHORT3 - Three patients treated at 475 and no DLTs
#data
dataBLRMmodelcohort3 <- list(Ncohorts=3, DoseRef=575,
                             DosesAdm=c(375,425,475),Npat=c(3,3,3), Ntox=c
(0,0,0), Ndoses=5,Doses=c(375, 425, 475, 525, 575),
                             Prior=c(-0.847, 0.381, 2.015, 1.027, 0),Pint=
c(0.2,0.30,0.6), #Interval definition
                             LossFunction1=c(1,0,1,1),LossFunction2=c(1,0,
1,2),LossFunction3=c(1,0,2,4)) #Ignore loss functions N/A

#WinBugs code to run after the first cohort
BLRMmodelcohort3 <-bugs(dataBLRMmodelcohort3, inits=NULL, parameters.to.sa
ve =c("Pr.Tox","Pr.Cat[,2]"),
                        "BLRMmodelFINAL.txt", n.chains=3,n.iter=20000, n.b
urnin=5000, n.thin=1,
                        bugs.directory=dir.bugs) #debug=T
print(BLRMmodelcohort3, digits.summary = 3)
#Calculation of the overdose control
#The dose chosen for the next cohort will be, among the doses that fulfill
the overdose control criterion,
#the dose that has a maximal posterior toxicity probability for the target
toxicity interval, i.e.,
Pr.Tox.samples.BLRM.cohort3.375 <- BLRMmodelcohort3$sims.list$Pr.Tox[,1]
Pr.Tox.samples.BLRM.cohort3.425 <- BLRMmodelcohort3$sims.list$Pr.Tox[,2]
Pr.Tox.samples.BLRM.cohort3.475 <- BLRMmodelcohort3$sims.list$Pr.Tox[,3]
Pr.Tox.samples.BLRM.cohort3.525 <- BLRMmodelcohort3$sims.list$Pr.Tox[,4]
Pr.Tox.samples.BLRM.cohort3.575 <- BLRMmodelcohort3$sims.list$Pr.Tox[,5]
Prob.Tox.BLRM.cohort3.375.larger0.25 <-sum(Pr.Tox.samples.BLRM.cohort3.375
>0.25)/length(Pr.Tox.samples.BLRM.cohort3.375)
Prob.Tox.BLRM.cohort3.425.larger0.25 <-sum(Pr.Tox.samples.BLRM.cohort3.425
>0.25)/length(Pr.Tox.samples.BLRM.cohort3.425)
Prob.Tox.BLRM.cohort3.475.larger0.25 <-sum(Pr.Tox.samples.BLRM.cohort3.475
>0.25)/length(Pr.Tox.samples.BLRM.cohort3.475)
Prob.Tox.BLRM.cohort3.525.larger0.25 <-sum(Pr.Tox.samples.BLRM.cohort3.525
>0.25)/length(Pr.Tox.samples.BLRM.cohort3.525)
Prob.Tox.BLRM.cohort3.575.larger0.25 <-sum(Pr.Tox.samples.BLRM.cohort3.575
>0.25)/length(Pr.Tox.samples.BLRM.cohort3.575)
round(Prob.Tox.BLRM.cohort3.375.larger0.25,3)
round(Prob.Tox.BLRM.cohort3.425.larger0.25,3)
round(Prob.Tox.BLRM.cohort3.475.larger0.25,3)
round(Prob.Tox.BLRM.cohort3.525.larger0.25,3)
round(Prob.Tox.BLRM.cohort3.575.larger0.25,3)
# Doses 1 to 5 fulfill the overdose control criterion. Among them dose lev
el number 5 (575) maximizes the posterior
#probability of the target interval. Dose 575 is the dose that would have
been given to the next cohort if
#we were not restricting the dose escalation to one level. Next cohort tre
ated with dose level 4 (525)
##############################################################################
#COHORT4 - Six patients treated at 525 and no DLTs
```

```
#data
dataBLRMmodelcohort4 <- list(Ncohorts=4, DoseRef=575,
                             DosesAdm=c(375,425,475,525),Npat=c(3,3,3,6),
Ntox=c(0,0,0,0), Ndoses=5,Doses=c(375, 425, 475, 525, 575),
                             Prior=c(-0.847, 0.381, 2.015, 1.027, 0),Pint=
c(0.2,0.30,0.6), #Interval definition
                             LossFunction1=c(1,0,1,1),LossFunction2=c(1,0,
1,2),LossFunction3=c(1,0,2,4)) #Ignore loss functions N/A

#WinBugs code to run after the first cohort
BLRMmodelcohort4 <-bugs(dataBLRMmodelcohort4, inits=NULL, parameters.to.sa
ve =c("Pr.Tox","Pr.Cat[,2]"),
                        "BLRMmodelFINAL.txt", n.chains=3,n.iter=20000, n.b
urnin=5000, n.thin=1,
                        bugs.directory=dir.bugs) #debug=T
print(BLRMmodelcohort4, digits.summary = 3)
#Calculation of the overdose control
#The dose chosen for the next cohort will be, among the doses that fulfill
the overdose control criterion,
#the dose that has a maximal posterior toxicity probability for the target
toxicity interval, i.e.,
Pr.Tox.samples.BLRM.cohort4.375 <- BLRMmodelcohort4$sims.list$Pr.Tox[,1]
Pr.Tox.samples.BLRM.cohort4.425 <- BLRMmodelcohort4$sims.list$Pr.Tox[,2]
Pr.Tox.samples.BLRM.cohort4.475 <- BLRMmodelcohort4$sims.list$Pr.Tox[,3]
Pr.Tox.samples.BLRM.cohort4.525 <- BLRMmodelcohort4$sims.list$Pr.Tox[,4]
Pr.Tox.samples.BLRM.cohort4.575 <- BLRMmodelcohort4$sims.list$Pr.Tox[,5]
Prob.Tox.BLRM.cohort4.375.larger0.25 <-sum(Pr.Tox.samples.BLRM.cohort4.375
>0.25)/length(Pr.Tox.samples.BLRM.cohort4.375)
Prob.Tox.BLRM.cohort4.425.larger0.25 <-sum(Pr.Tox.samples.BLRM.cohort4.425
>0.25)/length(Pr.Tox.samples.BLRM.cohort4.425)
Prob.Tox.BLRM.cohort4.475.larger0.25 <-sum(Pr.Tox.samples.BLRM.cohort4.475
>0.25)/length(Pr.Tox.samples.BLRM.cohort4.475)
Prob.Tox.BLRM.cohort4.525.larger0.25 <-sum(Pr.Tox.samples.BLRM.cohort4.525
>0.25)/length(Pr.Tox.samples.BLRM.cohort4.525)
Prob.Tox.BLRM.cohort4.575.larger0.25 <-sum(Pr.Tox.samples.BLRM.cohort4.575
>0.25)/length(Pr.Tox.samples.BLRM.cohort4.575)
round(Prob.Tox.BLRM.cohort4.375.larger0.25,3)
round(Prob.Tox.BLRM.cohort4.425.larger0.25,3)
round(Prob.Tox.BLRM.cohort4.475.larger0.25,3)
round(Prob.Tox.BLRM.cohort4.525.larger0.25,3)
round(Prob.Tox.BLRM.cohort4.575.larger0.25,3)
# Doses 1 to 5 fulfill the overdose control criterion. Among them dose lev
el number 5 (575) maximizes the posterior
#probability of the target interval. Dose 575 is the dose that would have
been given to the next cohort if
#we were not restricting the dose escalation to one level. Next cohort tre
ated with dose level 5 (575)
##############################################################################
#COHORT5 - Four patients treated at 575 and 2 DLTs
#data
dataBLRMmodelcohort5 <- list(Ncohorts=5, DoseRef=575,
                             DosesAdm=c(375,425,475,525,575),Npat=c(3,3,3,
```

```
6,4), Ntox=c(0,0,0,0,2), Ndoses=5,Doses=c(375, 425, 475, 525, 575),
                              Prior=c(-0.847, 0.381, 2.015, 1.027, 0),Pint=
c(0.2,0.30,0.6), #Interval definition
                              LossFunction1=c(1,0,1,1),LossFunction2=c(1,0,
1,2),LossFunction3=c(1,0,2,4)) #Ignore loss functions N/A

#WinBugs code to run after the first cohort
BLRMmodelcohort5 <-bugs(dataBLRMmodelcohort5, inits=NULL, parameters.to.sa
ve =c("Pr.Tox","Pr.Cat[,2]"),
                        "BLRMmodelFINAL.txt", n.chains=3,n.iter=20000, n.b
urnin=5000, n.thin=1,
                        bugs.directory=dir.bugs) #debug=T
print(BLRMmodelcohort5, digits.summary = 3)
#Calculation of the overdose control
#The dose chosen for the next cohort will be, among the doses that fulfill
the overdose control criterion,
#the dose that has a maximal posterior toxicity probability for the target
toxicity interval, i.e.,
Pr.Tox.samples.BLRM.cohort5.375 <- BLRMmodelcohort5$sims.list$Pr.Tox[,1]
Pr.Tox.samples.BLRM.cohort5.425 <- BLRMmodelcohort5$sims.list$Pr.Tox[,2]
Pr.Tox.samples.BLRM.cohort5.475 <- BLRMmodelcohort5$sims.list$Pr.Tox[,3]
Pr.Tox.samples.BLRM.cohort5.525 <- BLRMmodelcohort5$sims.list$Pr.Tox[,4]
Pr.Tox.samples.BLRM.cohort5.575 <- BLRMmodelcohort5$sims.list$Pr.Tox[,5]
Prob.Tox.BLRM.cohort5.375.larger0.25 <-sum(Pr.Tox.samples.BLRM.cohort5.375
>0.25)/length(Pr.Tox.samples.BLRM.cohort5.375)
Prob.Tox.BLRM.cohort5.425.larger0.25 <-sum(Pr.Tox.samples.BLRM.cohort5.425
>0.25)/length(Pr.Tox.samples.BLRM.cohort5.425)
Prob.Tox.BLRM.cohort5.475.larger0.25 <-sum(Pr.Tox.samples.BLRM.cohort5.475
>0.25)/length(Pr.Tox.samples.BLRM.cohort5.475)
Prob.Tox.BLRM.cohort5.525.larger0.25 <-sum(Pr.Tox.samples.BLRM.cohort5.525
>0.25)/length(Pr.Tox.samples.BLRM.cohort5.525)
Prob.Tox.BLRM.cohort5.575.larger0.25 <-sum(Pr.Tox.samples.BLRM.cohort5.575
>0.25)/length(Pr.Tox.samples.BLRM.cohort5.575)
round(Prob.Tox.BLRM.cohort5.375.larger0.25,3)
round(Prob.Tox.BLRM.cohort5.425.larger0.25,3)
round(Prob.Tox.BLRM.cohort5.475.larger0.25,3)
round(Prob.Tox.BLRM.cohort5.525.larger0.25,3)
round(Prob.Tox.BLRM.cohort5.575.larger0.25,3)
# Doses 1 to 4 fulfill the overdose control criterion. Among them dose lev
el number 4 (525) maximizes the posterior
#probability of the target interval. Dose 525 is the dose that would have
been given to the next cohort if
#we were not restricting the escalation to one level. Next cohort treated
the next lower with dose level 4 (525)
##############################################################################
##############################################################################
#Testing toxicity at first dose level - Assuming 3 patients treated at 375
and 1 DLTs
#data
dataBLRMmodel1DLTcohort1 <- list(Ncohorts=1, DoseRef=375,
                                 DosesAdm=c(375,375),Npat=c(1,2), Ntox=c(0
,1), Ndoses=5,Doses=c(375, 425, 475, 525, 575),
```

```
                                  Prior=c(-0.847, 0.381, 2.015, 1.027, 0),P
int=c(0.2,0.30,0.6), #Interval definition
                                  LossFunction1=c(1,0,1,1),LossFunction2=c(
1,0,1,2),LossFunction3=c(1,0,2,4)) #Ignore loss functions N/A

#WinBugs code to run after the first cohort
BLRMmodel1DLTcohort1 <-bugs(dataBLRMmodel1DLTcohort1, inits=NULL, parameter
rs.to.save =c("Pr.Tox","Pr.Cat[,2]"),
                            "BLRMmodelFINAL.txt", n.chains=3,n.iter=20000,
n.burnin=5000, n.thin=1,
                            bugs.directory=dir.bugs) #debug=T
print(BLRMmodel1DLTcohort1, digits.summary = 3)
#Calculation of the overdose control
#The dose chosen for the next cohort will be, among the doses that fulfill
the overdose control criterion,
#the dose that has a maximal posterior toxicity probability for the target
toxicity interval, i.e.,
Pr.Tox.samples.BLRM.1DLTcohort1.375 <- BLRMmodel1DLTcohort1$sims.list$Pr.T
ox[,1]
Pr.Tox.samples.BLRM.1DLTcohort1.425 <- BLRMmodel1DLTcohort1$sims.list$Pr.T
ox[,2]
Pr.Tox.samples.BLRM.1DLTcohort1.475 <- BLRMmodel1DLTcohort1$sims.list$Pr.T
ox[,3]
Pr.Tox.samples.BLRM.1DLTcohort1.525 <- BLRMmodel1DLTcohort1$sims.list$Pr.T
ox[,4]
Pr.Tox.samples.BLRM.1DLTcohort1.575 <- BLRMmodel1DLTcohort1$sims.list$Pr.T
ox[,5]
Prob.Tox.BLRM.1DLTcohort1.375.larger0.25 <-sum(Pr.Tox.samples.BLRM.1DLTcoh
ort1.375>0.25)/length(Pr.Tox.samples.BLRM.1DLTcohort1.375)
Prob.Tox.BLRM.1DLTcohort1.425.larger0.25 <-sum(Pr.Tox.samples.BLRM.1DLTcoh
ort1.425>0.25)/length(Pr.Tox.samples.BLRM.1DLTcohort1.425)
Prob.Tox.BLRM.1DLTcohort1.475.larger0.25 <-sum(Pr.Tox.samples.BLRM.1DLTcoh
ort1.475>0.25)/length(Pr.Tox.samples.BLRM.1DLTcohort1.475)
Prob.Tox.BLRM.1DLTcohort1.525.larger0.25 <-sum(Pr.Tox.samples.BLRM.1DLTcoh
ort1.525>0.25)/length(Pr.Tox.samples.BLRM.1DLTcohort1.525)
Prob.Tox.BLRM.1DLTcohort1.575.larger0.25 <-sum(Pr.Tox.samples.BLRM.1DLTcoh
ort1.575>0.25)/length(Pr.Tox.samples.BLRM.1DLTcohort1.575)
round(Prob.Tox.BLRM.1DLTcohort1.375.larger0.25,3)
round(Prob.Tox.BLRM.1DLTcohort1.425.larger0.25,3)
round(Prob.Tox.BLRM.1DLTcohort1.475.larger0.25,3)
round(Prob.Tox.BLRM.1DLTcohort1.525.larger0.25,3)
round(Prob.Tox.BLRM.1DLTcohort1.575.larger0.25,3)
# Doses 1 to 4 fulfill the overdose control criterion. Among them dose lev
el number 4 (525) maximizes the posterior
#probability of the target interval. Dose 525 is the dose that would have
been given to the next cohort if
#we were not restricting the escalation to one level. Next cohort treated
the next lower with dose level 4 (525)
##########################################################################
##########################################################################
##########################################################################
##################### mTPI ############
```

```
##########################################################################
##########################################################################
#### Deflexifol clinical trial (bolus shcedule)####
# Doses x: 375, 425, 475, 525, 575 mg/m2
### Dose levels 1- 3 ###
### 3 patients treated and 0 DLTs ###
# Same outcomes for the first three cohorts
### Posterior beta distribution is Beta(1,4)
#mean posterior probability of DLT
#p.hat=a/(a+b)
p.hat.1to3=1/(1+4)
p.hat.1to3
##Credible interval
credible.interval.cohort1to3 <- qbeta(c(0.025,0.975), 1, 4)
credible.interval.cohort1to3
round(credible.interval.cohort1to3, 3)
## Calculation of the UPMs
# Calculate the probability P(p.cur belong [0, 0.20]) -> UI
probability.UI.cohort1to3 <- pbeta(0.20, 1, 4)
probability.UI.cohort1to3
#Calculating UPM.UI (divided by the length of the interval)
UPM.UI.cohort1to3 <-probability.UI.cohort1to3/(0.25-0.05)
UPM.UI.cohort1to3
# Calculate the probability P(p.cur belong [0.20, 0.30])
probability.EI.cohort1to3 <- pbeta(0.30, 1, 4)-pbeta(0.20, 1, 4)
probability.EI.cohort1to3
round(probability.EI.cohort1to3,3)
#Calculating UPM.EI (divided by the length of the interval)
UPM.EI.cohort1to3 <-probability.EI.cohort1to3/(0.05+0.05)
UPM.EI.cohort1to3
# Calculate the probability P(p.cur belong [0.30, 0.1])
probability.OI.cohort1to3 <-1- pbeta(0.30, 1, 4)
probability.OI.cohort1to3
#Calculating UPM.EI (divided by the length of the interval)
UPM.OI.cohort1to3 <-probability.OI.cohort1to3/(1-0.30)
UPM.OI.cohort1to3
#Safety Rule
#P(p.cur>0.25|y)<0.95
mTPI.safety.cohort1to3 <- 1-pbeta(0.25, 1, 4)
mTPI.safety.cohort1to3
##############################################################
### Dose level 4 ###
### 6 patients treated and 0 DLTs ###
### Posterior beta distribution is Beta(1,7) ###
#mean posterior probability of DLT
#p_hat=a/(a+b)
p.hat.cohort4=1/(1+7)
p.hat.cohort4
##Credible interval
credible.interval.cohort4 <- qbeta(c(0.025,0.975), 1, 7)
credible.interval.cohort4
round(credible.interval.cohort4, 3)
```

```
## Calculation of the UPMs
# Calculate the probability P(p.cur belong [0, 0.20]) -> UI
probability.UI.cohort4 <- pbeta(0.20, 1, 7)
probability.UI.cohort4
#Calculating UPM.UI (divided by the length of the interval)
UPM.UI.cohort4 <-probability.UI.cohort4/(0.25-0.05)
UPM.UI.cohort4
# Calculate the probability P(p.cur belong [0.20, 0.30])
probability.EI.cohort4 <- pbeta(0.30, 1, 7)-pbeta(0.20, 1, 7)
probability.EI.cohort4
#Calculating UPM.EI (divided by the length of the interval)
UPM.EI.cohort4 <-probability.EI.cohort4/(0.05+0.05)
UPM.EI.cohort4
# Calculate the probability P(p.cur belong [0.30, 0.1])
probability.OI.cohort4 <-1- pbeta(0.30, 1, 7)
probability.OI.cohort4
#Calculating UPM.EI (divided by the length of the interval)
UPM.OI.cohort4 <-probability.OI.cohort4/(1-0.30)
UPM.OI.cohort4
#Safety Rule
#P(p.cur>0.25|y)<0.95
mTPI.safety.cohort4 <- 1 - pbeta(0.25, 1, 7)
mTPI.safety.cohort4
############################################################
### Dose level 5 ###
### 4 patients treated and 2 DLTs ###
### Posterior beta distribution is Beta(3,3) ###
#mean posterior probability of DLT
#p_hat=a/(a+b)
p.hat.cohort5=3/(3+3)
p.hat.cohort5
##Credible interval
credible.interval.cohort5 <- qbeta(c(0.025,0.975), 3, 3)
credible.interval.cohort5
round(credible.interval.cohort5, 3)
## Calculation of the UPMs
# Calculate the probability P(p.cur belong [0, 0.20]) -> UI
probability.UI.cohort5 <- pbeta(0.20, 3, 3)
probability.UI.cohort5
#Calculating UPM.UI (divided by the length of the interval)
UPM.UI.cohort5 <-probability.UI.cohort5/(0.25-0.05)
UPM.UI.cohort5
round(UPM.UI.cohort5, 3)
# Calculate the probability P(p.cur belong [0.20, 0.30])
probability.EI.cohort5 <- pbeta(0.30, 3, 3)-pbeta(0.20, 3, 3)
probability.EI.cohort5
#Calculating UPM.EI (divided by the length of the interval)
UPM.EI.cohort5 <-probability.EI.cohort5/(0.05+0.05)
UPM.EI.cohort5
# Calculate the probability P(p.cur belong [0.30, 0.1])
probability.OI.cohort5 <-1- pbeta(0.30, 3, 3)
probability.OI.cohort5
```

```
#Calculating UPM.EI (divided by the length of the interval)
UPM.OI.cohort5 <-probability.OI.cohort5/(1-0.30)
UPM.OI.cohort5
#Safety Rule
#P(p.cur>0.25|y)<0.95
mTPI.safety.cohort5 <- 1 - pbeta(0.25, 3, 3)
mTPI.safety.cohort5
##############################################################################
##############################################################################
##############################################################################
#########################
##################### BOIN ############
##############################################################################
##############################################################################
##### BOIN statistics ######
#### Deflexifol clinical trial (bolus shcedule)####
#The safety rule applied is the same as the mTPI. Safety Rule P(p.cur>0.25
|y)<0.95
#The values are the same and were copied.
#95% IC
#Dose levels 1 to 3
0-1.96*sqrt((0*(1-0))/3)
0+1.96*sqrt((0*(1-0))/3)
#Dose level 4
#95% IC
0-1.96*sqrt((0*(1-0))/6)
0+1.96*sqrt((0*(1-0))/6)
#Dose level 5
#95% IC
0.5-1.96*sqrt((0.5*(1-0.5))/4)
0.5+1.96*sqrt((0.5*(1-0.5))/4)
#Obtaining the pre-tabulated dose escalation actions
library(BOIN)
get.boundary(0.25,30,1)
```

2023

DIOGO ALEXANDRE
SILVÉRIO FERREIRA

MODEL-BASED AND MODEL-ASSISTED DESIGNS FOR DOSE-ESCALATION IN CANCER SINGLE-AGENT PHASE I CLINICAL TRIALS: A SIMULATION STUDY

NOVA SCHOOL OF SCIENCE & TECHNOLOGY