Enhancing Clinical Development - Bayesian Methods and Mathematical Modeling for Leveraging Available Information

Moreno Ursino

Université Paris Cité

Mémoire en vue de l'obtention de

l'Habilitation à Diriger des Recherches

Specialité : Mathématiques Appliquées

Présentée par

Moreno URSINO

Enhancing Clinical Development - Bayesian Methods and Mathematical Modeling for Leveraging Available Information

---

Soutenue publiquement le 30 Septembre 2024

Composition du jury :

Rapporteurs :	Hélène JACQMIN-GADDA	Directrice de recherche, Inserm
	Annette KOPP-SCHNEIDER	Professeur, German Cancer Research Center
	Adeline SAMSON	Professeur, Université Grenoble Alpes
Président :	Matthieu RESCHE-REGON	PU-PH, Université Paris Cité
Garant :	Antoine CHAMBAZ	Professeur, Université Paris Cité

Acknowledgments

The path to becoming a researcher is not always linear, but no matter how winding it gets, it is never a journey made alone. I certainly have not walked it solo, and for that I would like to thank all the people I have crossed in my academic journey. From each of them I have learned something, taken inspiration or understood what kind of researcher I want to become. And I certainly would not be here if Mauro had not taken me as a PhD student and Sarah and Emmanuelle had not dragged me away from Italy. I have to admit that I was a little hesitant at first, but after seeing and savoring the picture of European, indeed worldwide, research and seeing how large hospitals and clinical trials work, I doubt I will ever want to go back (except for the occasional vacation or pasta binge).

Of course, I could not have come this far without the support of those who have walked alongside me: my family, my friends (near and far, old and new, I am not quoting you because there are so many of you but you know who you are), and my cat (its "work" contributions, while mostly in the form of sitting on my keyboard or photobombing video conferences, were invaluable). I was about to forget an important thanks, the one to Nutella. It too has always been there in times of need.

So, to everyone who helped me reach this point, whether through advice, companionship, or delicious snacks, thank you.

P.s. If you like the template used in this manuscript, I got it from here: https://olivier.commowick.org/thesis_template.php

Contents

Avant-propos

Ce mémoire d'habilitation à diriger des recherches résume les recherches que j'ai menées depuis l'obtention de mon doctorat en mars 2014 à l'École Polytecnique de Turin (Politecnico di Torino). Mes recherches en biostatistique se sont principalement concentrées sur les méthodes bayésiennes et la modélisation mathématique pour les petits échantillons et les essais cliniques, avec un intérêt particulier pour les phases précoces. Mon objectif principal est d'utiliser davantage d'informations disponibles dans la conception et l'analyse des essais cliniques, afin de mieux décrire et expliquer la relation dose-réponse (toxicité ou efficacité).

Mes contributions à l'amélioration des essais cliniques grâce à des informations supplémentaires peuvent être classées en deux catégories principales : d'une part en intégrant davantage d'informations recueillies pendant les essais eux-mêmes (informations internes) et d'autre part en considérant des sources de données externes supplémentaires (informations externes).

Les informations internes font référence aux données générées au sein de l'essai lui-même. Cela inclut mes recherches sur l'incorporation de la pharmacocinétique et de la pharmacodynamique dans les essais d'escalade de dose de phase I [Ursino et al., 2017, Gerard et al., 2022]. En outre, j'ai étudié l'utilisation de données provenant de cycles multiples [Ursino et al., 2022] et de plusieurs variables de toxicité [Ursino et al., 2019] dans le même contexte recherche de dose, ainsi que la mise en œuvre d'un design bayésien qui couvre les essais "seamless" de phase II et III [Duputel et al., 2023].

Mes travaux sur l'incorporation de sources de données externes englobent le développement de distributions a priori par le biais de l'analyse de données provenant d'essais cliniques antérieurs [Ollier et al., 2020], l'élicitation d'experts [Thall et al., 2019], ainsi que mes travaux sur la méta-analyse des études de recherche de dose [Ursino et al., 2021] et la consolidation des résultats de multiples études (précliniques) sur de petits échantillons [Boulet et al., 2024].

Après un premier chapitre qui présente le contexte de mes travaux, les quatre derniers chapitres abordent ces catégories plus en détail. Pour garder ce document concis, il n'inclut ni résultats exhaustifs ni détails des simulations ; il est recommandé aux lecteurs de se référer aux publications originales pour une compréhension approfondie. Il est aussi important de souligner que ce résumé ne couvre pas l'intégralité de mes recherches précédentes. Néanmoins, dans chaque chapitre, j'ai décrit au moins un travail dont je suis le premier ou le co-premier auteur et un travail que j'ai supervisé.

Depuis novembre 2023, j'ai demandé une autorisation à diriger une thèse (ADT) et en collaboration avec Emmanuelle Comets, je co-supervise la thèse de Axel Vuorinen, qui se concentre sur l'utilisation de la modélisation PK/PD dans les essais de plateforme de phase I/II (travail financé par une subvention MESSIDORE 2022/2023 dont je suis l'investigateur principal). Ce travail, ainsi que d'autres que je mène actuellement dans d'autres financés par des appels à projet, où je suis responsable de tâche ou de groupe (work package leader), seront cités à la fin de chaque chapitre.

Publications

Publications after my PhD used in this manuscript:

[1] Boulet, S., Ursino, M., Michelet, R., Aulin, L., Kloft, C., Comets, E., & Zohar, S. (2024). Bayesian Framework for Multi-Source Data Integration–Application to Human Extrapolation From Preclinical Studies. Statistical Methods in Medical Research, 33(4), 574-588.

[2] Duputel, B., Stallard, N., Montestruc, F., Zohar, S., & Ursino, M. (2023). Using dichotomized survival data to construct a prior distribution for a Bayesian seamless Phase II/III clinical trial. Statistical Methods in Medical Research, 32(5), 963–977.

[3] Ursino, M., Biard, L., & Chevret, S. (2022). DICE: A Bayesian model for early dose finding in phase I trials with multiple treatment courses. Biometrical Journal, 64(8), 1486–1497.

[4] Gerard, E., Zohar, S., Thai, H.-T., Lorenzato, C., Riviere, M.-K., & Ursino, M. (2022). Bayesian dose regimen assessment in early phase oncology incorporating pharmacokinetics and pharmacodynamics. Biometrics, 78(1), 300–312.

[5] Ursino, M., Röver, C., Zohar, S., & Friede, T. (2021). Random-effects meta-analysis of phase I dose-finding studies using stochastic process priors. Annals of Applied Statistics, 15(1): 174-193

[6] Ollier, A., Morita, S., Ursino, M., & Zohar, S. (2020). An adaptive power prior for sequential clinical trials–Application to bridging studies. Statistical Methods in Medical Research, 29(8), 2282–2294.

[7] Ursino, M., Yuan, Y., Alberti, C., Comets, E., Favrais, G., Friede, T., ... Zohar, S. (2019). A dose finding design for seizure reduction in neonates. Journal of the Royal Statistical Society Series C: Applied Statistics, 68(2), 427–444.

[8] Thall, P. F., Ursino, M., Baudouin, V., Alberti, C., & Zohar, S. (2019). Bayesian treatment comparison using parametric mixture priors computed from elicited histograms. Statistical Methods in Medical Research, 28(2), 404–418.

[9] Ursino, M., Zohar, S., Lentz, F., Alberti, C., Friede, T., Stallard, N., & Comets, E. (2017). Dose-finding methods for phase I clinical trials using pharmacokinetics in small populations. Biometrical Journal, 59(4), 804–825.

Other methodological publications after my PhD cited in this manuscript:

[1] Calderazzo, S., Tarima, S., Reid, C., Flournoy, N., Friede, T., Geller, N., ... Others. (2023). Coping with Information Loss and the Use of Auxiliary Sources of Data: A Report from the NISS Ingram Olkin Forum Series on Unplanned Clinical Trial Disruptions. Statistics in Biopharmaceutical Research, 16(2), 141–157.

[2] Röver, C., Ursino, M., Friede, T., & Zohar, S. (2022). A straightforward meta-analysis approach for oncology Phase I dose-finding studies. Statistics in Medicine, 41(20), 3915–3940.

[3] Gerard, E., Zohar, S., Lorenzato, C., Ursino, M., & Riviere, M.-K. (2021). Bayesian modeling of a bivariate toxicity outcome for early phase oncology trials evaluating dose regimens. Statistics in Medicine, 40(23), 5096–5114.

[4] Ollier, A., Zohar, S., Morita, S., & Ursino, M. (2021). Estimating similarity of dose–response relationships in phase I clinical trials—case study in bridging data package. International Journal of Environmental Research and Public Health, 18(4), 1639.

Chapter 1

Introduction

Contents

The main goal of this manuscript is to clearly show the work I have done since completing my PhD, sharing the journey, the progression and evolution of my research interests but also highlights the contribution to the academic community. My research in biostatistics has primarily focused on Bayesian methods and mathematical modeling for small samples and clinical trials, with a special interest in the early phases. I aim to leverage additional available data in both designing and analyzing clinical studies to better describe and explain the dose-response relationship (toxicity or efficacy).

In this chapter, I will provide a succinct introduction to the topics that form the foundation of my research works. The aim is not to delve deeply into each subject, but rather to outline the essential elements, paving the way for a more detailed introduction in subsequent chapters.

1.1 Drug development

Drug development is a complex and rigorous process, designed to ensure that any new medication is both safe and effective before it reaches the public. It is conducted under the strict oversight of health regulatory agencies. In the United States, this role is fulfilled by the Food and Drug Administration (FDA), while in the European Union, it is managed by the European Medicines Agency (EMA). This journey from laboratory to pharmacy shelves is meticulously structured into preclinical and subsequent clinical phases (Figure 1.1), each with distinct objectives and methodologies.

Preclinical Phase: Before human trials can begin, the preclinical phase serves as a crucial initial step, where the safety and efficacy of a compound are evaluated in vitro (e.g., petri

Figure 1.1 – Scheme of drug development.

dishes) and in vivo (in living organism, e.g., animal models). This stage aims to understand the pharmacokinetics (PK, how the body affects a drug) and pharmacodynamics (PD, how a drug affects the body) of the compound. Researchers assess the toxicity, dosing, and potential side effects to ensure the compound safety profile warrants further study in humans.

Phase I: The first phase of clinical trials involves a small group (20-100) of healthy volunteers, or patients in oncology or in vulnerable populations, and focuses primarily on assessing the safety of the drug, determining safe dosage ranges, and identifying side effects. Another goal is to establish the drug PK and PD in humans.

Phase II: This phase aims to evaluate the efficacy of the drug, alongside further safety assessment. It involves a larger group of participants (100-300) who have the condition that the drug is intended to treat. Phase II helps to determine the optimal dose and provide preliminary evidence on the drug effectiveness.

Phase III: Involving an even larger group of participants (1000-3000), this phase aims to confirm the drug's efficacy, monitor side effects, compare it to commonly used treatments, and collect information that will allow the drug to be used safely. This phase is critical for the drug approval process, providing the detailed data regulatory authorities need to approve or reject a drug for public use.

Phase IV: Conducted after the drug has been approved and is on the market, this phase involves the post-marketing surveillance of the drug performance in a real-world setting. It aims to detect any rare or long-term adverse effects over a larger population and longer time period, ensuring the drug ongoing safety and effectiveness.

Each phase of clinical research is crucial for the assessment of new drugs, ensuring that only safe and effective treatments reach the public. Typically, each phase is conducted sequentially, and a drug advances to the next phase only upon successful completion of the current one. However, as discussed later in this chapter, it is possible to integrate multiple phases into a single trial or protocol.

1.2 Early phase clinical trials

In early-stage clinical trials, including Phase I and Phase I/II (that is combining Phases I and II simultaneously or sequentially), researchers conduct dose-finding studies to assess drug safety and PK in humans. These initial trials, often with healthy volunteers, seek to determine the safe dosage range and toxicity relationship of promising new treatments. In cases of severe diseases like cancer or for specific populations such as children, patients with limited treatment options may participate in Phase I trials due to the potential high toxicity of therapies. These trials focus on identifying the maximum tolerated dose (MTD), typically defined by the likelihood of dose-limiting toxicity (DLT), which is often based on initial treatment cycle data, especially in oncology. Phase I trials are usually structured with predetermined doses, cohort sizes, and stopping rules. The dose closest to the desired toxicity probability is deemed the MTD, upon reaching the maximum sample size. This MTD then helps establish a recommended dose for further trials. Then, the Phase II stages assess a drug's efficacy in patients, often using PD markers to understand the dose-response relationship based on specific biomarkers.

Ethical considerations prevent simple random designs; instead, trials use response-adaptive sequential designs to find the appropriate dose with minimal participants, avoiding excessive toxicity or ineffective doses. These trials sequentially enroll patients into small cohorts, with adaptive dose-escalation strategies often relying on Bayesian methods to incorporate existing data and adjust the trial design in real-time. The goal is to quickly pinpoint drugs with therapeutic potential and identify the patient groups they benefit the most. The process starts with the first cohort receiving the lowest dose level. Subsequent cohorts are then given increasing doses, advancing until the MTD is identified or the maximum sample size is reached. Dose decisions for future cohorts are based on a statistical model that integrates data from all prior cohorts to optimize dose-escalation and minimize the risk of harmful or ineffective dosing. The three primary categories of these early-phase trial designs are: rule-based designs, model-based designs and model-assisted designs.

Rule-based designs, like the traditional "3+3" approach, have been standard for years in escalating patient doses [Storer, 1989]. In this memoryless design, three patients form a cohort; if no severe toxicity is observed, the dose increases. If two patients experience severe toxicity, the trial halts, considering the previous dose as the MTD. If one patient has severe toxicity, three more are added; if another has severe toxicity, the trial stops, or else it continues. Despite their simplicity, these designs are less effective at finding the correct MTD compared to model-based designs, which are slowly gaining popularity.

Model-based designs, such as the Continual Reassessment Method (CRM) [O'Quigley et al., 1990] or the Bayesian Linear Regression Model (BLRM) [Neuenschwander et al., 2008], use a parametric model for dose-toxicity relationships and are more efficient in dose escalation. The CRM employs a model-based strategy that continuously updates the probability of DLT based on patient responses. The dose-DLT function is typically parametrized as $P(DLT|x_k) = F(x_k, \boldsymbol{\theta})$, where F is a monotonically increasing function, often chosen to be a logistic function, $\boldsymbol{\theta}$ represents the parameters of this function (only one in the simplest case), which are updated as the trial progresses, and x_k the pseudo-doses, also called effective doses, which represents the prior believes on toxicity probability for doses $x_1, \cdots, x_K$ (on a panel of K doses). As each cohort of patients is treated and observed for DLTs, the data are used to update $\boldsymbol{\theta}$ using Bayes' theorem, refining the estimate of the MTD. The primary goal of CRM is to identify the dose level closest to a target toxicity rate, typically set between 20% and 33%. This is done by calculating the posterior distribution of $\boldsymbol{\theta}$, used to compute the toxicity probability for each dose level, and selecting the dose for the next cohort of patients that is estimated to be closest to the target toxicity rate.

Model-assisted designs blend the straightforwardness of rule-based designs with the precision of model-based approaches. They use statistical models for decision-making but maintain simple pre-trial rules for dose adjustments. Examples include the modified toxicity probability interval (mTPI) [Ji et al., 2010] and Bayesian optimal interval (BOIN) designs [Liu and Yuan, 2015]. These models are user-friendly yet capitalize on detailed statistical insights for dose determination.

As previously mentioned, traditional methods for estimating the MTD typically focus on short-term DLTs (first therapy cycle only) and often simplify drug dosage in models as a pseudo-dose, overlooking multiple drug intakes. In the following chapter, I will present my work that expands on this conventional approach. Additionally, while it was once assumed that both tox-

icity and efficacy of cytotoxic drugs in oncology increase with dosage, this paradigm does not apply to certain contemporary drugs, such as those used in immunology. The recent Optimus project illustrates that, in some cases, "less is more" [Shah et al., 2021]. Therefore, the optimization processes in drug development aimed at minimizing toxicity and maximizing benefit should begin in the early phases [Fourie Zirkelbach et al., 2022], and the drug's PK/PD profile plays a crucial role.

1.3 Pharmacokinetics/Pharmacodynamics

Clinical pharmacology hinges on the principle that the potency of drug effects, including both efficacy and toxicity, correlates with its concentration in the body. PK explores how drug levels fluctuate over time following administration, while PD examines how these concentrations influence biological or physiological responses via biomarkers. In essence, PK describes the body effects on the drug, and PD details the drug effects on the body. PK/PD modeling is employed to define the connection between dosage, scheduling, and the drug overall impact. Typically, drug levels for PK models are tracked through regular blood sampling post-administration. For PD analysis, corresponding biomarkers are monitored to assess the drug effect.

PK examines the journey of a drug through the body over time, involving absorption, distribution, metabolism, and excretion (ADME). Absorption is how a drug enters the bloodstream from its administration site, affected by the route of administration which determines its bioavailability, the proportion of the drug that reaches the systemic circulation. Once absorbed, the drug circulates to body tissues, with the volume of distribution reflecting the theoretical volume needed to achieve the observed plasma concentrations. Metabolism, primarily occurring in organs like the liver, transforms the drug to a more water-soluble form for easier excretion, typically through the kidneys or feces. Excretion rate is quantified by the drug clearance rate, indicating the volume of blood from which the drug is removed per time unit. These ADME processes collectively determine the drug plasma levels and tissue exposure, which in turn influence its therapeutic and pharmacological effects. PK metrics such as the area under the curve (AUC) or peak plasma concentration (Cmax) are used to describe the drug exposure profile. Non-compartmental analysis (NCA) calculates PK parameters from actual concentration measurements without assuming any particular model. NCA is typically used to assess drug exposure levels, for instance, calculating the area under the curve (AUC) using the trapezoidal rule, as well as other parameters like clearance.

Conversely, PK compartmental analysis is a mathematical modeling approach where the body is conceptualized as a system of compartments that represent groups of tissues with similar blood flow and drug affinity characteristics. The most basic model is the one-compartment model, which considers the entire body as a single, homogeneous compartment where the drug concentration is uniformly distributed. More complex models can include multiple compartments, representing different tissue groups or organs with their specific drug absorption, distribution, and elimination characteristics [Gabrielsson and Weiner, 2007]. PK analysis usually hinges on differential equations that define the rates of change in drug concentration within these compartments. For PD analysis, the most common equations utilized to explain how drug concentrations impact biological systems include the fixed, linear, log-linear, Emax, sigmoid Emax,

and indirect PD response models [Gabrielsson and Weiner, 2007].

Population PK(/PD) (pop-PK/PD) modeling is a technique that aggregates individual patient data into a cohesive compartmental model, adept at handling variations both within and across patients. The essence of pop-PK modeling lies in its ability to delineate a population PK profile, essentially capturing the characteristics of an “average" patient. It aims to elucidate the variability in drug concentrations (/effects) experienced among patients over time. Incorporating covariate information such as age, gender, body weight, and renal or hepatic function into pop-PK models enables the identification of factors driving PK variability within a population. This understanding is vital, as PK variations can significantly influence drug safety profiles. Insights garnered from pop-PK modeling facilitate the selection of suitable dosages for specific populations or subgroups, enhancing therapeutic efficacy and safety. Although a population-level approach reduces the need for extensive data collection on each individual, it demands more computational resources. This is attributed to its reliance on nonlinear mixed-effect models, which are more complex than the simpler non-compartmental PK models.

One of the key outcomes of PK/PD analysis is the characterization of the dose-response relationship. This relationship helps in determining the optimal dosing regimen for a drug, aiming to maximize therapeutic effects while minimizing adverse effects. Additionally, PK/PD modeling can be used in drug development to predict outcomes of dosing regimens (as it will be shown in one of my work of this manuscript), to support clinical trial design, and to inform drug labeling and regulatory decisions.

1.4 Seamless trials

Contrasting with conventional clinical trials, seamless clinical trials integrate multiple phases into a single adaptive design study, streamlining the process. This approach, also known as a combined-phase study, simplifies the trial structure by blending phases with interim assessments, saving time and requiring fewer participants. It accelerates the evaluation of a drug efficacy and safety, offering a more streamlined route to drug development. While early discussions may reference Phase I/II trials, seamless designs are typically more relevant to Phase II/III stages. In this case, the seamless design begins like a standard Phase II trial, focusing on preliminary efficacy, optimal dose-finding, and safety profiling. Upon achieving certain predetermined criteria, the trial then transitions into the Phase III stage, expanding its scope to confirm efficacy and further evaluate safety in a larger patient population. This transition is often based on interim analyses, which are preplanned points where the accumulated data are reviewed to make decisions about the course of the trial.

Seamless trials can be classified as “operationally seamless” when their phases follow one another but remain distinct, and as “inferentially seamless” (or “adaptive seamless”) when information from all stages is incorporated into the final analysis [Maca et al., 2006]. Adaptive features allow modifications to the trial based on interim results without undermining its integrity or validity. These modifications can include dose adjustments, sample size re-estimation, or even the dropping of ineffective treatment arms.

However, seamless designs are not without challenges. They require careful planning, with clear criteria for transition and robust statistical methods to ensure that interim decisions do

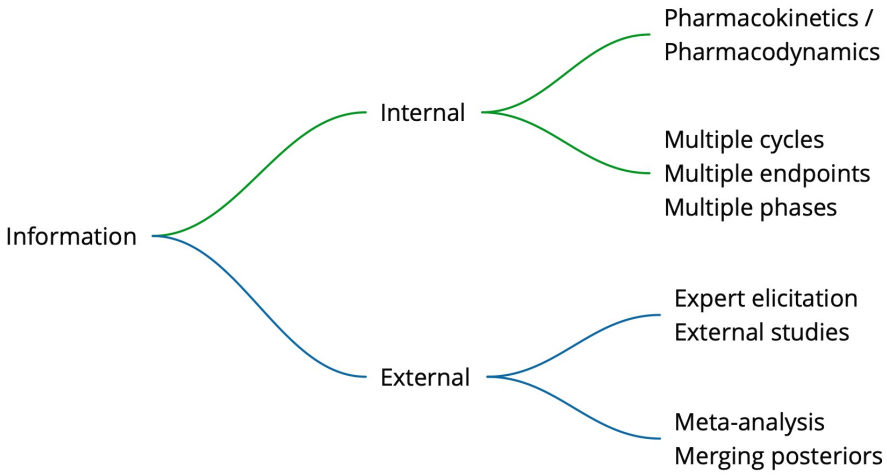


Figure 1.2 – Illustration of the classification scheme employed in this manuscript.

not introduce bias [Bretz et al., 2006, 2009]. Regulatory considerations must also be taken into account, as these designs can complicate the interpretation of results [U.S. Food and Drug Administration (FDA), 2018].

1.5 Outline of the manuscript

My contributions to enhancing clinical trials, whatever the phase, through additional information can be categorized into two main groups: the use of internal and external data sources, as illustrated in Figure 1.2.

Internal information refers to data generated within the trial itself. This includes my research on incorporating PK and PD measures in dose-finding Phase I trials [Ursino et al., 2017, Gerard et al., 2022]. Additionally, I have explored the utilization of data from multiple cycles [Ursino et al., 2022] and various toxicity endpoints [Ursino et al., 2019] within the same dose-finding context, as well as implementing a Bayesian seamless design that spans Phase II and III trials [Duputel et al., 2023].

On the other hand, external information encompasses the development of prior distributions via the analysis of data from previous clinical trials [Ollier et al., 2020] or through expert elicitation [Thall et al., 2019]. It also covers my work on meta-analysis of dose-finding studies [Ursino et al., 2021] and the consolidation of results from multiple small sample size (preclinical) studies [Boulet et al., 2024].

The following four chapters will delve into these categories in more detail. To maintain simplicity, comprehensive results or simulation setups are not provided here; readers are encouraged to consult the original papers for full insights. It is also worth noting that this summary does not encompass all my previous work, with additional papers mentioned in the discussions. Intentionally, I have chosen to highlight at least one piece of work where I serve as the primary or co-primary author, alongside one project under my supervision.

Chapter 2

Pharmacokinetics/Pharmacodynamics in dose-finding trials

Contents

In this chapter I will firstly illustrate the work I have done at the beginning of my first postdoc [Ursino et al., 2017]. My aim was to assess whether incorporating PK measurements into the dose-escalation process improves the final estimation or the decision-making in dose-finding studies. Therefore, the first source of internal information I worked on is PK measurements. This initial study was further developed by Emma Gerard, a PhD student under my co-supervision, to explore drug scheduling by integrating PD analysis [Gerard et al., 2022]. Unlike the first work, in this one we focused on the use of PK/PD only at the end of the clinical study, to make sure we could estimate complex models. In the remainder of this chapter, I will write a brief introduction to complement that of the previous chapter and show the reason behind the two research papers. Afterwards, I will summarize the two papers, focusing mainly on methods and results, and the last section will be devoted to discussion.

2.1 Introduction

In Phase I clinical trials, dose-finding and PK/PD assessments often occur simultaneously but are usually reported separately. A 2009 review of over 300 published Phase I oncology trials [Comets and Zohar, 2009] revealed that 84% acknowledged PK as a primary goal alongside determining the MTD, yet conducted and presented these analyses independently. Only 12% retrospectively linked toxicities or MTD to PK data, with no trials using PK data to influence dose allocation in real-time. Retrospective analyses showed a significant correlation between

toxicity and PK metrics in some studies [Broker et al., 2006, Ajani et al., 2005], but others found no clear relationship [Fracasso et al., 2005]. This suggests the potential, yet unexplored, impact of integrating PK data into dose allocation decisions on MTD determination. When I started to work in this topic, only few methods have been developed to integrate pharmacokinetics information into dose-finding studies. For instance, pharmacologically guided Phase I trials, as discussed in Collins et al. [1990], leverage preclinical PK data to determine the range of doses to be tested in dose-escalation studies, but not the dose-escalation procedure. A modification of the CRM, where the parametric dose-response function accounts for both drug dosage and a PK measure of exposure, were proposed by Piantadosi and Liu [1996]. An two-levels approach was suggested by Patterson et al. [1999] and Whitehead et al. [2007], who advocated for a Bayesian procedure with a nested hierarchical structure. This method utilizes PK/PD data to establish an overall dose-response relationship, treating PK/PD measurements as dependent variables in one model. Another application is found in O'Quigley et al. [2010], where, within the context of bridging studies, the dose corresponding to an average PK response was determined using linear regression models.

As briefly mentioned in the previous chapter, PK/PD analysis could inform the refinement of drug schedules. The National Cancer Institute (NCI) defines a drug schedule as a precise plan that outlines the method and timing for administering treatment, aiming to maximize safety while preserving the treatment potential efficacy for future use. Although it is logical to assume that higher doses may result in increased toxicity, determining the exact relationship between toxicity and various schedules for the same cumulative dose remains a challenge [Musuamba et al., 2017, Bullock et al., 2017]. Physicians can examine how changes in dose-sequence, the timing, and amount of dosages over time, impact patient outcomes. Research in immunology field suggests starting with a lower dose and gradually increasing it may lessen acute toxicities [Chen et al., 2019], while a higher initial loading dose could improve effectiveness. Therefore, dose-finding studies may investigate different dose sequences, even with identical cumulative doses, to find the optimal sequence. At the time of the starting of our work, several methodological works have focused on developing prospective approaches for determining the optimal schedule across various treatment regimens. Notably, works by Braun et al. [2005], Braun et al. [2007], Liu and Braun [2009], and Zhang and Braun [2013] have suggested focusing the evaluation of time to toxicity over the more traditional binary outcome, to refine both the dose and its timing. Wages et al. [2014] introduced the concept of viewing dose-schedule optimization as a two-dimensional challenge, adapting the partial order continual reassessment method (POCRM), which was originally designed for combination drug trials. Furthermore, Lyu et al. [2018] proposed a hybrid approach that combines algorithmic and model-based strategies for determining the sequence of dosages across multiple treatment cycles.However, only few approaches incorporate PK/PD data directly into the design for prospective dose allocation. For example, Günhan et al. [2020] introduced a Bayesian adaptive model based on time-to-event and a latent PK principle.

2.2 PK and dose-finding

In this work, I explored the influence of adding PK measurement for calculating the probabilities of toxicity (p_T) [Ursino et al., 2017]. The traditional approach directly links p_T with

the administered dose. This method requires a statistical model that establishes a relationship between p_T and the dose level. When adding the PK measures, two paths can be followed. In the first, PK measurements are added as covariates for p_T along with the dose. In the second, two separate estimation processes are involved. Initially, one model estimates the probability of toxicity given a specific PK measure of exposure or parameter (e.g., $P(Y = 1|AUC = x)$). Concurrently, another model predicts the distribution of the PK measure for any given dose ($P(AUC = x|d)$). The overall probability of toxicity for a dose is then derived by integrating these two models, formulated as $p_T(d) = \int_x P(Y = 1|AUC = x)P(AUC = x|d)$. Different models and variations have been tested in the cited article, however, for the sake of simplicity, only two main models are presented in this manuscript and compared to the standard CRM. Therefore, we are focusing at the estimation of the DLT probability at the end of the first therapy cycle.

Before describing the models, let us introduce the notation used in this work. Let us define d_k as a specific dose within the set D comprising K doses $(d_1, d_2, \ldots, d_K)$, and let d_i from D be the dose administered to the i-th participant. The variable y_i represents a binary outcome, taking the value 1 if the i-th participant experiences a DLT and 0 otherwise. Additionally, z_i denotes the logarithm of the AUC for the i-th patient's drug exposure. Then, p_T refers to the probability of experiencing toxicity, while $\boldsymbol{\beta}$ indicates the vector of regression coefficients (expressed in non-bold format when referring to a single parameter). The term n specifies the total number of participants in the study. The target probability of toxicity is represented by δ_T. When the symbol $\hat{}$ is placed above a parameter, it generally denotes the parameter posterior mean. Any additional symbols or notations specific to a particular model will be introduced as needed in the description of that model.

2.2.1 PKCOV

The first model, referred to as PKCOV, adapts the framework initially presented by Piantadosi and Liu [1996], maintaining the concept of incorporating PK as a covariate of the toxicity probability using the logit link. In this model, the dose-toxicity relationship is defined as:

$$\text{logit}(p_T(d_k, \Delta z_{d_k}, \boldsymbol{\beta})) = -\beta_0 + \beta_1 \log(d_k) + \beta_2 \Delta z_{d_k} \quad \forall d_k \in D,$$

where $\boldsymbol{\beta} = (\beta_1, \beta_2)$, β_0 is a constant, and Δz_{d_k} represents the difference between the population's logarithmic AUC at dose d_k and z, the individual's logarithmic AUC at the same dose. Assuming a underlying linear PK model, the two covariates, dose and Δz, should be uncorrelated as Δz depends on PK parameters like the clearance, not on the dose. Uniform priors have been assigned to β_1 and β_2, treating them as independent variables where the joint density function is expressed as $f(\beta_1, \beta_2) = (u_1 - l_1)^{-1}(u_2 - l_2)^{-1}$, with u_1 and l_1 denoting the upper and lower bounds of β_1's uniform distribution $U(l_1, u_1)$, and similarly, u_2 and l_2 for β_2, $U(l_2, u_2)$. The binomial likelihood after enrolling n patients is given by:

$$L_n(\boldsymbol{\beta}|\mathbf{y}) = \prod_{j=1}^{n} [p_T(d_j, \Delta z_{d_j}, \boldsymbol{\beta})]^{y_j} [1 - p_T(d_j, \Delta z_{d_j}, \boldsymbol{\beta})]^{1-y_j},$$

where p_T is substituted with its specific expression derived from inverting the equation above.

Dose Allocation Rules: The selection of the dose for the subsequent cohort is based on identifying the dose d_k from the set D whose posterior probability of toxicity,

$$p_T(d_k, \hat{\beta}_1) = \frac{1}{1 + e^{\beta_0 - \hat{\beta}_1 \log(d_k)}} \quad d_k \in D, \tag{2.1}$$

is nearest to the predetermined target probability. In other words, the next dose to be administered, d_{i+1}, is the one that minimizes the absolute difference between $p_T(d_k, \hat{\beta}_1)$ and the target probability θ, formalized as $d_{i+1} = \text{argmin}_{d_k} |p_T(d_k, \hat{\beta}_1) - \delta_T|$. Following the original methodology, Δz_{d_k} is set to zero, its expected value, in Eq. 2.2.1. As an additional rule, the 'no-skipping dose' guideline ensures that d_{i+1} is chosen from $D^* \subset D$, a subset of D that includes all previously tested doses plus the next level up, assuming not all doses have been explored yet. The final estimate for the MTD is represented by d_{n+1}, indicating the dose recommended for the (n+1)th participant if they were to be enrolled in the trial.

2.2.2 PKLOGIT

The second approach, which we refer to as PKLOGIT, adapts a model originally developed by Patterson et al. [1999] and Whitehead et al. [2001] for hierarchical modeling in cross-over trials. By modifying this model for use in non-cross-over trials, specifically by eliminating random effects and implementing an underlying one-compartment PK model, we establish a Bayesian linear regression framework for z, that is

$$z_i | \boldsymbol{\beta}, \nu \sim N\left(\beta_0 + \beta_1 \log d_i, \nu^2\right).$$

Here, $\boldsymbol{\beta} = (\beta_0, \beta_1)$, and the prior distributions for $\boldsymbol{\beta}|\nu$ and ν are given by $\boldsymbol{\beta}|\nu \sim N_2\left(\mathbf{m}, \nu^2 \mathbf{G}\right)$ and $\nu \sim \text{Beta}\,(a, b)$, respectively. The matrix $\mathbf{G}$ is a diagonal matrix, and the constants $\mathbf{m}$, $\mathbf{G}$, a, and b are selected based on prior knowledge. Diverging from the original formulation, a beta distribution is chosen for ν to account for its association with the inter-individual variability in clearance, a key PK parameter, with the assumption that its value lies between 0 and 1, reflecting the expected range of variability. Then, drawing inspiration from Whitehead et al. [2007], we utilize z_i instead of the dose d_i as the covariate in the logit function for the probability of toxicity (p_T), that is,

$$\text{logit}(p_T(z, \boldsymbol{\beta})) = -\beta_3 + \beta_4 z,$$

where the prior distribution for $\boldsymbol{\beta} = (\beta_3, \beta_4)$ is a bivariate uniform distribution, defined as $f(\beta_3, \beta_4) = (u_3 - l_3)^{-1}(u_4 - l_4)^{-1}$, with u_3 and l_3 being the upper and lower bounds of the uniform distribution for β_3, $U(l_3, u_3)$, and u_4 and l_4 for β_4, $U(l_4, u_4)$.

Dose Allocation Rules: the dose for the next patient is selected based on the predictive probability that closely matches the target probability δ_T, as defined by:

$$P(y_{i+1} = 1 | \hat{\boldsymbol{\beta}}_i) = E\left[\frac{1}{1 + e^{\hat{\beta}_3 - \hat{\beta}_4 z}}\right] = \int \frac{1}{1 + e^{\hat{\beta}_3 - \hat{\beta}_4 z}} g(z)\, dz,$$

where g represents the predictive normal density of z given $\hat{\boldsymbol{\beta}}$ and now $\boldsymbol{\beta} = (\beta_1, \beta_2, \beta_3, \beta_4)$. Similar to the previous method, a no-skipping dose rule is implemented, and the MTD is identified as the dose $d_k \in D$ recommended for the $(n+1)$th subject.

2.2.3 Results

To assess the performance of the methods, we conducted a simulation study. In our simulation framework, the relationship between toxicity and PK exposure, specifically the AUC, is a focal point. This necessitates a departure from the conventional method of simulating toxicity in dose-finding studies, where probabilities of experiencing toxic effects are pre-defined for each dose level and patient responses are simulated using Bernoulli distributions. Instead, our approach requires the simulation of PK data as a preliminary step, subsequently linking this data to toxicity outcomes. To implement this method, we utilized a case study based on the PK model for the TGF-β inhibitor LY2157299 in patients with glioma, as detailed in previous research by Gueorguieva et al. [2014], Bueno et al. [2008] and Lestini et al. [2015]. We simulated several scenarios with 6 doses and different position of the MTD. It was achieved modifying the threshold on AUC to define the DLT occurrence or increasing and decreasing the inter-individual variability in PK parameters. More details are given in Ursino et al. [2017]. As anticipated, the classical CRM was included for comparison.

In the simulation study, 1000 trials were conducted for each scenario. Each trial involved a cohort size of a single patient, a total fixed sample size of 30 patients and a target toxicity of 20%. The trial started with a one-by-one dose escalation, moving sequentially from the lowest to the next dose level for each new patient. This escalation continued until the occurrence of the first instance of toxicity. At this point, the trial transitioned to a two-stage design, incorporating the chosen model for further dose evaluation and decisions. If toxicity is observed at any point during the trial, the specified methods are applied at the conclusion of the trial to make comprehensive assessments and recommendations.

Table 2.1 shows the results in terms of percentage of final selection, percentage of allocation and number of DLTs in four scenarios extracted from the original article Ursino et al. [2017]. PKCOV and PKLOGIT demonstrate comparable performance metrics in terms of their ability to identify MTD. PKCOV shows a slight tendency towards recommending higher doses, which could increase the risk of overdosing. However, they are both outperformed by the CRM. In further analyses, we explored the precision of each method in estimating the dose-toxicity relationship. This was conducted by examining their estimates of the probability of toxicity across various tested doses, with sample sizes incrementally increasing from 20 to 100. This approach was designed to assess how effectively and accurately each method converges to the true dose-toxicity relationship as more data become available. For scenario D, the findings are illustrated in Figure 2.1 where the true probabilities of toxicity, utilized in the simulations, are marked by black horizontal lines. Each curve within the figure represents the median estimate of the probability of toxicity for each method across 1000 simulations, offering a visual comparison of their performance. Figure 2.1 demonstrates that all methods accurately estimate the probability of toxicity at the MTD, aligning with earlier findings regarding their efficacy in MTD selection. An interesting observation from the analysis is the distinct capability of PKLOGIT to accurately estimate the probability of toxicity across a broader range of dose levels, beyond just the MTD and its adjacent doses. This is evidenced by the true probability of toxicity falling within the first and third quartiles of the estimated probability distribution.

Table 2.1 – Percentage of dose selection at the end of the trials, percentage of dose allocation in parenthesis and median, minimum and maximum number of DLT for several scenarios. Real percentage of toxicity of each dose is written in italics.

Method	% dose selection						number of DLTs	
	1	2	3	4	5	6	median (n)	min - max
Scenario A	*0.001*	*0.05*	*0.10*	*0.20*	*0.35*	*0.45*		
PKCOV	0.054	0.015	0.177	**0.550**	0.163	0.041	6	1 - 11
	(0.087)	(0.067)	(0.188)	(0.370)	(0.172)	(0.116)		
PKLOGIT	0.066	0.032	0.276	**0.530**	0.088	0.008	5	1 - 10
	(0.117)	(0.105)	(0.251)	(0.350)	(0.112)	(0.065)		
CRM	0.020	0.014	0.196	**0.600**	0.161	0.009	6	1 - 11
	(0.058)	(0.076)	(0.215)	(0.410)	(0.176)	(0.065)		
Scenario B	*0.021*	*0.139*	*0.199*	*0.29*	*0.40*	*0.467*		
PKCOV	0.190	0.161	**0.358**	0.243	0.040	0.008	7	1 - 13
	(0.207)	(0.166)	(0.266)	(0.226)	(0.081)	(0.054)		
PKLOGIT	0.169	0.283	**0.336**	0.172	0.029	0.011	6	1 - 12
	(0.235)	(0.257)	(0.239)	(0.165)	(0.056)	(0.048)		
CRM	0.031	0.239	**0.436**	0.262	0.029	0.003	6	1 - 12
	(0.096)	(0.239)	(0.313)	(0.245)	(0.072)	(0.034)		
Scenario C	*0*	*0*	*0.001*	*0.025*	*0.184*	*0.385*		
PKCOV	0	0	0	0.080	**0.672**	0.248	6	2 - 10
	(0.033)	(0.033)	(0.038)	(0.164)	(0.424)	(0.308)		
PKLOGIT	0	0	0	0.143	**0.747**	0.110	5	1 - 10
	(0.034)	(0.036)	(0.055)	(0.240)	(0.443)	(0.193)		
CRM	0	0	0	0.093	**0.762**	0.145	5	2 - 9
	(0.033)	(0.033)	(0.038)	(0.183)	(0.492)	(0.220)		
Scenario D	*0.019*	*0.135*	*0.195*	*0.286*	*0.398*	*0.466*		
PKCOV	0.193	0.144	**0.322**	0.287	0.048	0.006	6.5	1 - 13
	(0.218)	(0.163)	(0.242)	(0.239)	(0.085)	(0.053)		
PKLOGIT	0.197	0.212	**0.330**	0.208	0.040	0.013	6	1 - 12
	(0.242)	(0.213)	(0.250)	(0.185)	(0.067)	(0.043)		
CRM	0.015	0.240	**0.429**	0.275	0.038	0.003	6	1 - 12
	(0.084)	(0.249)	(0.308)	(0.248)	(0.077)	(0.033)		

2.3 PK/PD and drug schedule

Since we saw that incorporating PK into the dose-allocation procedure in dose-finding clinical studies enhances understanding of the relationship between dosage and toxicity, we explore if it could help in understanding more complex relationship. The inspiration for this study came from the first-in-human dose escalation trial of SAR440234, given as a monotherapy to patients with relapsed or refractory acute myeloid leukemia, high-risk myelodysplastic syndrome, or B-cell acute lymphoblastic leukemia (NCT03594955). The therapy induces the release of inflammatory cytokines, which can lead to cytokine release syndrome (CRS). CRS, a systemic inflammatory response, represents a frequent adverse effect of T-cell engaging bispecific antibodies, like blinatumomab, a dual-targeting anti-CD19/CD3 antibody [Shimabukuro-Vornhagen et al., 2018]. In literature, the correlation between cytokine peaks and CRS severity has been explored [Teachey et al., 2016], with findings indicating that adjusting the drug dosage can mitigate CRS effects, especially when the initial dose is administered in progressively smaller steps [Chen et al., 2019]. Consequently, this study adopted an intrapatient dose escalation strategy, starting with lower

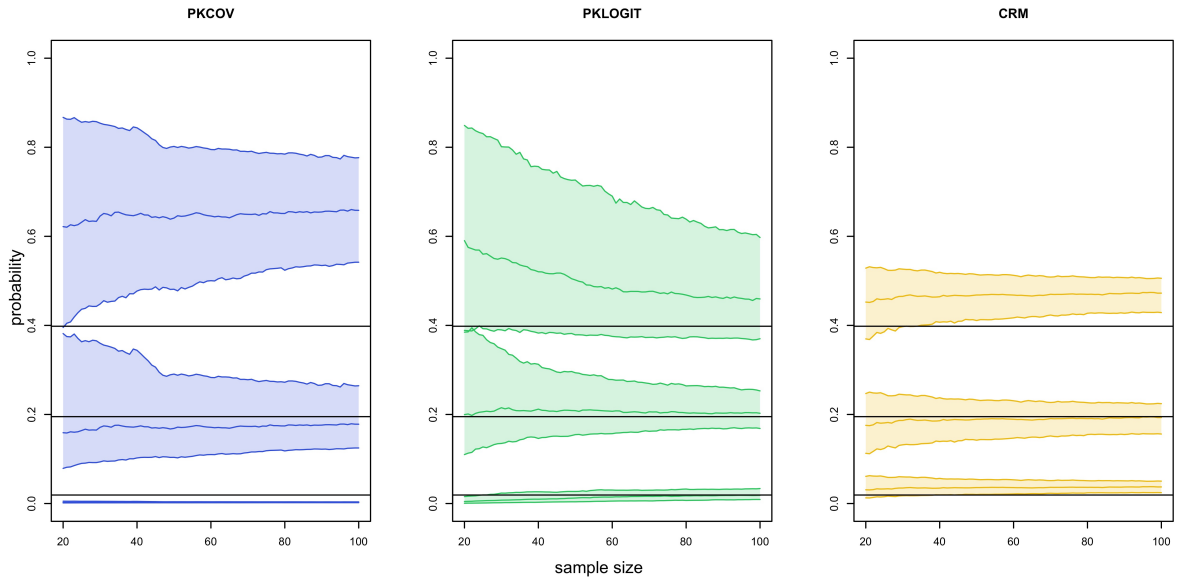


Figure 2.1 – Plot of three estimated probabilities (p1, p3 and p5) of toxicity versus sample size for scenario D for PKCOV, PKLOGIT and CRM. Each quadrant shows the three estimated probabilities using a method. Median, first and third quartile over 1000 simulations of the corresponding estimated probabilities are plotted. The black horizontal lines represent the true probability. From below, the first is for p1, the second for p3 and the last for p5.

doses and gradually increasing to a higher maintenance dose, to minimize CRS occurences [Boissel et al., 2018]. Figure 2.2 shows how the intrapatient dose escalation could reduce the cytokine peak. We suggest retrospectively modeling the binary toxicity endpoint alongside the cytokine profile, which serves as the continuous PD response, to understand the relationship between dose sequence and PD. This approach aims to identify the maximum tolerated dose sequence (MTDS) by the trial conclusion. The work was performed by Emma Gerard, a PhD student (CIFRE) I have co-supervised [Gerard et al., 2022]. In the following, a summary of the work and the results.

Similar to the earlier section, we begin by introducing new notation for this part. Define $\mathcal{D} = \{d_1, ..., d_L\}$ as the set of possible doses for patient administration, with $d_l < d_{l+1}$. The subset $\mathcal{S} = \{S_1, ..., S_K\}$ represents the set of dose sequences to be evaluated in the trial. Each dose sequence, $S_k \in \mathcal{S}$, for $k \in \{1, ..., K\}$, consists of a series of J doses, $S_k = (d_{k,1}, d_{k,2}, ..., d_{k,J})$, administered at specific times $t = (t_1, t_2, ..., t_J)$, where each $d_{k,j}$ belongs to $\mathcal{D}$ for $j \in 1, ..., J$. The subsequence $S_{k,j}$ includes doses in S_k up to the j^{th} dose, $S_{k,j} = (d_{k,1}, d_{k,2}, ..., d_{k,j})$, for $j < J$. Let the trial include $n \in \mathbb{N}$ patients. The binary toxicity response for patient i after the j^{th} dose is $Y_{i,j}$, and Y_i represents the overall toxicity response for patient i across all administrations. Consider $\tilde{s}_i = (d_{i,1}, d_{i,2}, ..., d_{i,J}) \in \mathcal{S}$ as the designated dose-sequence for the i^{th} patient. It is assumed that the treatment is discontinued upon the occurrence of toxicity, with j_i marking the final dose administered to patient i. The actual sequence administered, denoted as $s_i = (d_{i,1}, d_{i,2}, ..., d_{i,j_i})$, is a subset of $\tilde{s}_i$, where s_i equals $\tilde{s}_i$ if no toxicity is observed. The subsequence $s_{i,j}$ represents the portion of s_i up to the j^{th} dose, with $j \leq j_i$.

The objective is to identify the MTDS by the trial end, defined as the dose-sequence with a

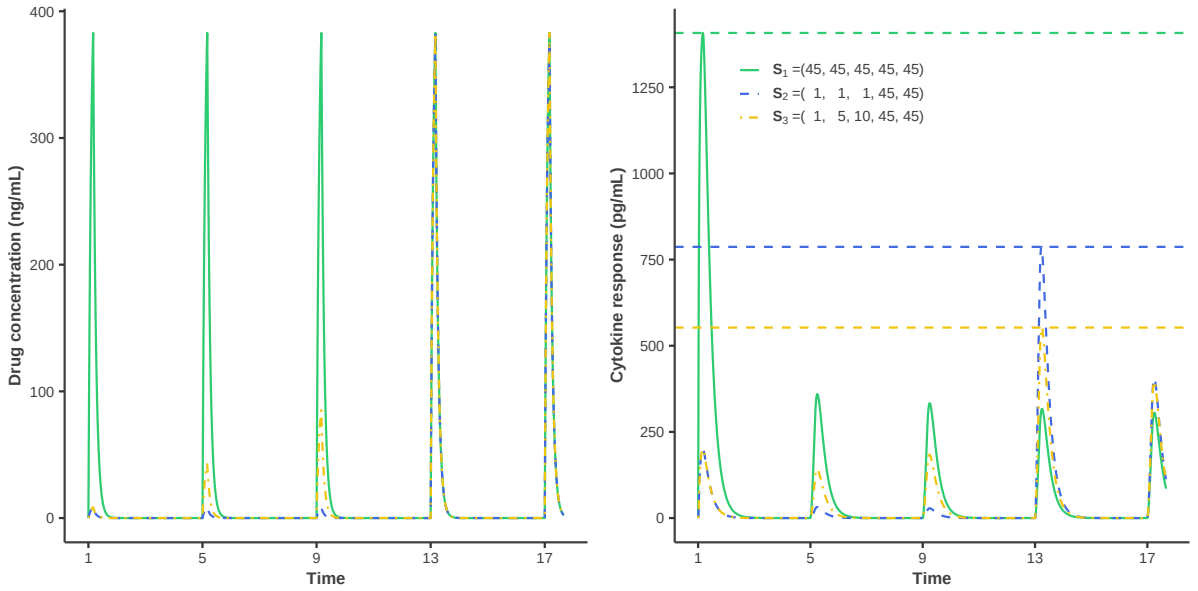


Figure 2.2 – Concentration (left) and cytokine (right) profiles of three patients, each one receiving a different dose-sequence with intrapatient escalation administered on days 1, 5, 9, 13, and 17. Horizontal lines represent the maximum peak of cytokine observed after each dose-sequence.

toxicity probability nearest to a predefined target toxicity rate δ_T. Thus, MTDS $= S_{k^\star}$, where $k^\star = \arg\min_k |p_T(S_k) - \delta_T|$.

We assumed that a PD endpoint, computed from a continuous PD profile of a biomarker, serves as a link in the dose-toxicity relationship. A first model will focus on the relationship between the dose-sequence and the PD endpoint, and a second model will describe the connection between the PD endpoint and the toxicity response. Combining these models will create a pathway from the dose-sequence to the toxicity response, aiding in the determination of the MTDS.

2.3.1 Dose sequence - PD response model

This model represents a classical population PK/PD model. Let $C(t)$ represent the continuous concentration of the drug, and $E(t)$ denote the continuous PD response related to toxicity at time t. These can be described by nonlinear mixed models as follows:

$$\begin{aligned} C(t) &= f^{(1)}\left(\theta_i^{(1)}, t\right) + g^{(1)}\left(\theta_i^{(1)}, t, \xi_1\right)\varepsilon^{(1)}, \\ E(t) &= f^{(2)}\left(\theta_i^{(2)}, t\right) + g^{(2)}\left(\theta_i^{(2)}, t, \xi_2\right)\varepsilon^{(2)}, \end{aligned} \tag{2.2}$$

where $f^{(1)}$ and $f^{(2)}$ are the structural models that typically arise from differential equations grounded in biological principles. The parameter vector specific to the ith patient, $\theta_i = \left(\theta_i^{(1)}, \theta_i^{(2)}\right)$, often takes the form $\theta_i = \mu e^{\eta_i}$, incorporating both fixed effects μ and random effects η_i. The random effects η_i follow a normal distribution with mean zero and variance-covariance matrix Ω, denoted as $\eta_i \sim \mathcal{N}(0, \Omega)$. This framework allows for the modeling of individual patient variability in drug concentration and PD response over time.

The functions $g^{(1)}$ and $g^{(2)}$ represents the error models, which are influenced by additional parameters ξ_1 and ξ_2, respectively. These error models account for the variability in the data that is not explained by the structural models. The variables $\varepsilon^{(1)}$ and $\varepsilon^{(2)}$ represent standard Gaussian noise factors that introduce randomness into the system. Common types of error models include: (i) the constant model, where the error function is defined as $g^{(l)}\left(\theta_i^{(l)}, t, \xi_l = a\right) = a$; (ii) the proportional model, where the error function is $g^{(l)}\left(\theta_i^{(l)}, t, \xi_l = b\right) = bf^{(l)}\left(\theta_i^{(l)}, t\right)$; and (iii) the combinations of the constant and proportional models. The models selected for this study, which also generated Figure 2.2, are detailed in the original manuscript.

2.3.2 PD endpoint - toxicity model

We proposed two ways to model the link between the PD endpoint and the toxicity. We use $r : r(\theta_i, s_{i,j})$ to represent a function emerging from the PK/PD models, which calculates the value of the PD endpoint (such as the peak level of a biomarker) following the administration of the dose sequence $s_{i,j}$, considering the individual's PK/PD parameters θ_i. Additionally, $R : R(\theta_i, s_{i,j})$ is defined as a function that outputs a vector containing all PD endpoints (like all biomarker peaks) observed after the administration of the sequence $s_{i,j}$, tailored to the individual's PK/PD parameters θ_i. For convenience, for patient i, these notions are simplified as $r_{i,j} = r\left(\theta_i, s_{i,j}\right)$ and $R_{i,j} = R\left(\theta_i, s_{i,j}\right)$, with R_i denoting the vector of all PD endpoints up to R_{i,j_i}. Furthermore, let $r_i^M = \max_{l\in\{1,\ldots,j_i\}}(r_{i,l})$ symbolize the summary PD endpoint (such as the highest biomarker peak), presumed to correlate with toxicity, recorded for patient i.

To establish the prior distributions, we use $(\bar{r}_1^M, \bar{r}_2^M, \ldots \bar{r}_K^M)$ to represent the reference values of the maximum PD endpoint across all dose sequences in the trial, denoted by $(S_1, \ldots, S_K)$. These reference values can be derived from population averages, where $\bar{r}_k^M = \max\left(r\left(\mu, S_{k,1}\right), \ldots, r\left(\mu, S_k\right)\right)$, with μ representing the vector of fixed effects in the PK/PD model. In the next sections, the two proposed statistical models are presented.

Bayesian PD logistical model

We suggest a Bayesian logistic model to establish a connection between the overall binary toxicity response of patient i, who receives the sequence s_i, and their summary PD endpoint indicative of toxicity. The relationship is formulated as follows:

$$\text{logit}\left(\mathbb{P}\left(Y_i = 1\right)\right) = \beta_0 + \beta_1 \log\left(\frac{r_i^M}{\bar{r}_{k_T}^M}\right) = \text{logit}\left(\pi_1\left(\left(\beta_0, \beta_1\right), r_i^M\right)\right). \tag{2.3}$$

In this model, a positive β_1 indicates that the probability of toxicity increases with the value of the summary PD endpoint. The PD endpoint is normalized for the purpose of prior elicitation using $\bar{r}_{k_T}^M$, which is the reference value for the dose-sequence S_{k_T} anticipated to have a toxicity probability of δ_T. This approach simplifies the complexity by not accounting for the longitudinal biomarker values, operating under the assumption that toxicity is directly related to the summary PD endpoint rather than a cumulative effect of the biomarker profile. Nevertheless, the influence of previous drug administrations is implicitly incorporated in the determination of the biomarker through the PK/PD model.

In setting up the prior distributions for our Bayesian logistic model, we adopt a normal distribution for the intercept, β_0, expressed as $\beta_0 \sim \mathcal{N}(\bar{\beta}_0, \sigma^2_{\beta_0})$, and a gamma distribution for the slope, β_1, to guarantee its positivity, formulated as $\beta_1 \sim \gamma(\alpha_1, \frac{\alpha_1}{\bar{\beta}_1})$. Here, α_1 is the shape parameter of the gamma distribution, with $\bar{\beta}_0$ and $\bar{\beta}_1$ representing the expected values (means) of β_0 and β_1, respectively. The mean of the intercept, $\bar{\beta}_0$, is determined to be the logit transformation of the target toxicity rate δ_T, derived from the logistic model equation with $r_i^M = \bar{r}_{k_T}^M$. This sets a baseline around which the model intercept is centered. To estimate $\bar{\beta}_1$, the slope's mean, initial guesses of the toxicity probabilities for each dose sequence $(S_1, ..., S_K)$ are utilized, with $p_{k_T} = \delta_T$ for the reference sequence S_{k_T}. The estimation of $\bar{\beta}_1$ can proceed in two ways:

1. Using only one dose-sequence, not including the reference sequence S_{k_T}, by setting $\pi_1\left(\left(\bar{\beta}_0, \bar{\beta}_1\right), \bar{r}_k^M\right) = p_k$ for a single $k \neq k_T$.
2. Employing multiple sequences, specifically those adjacent to the reference sequence (e.g., S_{k_T-1} and S_{k_T+1}), to minimize the squared difference between the estimated and actual toxicity probabilities for these sequences:

$$\bar{\beta}_1 = \arg\min_{\beta_1} \sum_{k=k_T-1}^{k_T+1} \left(p_k - \pi_1\left(\left(\bar{\beta}_0, \beta_1\right), \bar{r}_k^M\right)\right)^2.$$

Bayesian PD hierarchical model

In this methodology, we posit that toxicity occurs in patients when their PD response surpasses a patient-specific, yet unidentified, threshold. To accommodate the interindividual variability in susceptibility to toxicity, we introduce a continuous latent variable, Z_i, for each patient. This variable symbolizes the unique toxicity threshold corresponding to the PD response for that individual. Diverging from earlier strategies, we now approach the modeling of toxicity following each drug administration through a Bayesian hierarchical model:

$$\begin{cases} Y_{i,j} = \begin{cases} 0 & \text{if } Z_i > \log\left(\dfrac{r_{i,j}}{\bar{r}_{k_{50}}^M}\right) \\ 1 & \text{if } Z_i \leq \log\left(\dfrac{r_{i,j}}{\bar{r}_{k_{50}}^M}\right) \end{cases} \\ Z_i \sim \mathcal{N}(\mu_z, \tau_z^2) \end{cases} \tag{2.4}$$

where $\bar{r}_{k_{50}}^M$ serves as the reference value for the dose-sequence $S_{k_{50}}$, which is initially estimated to have a 50% probability of inducing toxicity in patients. Although this model incorporates elements reminiscent of a probit model, primarily through its handling of variability and threshold effects, it diverges in its core mechanism. Specifically, the initial part of the equation operates as a deterministic step function, indicating a direct, non-probabilistic transition once the PD response exceeds a certain threshold. The inclusion of a random effect introduces a layer of complexity that aligns the model with certain aspects of probit models. Here, τ_z^2 symbolizes the variance between subjects, reflecting the diversity in individual toxicity thresholds. This parameter controls how information is shared or "borrowed" across the patient population.

When evaluating a new patient i with a vector of biomarker endpoints R_i, the probability of experiencing toxicity can be predicted using the formula $\mathbb{P}(Y_i = 1) = F_z\left(\log\left(\frac{r_i^M}{\bar{r}_{k_{50}}^M}\right)\right) = \pi_2\left(\left(\mu_z, \tau_z^2\right), r_i^M\right)$, where F_z represents the cumulative distribution function (CDF) of a normal distribution with mean μ_z and variance τ_z^2. Regarding prior distributions, for μ_z, a normal distribution $\mu_z \sim \mathcal{N}(0, \sigma_{\mu_z}^2)$ is chosen, reflecting a neutral initial stance regarding the central tendency of the toxicity threshold distribution across the population. The choice of a half-Cauchy distribution for τ_z, $\tau_z \sim$ half-Cauchy$(0, \sigma_{\tau_z}^2)$, follows the guidance of Gelman [2006], acknowledging the potential for τ_z to be close to zero.

2.3.3 Dose sequence-toxicity model

The posterior toxicity probability for a given dose sequence S_k is calculated by integrating over all potential values of the PD endpoint within the toxicity model. Due to the complexity of this integration, which often precludes an analytical solution, the posterior toxicity probability for sequence S_k is approximated through simulation. This involves generating an M-sized hypothetical cohort of patients, each assigned a posterior toxicity probability, resulting in an M-vector of probabilities $(p_T(S_k)^{(1)}, ..., p_T(S_k)^{(M)})$. The overall posterior toxicity probability for sequence S_k is then determined by averaging these individual probabilities, yielding the posterior mean $\widehat{p_T}(S_k) = \dfrac{1}{M}\displaystyle\sum_{m=1}^{M} p_T(S_k)^{(m)}$.

2.3.4 Results

In our study, we explored four distinct toxicity scenarios altering the positioning of the MTDS within each scenario. We analyzed six dose-sequences across all scenarios, with each sequence consisting of seven dose administrations scheduled on days 1, 5, 9, 13, 17, 21, and 25. More details on PK/PD and toxicity data generation are given in the published article [Gerard et al., 2022]. For the first three scenarios, the MTDS was identified at different sequences: in scenario 1, the MTDS was found at sequence S_4; scenario 2 placed the MTDS at sequence S_2; in scenario 3, the MTDS shifted to sequence S_6. Scenario 4 presented a unique case where two sequences, S_4 and S_5, were both viable candidates for the MTDS designation.

In our analysis, we simulated 1000 trials for each of the four toxicity scenarios, targeting a toxicity probability (δ_T) of 0.3. We focused on assessing the impact of two traditional dose escalation designs: the 3+3 design and the CRM with a 2-parameter logistic regression model. These methods were applied with cohorts of size 3, aiming for a total sample size of 30 patients in each trial. For the CRM, we based the working model (the initial guesses of the toxicity probabilities) on data from the first scenario. Table 2.2 shows the results in terms of proportions of correct selection (PCS) and the mean sample size. Across all scenarios, the PCS for both the Bayesian PD logistical model and the Bayesian PD hierarchical model showed similarity, indicating that both proposed models are effective in identifying the MTDS. Moreover, these methods demonstrated a superior performance over traditional dose escalation designs in the majority of scenarios tested. For scenarios 1, 2, and 4, our methods were able to accurately select the MTDS in over 10% more trials than the 3+3 design. Particularly noteworthy was scenario

Table 2.2 – Proportions that each sequence is being selected as the MTDS over the 1000 trials for the 4 toxicity scenarios and the 2 dose allocation designs, either the 3+3 design or the CRM. For each scenario, the PCS on the true MTDS are represented in bold. For each dose allocation design, the mean sample size at each dose-sequence is displayed.

		S_1	S_2	S_3	S_4	S_5	S_6
Scenario 1		**0.08**	**0.11**	**0.15**	**0.3**	**0.44**	**0.52**
3+3	Mean sample size	3.6	3.5	3.5	**3**	1.6	0.4
	Bayesian PD logistical model	8.6	5.9	19	**42.2**	19.6	4.7
	Bayesian PD hierarchical model	7.5	7.6	19.1	**43.8**	18.6	3.4
	3+3	13.9	16.1	32.2	**27.6**	8.6	1.6
CRM	Mean sample size	4.2	3.7	5.6	**8.8**	5.6	2.1
	Bayesian PD logistical model	0	1.2	15.5	**64.6**	15.5	3.2
	Bayesian PD hierarchical model	0	0.8	12.8	**64.3**	19.4	2.7
	Logistic CRM	0	1.4	15.1	**50.4**	27.1	6
Scenario 2		**0.15**	**0.3**	**0.44**	**0.52**	**0.69**	**0.83**
3+3	Mean sample size	4	**3.6**	1.8	0.5	0.1	0
	Bayesian PD logistical model	27.2	**42.5**	24.7	5.2	0.4	0
	Bayesian PD hierarchical model	29.3	**41.2**	24.3	4.8	0.4	0
	3+3	57.3	**31**	9.8	1.7	0.2	0
CRM	Mean sample size	8.7	**11.1**	7.5	2.3	0.3	0
	Bayesian PD logistical model	14.8	**65.9**	17.4	1.7	0.2	0
	Bayesian PD hierarchical model	12.3	**66.2**	18.9	2.6	0	0
	Logistic CRM	12.5	**56**	26.7	4.7	0.1	0
Scenario 3		**0.01**	**0.02**	**0.05**	**0.09**	**0.17**	**0.3**
3+3	Mean sample size	3.1	3.1	3.4	3.6	3.5	**3**
	Bayesian PD logistical model	0.3	0.6	2	9.9	21.9	**65.3**
	Bayesian PD hierarchical model	0.1	0.8	2	9.3	23.8	**64**
	3+3	0.9	2	9.3	17.5	33.9	**36.4**
CRM	Mean sample size	3.1	3	3.1	3.6	5.4	**11.8**
	Bayesian PD logistical model	0	0	0	0.8	21.2	**78**
	Bayesian PD hierarchical model	0	0	0	0.4	19.9	**79.7**
	Logistic CRM	0	0	0	1.1	19.2	**79.7**
Scenario 4		**0.04**	**0.08**	**0.15**	**0.26**	**0.32**	**0.43**
3+3	Mean sample size	3.4	3.6	3.6	**3.1**	**2**	0.9
	Bayesian PD logistical model	3.5	5.5	21.1	**30.2**	**23**	16.7
	Bayesian PD hierarchical model	2.7	6.2	19.6	**33.3**	**24**	14.2
	3+3	8.4	17.2	29.2	**24.1**	**15.4**	5.7
CRM	Mean sample size	3.5	3.5	5.2	**7**	**5.9**	4.8
	Bayesian PD logistical model	0	0.2	11.6	**34.7**	**30.2**	23.3
	Bayesian PD hierarchical model	0	0.2	9.2	**37.3**	**34**	19.3
	Logistic CRM	0	0.6	9.7	**28**	**34.4**	27.3

3, where our methods exceeded the 3+3 design's accuracy by more than 20%. Compared to the CRM, our methods showed an improvement of approximately 10% in PCS for scenarios 1 and 2, underscoring their robustness in these settings. In scenario 3, the performance of all methods was closely matched. Scenario 4 presented a nuanced outcome; the Bayesian PD logistical model and CRM achieved similar overall PCS rates (64.9% and 62.4%, respectively), but the Bayesian PD hierarchical model outperformed both (with a PCS of 71.3%), showcasing its particular strength in this scenario.

As expected, one of the primary advantages of our methods is their ability to model the comprehensive relationship between dose-sequences and toxicity. This feature is particularly valuable because it allows for the recommendation of dose-sequences not originally included in the trial for further investigation in subsequent expansion studies. Consider a hypothetical scenario within a CRM design framework where the panel of dose-sequences overlooks the actual MTDS. In this scenario, dose-sequence $S_3 = (5, 10, 25, 50, 50, 50, 50)$ μg/kg results in underdosing, while sequence $S_4 = (10, 25, 50, 100, 150, 150, 150)$ μg/kg leads to overdosing (Figure 2.3). The estimated toxicity probabilities highlight a discernible gap between sequences S_3 and S_4, suggesting the existence of an optimal dose-sequence that has not been tested yet, potentially lying between these two sequences. Our methods ability to predict the toxicity probability of a new, untested sequence, $S_{\text{new}} = (10, 25, 50, 100, 100, 100, 100)$ μg/kg, demonstrates a significant advantage over the CRM. The CRM limitation stems from its reliance on a predefined working model based on the set of tested sequences, rendering it incapable of estimating the toxicity probabilities for sequences that were not part of the original trial design. In contrast, both of our proposed Bayesian models, through their comprehensive modeling of the dose-toxicity relationship, enable the prediction of toxicity probabilities for new sequences. This predictive power is illustrated in Figure 2.3, where S_{new} is shown to have a toxicity probability aligning more closely with the target toxicity level. Consequently, this new sequence can be recommended for testing in expansion cohorts.

2.4 Discussion

Dose-finding methods can be broadly classified according to two different types of approaches, the "dose-estimators" and the "dose-finders" [Rosenberger and Haines, 2002]. In the first case, the aim is to estimate the entire dose-response curve, and to determine the dose associated with a given percentile of the dose-response curve. Model-based approaches combined with optimal design belong to this framework [Bornkamp et al., 2007]. The second case is the approach usually applied in the oncology setting. The aim is to hone in on the MTD between the doses tested in the trial, without estimating the entire dose-response (dose-toxicity) curve [O'Quigley et al., 1990, Whitehead and Brunier, 1995], implying that the results are mostly valid for the doses tested. The CRM, since it is built directly to take care of a discrete number of doses, is a "dose-finder" method. On the other hand, PK methods are more "dose-estimators" methods, since they try to estimate the entire dose-response curve, and only in the last step of the method do they focus on the discrete set of candidate doses. In that case, extrapolation (or interpolation) to the entire dose-toxicity curve is a natural extension. Our results show that dose-finder approaches hone in on the MTD in an efficient way, but that the probability of toxicity was estimated poorly

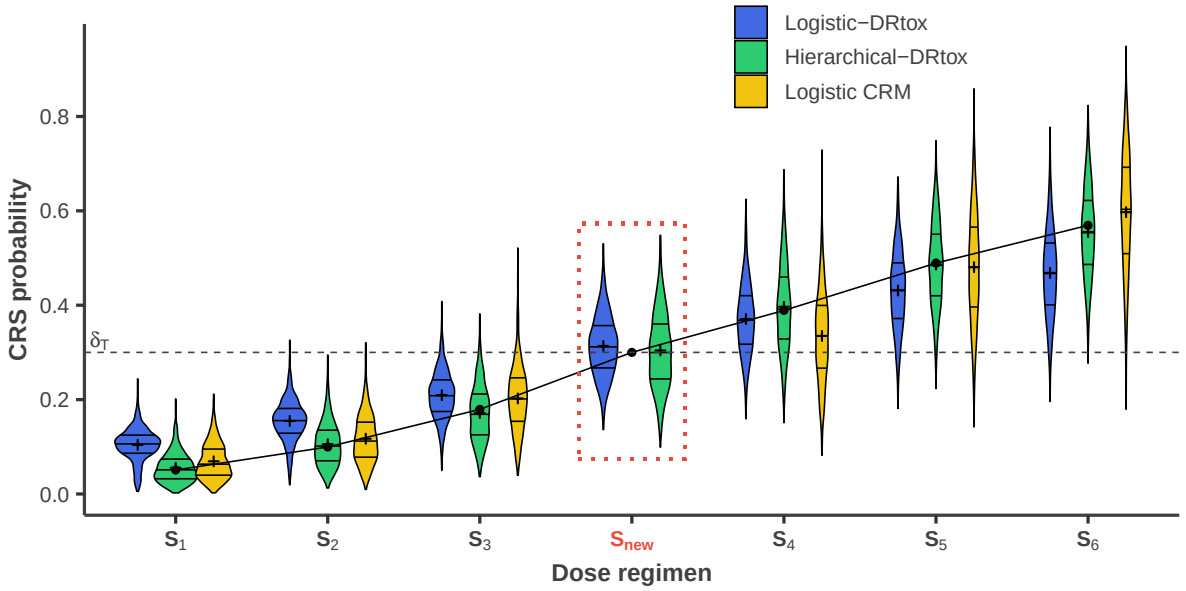


Figure 2.3 – Violin plots of the estimated toxicity probabilities, for an additional scenario where the panel of dose-sequences missed the true MTDS, on 1000 CRM trials with 30 patients each. Horizontal lines on the density estimates represent the median and first and third quantiles of the distributions and the plus sign represents the mean. The dashed line represents the toxicity target and the solid line represents the true toxicity probabilities.

for doses far from the MTD. Dose-estimators on the other hand allowed us to better estimate the probabilities of toxicities throughout the range of doses, which make them more useful when considering the entire drug development process.

It was confirmed in the second work (that was also extended to a bivariate case [Gerard et al., 2021]). This research demonstrated that incorporating PK/PD information significantly enhances the accuracy of proportion of correct selection when the toxicity-generating process is reasonably understood and approximated. This addition facilitates improved estimation of sequence-toxicity curves by enabling more extensive information sharing across sequences. Furthermore, these methodologies enable us to identify and recommend additional sequences for expansion cohorts.

The concept of integrating PK/PD modeling into dose-finding studies has been further explored by other researchers. The following are some examples. Günhan et al. [2020] introduced the Time-To-Event PK model, which I mentioned earlier in the introduction. Su et al. [2022] created the Semi-mechanistic Dose-Finding framework, which employs three interconnected models integrating PK and PD outcomes. Lastly, Micallef et al. [2022] proposed the Exposure Driven Escalation With Overdose Control method, an advancement of PKLOGIT that substitutes the AUC-dose relationship with a real pop-PK analysis.

Therefore, employing PK/PD analysis in the dose-escalation process or decision-making appears very promising. However, further research is necessary to address practical challenges, such as the unavailability of PK/PD samples after each patient's sampling, and technical issues, including the requirement for ample data for estimating PK/PD parameters and assessing model relevance.

Chapter 3

Design for multiple endpoints and phases

Contents

In this chapter, I will outline three contributions I have made to leveraging alternative internal data beyond PK/PD in clinical trials. These contributions are presented not in chronological order but in a sequence that builds upon the narrative from the previous chapter. Indeed, still within the context of dose-finding, I introduced a model designed to manage multiple treatment cycles [Ursino et al., 2022], addressing a concern akin to dose scheduling, explored in Chapter 2, but applied to several cycles (here with a stable dose). Remaining in the dose-finding framework, but transitioning to a Phase I/II context, I developed a dose-finding strategy that accounts for dual toxicity types along with an efficacy measure for treating newborns experiencing seizures [Ursino et al., 2019]. Interesting to say, this design was then immediately used for the clinical study, which was recently finished and the final analysis of which is underway. Finally, moving to seamless Phase II/III trials, I co-supervised a PhD student, Benjamin Duputel, in a project that aimed to integrate binary Phase II outcomes into the survival analysis of Phase III trials [Duputel et al., 2023]. Mirroring the structure of the previous chapter, the rest of this chapter will start with a brief introduction to set the stage for the three discussed papers, elucidating the motivations. Following that, I will summarize each of the three papers,

with a focus on the methodologies employed and the results obtained. The concluding section will be dedicated to a discussion.

3.1 Introduction

As stated in Chapter 1, in oncology, dose-finding trials for both cytotoxic and molecularly targeted drugs are crucial for establishing the MTD of a new medication, which is typically based on the incidence of patients experiencing short-term DLTs observed during the study. However, the emergence of DLTs may be delayed, possibly due to the drug pharmacokinetics or the accumulated dosage. For instance, research by Postel-Vinay et al. [2014] on Phase I trials of molecularly targeted agents revealed a higher frequency of DLTs after the first cycle (1087 vs. 936 in the first cycle), leading to the recommendation that "evaluation of the recommended Phase 2 dose should consider all available data from any cycle." The time-to-event continual reassessment method (TITE-CRM) [Cheung and Chappell, 2000], viewed as an adaptation of the CRM with adjustments for time, represents an early endeavor to incorporate DLT observations over a larger monitoring period. Nonetheless, this approach treats all patients who suffer DLTs as equally influential in the analysis, regardless of when the DLT occurs. More recently, a Bayesian approach has been introduced to estimate the conditional probability of encountering a DLT in each cycle, given the absence of DLTs in preceding cycles [Fernandes et al., 2016]. This model includes three parameters designed to capture the impact of the initial dose, the cumulative effect of the drug, and the body developed resistance over time. Nonetheless, to avoid biased estimates, it is crucial to apply a highly informative prior for the parameter associated with acquired resistance.

Another situation in early phase arise when distinct toxicity/efficacy types should be considered separately. This may be because toxicities have different generative mechanisms or occur or can be measured at different times. In this context, previous attempts include incorporating elicited numerical utilities for composite outcomes derived from two efficacy measures and one safety outcome [Thall et al., 2014]. Alternatively, modeling three outcomes - toxicity, efficacy, and a surrogate efficacy endpoint - through a trivariate continual reassessment method, offers a different approach to managing multiple trial outcomes [Zhong et al., 2012].

Moving to later clinical trial phases, seamless designs represent a notable class in clinical trial methodology, merging multiple trial phases within a single protocol, unlike traditional approaches where Phase II and III trials are conducted separately [Bretz et al., 2006]. Seamless trials can be categorized as "operationally seamless" when phases are sequential but distinct, and "adaptive seamless" when data from all phases contribute to the final analysis [Maca et al., 2006]. This adaptive approach involves making adjustments based on interim data reviews, posing statistical challenges related to multiple comparisons and the need for controlling the familywise error rate due to selection processes and interim analyses [Stallard, 2011, Stallard et al., 2015, Quan et al., 2020, Friede et al., 2012, 2020]. While Bayesian inference is well-suited for these seamless trial designs, there have been limited endeavors to integrate Bayesian inference into Phase II/III seamless designs [Chapple and Thall, 2019]. Examples include the use of Bayesian methodologies for selecting treatments [Schmidli et al., 2007, Kimani et al., 2009], or for choosing sub-populations [Brannath et al., 2009].

3.2 DICE: Dose-fInding CumulativE

In this work, we suggest a cumulative modeling strategy for Phase I adaptive clinical trials aiming to model the probability of experiencing DLTs by considering the cumulative impact of the administered drug. This approach seeks to identify the most suitable MTDS (maximum tolerated dose sequence) from a range of predefined dose sequences [Ursino et al., 2022]. As a reminder, even if I use the same acronym as before, here we are seeking to find the MTDS over several cycles and without PK/PD measurements, whereas in Chapter 2, we were interested in the sequence within the same cycle.

As usual, I will review the specific notation used in this work, which is very similar to that presented in the second work of Chapter 2. Let us consider K as the total number of cycles planned for the trial. Define $\mathbf{S} = \{\bar{\mathbf{s}}_1, \ldots, \bar{\mathbf{s}}_J\}$ as the set comprising all possible J predetermined sequences of doses to be administered throughout the trial. Each sequence, $\bar{\mathbf{s}}_j = \{d_{[j,1]}, \ldots, d_{[j,K]}\}$, is ordered such that $d_{[j,k]} \leq d_{[j+1,k]}$ for every cycle k from 1 to K and for all sequence indices j from 1 to $J-1$, ensuring no two sequences $\bar{\mathbf{s}}_j$ and $\bar{\mathbf{s}}_{j+1}$ are identical. For the i-th participant, $d_{i,k}$ represents the planned dosing sequence across cycles $k = 1, \ldots, K$. It is important to note that a participant may not receive all planned dosages but only up to the point of experiencing a DLT, after which treatment is discontinued. For this participant, let $Y_{i,k}$ denote the binary outcome indicating the occurrence of a DLT at cycle k. We propose modeling the cumulative probability of toxicity by cycle k as

$$P\left(\sum_{m=1}^{k} Y_{i,m} \geq 1\right) = f_{\boldsymbol{\theta}}(d_{i,1}, k, D_{i,k}), \tag{3.1}$$

where $f_{\boldsymbol{\theta}}(.)$ is a specific parametric link function and $D_{i,k} = \sum_{m=2}^{k} d_{i,m}$ for $k > 1$, and is 0 for $k = 1$. The likelihood contribution of the i-th participant is then given by

$$\mathcal{L}_i(\boldsymbol{\theta}|y_{i,k_i}) = \left(1 - f_{\boldsymbol{\theta}}(d_{i,1}, k_i, D_{i,k_i-1})\right)^{1-y_{i,k_i}} \times \left(f_{\boldsymbol{\theta}}(d_{i,1}, k_i, D_{i,k_i}) - f_{\boldsymbol{\theta}}(d_{i,1}, k_i - 1, D_{i,k_i-1})\right)^{y_{i,k_i}},$$

where k_i is the last cycle completed by the i-th participant and y_{i,k_i} reflects the outcome at that cycle. For n patients enrolled in the study, the overall likelihood is expressed as $\mathcal{L}(\boldsymbol{\theta}|\mathbf{y}) = \prod_{i=1}^{n} \mathcal{L}_i(\boldsymbol{\theta}|y_{i,k_i})$, integrating the individual likelihood contributions from each patient. To choose the parametric function of Eq. 3.1, we suggest modifying the Bayesian logistic regression model as described by Neuenschwander et al. [2008], to accommodate the cumulative dose effects with the following formulation:

$$f_{\boldsymbol{\theta}}(d_{i,1}, k, D_{i,k}) = \text{logit}^{-1}\left[\alpha + \exp(\beta)\log\left(\frac{d_{i,1}}{d^*}\right) + \exp(\gamma)\log\left(\frac{D_{i,k}}{D^*} + 1\right)\right],$$

where d^* signifies the initial dose in the sequence $\bar{\mathbf{s}}_* \in \mathbf{S}$, identified as the most probable MTDS prior to the trial, and D^* represents the total cumulative dose across the sequence $\bar{\mathbf{s}}_*$. The term $+1$ alongside $\exp(\gamma)$ ensures a non-negative effect of cumulative dosing: a negative log value would erroneously suggest a decrease in the cumulative toxicity probability relative to the initial dose impact, which is conceptually flawed as the cumulative toxicity risk should logically increase or remain constant. The parameters $\boldsymbol{\theta} = \{\alpha, \beta, \gamma\}$ serve distinct roles: α correlates with the toxicity probability at the reference dose d^* in the first cycle, $\exp(\beta)$ shapes the initial

toxicity curve, and $\exp(\gamma)$ quantifies the incremental toxicity risk attributed to cumulative dosing across cycles. Independent normal prior distributions are assigned to the three parameters: $\alpha \sim \mathcal{N}(\mu_\alpha, \sigma_\alpha^2)$, $\beta \sim \mathcal{N}(\mu_\beta, \sigma_\beta^2)$, and $\gamma \sim \mathcal{N}(\mu_\gamma, \sigma_\gamma^2)$.

3.2.1 Dose-finding design

The allocation of dose sequences to patients is dynamically adjusted after each cohort of L patients joins the trial. Initially, the first cohort of L patients is assigned the $\bar{\mathbf{s}}_1$ dose sequence. The toxicity probability for a given theoretical dose sequence j within the panel $\mathbf{S}$ is mathematically computed as

$$p_T(\bar{\mathbf{s}}_j) = f_{\boldsymbol{\theta}}\left(d_{[j,1]}, K, \sum_{m=2}^{K} d_{[j,m]}\right). \tag{3.2}$$

Given that $p_T(.)$ in Eq. 3.2 relies on the unknown parameters $\boldsymbol{\theta}$ which can be estimated through Bayesian inference, we use its estimated version $\hat{p}_T(.)$. After the first patient of a subsequent cohort is enrolled, we update the posterior distribution of the parameters $\boldsymbol{\theta} = \{\alpha, \beta, \gamma\}$ utilizing all collected data, including from those patients who have not encountered a DLT and those who have not completed the full number of cycles ($k_i < K$). Then, $\hat{p}_T(.)$ is recalculated using these updated posterior distributions $p_{post}(\alpha, \beta, \gamma)$. For estimating $\hat{p}_T$, various approaches can be considered. We recommend employing the predictive probability of toxicity, namely, the expected value of $p_T(\bar{s}_j)$ according to the posterior distribution $E_{p_{post}(\alpha,\beta,\gamma)}[p_T(\bar{s}_j)]$, or the median of these predicted probabilities, which corresponds to the median of the posterior distribution of $p_T(\bar{s}_j)$.

The dose sequence for the forthcoming cohort is selected based on the estimated toxicity probability that most closely aligns with the predefined target δ_T, as defined by

$$\bar{\mathbf{s}}_{next} = \arg\min_j |\hat{p}_T(\bar{\mathbf{s}}_j) - \delta_T|.$$

Upon the trial conclusion, the estimated MTDS for up to cycle k, denoted as MTDS_k, is identified as the sequence truncated at cycle k whose estimated probability of toxicity is nearest to the predetermined target δ_T. This is mathematically expressed as

$$\text{MTDS}_k = \arg\min_j |\hat{p}_T(\bar{\mathbf{s}}_{j,k}) - \delta_T|,$$

where $\bar{\mathbf{s}}_{j,k}$ represents the sequence $\bar{\mathbf{s}}_j$ shortened to cycle k. Consequently, the estimated MTDS accounting for all predefined cycles is represented as MTDS_K. For simplicity, we will refer to MTDS_K simply as MTDS from this point forward. To mitigate the risk of unnecessarily subjecting patients to highly toxic dose sequences, we propose two precautionary measures. The first is the implementation of a "no-skipping" rule, which stipulates that a new dose sequence j cannot be allocated to a cohort unless the immediately preceding dose sequence $j-1$ has been administered to at least one patient before. The second safety measure hinges on the posterior probability of toxicity associated with the initial dose sequence. The trial is halted at any phase if the following condition is met:

$$P_{post}\left(p_T(\bar{s}_1) > \delta_T(1+\varepsilon)\right) > \tau_T,$$

where τ_T denotes a predetermined probability threshold, typically set between 0.8 and 0.99, and ε represents the maximum allowable relative error in the targeted toxicity probability.

This methodology, henceforth referred to as DICE (Dose-fInding CumulativE), leverages Bayesian estimation techniques and allows for its application from the outset of the trial, immediately after the first participating patient completes the initial cycle. As usual, early in the trial the choice of prior distributions significantly influences the escalation of dose sequences.

3.2.2 Results

We assessed the performance of our method across six distinct scenarios, each characterized by variations within five dose sequences ($J = 5$), spanning five cycles ($K = 5$). These dose sequences were established by maintaining consistent dose levels throughout. The six scenarios yielded different MTDS across panel positions according to the target $\delta_T = 0.3$. We modeled patient accrual at a steady rate of one per month and assumed that each cycle would precisely last one month, thereby limiting patient observation to a maximum of five months. For comprehensive evaluation, we conducted 5000 independent simulations for each scenario. For further information and in-depth analysis, readers are encouraged to consult the original published paper.

Table 3.1 shows the results of simulations comparing our proposed method, the DICE, with the TITE-CRM and a benchmark standard. The metrics used for this comparison include the frequency at which the MTDS is correctly selected, the distribution of dose allocations among participants, and the average number of DLTs per trial. For both the DICE and TITE-CRM, data are provided for scenarios involving cohorts of 1 and 3 patients respectively.

On average, DICE demonstrated operating characteristics similar to those of TITE-CRM regarding the selection of dose sequences. However, DICE exhibited enhanced performance in scenarios where the MTDS was positioned in the second half of the sequence panel, showing a proportion of correct selection (PCS) rate increase of about 10% in Scenario 1 and nearly double the PCS in Scenario 3 for cohorts of 3 patients. Conversely, TITE-CRM had a slight advantage over DICE when the MTDS was in the first half of the panel. This pattern was also reflected in the distribution of dose sequence allocations. Furthermore, although the number of DLTs recorded was within a similar range for both methods, DICE reported an average of 1 or 2 more DLTs. This result is expected, as DICE is based on a larger number of parameters that need to be estimated compared to TITE-CRM. Moreover, utilizing the DICE methodology at the trial end allows clinicians to forecast the toxicity probabilities for sequences that comprise fewer cycles than initially planned, specifically MTDS_k for $k < K$. As illustrated in Figure 3.1, the selection frequencies of the MTDS for a reduced number of cycles are displayed, for example, MTDS_4 in Scenario 2, which involves only 4 cycles, and MTDS_3 in Scenario 5, with just 3 cycles. In the shortened Scenario 2, the accurate MTDS_4 corresponds to sequence 5, with DICE achieving a PCS of around 70%, irrespective of the cohort size. Meanwhile, for the shortened Scenario 5, where the true MTDS_3 aligns with sequence 2, the PCS is about 60%.

Table 3.1 – Results of the 5000 simulated trials for the six proposed scenarios for each method. The proportion of the selected MTDS, proportion of dose allocation and number of DLTs are shown. In bold, the results regarding the true MTDS in each scenario are shown.

Method		**Dose sequence selection**					**Dose sequence allocation**					**DLTs**
- cohort	None	1	2	3	4	5	1	2	3	4	5	M (Q1, Q3)
Scenario 1												
p_{tox}		0.11	0.20	**0.30**	0.50	0.65						
DICE - 1	0.002	0.009	0.157	**0.665**	0.165	0.003	0.064	0.159	0.448	0.221	0.107	11 (9, 12)
TITE-CRM - 1	0.004	0.023	0.3	**0.571**	0.1	0.002	0.143	0.271	0.363	0.153	0.069	9 (7, 11)
DICE - 3	0.001	0.012	0.175	**0.624**	0.181	0.008	0.137	0.23	0.401	0.179	0.053	9 (8, 10)
TITE-CRM - 3	0.003	0.026	0.352	**0.533**	0.085	0.001	0.211	0.333	0.334	0.106	0.016	7 (6, 9)
benchmark	0	0.043	0.336	**0.578**	0.043	0.001						
Scenario 2												
p_{tox}		0.03	0.06	0.15	**0.30**	0.50						
DICE - 1	0	0	0.001	0.126	**0.73**	0.143	0.036	0.04	0.153	0.431	0.341	10 (9, 11)
TITE-CRM - 1	0	0	0.004	0.208	**0.68**	0.108	0.046	0.066	0.211	0.397	0.28	9 (8, 10)
DICE - 3	0	0	0.001	0.139	**0.677**	0.183	0.104	0.113	0.204	0.337	0.242	8 (7, 9)
TITE-CRM - 3	0	0	0.004	0.236	**0.644**	0.116	0.113	0.13	0.255	0.341	0.161	7 (6, 8)
benchmark	0	0	0.008	0.183	**0.756**	0.053						
Scenario 3												
p_{tox}		0.06	0.09	0.15	0.22	**0.30**						
DICE - 1	0	0	0.006	0.142	0.201	**0.651**	0.043	0.05	0.177	0.182	0.547	7 (6, 8)
TITE-CRM - 1	0.001	0.001	0.025	0.145	0.33	**0.499**	0.075	0.098	0.176	0.25	0.401	6 (5, 8)
DICE - 3	0	0	0.003	0.082	0.238	**0.677**	0.108	0.116	0.201	0.208	0.367	6 (5, 7)
TITE-CRM - 3	0.001	0.001	0.046	0.229	0.393	**0.33**	0.147	0.184	0.261	0.238	0.17	5 (4, 6)
benchmark	0	0.004	0.017	0.115	0.373	**0.492**						
Scenario 4												
p_{tox}		0.20	**0.30**	0.45	0.65	0.80						
DICE - 1	0.011	0.188	**0.452**	0.339	0.009	0	0.221	0.326	0.368	0.062	0.023	11 (9, 12)
TITE-CRM - 1	0.038	0.26	**0.54**	0.16	0.003	0	0.411	0.374	0.169	0.033	0.012	9 (7, 10)
DICE - 3	0.007	0.197	**0.47**	0.311	0.015	0.001	0.261	0.377	0.299	0.053	0.01	10 (9, 11)
TITE-CRM - 3	0.029	0.284	**0.557**	0.128	0.002	0	0.483	0.383	0.12	0.013	0.001	8 (7, 9)
benchmark	0	0.339	**0.558**	0.103	0	0						
Scenario 5												
p_{tox}		**0.30**	0.50	0.70	0.85	0.95						
DICE - 1	0.098	**0.707**	0.186	0.009	0	0	0.558	0.287	0.117	0.027	0.012	12 (11, 14)
TITE-CRM - 1	0.158	**0.721**	0.121	0	0	0	0.723	0.201	0.048	0.018	0.009	10 (9, 12)
DICE - 3	0.083	**0.733**	0.179	0.005	0	0	0.591	0.313	0.087	0.008	0	12 (10, 13)
TITE-CRM - 3	0.131	**0.737**	0.131	0.001	0	0	0.743	0.222	0.033	0.002	0	10 (8, 12)
benchmark	0	**0.929**	0.07	0	0	0						
Scenario 6												
p_{tox}		0.10	0.20	**0.30**	0.40	0.50						
DICE - 1	0	0.015	0.163	**0.523**	0.24	0.059	0.084	0.174	0.438	0.198	0.106	9 (8, 11)
TITE-CRM - 1	0.005	0.022	0.337	**0.445**	0.171	0.021	0.188	0.321	0.307	0.14	0.045	7 (6, 9)
DICE - 3	0	0.012	0.153	**0.503**	0.247	0.085	0.128	0.227	0.397	0.165	0.083	9 (8, 10))
TITE-CRM - 3	0.003	0.022	0.403	**0.434**	0.127	0.011	0.248	0.384	0.275	0.081	0.012	7 (5, 8)
benchmark	0	0.015	0.294	**0.489**	0.172	0.031						

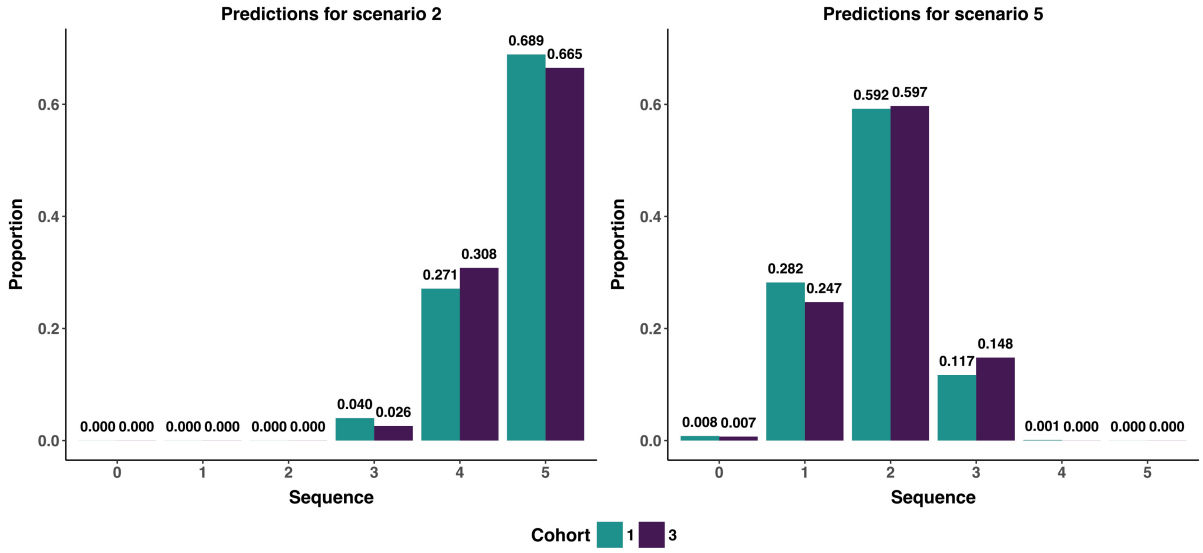


Figure 3.1 – Simulation study: Probability of sequence selection using the model predictions at earlier cycles than the predefined window. Left panel: selection based on predicted cumulative probability of toxicity at cycle 4 in scenario 2; Right panel: selection based on predicted cumulative probability of toxicity at cycle 3 in scenario (x-axis 0 indicating no sequence recommendation at the end of the trial)

3.3 Dose-finding design with two toxicities and an efficacy endpoint

In this work, I proposed a Bayesian design for Phase I-II of the LEVNEONAT trial (Levetiracetam Treatment of Neonatal Seizures: Safety and Efficacy Phase II Study, with registration number NCT 02229123 listed on ClinicalTrials.gov) aimed at determining the optimal dosage of Levetiracetam for first line neonatal seizure treatment. The optimal dosage is identified as one associated with a toxicity threshold below 10% and an efficacy exceeding 60%. Here, efficacy is defined as achieving at least an 80% decrease in seizure burden subsequent to the administration of the loading dose. This trial poses several challenges: (i) besides the standard short-term DLT measurements, we aim to assess hearing loss, which cannot be evaluated in real-time and is only confirmable on day 30; (ii) there may be a need to administer an additional medication (A2) during treatment (maintenance doses) if Levetiracetam proves to be insufficiently effective, prior to the DLT evaluation; (iii) as newborns can be enrolled at any time, our models must rapidly generate results, enabling the pharmacy to prepare the appropriate dose level; (iv) the low target toxicity threshold of 10% may result in the trial persistently allocating low doses if DLTs appear early in the trial. To address these challenges, we develop a model that includes one endpoint for efficacy and two endpoints for toxicity (see Figure 3.2 for a simplified trial overview), employing a pseudo-likelihood method for inference.

In the following, three statistical models that establish the link between dose levels and their effects on efficacy, as well as short-term (denoted $T1$) and long-term toxicity (denoted $T2$) are outlined. These frameworks will serve as the basis for determining the appropriate dose allocation and selection. It is important to note that the association between the drug

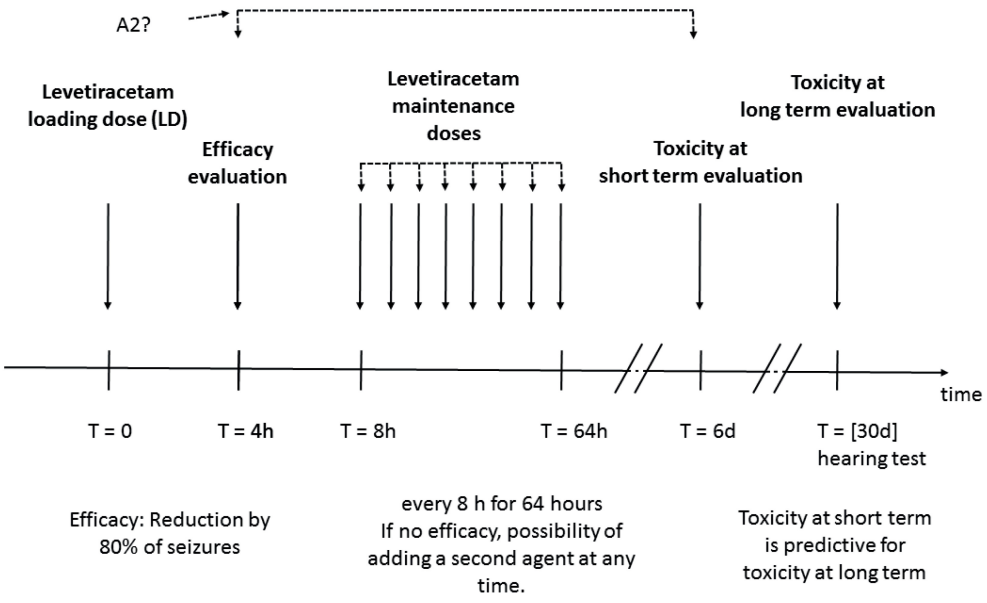


Figure 3.2 – LEVNEONAT clinical trial: doses and endpoint measurements scheme. The loading dose (LD) is given at time 0 and after four hours the efficacy endpoint is evaluated. Up to 8 maintenance doses (one quarter of LD), are administrated between hours 6 and 64. The investigators have the option to add a second agent $A2$ as a rescue medication. After 6 days, the first toxicity endpoint (short-term toxicity) is measured while the long-term toxicity endpoint (i.e., hearing loss) is assessed after 30 days or when the neonate is released from the hospital, whatever occurs first.

effectiveness and toxicity has been excluded from consideration due to its minimal impact in prior research. Consider d_k as the initial dose, with d_k belonging to the set $\{d_1, d_2, \ldots, d_K\}$, and $d_{[i]}$ representing the dose given to the i-th participant. Let $y_{E,i}$ denote a binary indicator of efficacy, assigning a value of 1 when the i-th participant shows positive effects and 0 otherwise. Similarly, $y_{T1,i}$ is a binary indicator for short-term toxicity $T1$, and $y_{T2,i}$ for long-term toxicity $T2$, each taking a value of 1 when the respective adverse effect is observed and 0 if not.

3.3.1 Dose-efficacy model

Levetiracetam is introduced with an initial loading dose d_k, succeeded by maintenance doses that are a quarter of the loading dose. The drug effectiveness is assessed prior to the starting of maintenance dosing, relying solely on d_k. The probability of efficacy for a patient receiving a specific dose x is represented by $p_E(x) = P(y_{E,i} = 1 | d_{[i]} = x)$. To describe the relationship between dose and efficacy, we employ essentially the CRM adapted for efficacy, that is a logistic model as follows:

$$\text{logit}(p_E(x)) = \alpha_1 + e^{\beta_1} x, \quad x \in \{\tilde{d}_1, \ldots, \tilde{d}_K\}; \tag{3.3}$$

where α_1 and β_1 symbolize the intercept and slope parameters, respectively. The term $\tilde{d}_k$ refers to the "effective" dose, conceived as the preliminary estimate of the probability of efficacy tied to dose d_k. This effective dose is determined by setting values for α_1, β_1, and p_E and then reversing

Eq. 3.3. The intercept α_1 is set at a value of 3, though alternative values can be selected for different applications. The parameter β_1 is assigned a normal prior distribution with a mean of 0 and a standard deviation of 1.34 [Cheung, 2011].

3.3.2 Short-term toxicity model

The evaluation of short-term toxicity ($T1$) occurs within six days of beginning treatment. A notable challenge, as depicted in Figure 3.2, arises when clinicians adjust the treatment regimen due to Levetiracetam perceived insufficient efficacy in reducing seizures. This adjustment might include decreasing or discontinuing the maintenance dose of Levetiracetam and introducing an additional medication ($A2$) to enhance therapeutic outcomes. Consequently, modeling $T1$ becomes more intricate than in typical dose-finding studies due to the potential confounding effects of $A2$ on the assessment of Levetiracetam toxicity. This complication arises because, upon observing toxicity following the addition of $A2$, it is unclear whether the adverse effects are attributable to Levetiracetam, $A2$, or a combination of both. Additionally, the relatively brief assessment window for $T1$ (six days) is further complicated by the continuous admission of new patients who require immediate care. This situation leads to what is known as the "late-onset outcome" dilemma, where the ongoing enrollment and treatment initiation for new patients can disrupt the timely and adaptive dose assignment based on the incomplete toxicity evaluations of previously admitted neonates.

We firstly tackle the issue of possible late-onset outcomes (still referred to $T1$). We will now refer to $y_{T1,i}$ as y_i. Based on discussions with researchers, it was determined that $T1$ is more likely to manifest at the onset of the 0-6 day evaluation period. Consequently, a modified TITE-CRM [Cheung and Chappell, 2000, Braun, 2006] was developed to manage this challenge of late-onset toxicity. This approach defines T_i as the time until toxicity occurrence for the i-th patient, with T_{max} representing the maximum duration of the toxicity assessment period for $T1$, which is 6 days. Starting from the principle of conditional distribution for $t \leq T$, we derive:

$$P(T_i \leq t|x_i) = P(T_i \leq t|x_i, y_i = 1) \times P(y_i = 1|x_i).$$

Here, $P(y_i = 1|x_i)$ signifies the probability of experiencing short-term DLT for a patient administered dose x, denoted as $p_{T1}(x)$. This probability is modeled using a one-parameter logistic model, as follows:

$$\text{logit}(p_{T1}(x)) = \alpha_2 + e^{\gamma_1}x, \quad x \in \{\bar{d}_1, \ldots, \bar{d}_K\}.$$

In this model, α_2 is a fixed intercept, γ_1 represents the unknown slope parameter, and $\bar{d}_k$ refers to the "effective" dose, which are the prior estimates of short-term DLT for dose d_k. Similar to the efficacy model, this formulation presupposes that toxicity $T1$ increases in a dose-dependent manner. For γ_1, a normal prior distribution is assumed with $\gamma_1 \sim N(0, 1.34)$, and α_2 is set at 3.

For the maximum evaluation period, denoted as T_{max}, which is 6 days, we adopt a methodology akin to Braun [2006], presuming the normalized time until toxicity, t/T_{max}, adheres to a Beta distribution, $Beta(1, \zeta)$. This leads to the following expression:

$$P(T_i \leq t|x_i, Y_i = 1) = 1 - \left(1 - \frac{t}{T_{max}}\right)^{\zeta}, \quad \zeta > 0.$$

By setting the first parameter of the Beta distribution to 1, we simplify the model, yet retain sufficient flexibility to depict various patterns of time until toxicity occurs. Moreover, ensuring the sum of the Beta distribution parameters exceeds 1 eliminates the possibility of a U-shaped distribution for the time to toxicity, which is considered implausible in this context. Given the small sample size and the even smaller number of observed toxicities in the trial, selecting a suitable prior for ζ is essential to mitigate the risk of highly uncertain estimates. The prior distribution for ζ is elicited through consultation with clinicians. In the LEVNEONAT clinical trial, ζ was treated uniformly across all doses to simplify the model. However, based on the principle of monotonicity, that higher doses lead to earlier toxicity, it is feasible to make ζ dose-dependent by defining $\zeta_i = z_k^\lambda$, where λ is a negative value. This approach entails transforming d_k into z_k, a value constrained within the interval $[0, 1]$.

To address the confounding problem arising from the potential introduction of a new agent $A2$ during the maintenance dose phase, we face the challenge of distinguishing whether observed toxicities are due to Levetiracetam, $A2$, or a combination of both. Our approach involves the creation of a pseudo observation y_i^*, which estimates the likelihood that the toxicity can be attributed to the maintenance doses of Levetiracetam. Specifically, n_i represents the number of maintenance doses administered to the i-th patient, and $x_{i,m}$ refers to the actual dosage received during these maintenance phases (or the corresponding loading dose, given their direct correlation). Considering the observed toxicity outcome y_i, we introduce a pseudo observation defined as:

$$y_i^* = w(n_i, x_{i,m}) y_i,$$

where the weighting function is given by:

$$w(n_i, x_{i,m}) = \begin{cases} f_w(n_i, x_{i,m})\tau, & \text{if A2 is introduced} \\ 1, & \text{otherwise,} \end{cases} \tag{3.4}$$

with $f_w(n_i, x_{i,m})$ being a predetermined function and the constant $\tau < 1$ reflecting the probability that the toxicity results from Levetiracetam, assuming the full regimen of maintenance doses is administered. The value of τ should be determined in consultation with clinical experts. Essentially, this formula suggests that if $A2$ is not added, the actual outcome y_i is entirely attributed to Levetiracetam, warranting a weight of 1. Conversely, if $A2$ is incorporated into the regimen, the attributed fraction of toxicity, $f_w(n_i, x_{i,m})\tau$, is considered to be due to Levetiracetam. We propose utilizing the function:

$$f_w(n_i, x_{i,m}) = \frac{e^{\gamma n_i x_{i,m}} - 1}{e^{\gamma N_M x_K} - 1},$$

where N_M represents the maximum number of maintenance doses outlined in the study protocol, x_K indicates the dosage for the highest dose level, and γ serves as a calibration parameter. This formula is designed to progressively increase the proportion of toxicity attributed to Levetiracetam in relation to the cumulative maintenance doses received by a patient. Specifically, when the new agent $A2$ is incorporated without any maintenance doses having been administered ($n_i = 0$), the function $f_w(n_i = 0, x_{i,m})$ equals zero, leading to $y_i^* = 0$, implying that toxicity is primarily associated with $A2$. Conversely, when the full regimen of maintenance doses is delivered at the highest dosage ($n_i = N_M$, $x_{i,m} = x_K$), the function $f_w(n_i = N_M, x_{i,m} = x_K)$ equals

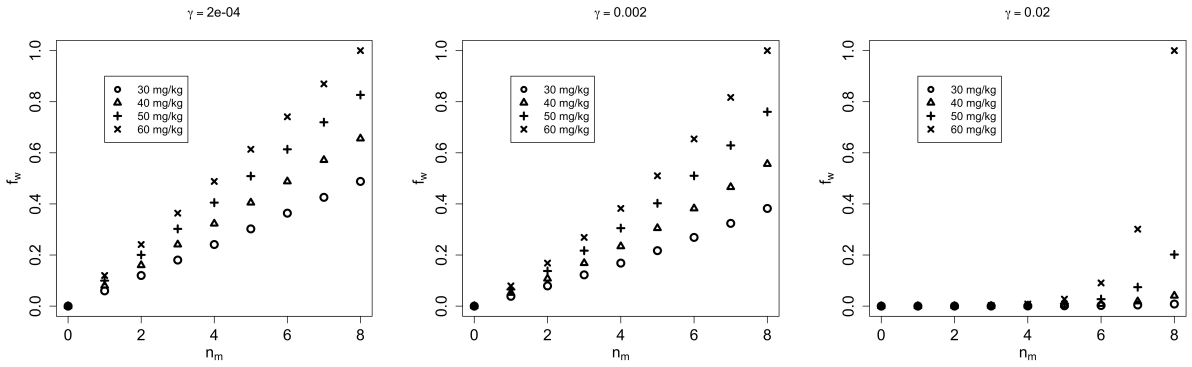


Figure 3.3 – Values of f_w for several doses, 30, 40, 50 and 60 mg/Kg , versus the number of maintenance doses and for three γ values.

one, making $y_i^* = \tau y_i$, which reflects the scenario where the observed toxicity is fully considered to be attributable to Levetiracetam, assuming the influence of $A2$ is proportionally adjusted by τ. In the LEVNEONAT trial, informed by prior research and discussions with medical experts, τ was established at 0.8, and γ was determined to be 0.002. Ideally, these parameters, τ and γ, would be derived from the trial data; however, they cannot be distinctly identified due to the complete overlap in the effects of Levetiracetam and $A2$. Therefore, sensitivity analysis is employed as a method to evaluate the effectiveness and stability of the selected values for τ and γ. Exploring different values of γ reveals its impact on the weighting function f_w across the four dosing levels. An increase in γ tends to reduce the weights to zero except for the last few maintenance doses, whereas a decrease in γ results in a more linear distribution of weights. The initial formula for f_w assigns a zero weight to toxicities when no maintenance doses are administered. If there is a preference to allocate a weight greater than zero under such circumstances, an alternative formulation could be employed:

$$w(n_i, x_{i,m}) = \begin{cases} \tau_L + f_w(n_i, x_{i,m})(\tau - \tau_L), & \text{if A2 is added} \\ 1, & \text{otherwise,} \end{cases}$$

where τ_L signifies the weight assigned at the loading dose (uniform across all doses for simplicity). Another approach, not covered in this work, involves directly obtaining each weight from clinical practitioners.

Combining all the elements, for n neonates treated in the trial, the pseudo likelihood for $T1$ is formulated as follows:

$$QL_1(\gamma_1|\mathbf{y}, \alpha_1, \zeta) \propto \prod_{i=1}^{n} \left(\left[1 - \left(1 - \frac{u_i}{T_{max}} \right)^{\zeta} \right] p_{T1} \right)^{y_i^*} \left(1 - \left[1 - \left(1 - \frac{u_i}{T_{max}} \right)^{\zeta} \right] p_{T1} \right)^{(1-y_i^*)}.$$

This equation represents the pseudo likelihood, termed so because the likelihood derived from it does not precisely reflect the true likelihood due to the empirical weighting of the toxicity probability p_{T1}.

3.3.3 Long-term toxicity model

Contrary to short-term toxicity $T1$, which may manifest at any point from day 0 to 6, long-term toxicity $T2$ (e.g., hearing loss) is only assessable on day 30, even though it could develop well before this date, as illustrated in Figure 3.2. Given the potential of $T1$ to predict $T2$, we express the probability of $T2$ given the dose and the occurrence of $T1$ as $p_{T2}(x, y_{T1}) = P(y_{T2} = 1|x, y_{T1})$. This relationship is stated using the logistic model:

$$\text{logit}(p_{T2}(x, y_{T1})) = \alpha_3 + e^{\delta_1} x + e^{\delta_2} y_{T1}, \quad x \in \{\dot{d}_1, \ldots, \dot{d}_K\}. \tag{3.5}$$

Here, $\{\dot{d}_1, \ldots, \dot{d}_K\}$ represent the "effective" doses, corresponding to prior estimates of the probability of $T2$ toxicity for each dose level. α_3 is a fixed parameter, while δ_1 and δ_2 are parameters to be determined. The model assumes both e^{δ_1} and e^{δ_2} are positive, indicating an increased likelihood of experiencing $T2$ for patients who have encountered short-term toxicity $T1$ or have been administered a higher dose. A time-to-event model is not utilized for $T2$ because its occurrence cannot be monitored in real-time; it is only measurable on day 30. Like $T1$, the measurement of $T2$ is confounded by the potential introduction of agent $A2$. To address this, we apply the same pseudo observation technique, substituting the actual observed outcome y_{T2} with $y^*_{T2} = w(n_m, x_m) y_{T2}$, where $w(n_m, x_m)$ follows the formula provided by Eq. 3.4. For consistency, we also replace y_{T1} in Eq. (3.5) with y^*_{T1}, utilizing the same adjusted value as in the $T1$ model. Let n^* represent the number of patients who have successfully undergone the evaluation for long-term toxicity $T2$. The pseudo-likelihood for this scenario is expressed as:

$$QL_2(.) = \prod_{i=1}^{n^*} \left[p_{T2}(x_{[i]}, y^*_{T1,i})\right]^{y^*_{T2,i}} \left[1 - p_{T2}(x_{[i]}, y^*_{T1,i})\right]^{1-y^*_{T2,i}}.$$

The parameters δ_1 and δ_2 are assumed to follow a normal distribution with mean 0 and standard deviation 1.34, and α_3 is fixed at 3. The model specified in Eq. 3.5 outlines the conditional probability of observing y_{T2} given y_{T1}. The overall likelihood of y_{T2}, denoted as p_{T2}, is determined through the law of total probability.

3.3.4 Avoiding stickiness

It is recognized that premature occurrences of short-term toxicity can significantly influence the dose allocation strategy, particularly when the targeted probability lies within the distribution tails. A "greedy" algorithm may become entrenched in a sub-optimal action because it continuously selects this sub-optimal choice. Adopting the methodology of Resche-Rigon et al. [2008, 2010], we apply adaptive "relevance weights" (denoted as $wr_i(.)$) to the previously discussed pseudo-likelihood for $T1$. These weights are dynamically adjusted based on the current patient accrual within the trial and the incidence of short-term toxicities observed at each dosage level. Specifically, for any given dose k, let n_{alloc,d_k} represent the number of patients assigned to that dose, and n_{DLT,d_k} denote the number of patients who have experienced toxicity. Within the context of the LEVNEONAT clinical trial, a tailored weighting scheme is implemented as follows:

$$wr_i(y_i, d_i) = \begin{cases} 1, & \text{if } y_i = 0 \\ \min\left(\max\left(\pi \frac{n_{alloc,d_i}}{n_{max}} + (1-\pi)\frac{n_{DLT,d_i}}{n_{alloc,d_i}}, \frac{n_{alloc,d_i}}{n_{max}}\right), 1\right), & \text{if } y_i = 1. \end{cases}$$

Here, n_{max} is a preset constant often related to the desired probability target, and π is a mixture constant that integrates the proportion of patients allocated to dose i with the observed toxicity rate at that dose. Once n_{max} patients have been allocated to each dose level, all weights adjust to 1.

3.3.5 Dose allocation rule

The objective is to allocate a dose to each new cohort patient(s) that balances maximal efficacy with acceptable safety. If all doses are too toxic or not effective enough, the trial will stop. Upon enrolling n neonates, out of which n_2, $n_2 \leq n$, have completed the full follow-up until the evaluation of $T2$, the dose for the subsequent cohort is chosen from a pool of doses that satisfy a set of predefined criteria. These criteria are based on current estimates of the probabilities and include:

- $P(p_{T1} > \delta_{T1} + \varepsilon_1) < g(n)$
- $P(p_{T2} > \delta_{T2} + \varepsilon_2) < g(n_2)^{1_{n_2>1}}$
- $P(p_E < \delta_E - \varepsilon_E) < g_2(n)^{1_{n>11}}$.

Here, ε_1, ε_2, and ε_E are constants that are predetermined and discussed further in the trial protocol. In conclusion, the selection process prioritizes the highest effective dose within the bounds of toxicity constraints. The notation $1_{(...)}$ denotes the indicator function, which takes the value of one when the specified condition is satisfied and zero otherwise. This mechanism ensures that constraints related to $T2$ and efficacy only impact dose selection decisions when relevant data is available. To accommodate the evolving nature of the trial, adaptive thresholds $g(n)$ and $g_2(n)$ are suggested. These thresholds are dynamically adjusted based on the number of participants already enrolled in the trial and for whom data has been collected:

$$g(n) = \max\left(0.5, 0.9\frac{1}{1+0.025n}\right)$$
$$g_2(n) = \max\left(0.5, 0.9\frac{1}{1+0.02n}\right).$$

The parameters ε_E, ε_1, and ε_2 were determined to be 0.02 each, based on a sensitivity analysis conducted in the context of the LEVNEONAT clinical trial, where the thresholds $\delta_{T1} = \delta_{T2} = 0.1$ for toxicity and $\delta_E = 0.6$ for efficacy were established. In situations where no dose meets the eligibility criteria the trial is discontinued. As extra conditions, which overlap with the previous set as the trial evolves, the trial is stopped if the likelihood of the first dose being excessively toxic ($P(p_{T1} > \delta_{T1}|d_1) > 0.9$ or $P(p_{T2} > \delta_{T2}|d_1) > 0.9$) or the last dose being insufficiently efficacious ($P(p_E < \delta_E|d_K) > 0.9$) exceeds 90%. Additionally, the no-skipping rule is enforced, requiring that at least one patient is treated with every preceding dose level before advancing to a higher dose.

Upon concluding the trial, the minimum effective dose ($d_{e,\min}$) is identified as the dose that minimizes the absolute difference between the estimated probability of efficacy ($\hat{p}_E(d)$) and the efficacy threshold (δ_e), across all considered doses (D). Similarly, the MTD ($d_{t,\max}$) is determined as the minimum of the doses that minimize the absolute differences between the estimated probabilities of toxicities ($\hat{p}_{T1}(d)$, $\hat{p}_{T2}(d)$) and the toxicity thresholds (δ_{T1}, δ_{T2}). The

recommended dose at the trial conclusion equals $d_{t,\max}$ provided $d_{e,\min} \leq d_{t,\max}$; otherwise, no dose is recommended. The estimated probabilities $\hat{p}_E(d)$, $\hat{p}_{T1}$, and $\hat{p}_{T2}(d)$ are calculated as the posterior median values given the dose d.

3.3.6 Results

Design performances have been tested via extensive simulation studies. For each scenario, we simulated 1000 trials with four dose levels. We established a cohort size of two newborns per dose level and total sample sizes of 30, 40, and 50 neonates for each trial, with the assumption that one newborn would be enrolled every 15 days. For detailed information on how the scenarios were generated, please consult the published paper. We tested either the approach with relevance weights (M_1) or the one without them (M_2) within the pseudo-likelihood framework. To assess the effectiveness of our design proposal, we compared the percentage of correct dose selection (PCS) upon trial conclusion. Obtained results in six relevant scenarios are shown in Tables 3.2 and 3.3.

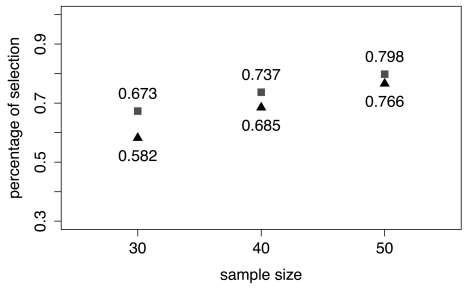
In Scenario 1, without the addition of drug $A2$, Model M_1 (represented by squares in the table) achieved a higher PCS than Model M_2 (indicated by triangles in the table), with a PCS exceeding 67% for groups of 30 or more patients. This scenario examined the impact of relevance weights. In Scenario 2, which was similar to Scenario 1 but included the use of $A2$ with a probability of $p_a = 0.5$, once again the PCS for M_1 surpassed that of M_2. Scenario 3, where the optimal dose under toxicity constraints was the last one in the sequence and included $A2$, saw PCS values exceeding 80% with M_1. Scenario 4 suggested that while all administered doses were deemed safe, only the last dose achieved the desired efficacy goal of 60%. Here, PCS values were above 71% for both models across all sample sizes. Scenario 5 aimed to assess a situation where the probability of $T1$ remained constant, but the probability for $T2$ increased with the addition of $A2$. In this scenario, M_2 showed higher PCS for groups of 30 patients. Scenario 6 modeled $T1$ and $T2$ as independent from each other, resulting in observed PCS rates around 60% for both models across all sample sizes.

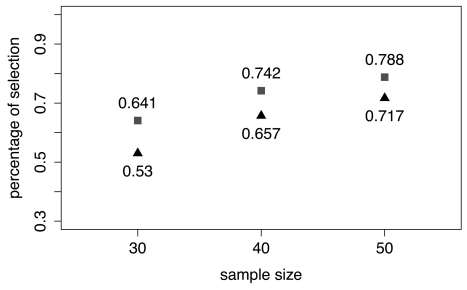
3.4 Bayesian seamless Phase II/III trials with correlated survival endpoints

In this section, I transition from dose-finding to seamless Phase II/III trials. The focus is on examining how information can be transferred between Phase II and Phase III when employing two distinct but related efficacy endpoints (as for T2 and T1 in my earlier work within the context of toxicity, but now in the context of efficacy). In particular, in this work, done by a PhD student I have co-supervised, the objective was to develop and evaluate Bayesian operationally seamless Phase II/III trial designs that employ a binary endpoint for Phase II and a time-to-event endpoint for Phase III [Duputel et al., 2023]. Upon concluding Phase II, arm selection relies on posterior probabilities for assessing futility and on predictive probabilities for making selection decisions. The results from Phase II are subsequently integrated into the prior distributions, this is why it continues to be operationally seamless, for a time-to-event model in Phase III.

Table 3.2 – Results for the first three scenarios in term of right dose selection are shown for sample size of 30, 40 and 50 neonates. In the second column, values for p_{T1}, p_{T2}, p_E, along with p_a, $p_{T1|A2}$ and $p_{T2|A2}$ used in simulations are summarized for each dose. In the third column, the percentage of correct selection (PCS) is given through a plot where the PCS is plotted versus sample size. Squares refer to M_1 results, triangles refer to M_2 ones.

	dose				PCS
	1	2	3	4	
Scenario 1					
$p_{T1,true}$	0.001	0.01	**0.1**	0.2	
$p_{T2,true}$	0.001	0.01	**0.1**	0.2	
$p_{E,true}$	**0.6**	0.7	0.8	0.9	
$p_a = 0$					
Recommended dose			**3**		
Scenario 2					
$p_{T1,true}$	0.001	0.01	**0.1**	0.2	
$p_{T2,true}$	0.001	0.01	**0.1**	0.2	
$p_{E,true}$	**0.6**	0.7	0.8	0.9	
$p_a = 0.5$					
$p_{T1\|A2,true}$	0.005	0.05	0.15	0.25	
$p_{T2\|A2,true}$	0.005	0.05	0.15	0.25	
Recommended dose			**3**		
Scenario 3					
$p_{T1,true}$	0.001	0.001	0.01	**0.1**	
$p_{T2,true}$	0.001	0.006	0.026	**0.09**	
$p_{E,true}$	0.5	**0.6**	0.7	0.8	
$p_a = 0.5$					
$p_{T1\|A2,true}$	0.005	0.005	0.05	0.15	
$p_{T2\|A2,true}$	0.005	0.005	0.05	0.15	
Recommended dose			**4**		





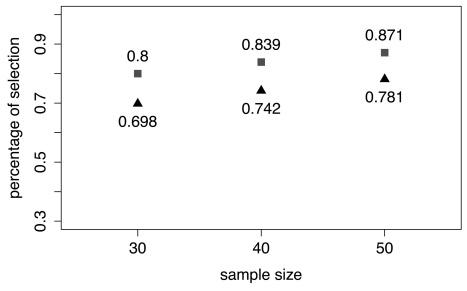
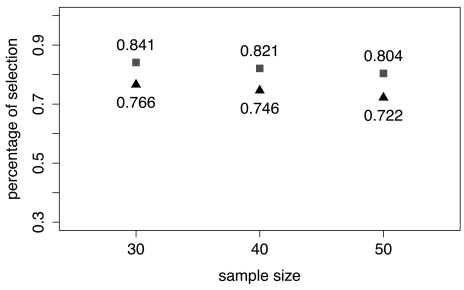
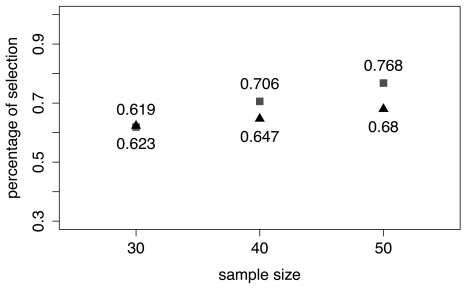
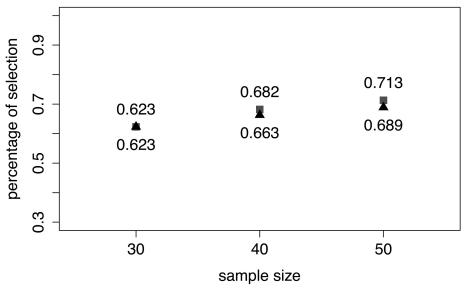


Table 3.3 – Results for the last three scenarios in term of right dose selection are shown for sample size of 30, 40 and 50 neonates.In the second column, values for p_{T1}, p_{T2}, p_E, along with p_a, $p_{T1|A2}$ and $p_{T2|A2}$ used in simulations are summarized for each dose. In the third column, the percentage of correct selection (PCS) is given through a plot where the PCS is plotted versus sample size. Squares refer to M_1 results triangles refer to M_2 ones.

	dose				PCS
	1	2	3	4	
Scenario 4					
$p_{T1,true}$	0.001	0.005	0.01	0.05	
$p_{T2,true}$	0.001	0.007	0.015	0.05	
$p_{E,true}$	0.3	0.4	0.5	**0.6**	
$p_a = 0.5$					
$p_{T1\|A2,true}$	0.005	0.009	0.012	0.06	
$p_{T2\|A2,true}$	0.005	0.009	0.012	0.06	
Recommended dose		**4**			
Scenario 5					
$p_{T1,true}$	0.01	**0.1**	0.25	0.35	
$p_{T2,true}$	0.009	**0.1**	0.18	0.26	
$p_{E,true}$	0.6	0.7	0.8	0.9	
$p_a = 0.5$					
$p_{T1\|A2,true}$	0.01	0.1	0.25	0.35	
$p_{T2\|A2,true}$	0.01	0.1	0.25	0.35	
Recommended dose		**2**			
Scenario 6					
$p_{T1,true}$	0.001	0.01	**0.1**	0.2	
$p_{T2,true}$	0.01	**0.1**	0.2	0.3	
$p_{E,true}$	**0.6**	0.7	0.8	0.9	
$p_a = 0.5$					
$p_{T1\|A2,true}$	0.005	0.05	0.15	0.25	
$p_{T2\|A2,true}$	0.005	0.05	0.15	0.25	
Recommended dose		**2**			

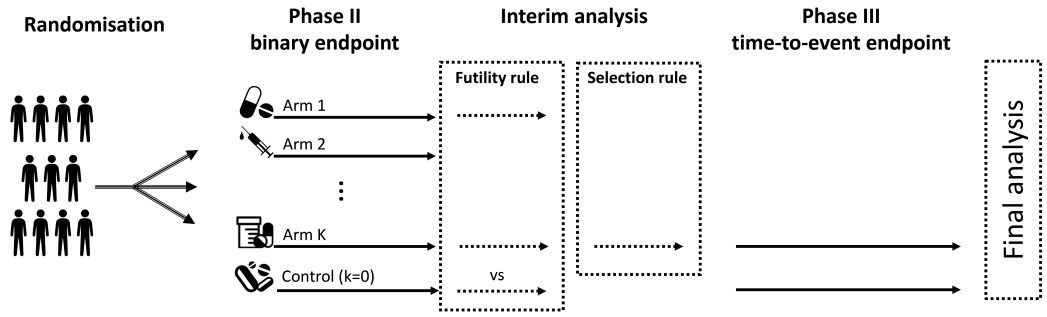


Figure 3.4 – Scheme of the seamless design, with a Phase II involving binary outcomes, that is, mortality rate at a prespecified time, and a Phase III with a time-to-event endpoint.

The binary outcomes for the Phase II stage and the survival outcomes for the Phase III stage are derived from Atalante-1 (NCT02654587), our motivational study that features only two arms. However, we suggest expanding this design framework to accommodate multiple treatment arms during the initial stage, as depicted in Figure 3.4. For ease of explanation and analysis, we have chosen to adopt a 1:1 randomization ratio across both stages.

Let us consider k as the index for the K treatments in the trial, with $k = 0$ representing the control group. The maximum sample size for the Phase II stage for each treatment group is denoted by N_k, and for the Phase III stage, the maximum sample size for the selected treatment arm ($\tilde{k} = 1$) and the control arm ($\tilde{k} = 0$) is represented by $M_{\tilde{k}}$. The time to event (e.g., death) for individuals in Phase II and Phase III are indicated by $t_{i,k}$, with $i \in \{1, \ldots, N_k\}$, and $t_{i,\tilde{k}}$, with $i \in \{1, \ldots, M_{\tilde{k}}\}$, respectively. Similarly, $c_{i,k}$ and $c_{i,\tilde{k}}$ stand for the censoring times for each participant. For each participant, the observation is either the event time or the censoring time (indicating loss to follow-up or being alive at the study conclusion). For individuals in Phase II, $y_{i,k}$ marks the earlier of the two times (event or censor), and $\nu_{i,k}$ is the event indicator, that is $y_{i,k} = \min(t_{i,k}, c_{i,k})$ with $t_{i,k}, c_{i,k} > 0$ and

$$\nu_{i,k} = \begin{cases} 1 \text{ if } y_{i,k} = t_{i,k} \iff t_{i,k} \leq c_{i,k} \\ 0 \text{ if } y_{i,k} = c_{i,k} \iff t_{i,k} > c_{i,k} \end{cases}.$$

12Let $\mathcal{D}_k$ represent the dataset from group k, where $\mathcal{D}_k = \{N_k, \mathbf{y}_k, \boldsymbol{\nu}_k\}_{k=0,\ldots K}$. In this representation, $\mathbf{y}_k$ and $\boldsymbol{\nu}_k$ are vectors of length N_k that group all observed values of $y_{i,k}$ (the time of the first occurrence between event or censor for each participant) and $\nu_{i,k}$ (the event indicator, signifying whether the event of interest or censoring occurred first), respectively, for each treatment group indexed by k from 0 to K. Similarly, $\mathcal{D}_{\tilde{k}}$ is defined for $\tilde{k} = 0, 1$ in the Phase III.

3.4.1 Stage 1 - Phase II

At this stage, a binary primary endpoint is defined as survival status at a specific time point, t^*, which for our case is set at 12 months. The outcomes are categorized into two groups: those who die before reaching t^* and those who are still alive at that time. To simplify, patients who

are censored before t^* are not included in the calculation of the survival rate, though imputation techniques can be applied to address this issue. The analysis for Phase II is scheduled to occur once N_k patients in each group k have been enrolled and completed the follow-up period. The survival indicator at time t^* for patient i in group k, denoted as $y^*_{i,k}$, is defined as follows:

$$y^*_{i,k} = \begin{cases} 1 & \text{if } y_{i,k} \geq t^* \\ 0 & \text{if } y_{i,k} < t^*, \end{cases} \tag{3.6}$$

where $i \in \{1, \ldots, n_k\}$, and $n_k \leq N_k$ represents the count of noncensored patients in group k. Let p_k represent the probability of survival at time t^* for group k, which lies within the interval [0,1]. The sum of survival indicators $y^*_{i,k}$ follows a Binomial distribution, $\sum y^*_{i,k} \sim \text{Binomial}(n_k, p_k)$. To facilitate inference on p_k, we employ a logit link function, that is, $p_k = \text{logit}^{-1}(\theta_k)$, where $\theta_k = \theta_0 + \mu_k$ for $k = 1, \ldots, K$, and θ_0 is the parameter for the control arm, with $p_0 = \text{logit}^{-1}(\theta_0)$. For the Bayesian framework, normal prior distributions are specified, with μ_k following $\mathcal{N}(\tilde{\mu}_k, \sigma_k^2)$, and θ_0 following $\mathcal{N}(\mu_0, \sigma_0^2)$. Historical data pertaining to the control group can inform the hyperparameters, and $\tilde{\mu}_k$ might be set to 0 to adopt a conservative stance assuming no treatment effect. While a beta-binomial model could alternatively be used for each arm analysis, the suggested model not only facilitates quicker interpretation of treatment effects but also allows for further model complexity enhancements. Such enhancements might include hierarchical structures, incorporation of dosage levels within treatment groups, or establishing a dose-response relationship for specific groups.

At the conclusion of Phase II, the objective is to identify the most promising treatment arm. This is achieved through a two-step algorithm:

1. Futility rule: Initially, any treatment arms that fail to surpass a specified threshold, τ_1, for the posterior probability of demonstrating a superior survival rate compared to the control arm, expressed as $\mathbb{P}(p_k - p_0 > 0 | \mathcal{D}_k) < \tau_1$, are eliminated from further consideration.
2. Selection of the most promising arm: From the remaining arms, the one exhibiting the highest predictive probability of success is chosen for the Phase III survival study. Success here is defined by the likelihood that a future patient will survive at least one year under the treatment. In the context of a binomial distribution, this predictive probability aligns with the expected value of the posterior distribution of p_k, calculated as $\int p_k \pi_{post}(p_k|y_k)\, dp_k$.

If the futility rule leads to the disqualification of all treatment arms, the trial is concluded. Interchanging the order of the two steps does not affect the outcome.

3.4.2 Stage 2 - Phase III

Upon the identification of the most promising treatment arm, the study progresses to Phase III, focusing on a survival endpoint. For modeling the survival times, we opt for the Weibull distribution, though it is worth noting that alternative parametric distributions could also be applicable. The chosen parametrization for regression analysis involves the shape parameter α and scale parameter $\gamma_{\tilde{k}}$, using the notation $W(\alpha, \gamma_{\tilde{k}})$. The probability density function is thus defined as:

$$f(t|\alpha, \gamma_{\tilde{k}}) = \frac{\alpha}{\gamma_{\tilde{k}}} \left(\frac{t}{\gamma_{\tilde{k}}} \right)^{\alpha - 1} \exp\left(- \left(\frac{t}{\gamma_{\tilde{k}}} \right)^{\alpha} \right),$$

with the condition that $\alpha > 0$ and $\gamma_{\tilde{k}} > 0$, where $\tilde{k} = 0$ for the control arm and $\tilde{k} = 1$ for the selected treatment arm. In this Weibull survival regression framework, the shape parameter α is shared between the two arms, indicating a common survival pattern curvature, whereas the scale parameter $\gamma_{\tilde{k}}$ varies across arms, reflecting differences in the survival times scaling between the control and the treatment groups. For each arm in the study, the survival function associated with the time to event or censoring is expressed as:

$$S(y_{i,\tilde{k}}|\alpha, \gamma_{\tilde{k}}) = \exp\left(-\left(\frac{y_{i,\tilde{k}}}{\gamma_{\tilde{k}}}\right)^{\alpha}\right),$$

and the likelihood for $M_{\tilde{k}}$ patients accrued in arm $\tilde{k}$, which accommodates censored data, is given by:

$$L(\alpha, \gamma_0, \gamma_1|\mathcal{D}) = \prod_{\tilde{k}=0}^{1}\prod_{i=1}^{M_{\tilde{k}}} f(y_{i,\tilde{k}}|\alpha, \gamma_{\tilde{k}})^{\nu_{i,\tilde{k}}} S(y_{i,\tilde{k}}|\alpha, \gamma_{\tilde{k}})^{(1-\nu_{i,\tilde{k}})},$$

where $y_{i,\tilde{k}}$ denotes the observed time (event or censoring) for individual i in the $\tilde{k}$-th group, and $\nu_{i,\tilde{k}}$ is the indicator of whether this observed time is due to an event (1) or censoring (0). To facilitate more stable inference, the log transformation of the scale parameter, $\beta_{\tilde{k}} = \log(\gamma_{\tilde{k}})$, is estimated for each arm $\tilde{k}$. To close the Bayesian model for Phase III, incorporating prior distributions for the parameters α, β_0, β_1 is essential. Here, we discuss two methodologies that utilize outcomes from Phase II for setting these priors.

In the first approach, called the ESS (Effective Sample Size) approach, the prior for the shape parameter α is chosen to be an Inverse Gamma distribution, $\pi(\alpha) = \text{InverseGamma}(\rho, \kappa)$, with both ρ and κ being greater than zero. For the scale parameters $\beta_{\tilde{k}}$, normal distributions are selected, $\pi(\beta_{\tilde{k}}) = \mathcal{N}(\mu_{\tilde{k}}, \sigma^2_{\tilde{k}})$, with $\sigma_{\tilde{k}} > 0$ indicating the standard deviation. The hyperparameter $\mu_{\tilde{k}}$ for each $\beta_{\tilde{k}}$ is derived from the survival rates estimated at the end of Phase II, specifically at time t^*. This estimation leverages the survival function relationship:

$$\tilde{p}_{\tilde{k}} = \exp\left(-\left(\frac{t^*}{\exp(\beta_{\tilde{k}})}\right)^{\alpha}\right),$$

where $\tilde{p}_{\tilde{k}}$ represents the estimated survival rate for the $\tilde{k}$-th arm, typically determined by the posterior mean or median of the survival probability at time t^*. Assuming a fixed value for α (denoted as $\tilde{\alpha}$, often set to 1 for simplicity), the mean $\mu_{\tilde{k}}$ for the prior distribution of each $\beta_{\tilde{k}}$ can be calculated using the formula:

$$\mu_{\tilde{k}} = \log\left(\frac{-t^*}{\log\left(\tilde{p}_{\tilde{k}}\right)^{1/\tilde{\alpha}}}\right).$$

The standard deviation parameters, $\sigma_{\tilde{k}}$, are determined to correspond with a desired ESS, a concept that reflects the amount of information encapsulated within the prior distribution, akin to the inclusion of an equivalent number of additional patients in the analysis. Different methodologies for calculating ESS are discussed in the literature [Morita et al., 2008, Neuenschwander et al., 2020, Wiesenfarth and Calderazzo, 2020], however, applying these ESS concepts directly to time-to-event models poses certain challenges. Therefore, we propose applying a novel ESS

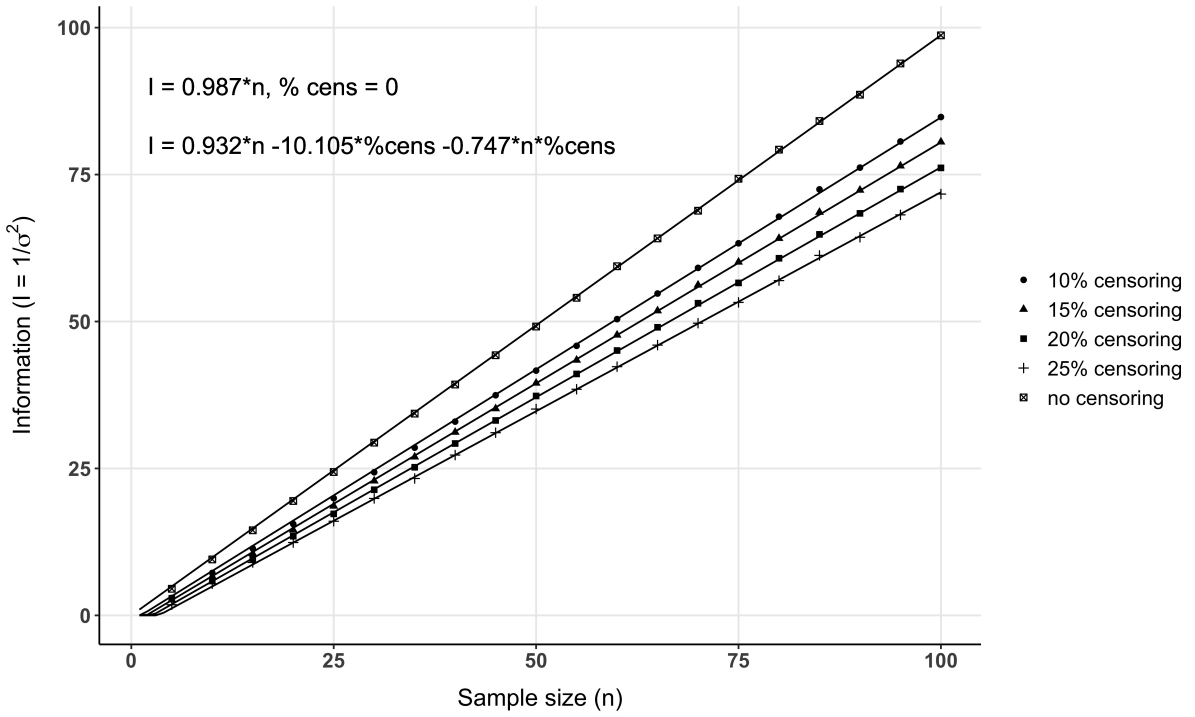


Figure 3.5 – Information vs. sample size and censoring rate and the estimated regression lines.

concept based on the unit information principle [Liang et al., 2008], a concept that originates from frequentist statistics but is equally applicable in Bayesian settings. The unit information, $\mathcal{I}_u$, represents the information contributed by a single subject, often defined as the Fisher information matrix normalized by the sample size. In Bayesian terms, it can be thought of as the inverse of the variance of the posterior distribution per sample unit. To achieve an ESS of n_*, we adjust the standard deviation parameter, $\sigma_{\tilde{k}}$, such that it aligns with the unit information scaled by n_*, specifically: $\sigma_{\tilde{k}} = (n_* \mathcal{I}_u)^{-1/2}$. This scaling ensures that the prior influence is equivalent to adding n_* hypothetical subjects to the analysis. The calculation of $\mathcal{I}_u$ is linked to the statistical model in use and is derived from examining the expected relationship between sample sizes and the variance of posterior estimates. In the context of survival analysis, the censoring rate (c_r) also affects the unit information, necessitating a formulation where $\sigma_{\tilde{k}}$ accounts for both the sample size and the censoring rate, rendering it as: $\sigma_{\tilde{k}} = (\mathcal{I}(n^*, c_r))^{-1/2}$ (Figure 3.5).

In the second approach, called the likelihood approach, for determining the joint prior distribution of α, β_0, β_1 we leverage the binary likelihood derived from Phase II data, employing the Weibull distribution to model survival times. This method integrates the survival function $S(t^*|\alpha, \exp(\beta_{\tilde{k}}))$ for the binary outcome observed at time t^* (the predefined survival threshold), across both treatment arms (control and selected treatment arm). The prior distribution is formulated as follows:

$$\pi(\alpha, \beta_0, \beta_1) \propto \prod_{\tilde{k}=0}^{1} S(t^*|\alpha, \exp(\beta_{\tilde{k}}))^{y_{\tilde{k}}} \left(1 - S(t^*|\alpha, \exp(\beta_{\tilde{k}}))\right)^{n_{\tilde{k}} - y_{\tilde{k}}} \pi(\alpha)\pi(\beta_{\tilde{k}}),$$

where $\pi(\alpha)$ and $\pi(\beta_{\tilde{k}})$ are chosen to be noninformative priors, potentially following the same distributions as those used in the ESS approach. Here, $n_{\tilde{k}}$ represents the total number of patients analyzed at Phase II in arm $\tilde{k}$, and $y_{\tilde{k}} = \sum y^*_{i,\tilde{k}}$ signifies the number of patients who survived until t^* in the same arm.

The determination of the treatment superiority over the control arm in the final analysis hinges on the parameter denoting the treatment effect, $\Delta = \beta_1 - \beta_0$. The criterion for asserting the treatment superiority is that the posterior probability of Δ being greater than zero exceeds a predetermined threshold, τ_2: $P(\Delta > 0) > \tau_2$. This approach mirrors the rationale behind one-sided hypothesis testing in frequentist statistics, where τ_2 is commonly set to 0.975 to align with a 95% confidence level for the superiority margin. In scenarios requiring a more pronounced treatment effect or where a noninferiority margin (ζ_u) is permissible, the criterion can be adjusted to: $P(\Delta > \zeta_u) > \tau_2$. This adjustment allows for flexibility in defining what constitutes a clinically meaningful difference between the treatment and control arms, taking into account both statistical significance and clinical relevance.

3.4.3 Results

In our comprehensive simulation study, we assessed the performance of five distinct trial designs, including a frequentist operationally seamless method (F-OP) as the primary benchmark. In the F-OP design, both stages of the trial, Phase II and Phase III, are outlined within the same protocol and conducted sequentially. However, only the data collected during Phase III are utilized to make the final determination regarding the treatment efficacy. Importantly, this approach does not necessitate an adjustment for Type I error, given that the interim analysis at the end of Phase II does not influence the statistical inference for the final claim. The Bayesian operational counterpart, identified as B-OPwinf, employs a weakly informative prior for the second stage of the trial. This approach is characterized by an ESS of 1, with specific hyperparameters for the prior distribution set as $\rho = 10$ and $\kappa = 9$. The alternative Bayesian method, denoted as B-OPinfESS, incorporates an informative prior, tailored through the ESS approach. This method calculates the standard deviation parameter, $\sigma_{\tilde{k}}$, as: $\sigma_{\tilde{k}} = \left(\mathcal{I}(n_{\tilde{k}}, \hat{c}_{r,\tilde{k}})\right)^{-1/2}$, leveraging all $n_{\tilde{k}}$ Phase II patients in arm $\tilde{k}$ who were not censored before 12 months, alongside the empirical censoring rate $\hat{c}_{r,\tilde{k}}$ observed in Phase III for the same treatment arm. The fourth method, labeled B-OPinLIK, adopts the same foundational principles as B-OPinfESS but distinguishes itself by constructing the informative prior through a likelihood-based approach. The fifth method introduced for comparative analysis is a purely Bayesian inferential design, denoted as B-INFER. Unlike the operational designs, B-INFER undertakes a comprehensive analysis at the trial conclusion, utilizing the actual survival times from all patients across both phases. This means that for B-INFER, detailed survival data, including the time of death for each patient from both the control and selected treatment arms from Phase II, are followed up, recorded, and analyzed, beyond the binary outcomes initially used for selection decisions at the end of Phase II. Weakly informative priors, similar to those employed in B-OPwinf, are utilized in B-INFER to ensure a minimal prior influence. For all Bayesian designs discussed, including B-INFER, the efficacy threshold τ_2 is set at 0.975. This threshold is selected to align with the traditional 2.5% significance level used in frequentist hypothesis testing (although we know that they are not the

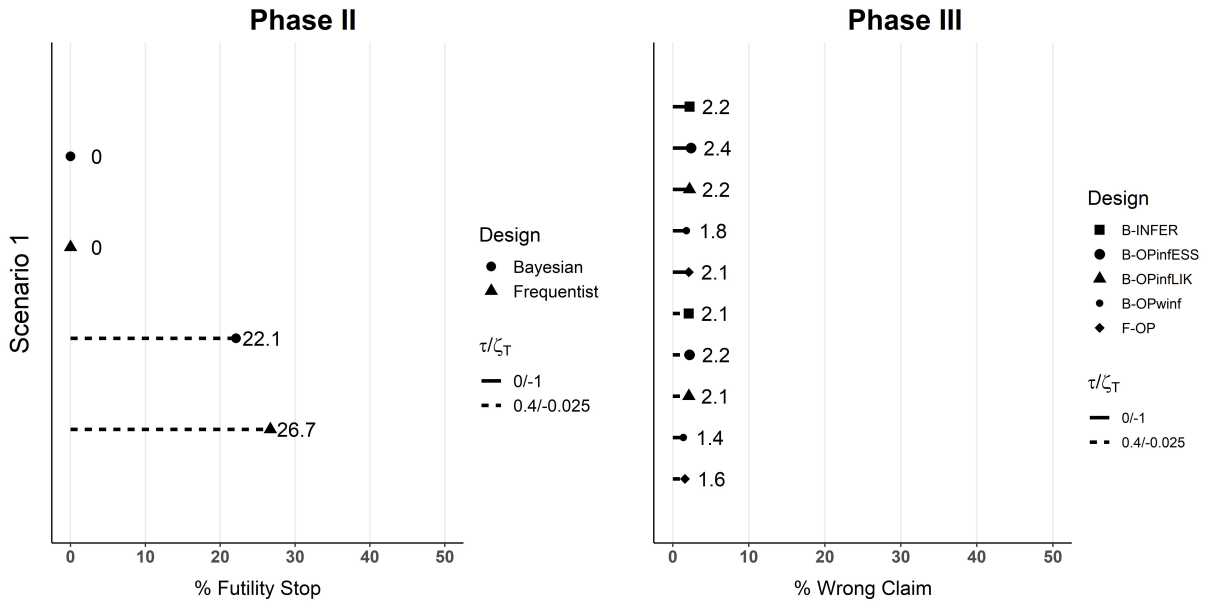


Figure 3.6 – Results in terms of simulated type one error. On the left-hand side, the percentage of trials that were stopped at the futility analysis at the end of Phase II in scenario 1 is shown. On the right-hand side, the percentage of trials where the treatment arm was erroneously claimed to be superior is shown. Straight lines refer to no futility rule applied, that is, $\tau = 0$ and $\zeta_T = -1$, while dashed lines refer to $\tau = 0.4$ and $\zeta_T = -0.025$.

same theoretically). To assess the operating characteristics, including simulated Type I error rates and power, of the various trial designs discussed, four principal scenarios were identified for detailed analysis. For each scenario, 1000 trial simulations were conducted.

Figure 3.6 illustrates the outcomes for five trial designs under scenario 1. Applying a futility threshold $\tau = 0.4$ or $\zeta_T = -0.025$ leads to the cessation of 22.1% and 26.7% of trials at the interim stage for Bayesian and frequentist designs, respectively. Without futility analysis ($\tau = 0$ or $\zeta_T = -1$), the Type I error rate varies between 1.8% (B-OPwinf) and 2.4% (B-OPinfESS). With futility, errors range from 1.4% to 2.2%.

Figure 3.7 presents power results in scenarios 2, 3, and 4, highlighting the frequency of correct arm selection for both Bayesian and frequentist methods under identical futility thresholds. Futility generally decreases correct selections by about 5 points, correlating with the rate of early trial terminations. B-OPwinf performs comparably to F-OP. B-OPinfESS shows similar or marginally lower correct claim rates than B-INFER, with B-OPinfLIK effectiveness positioned between the two other Bayesian operational designs.

3.5 Discussion

The first study, DICE was designed to enable the prediction of outcomes across different treatment cycles assessed in the trial. Making reliable predictions, which involves accurately estimating the likelihood of toxicity for various sequences and cycles, is a crucial aspect of this new approach. For example, if it turns out that the treatment is effective with fewer cycles at later evaluations, DICE could adjust its toxicity probability estimates and update the recommended

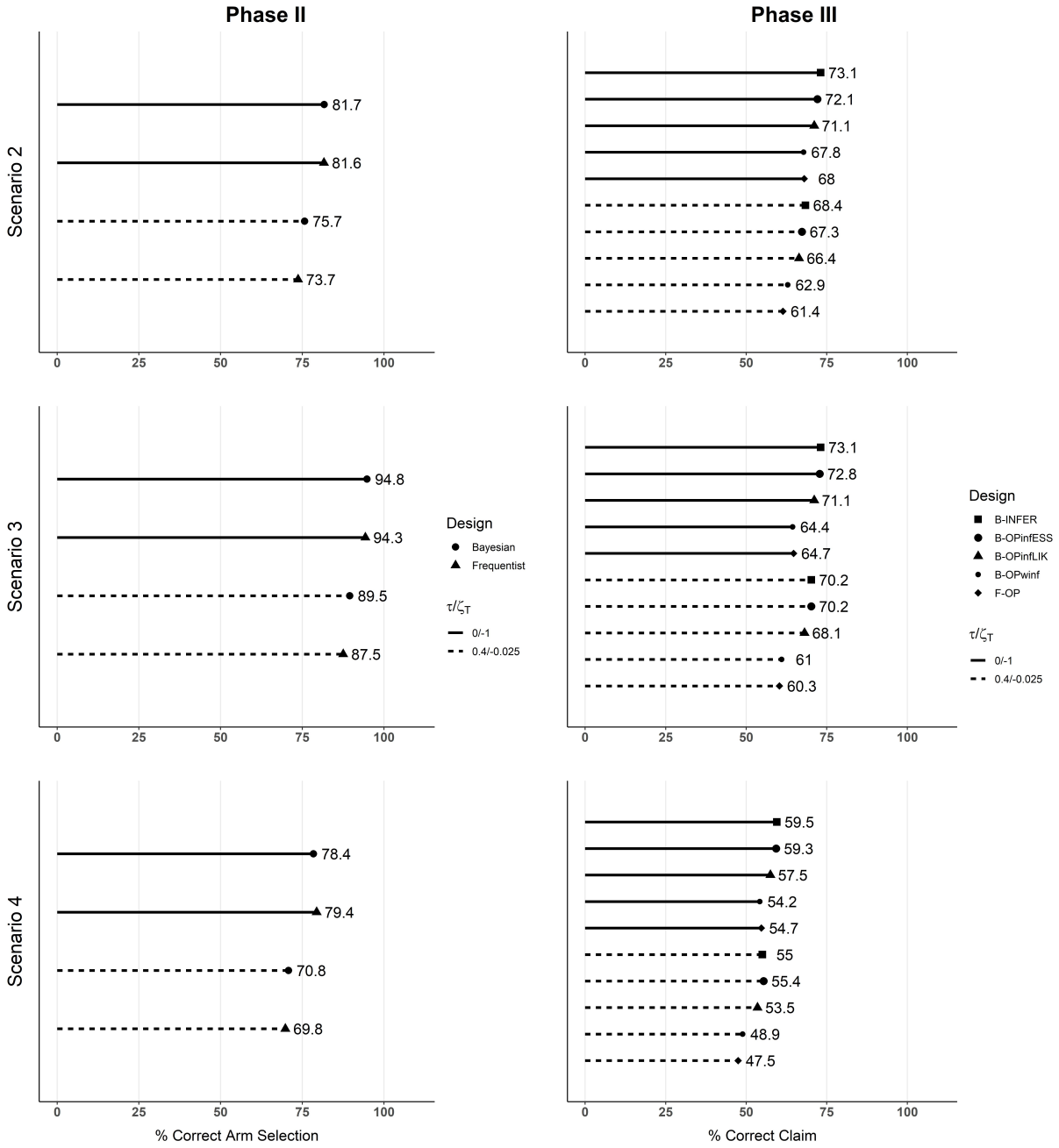


Figure 3.7 – Results in terms of simulated power. Each plot line represents one scenario. In the first column, the percentages of correct arm selection at the end of Phase II are shown for the Bayesian methods and the frequentist one. In the second column, the percentages of final Phase III correct claims associated with each design are given. Straight lines refer to no futility rule applied, that is, $\tau = 0$ and $\zeta_T = -1$, while dashed lines refer to $\tau = 0.4$ and $\zeta_T = -0.03$.

MTDS accordingly. While DICE has the potential to predict outcomes for untested sequences, like different doses or more cycles, we must be very cautious about these predictions, especially when they involve doses or cycles not previously tested (this would be a form of extrapolation). Unlike the methods that used PK/PD data, this approach has limited information sharing. Another challenge is the requirement for a predefined order of dose sequences based on toxicity risk. If this order is not clear, we may need to reconsider the no-skipping rule. The DICE forces and weakness are related to the discussion on the "dose-estimators" and the "dose-finders" done in the previous chapter. It is challenging to outperform simpler models like the TITE-CRM, which are specifically designed for a single purpose (identifying the MTD). However, more complex models could facilitate future optimization and still deliver performance comparable to the simpler ones on MTD (or MTDS) finding.

The second work represents a modeling exercise tailored to a particular case study. The simulation study for the proposed method suggests it offers a balanced compromise between model complexity, achieving high probability of correct selection, and maintaining a reasonable number of both short and long-term toxicities, even with the limitations of a small sample size. The incorporation of clinical relevance weights has proven beneficial in ensuring the dose allocation process progresses smoothly without being stuck. Even if the work was tailored for the LEVNEONAT trial, the methodology itself is versatile and can be adapted to various scenarios by modifying different model components, such as the time-to-event distribution and the relevance weights. To effectively customize this method, it is essential to foster close cooperation among statisticians, physicians, and other collaborators involved in the trial. Expert elicitation played a crucial role in this work, particularly in setting the shape of the time-to-event distributions and other design parameters, allowing the model to be estimated. This approach simplified the final model to a level that it could be estimated within a timeframe acceptable for practical hospital procedures.

Finally, in the third work, we delved into the application of the Bayesian approach in seamless Phase II/III study designs, specifically focusing on how information transfer can occur between different yet related endpoints. We introduced two methods for establishing informative priors. The first method, known as the ESS approach, assesses the appropriate amount of unit information for a Weibull survival regression, finding a linear correlation in our setting, although the ESS might change with the survival rate in other models. The second method, the likelihood approach, integrates information at the binomial level instead of the survival level, leading to less informative prior distributions. However, using only Phase II data to estimate potential dropout rates might deliver more substantial information than that obtained from actual Phase III participants, given that a censored patient provides less data than one who experiences the event in the survival analysis. While enhancing study power, this method risks increasing Type I errors in inferentially seamless designs. A more cautious strategy might involve reducing the number of patients used to create the prior distribution to mitigate this risk. As it is standard with such designs, conducting a simulation study is necessary to examine the operating characteristics.

Chapter 4

Building prior distributions

Contents

This chapter marks our step into exploring the use of external data, contrasting with the internal sources discussed in Chapters 2 and 3. I will detail two studies: the first focuses on Phase I dose-finding studies where external information comes from an external clinical trial; the second delves into the context of a randomized Phase II study, incorporating experts' elicitation as external information. In the first work, I briefly present the work I have conducted on building prior distributions in Bayesian inference, incorporating external data into the actual study. In Ollier et al. [2020], a project conducted by a PhD student under my co-supervision, we introduced a power prior approach for utilizing data from a previous clinical trial within the context of Phase I bridging studies, specifically from Caucasians to Asians, where there is a study already performed with Caucasian participants and a new prospective study is planned with Asian participants. Then, in the second study where I am a co-first author, we explore the use of expert elicitation to build prior distributions for analyzing a randomized Phase II trial [Thall et al., 2019]. This approach is especially beneficial for rare diseases with limited sample sizes, where seeking external opinions can aid in building appropriate prior distributions. Maintaining the organization of the priors chapters, this section begins with a succinct introduction to frame the discussion around the two works. Subsequently, I will provide summaries of these papers, concentrating on the methodologies utilized and the findings achieved. The chapter will conclude with a brief discussion.

4.1 Introduction

The use of Bayesian inference in the design, conducting, and analysis of clinical trials is becoming more common. One of its key benefits is the ability to incorporate external information, such as data from prior clinical trials, electronic health records, medical publications, and experts' opinions into the statistical analysis. This capability can lead to smaller required sample sizes while enhancing the study statistical power [Hobbs et al., 2011, Röver and Friede, 2020, van Rosmalen et al., 2018]. Nonetheless, the integration of external data sources into the prior distribution must be approached with caution, as these inputs can either contradict or strengthen the derived posterior outcomes. Several techniques for incorporating historical data into prior distributions are highlighted in the literature, including: the power prior (PP), the commensurate prior and the meta-analytic predictive (MAP) prior. When using the PP approach, the quantity of information to be utilized in the prior is determined by assigning a weight to the prior data [Ibrahim and Chen, 2000, Neuenschwander et al., 2009]. This weight acts as a power parameter and ranges from 0 to 1, where 0 signifies a non-informative prior and 1 indicates full borrowing from the prior. However, the process of determining this parameter is subjective and may result in incorrect specification of the prior if not executed properly [Ibrahim et al., 2015]. The commensurate priors aims at quantifying the level of similarity between the informative prior distribution and the likelihood [Hobbs et al., 2011, 2012]. In this setting, the researchers introduce a commensurability parameter, which is estimated after observing the trial data incorporated in the likelihood. Concerning commensurability, Schmidli et al. [2014] proposed a meta-analytic approach utilizing the MAP where the parameters are considered exchangeable across trials.

Another way to build prior distributions is considering expert opinions. Numerous researchers across a variety of disciplines have tackled the challenges of gathering expert opinions, formulating prior beliefs based on these opinions, and conducting Bayesian analysis with them. Methods for eliciting priors, as discussed by several researchers [Johnson et al., 2010, Savage, 1971, Chaloner and Duncan, 1983, Kadane and Wolfson, 1998, O'Hagan, 1998, Chaloner and Rhame, 2001, Kuhnert et al., 2010], focus on establishing foundational beliefs for Bayesian model-based clinical trials and medical applications [Spiegelhalter et al., 1994, Tan et al., 2003, Hiance et al., 2009], employing graphical techniques for prior elicitation [Johnson et al., 2010, DuMouchel, 1988, Chaloner et al., 1993], and integrating priors with expert insights [Clemen and Winkler, 1999, Moatti et al., 2013]. A comprehensive overview of these discussions is presented by O'Hagan et al. [2006].

Regarding prior calibration, the Effective Sample Size (ESS) method enables the interpretation of the parametric prior as the number of hypothetical patients considered in developing the prior distribution [Morita et al., 2008, 2010, 2012]. ESS serves as a measure of the prior distribution informativeness, providing a way to quantify it. Other definitions and modification of the ESS notion can be found in recent literature [Neuenschwander et al., 2020, Wiesenfarth and Calderazzo, 2020].

4.2 Adaptive power prior for bridging study

In Ollier et al. [2020], our objective is to introduce a novel methodology that leverages data from a previously conducted fixed historical trial to aid in the design and dose-allocation for an upcoming clinical trial, such as in bridging studies. Bridging studies aim to close the gap on aspects like efficacy, comorbidities, safety, and dosing schedules across different populations. As per the guidelines of the International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use E5 (ICH-E5), a bridging study for a medication is characterized as an extra research conducted within a new group, such as a different ethnic demographic, to establish a link to new clinical data regarding safety, efficacy, and dose-response [ICH, 1998]. We suggest an adaptive prior approach (APP) that is formulated on a criterion combining the power prior, the ESS, and the Hellinger distance. The ESS is utilized to determine the maximum amount of information desired, the power prior for incorporating historical data, and the Hellinger distance for adjusting the extent of data borrowing in case of prior-data conflict.

As usual, we begin by establishing the notation. Let us consider θ as the parameter or group of parameters of interest. For the sake of clarity, we will present all notations as though dealing with a single dimension, yet the principles can be readily extended to multi-dimensional vectors and matrices. Let D_0 represent the historical data, denoted by $D_0 = \{y_j\}_{n_0}$, where n_0 is the sample size of D_0, and $L(\theta|D_0)$ the likelihood of θ given D_0. Similarly, let D represent the current data, $D = \{y_i\}_n$, with n being the sample size for D, and $L(\theta|D)$ the likelihood of θ based on D. We introduce an adaptive power prior, π^{APP}, defined as follows:

$$\pi^{APP}(\theta) = \frac{L(\theta|D_0)^{\alpha_0(1-\gamma)}\pi_0(\theta)}{\int L(\theta|D_0)^{\alpha_0(1-\gamma)}\pi_0(\theta)d\theta}, \tag{4.1}$$

where π_0 is a non-informative prior for θ, and the conventional power prior parameter [Ibrahim and Chen, 2000] is divided into two components, $0 \leq \alpha_0 \leq 1$ and $0 \leq \gamma \leq 1$. These two new parameters, α_0 and γ, referred to as the "quantity of information" and "commensurability" parameters respectively, are assigned specific meanings. Therefore, we recommend a two-step method to establish their values.

4.2.1 Quantity of information parameter value

In the initial phase, α_0 is selected to establish a cap on the maximum amount of information it is preferable to incorporate. During this step, γ is provisionally assigned a value of 0, and the ESS for the adaptive power prior, denoted by Eq. 4.1, is calculated as $s = ESS[\pi^{APP}(\theta|\gamma = 0)] = \alpha_0 ESS[L(\theta|D_0)] + ESS[\pi_0(\theta)]$, where $L(\theta|D_0)$ is considered as a distribution [Morita et al., 2008]. If a highly non-informative prior is opted for π_0, such as an improper uniform distribution when feasible, the second component of the sum, $ESS[\pi_0(\theta)] = s_0$, approaches zero. Given that $L(\theta|D_0)$ encompasses n_0 observations, it is logical to consider that the ESS could nearly equal n_0. The resultant ESS is thus expressed as $s \approx \alpha_0 n_0 + s_0$, showing a linear relation with α_0. Consequently, to achieve a specified desirable ESS, s^*, for the analysis ahead, α_0 can be determined by reversing the aforementioned equation, i.e., $\alpha_0 = (s^* - s_0)/n_0$. α_0 acts as a limit to the information borrowed, given that $ESS[\pi^{APP}(\theta|\gamma = 0)] > ESS[\pi^{APP}(\theta|\gamma > 0)]$. The optimal s^* is contingent upon the application and the sample size n of the current data. Except

in unusual circumstances, it is typically agreed that $s^* < n$ to prevent the prior distributions from overshadowing the actual data.

4.2.2 Commensurability parameter value

In the subsequent step, we determine the commensurability parameter, γ, which is crucial for addressing potential discrepancies between the historical and current datasets. When these datasets significantly diverge, adopting a non-informative prior is advisable; conversely, when they exhibit similarity, full data borrowing is favored. We recommend associating γ with a metric that evaluates the distance between the two datasets, namely, between D_0 and D. We look for a metric which yield a value within the range of 0 to 1, escalating to its peak when the datasets markedly differ, and diminishing to zero as D_0 and D become more aligned. Thus, we suggest employing an empirical Bayes technique that utilizes the Hellinger distance between normalized likelihoods, as follows

$$d^2(D_0, D) = \frac{1}{2} \int \left(\sqrt{\frac{L(\theta|D)}{\int L(\theta|D) d\theta}} - \sqrt{\frac{L(\theta|D_0)^{\frac{n}{n_0}}}{\int L(\theta|D_0)^{\frac{n}{n_0}} d\theta}} \right)^2 d\theta \tag{4.2}$$

where d^2 denotes the square of the Hellinger distance. In this formulation, $L(\theta|D_0)$ is adjusted by a factor of n/n_0 to make it comparable to $L(\theta|D)$, considering that the historical data likelihood, $L(\theta|D_0)$, is presumed to be more precise or exhibit less variance due to a larger sample size ($n_0 \geq n$). For the Bernoulli case, where each y_i follows a Bernoulli distribution, the likelihood function given the historical data, $L(\theta|D_0)$, is expressed as $p^{\sum_{n_0} y_i}(1-p)^{n_0 - \sum_{n_0} y_i}$. Upon incorporating the exponent, the adjusted likelihood becomes $L(\theta|D_0)^{n/n_0} = p^{n\bar{y}_{n_0}}(1-p)^{n-n\bar{y}_{n_0}}$, where $\bar{y}_{n_0} = (\sum_{n_0} y_i)/n_0$. This formulation ensures that the mean of the dataset is maintained, facilitating a comparison between the historical and current likelihoods in terms of their variability. Similarly, in the Gaussian scenario, where y_i is normally distributed with mean μ and variance σ^2, the likelihood function for the historical data is given by $L(\theta|D_0) = (2\pi\sigma^2)^{-n_0/2} \exp\left(-(2\sigma^2)^{-1} \sum_{n_0} (y_i - \mu)^2\right)$. Adjusting for the exponent, we obtain $L(\theta|D_0)^{n/n_0} = (2\pi\sigma^2)^{-n/2} \exp\left(-(2\sigma^2)^{-1} n \left(\sum_{n_0} (y_i - \mu)^2\right)/n_0\right)$. Here, by representing $\sum_{n_0}(y_i - \mu)^2$ as $n_0(\sum_{n_0}(y_i - \mu)^2)/n_0$, it becomes clear that the quantity $\sum_{n_0}(y_i - \mu)^2)/n_0$ is conserved, albeit scaled to reflect a sample size of n. It is important to note that Eq. 4.2 initially assumes $n_0 \geq n$, but it can be adapted for a more general scenario as follows:

$$d^2(D_0, D) = \frac{1}{2} \int \left(\sqrt{\frac{L(\theta|D)^{\min(1, \frac{n_0}{n})}}{\int L(\theta|D)^{\min(1, \frac{n_0}{n})} d\theta}} - \sqrt{\frac{L(\theta|D_0)^{\min(1, \frac{n}{n_0})}}{\int L(\theta|D_0)^{\min(1, \frac{n}{n_0})} d\theta}} \right)^2 d\theta. \tag{4.3}$$

Subsequently, γ is defined as the square root of $d^2(D_0, D)$, taking into account only the real root. This method can be applied universally for any c-roots. The parameter γ can be derived from any power of d, as $\gamma = d^c(D_0, D) \in [0, 1]$ for all $c \in \mathbb{R}^+$. Using values of c greater than 1 will diminish the computed distance, facilitating more extensive data borrowing. Conversely, values less than 1 will adopt a more cautious stance, augmenting the computed distance and thus minimizing the extent of data borrowing.

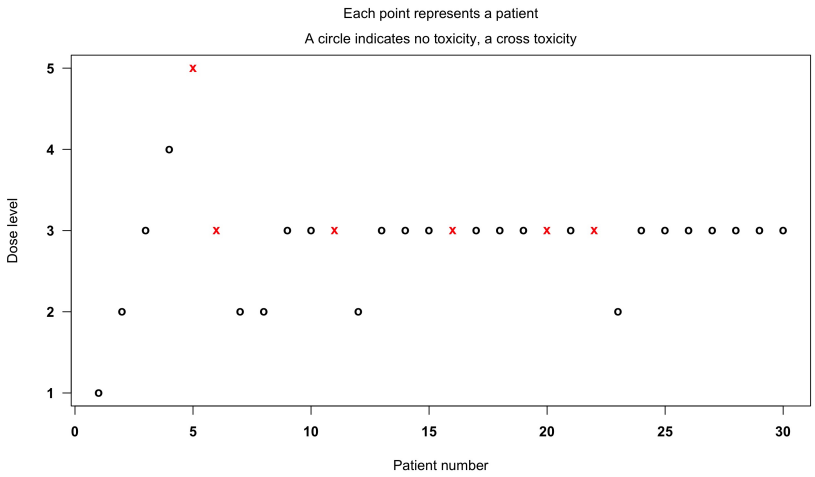


Figure 4.1 – Dose allocation and toxicity representation for the historical data. On the x-axis, the number given to the accrued patient is shown, while, on the y-axis, it is marked at which dose s/he was allocated. A circle denotes that the patient did not experience any DLT, while a cross indicates that the patient had at least a DLT. The historical trial followed a CRM design with 6 doses, $d_i, i = 1, \ldots, 6$, and generating probabilities of toxicity at each dose equals to $p_T(d_1) = 0.05$, $p_T(d_2) = 0.07$, $p_T(d_3) = 0.2$, $p_T(d_4) = 0.4$, $p_T(d_5) = 0.5$ and $p_T(d_6) = 0.55$. This specific trial was chosen since the estimated probabilities of toxicities were similar to the generating ones.

4.2.3 Results

We implement the APP approach in Phase I bridging studies that utilize the CRM for their design. Specifically, we select the logistic model for this purpose, $\text{logit}(p_i) = a + \exp(\beta)\tilde{d}_i$, where p_i denotes the toxicity probability at dose level i, $a = 3$ is a fixed parameter, $\tilde{d}_i$ represents the "effective" dose which is essentially the prior estimate of the toxicity probability associated with dose level i, and β is the parameter of interest that is to be estimated. Via simulations, we tested several version of the APP approach. The traditional CRM, where either $\gamma = 1$ or $\alpha = 0$ is applied across all cohorts to guarantee that no historical data is incorporated, will henceforth be referred to as P_NI. AP_L represents the model where γ is set to $\sqrt{d^2(D_0, D)}$ and features a linear ESS, denoted as $s^*(n) = n$. The variant termed AP_S is identical to the previous model but adjusts γ to $\sqrt[4]{d^2(D_0, D)}$, effectively setting the square root of the Hellinger distance as the value for γ. AP_SOC denotes the AP_S model coupled with Occam's window (that is $\alpha = \alpha I(\alpha > \tau_\alpha)$, where I is the indicator function), setting the threshold τ_α to 0.2, with the ESS term replaced with $s^*(n) = \min(n, 20)$. Other variants have been assessed in the original manuscript [Ollier et al., 2020]. Regarding the simulation setting, for each scenario, 1000 trials are simulated. In terms of design, each trial was structured to have a maximum of 30 patients, encompassing six dose levels, with each cohort consisting of a single patient. Additionally, the no-skipping rule was used. We built five scenarios with a target toxicity of 20% and Figure 4.1 shows the historical data D_0.

Table 4.1 shows the main results: the percentage of correct selection (PCS), the percentage of dose allocation and the number of DLTs, each for six different scenarios.

Method	% dose selection						DLTs	% dose selection							DLTs
	1	2	3	4	5	6	median (Q1, Q3)	0	1	2	3	4	5	6	median (Q1, Q3)
Scenario 1								Scenario 2							
p_{tox}	0.001	0.01	0.05	0.07	**0.2**	0.4			0.01	0.05	0.07	**0.2**	0.4	0.5	
P_NI	0	0	2	28	**54**	16	5 (5, 6)		0	1	25	**61**	12	1	6 (5, 7)
AP_L	0	0	8	37	**39**	16	4 (4, 5)		0	0	45	**48**	6	1	5 (4, 6)
AP_S	0	0	6	32	**47**	15	5 (4, 5)		0	0	37	**54**	8	0	5 (4, 6)
AP_SOC	0	0	3	28	**52**	17	5 (5, 6)		0	0	30	**58**	11	1	5 (5, 6)
Scenario 3								Scenario 4							
p_{tox}	0.05	0.07	**0.2**	0.4	0.5	0.55			0.07	**0.2**	0.4	0.5	0.55	0.65	
P_NI	1	18	**70**	11	0	0	6 (5, 7)		17	**68**	15	0	0	0	6 (6, 8)
AP_L	0	9	**87**	4	0	0	6 (5, 7)		11	**61**	28	0	0	0	7 (6, 8)
AP_S	0	11	**84**	5	0	0	6 (5, 7)		15	**63**	22	0	0	0	7 (6, 8)
AP_SOC	1	11	**80**	8	0	0	6 (5, 7)		18	**62**	19	0	0	0	7 (6, 8)
Scenario 5								Scenario 6							
p_{tox}	**0.2**	0.4	0.5	0.55	0.65	0.7			0.35	0.45	0.5	0.6	0.7	0.8	
P_NI	**86**	14	1	0	0	0	8 (7, 9)	**88**	10	2	0	0	0	0	9 (8, 10)
AP_L	**77**	22	1	0	0	0	8 (7, 9)	**86**	10	3	1	0	0	0	9 (8, 10.25)
AP_S	**79**	20	1	0	0	0	8 (7, 9)	**86**	10	2	1	0	0	0	9 (8, 10)
AP_SOC	**86**	13	1	0	0	0	8 (7, 9)	**88**	10	2	1	0	0	0	9 (8, 10)

Table 4.1 – Results for each method and each scenario in terms of the percentage of dose selection at the end of the trial and median number of DLTs, along with the first (Q1) and the third (Q3) quartiles. At the beginning of each scenario section, the true probabilities used for the scenario simulation are displayed. All methods were provided with stopping rules for scenario 6.

We also investigated the convergence trend of α for AP_SOC by expanding the trial sample size to up to 100 patients and utilizing Eq. 4.3 to calculate the distance for γ. The findings are illustrated in Figure 4.2 which displays the median α value along with the first and third quartiles for scenarios 1, 2, and 3. In the third scenario, which favors full data borrowing, the median α exhibits an upward trend, and the range between the quartiles narrows as the sample size increases. Beyond 70 patients, α consistently exceeds 0.4. For scenario 1, the median, first, and third quartiles all align, each being zero. In scenario 2, the median rises to 0.20 with around 20 patients but then sharply falls back to zero, with the third quartile also dropping to zero after 40 patients. Given that scenarios 4 and 5 mirrored the outcomes of scenarios 2 and 1, respectively, they were not included in the plot.

In Figure 4.3, we presented the dose allocation and toxicity outcomes of a trial simulated under scenario 3, comparing the use of a non-informative prior (illustrated on the left side) with that of the AP_SOC (shown on the right side). The dose allocation strategies for both approaches are identical up to the 14th patient. Beyond this point, AP_SOC maintains the MTD level, in contrast to P_NI, which adopts a more conservative approach by lowering the dose level.

4.3 Parametric mixture priors from elicited histograms

In this work, where I am co-first author, we introduce a Bayesian approach for building a parametric prior concerning two parameters of treatment effects, using graphical data obtained from expert physicians. This approach was developed to be applied in a study (NEPHROMYCY, NCT01092962) involving children with idiopathic nephrotic syndrome to assess the efficacy of two treatments, that is, cyclophosphamide versus Mycophenolate Mofetil (MMF). The approach involves creating histograms for the treatment parameters by each expert physician, based on the technique proposed by Johnson et al. [2010]. These histograms help in defining a marginal prior

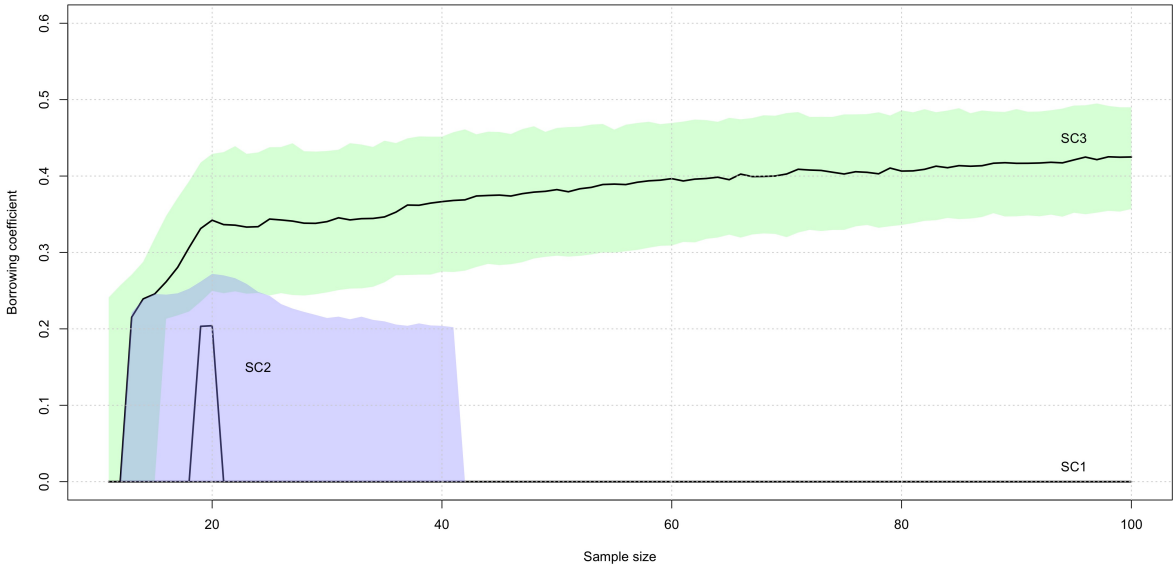


Figure 4.2 – Evolution of the α value in model AP_SOC for a sample size going up to 100 patients. The median and the first and third quartiles are plotted for scenario 1 (SC1), scenario 2 (SC2) and scenario 3 (SC3).

for each treatment effect, described by its central tendency and precision, for every physician. These marginals are then combined, taking into account a distribution of latent effects among the physicians, to form a bivariate prior. Finally, an aggregate prior is formed by merging the individual priors from all participating physicians. In the following, I will details all these steps, starting by setting the notation.

For each participant, labeled as the i^{th} subject in the dataset, where i ranges from 1 to n, the treatment received is represented by τ_i, the outcome observed is denoted as Y_i, and the covariates associated with each subject are indicated by $\boldsymbol{Z}_i = (Z_{i,1}, \ldots, Z_{i,q})$. The treatments are indexed by $j = 1, 2$, and $\theta_{j,i}$ represents the expected outcome Y_i given the treatment $\tau_i = j$ and the covariates $\boldsymbol{Z}_i$. The overall effect of treatment j is determined by calculating the average effect across the n_j subjects who received treatment j,

$$\bar{\theta}_j = \int \theta_{j,i}(\boldsymbol{z}) f_{\boldsymbol{Z}}(\boldsymbol{z}) d\boldsymbol{z} = \frac{1}{n_j} \sum_{i=1}^{n_j} \theta_{j,i},$$

where $f_{\boldsymbol{Z}}$ denotes the patient covariate distribution. For each treatment, indicated by $j = 1, 2$, the overall effect, denoted as $\bar{\theta}_j$, is calculated by averaging the outcomes across all subjects within the treatment group, treating each subject's covariates equally within their respective groups. If we do not have information on subject covariates and assume all subjects are similar, then the expected outcome for treatment j is uniform across all n_j subjects, meaning $\theta_{j,1} = \ldots = \theta_{j,n_j} = \bar{\theta}_j$.

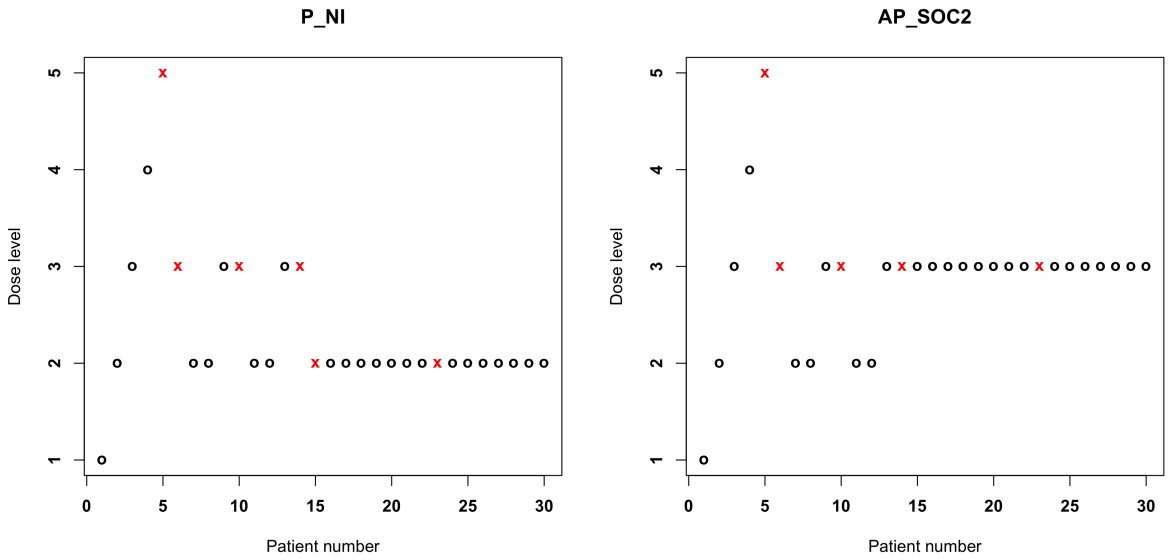


Figure 4.3 – Dose allocation representation for the same trial using P_NI (left) and AP_SOC2 (right). Each point represents a patient, a circle indicates no toxicity, and a cross indicates toxicity.

4.3.1 Probability models for the physicians' marginal priors

For experts numbered $k = 1, \ldots, K$, who provide the histograms for $\bar{\theta}_1$ and $\bar{\theta}_2$, we establish a marginal model for each k^{th} physician's prior regarding $\bar{\theta}_j$. This involves assuming a parametric distribution $p_{j,k}(\bar{\theta}_j|\mu_{j,k}, \gamma_{j,k})$, where $\mu_{j,k}$ signifies the location parameter, and $\gamma_{j,k} > 0$ represents the precision parameter, for both treatments $j = 1, 2$. The rationale behind defining the marginal distributions in terms of location and precision parameters is to simplify the examination of how the posterior inferences are influenced by prior biases and their informativeness. The marginal distributions of $[\bar{\theta}_1|\mu_{1,k}, \gamma_{1,k}]$ and $[\bar{\theta}_2|\mu_{2,k}, \gamma_{2,k}]$ are specified based on the frailties associated with the k^{th} physician. Following this, a bivariate prior is constructed by averaging the product of these conditional marginals across the distribution of frailties.

For binary outcomes Y, where the $\theta_{j,k}$ represent probabilities, the beta distribution serves as a suitable and adaptable choice for parametric priors. Temporarily setting aside the indices j, k, the beta probability density function with mean μ and variance $\mu(1-\mu)/(1+\gamma)$ is defined as:

$$p(x|\mu, \gamma) = \frac{x^{\mu\gamma-1}(1-x)^{(1-\mu)\gamma-1}}{B(\mu\gamma, (1-\mu)\gamma)}, \quad 0 < x < 1,$$

where $B(a, b) = \Gamma(a)\Gamma(b)/\Gamma(a+b)$ and $\Gamma(\cdot)$ is the gamma function. Here, a higher γ value indicates increased precision.

For continuous outcomes Y, the normal distribution, characterized by mean μ and precision parameter $\gamma = 1/\text{var}(\theta)$, is an appropriate prior family selection. In cases where Y represents an event time or another non-negative random variable, multiple two-parameter models can be formulated in terms of location and precision parameters. A versatile model for the $\theta_{j,k}$ prior is

the gamma distribution with mean μ and precision γ, whose pdf is expressed as:

$$p(x|\mu,\gamma) = \frac{(\mu\gamma)^{\mu^2\gamma} x^{\mu^2\gamma-1} e^{-\mu\gamma x}}{\Gamma(\mu^2\gamma)}, \quad x > 0.$$

To derive priors for $\bar{\theta}_1$ and $\bar{\theta}_2$, the parametric models $p_{1,k}(\bar{\theta}_1|\mu_{1,k},\gamma_{1,k})$ and $p_{2,k}(\bar{\theta}_2|\mu_{2,k},\gamma_{2,k})$ are matched to the histograms provided by the k^{th} physician. This process results in the determination of the hyperparameters $\mu_{1,k},\gamma_{1,k},\mu_{2,k},\gamma_{2,k}$ for each physician, indexed from $k=1$ to K. A numerical technique is described in the original manuscript [Thall et al., 2019]. Given that both marginal prior distributions $p_{2,k}$ and $p_{2,k}$ originate from the same physician, it is assumed from the outset that there is a correlation between $\bar{\theta}_1$ and $\bar{\theta}_2$ for each physician. To formally establish this association, we introduce two approaches that are similar yet distinct. Both methods utilize bivariate latent effects related to the physician (frailties) to encourage a prior correlation within-physician between $\bar{\theta}_1$ and $\bar{\theta}_2$.

4.3.2 First method for computing prior hyperparameters

In Method 1, to set up physician-specific priors for $(\bar{\theta}_1,\bar{\theta}_2)$, the parameters $\mu_{j,k}$ and $\gamma_{j,k}$ are each connected to a linear term. This term consists of the sum of a real-valued parameter and a latent physician effect. Specifically, let $\boldsymbol{\varepsilon}_k = (\varepsilon_{k,\mu},\varepsilon_{k,\gamma})$, for $k=1,\ldots,K$, represent pairs of real-valued latent physician effects. These pairs are independent and identically distributed (iid) and follow a bivariate normal distribution.

$$\boldsymbol{\varepsilon}_k \sim N(\mathbf{0},\boldsymbol{\Sigma}) = N\left(\mathbf{0}, \begin{bmatrix} \sigma^2_{\varepsilon,\mu} & \rho\sigma_{\varepsilon,\mu}\sigma_{\varepsilon,\gamma} \\ \rho\sigma_{\varepsilon,\mu}\sigma_{\varepsilon,\gamma} & \sigma^2_{\varepsilon,\gamma} \end{bmatrix}\right). \tag{4.4}$$

Let $\boldsymbol{\sigma} = (\sigma_{\varepsilon,\mu},\sigma_{\varepsilon,\gamma},\rho)$ represent the parameters of the bivariate normal distribution, denoted as $p_{\boldsymbol{\varepsilon}}(x_\mu,x_\gamma|\boldsymbol{\sigma})$ for any pair (x_μ,x_γ) within R^2. The functions g_μ and g_γ are identified as appropriate link functions to transform these parameters. When each $\bar{\theta}_j$ signifies a probability, options for g_μ include the logit, probit, or complementary log-log link, depending on the specific modeling needs. For $\mu_{j,k}$, which can either be real-valued or strictly positive, the identity link or the log link are suitable, respectively. Given that $\gamma_{j,k}$ must always be positive, the log link is utilized for g_γ.

In Method 1, the approach is based on the assumption that the transformations of the parameters $\mu_{j,k}$ and $\gamma_{j,k}$ through their respective link functions g_μ and g_γ are expressed as follows:

$$\begin{aligned} g_\mu(\mu_{j,k}) &= \upsilon_{j,k,\mu} + \varepsilon_{k,\mu} \\ g_\gamma(\gamma_{j,k}) &= \upsilon_{j,k,\gamma} + \varepsilon_{k,\gamma}, \end{aligned} \tag{4.5}$$

Here, $\upsilon_{j,k,\mu}$ represents the real-valued location parameters and $\upsilon_{j,k,\gamma}$ denotes the real-valued precision parameters for each treatment j and physician k. Denoting $\boldsymbol{\upsilon}_k = (\upsilon_{1,k,\mu},\upsilon_{2,k,\mu},\upsilon_{1,k,\gamma},\upsilon_{2,k,\gamma})$, the joint prior is

$$p_k(\bar{\theta}_1,\bar{\theta}_2|\boldsymbol{\upsilon}_k,\boldsymbol{\sigma}) = \int_{R^2} \left\{ \prod_{j=1,2} p_{j,k}(\bar{\theta}_j \mid g_\mu^{-1}(\upsilon_{j,k,\mu}+x_\mu), g_\gamma^{-1}(\upsilon_{j,k,\gamma}+x_\gamma)) \right\} p_{\boldsymbol{\varepsilon}}(x_\mu,x_\gamma|\boldsymbol{\sigma})\, dx_\mu dx_\gamma.$$

It is crucial to note that for each k^{th} physician expert, the information gathered from the two elicited histograms is limited to enabling the calculation of numerical values for $\mu_{1,k}$, $\gamma_{1,k}$, $\mu_{2,k}$, and $\gamma_{2,k}$ only. There is no elicited prior information on the parameters $\sigma = (\rho, \sigma_{\varepsilon,\mu}, \sigma_{\varepsilon,\gamma})$. Therefore, the numerical values for these hyperparameters must be pre-specified. As detailed below, our second method for calculating hyperparameters does yield numerical values for $\boldsymbol{\sigma}$, primarily by utilizing the information found in physician covariates. Therefore, to finalize the prior specification when applying Method 1, we adopt the numerical values of $\boldsymbol{\sigma}$ that are derived using Method 2.

4.3.3 Second method for computing prior hyperparameters

Method 2, another strategy for developing physician-specific priors for $(\bar{\theta}_1, \bar{\theta}_2)$, takes into account the covariate vectors of physicians, $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_K$, whenever such data is available. This method enhances the model for $p_{j,k}(\bar{\theta}_j|\mu_{j,k}, \gamma_{j,k})$ by incorporating a regression structure, making it suitable for cases where it is necessary for the priors to reflect the impact of physician covariates on both $\mu_{j,k}$ and $\gamma_{j,k}$. Unlike Method 1, Method 2 utilizes the values $\mu_{1,k}, \gamma_{1,k}, \mu_{2,k}, \gamma_{2,k}$ derived from fitting the parametric models $p_{1,k}(\bar{\theta}_1|\mu_{1,k}, \gamma_{1,k})$ and $p_{2,k}(\bar{\theta}_2|\mu_{2,k}, \gamma_{2,k})$ to the histograms provided by physicians, but it does so in a distinctly different manner. For Method 2, the latent physician effects are as before, but for each k, we assume that

$$\begin{aligned} g_\mu(\mu_{j,k}) &= \upsilon_{j,\mu} + \boldsymbol{\beta}_\mu \boldsymbol{X}_k + \varepsilon_{k,\mu} + e_\mu \\ g_\gamma(\gamma_{j,k}) &= \upsilon_{j,\gamma} + \boldsymbol{\beta}_\gamma \boldsymbol{X}_k + \varepsilon_{k,\gamma} + e_\gamma, \end{aligned} \tag{4.6}$$

where $\mathbf{e} = (e_\mu, e_\gamma) \sim N_2(\mathbf{0}, \left(\begin{smallmatrix} \sigma_0 & 0 \\ 0 & \sigma_0 \end{smallmatrix}\right))$ are error terms.

In Method 2, we represent the coefficients as $\boldsymbol{\beta} = (\boldsymbol{\beta}_\mu, \boldsymbol{\beta}_\gamma)$, and define the marginal distributions of $\bar{\theta}_1$ and $\bar{\theta}_2$, taking into account both the physician covariates $\boldsymbol{X}_k$ and the latent physician effects $\boldsymbol{\varepsilon}_k$, as follows:

$$p_{j,k}(\bar{\theta}_1|\upsilon_{j,\mu}, \upsilon_{j,\gamma}, \boldsymbol{\beta}, \boldsymbol{X}_k, \boldsymbol{\varepsilon}_k)$$

for $j = 1, 2$. In this regression-based approach, there are four constant intercepts, $\boldsymbol{\upsilon} = (\upsilon_{1,\mu}, \upsilon_{1,\gamma}, \upsilon_{2,\mu}, \upsilon_{2,\gamma})$, which are applied universally across all physicians to maintain the model identifiability. This contrasts with the method used in Method 1, where intercept parameters could vary for each physician $(k = 1, \ldots, K)$. Hence, Method 2 accounts for differences between physicians through their respective covariates, marking a significant departure from the approach in Method 1. In Method 2, the inclusion of physician covariate data enables the calculation of numerical values for the hyperparameters $(\boldsymbol{\upsilon}, \boldsymbol{\beta}, \boldsymbol{\sigma})$ within the physician's covariate regression model.

To apply Method 2, hyperparameter values are determined by treating the location and dispersion parameters, derived from the elicited histograms, as pseudo outcomes. The hyperparameter vector $(\boldsymbol{\upsilon}, \boldsymbol{\beta}, \boldsymbol{\sigma})$ is considered as pseudo parameters. By fitting a regression model and using the estimated pseudo parameters as the prior means for $(\boldsymbol{\upsilon}, \boldsymbol{\beta}, \boldsymbol{\sigma})$ in the marginal priors $\{p_{j,k}(\bar{\theta}_j|\boldsymbol{X}_k, \boldsymbol{\upsilon}, \boldsymbol{\beta}, \boldsymbol{\sigma}),\ j = 1, 2, k = 1, \ldots, K\}$, we obtain these hyperparameter values. This process can be conducted through various methods, which typically yield very similar numerical outcomes. In our case, we did it via Bayesian inference and details on prior distributions are given in Thall et al. [2019].

The posterior means derived from the nonlinear Bayesian regression model are utilized as the hyperparameters for the marginal priors specific to each physician. For the k^{th} physician, the joint prior of $(\bar{\theta}_1, \bar{\theta}_2)$ is established similarly to Method 1, by integrating over the distribution of bivariate physician effects. The joint prior can be represented as:

$$p_k(\bar{\theta}_1, \bar{\theta}_2 | \boldsymbol{\upsilon}, \boldsymbol{\beta}, \boldsymbol{\sigma}, \boldsymbol{X}_k) =$$

$$\int_{R^2} \left[\prod_{j=1,2} p_{j,k} \left\{ \bar{\theta}_j | g_\mu^{-1}(\upsilon_{j,\mu} + \boldsymbol{\beta}_\mu \boldsymbol{X}_k + x_\mu), g_\gamma^{-1}(\upsilon_{j,\gamma} + \boldsymbol{\beta}_\gamma \boldsymbol{X}_k + x_\mu) \right\} \right] p_{\boldsymbol{\varepsilon}}(x_\mu, x_\gamma | \boldsymbol{\sigma}) \; dx_\mu dx_\gamma.$$

In Method 2, the bivariate priors for all K physicians share the same hyperparameter vectors $(\boldsymbol{\upsilon}, \boldsymbol{\beta}, \boldsymbol{\sigma})$, making the prior p_k unique to the k^{th} physician through the physician's covariate vector $\boldsymbol{X}_k$ only.

4.3.4 Mixture priors

Once the K bivariate physician-specific parametric priors have been established through either Method 1 or Method 2, let $\boldsymbol{w} = (w_1, \ldots, w_K)$ represent the weights assigned to each physician, with these weights totaling 1. The aggregate prior for $(\bar{\theta}_1, \bar{\theta}_2)$ using Method 1 is formulated as a mixture model:

$$p(\bar{\theta}_1, \bar{\theta}_2 | \boldsymbol{\upsilon}_1, \cdots, \boldsymbol{\upsilon}_K, \boldsymbol{\sigma}) = \sum_{k=1}^{K} w_k \; p_k(\bar{\theta}_1, \bar{\theta}_2 | \boldsymbol{\upsilon}_k, \boldsymbol{\sigma}),$$

Similarly, the combined prior using Method 2 is defined as:

$$p(\bar{\theta}_1, \bar{\theta}_2 | \boldsymbol{\upsilon}, \boldsymbol{\beta}, \boldsymbol{\sigma}, \boldsymbol{X}_1, \cdots, \boldsymbol{X}_K) = \sum_{k=1}^{K} w_k \; p_k(\bar{\theta}_1, \bar{\theta}_2 | \boldsymbol{\upsilon}, \boldsymbol{\beta}, \boldsymbol{\sigma}, \boldsymbol{X}_k).$$

The determination of the physician weights, $\boldsymbol{w}$, can be approached through various methods, with three possibilities presented here. To empirically construct $\boldsymbol{w}$, one approach involves fitting a likelihood model for $[Y_i | \tau_i, \boldsymbol{Z}_i]$ to the dataset $\mathcal{D}_n$, where τ_i represents the treatment administered to patient i. The maximum likelihood estimates of the parameters for each pair (j, i) are denoted by $\hat{\theta}_{j,i}^{(like)}$, assembling these into a $2n$-vector $\hat{\boldsymbol{\theta}}^{(like)}$ comprising all such estimates. An alternative method involves deriving estimates as posterior means under a Bayesian framework, employing non-informative pseudo priors to guide the inference process without strong prior assumptions. For each treatment group, $j = 1, 2$, the empirical mean of θ_j derived from the elicited histogram of physician k is represented by $\hat{\theta}_{j,k}^{(elicited)}$. The correspondence between the mean vectors $\hat{\boldsymbol{\theta}}_k^{(elicited)} = (\hat{\theta}_{1,k}^{(elicited)}, \hat{\theta}_{2,k}^{(elicited)})$ calculated from each physician's provided histograms and the $2n$ vector of likelihood-based estimates, $\hat{\boldsymbol{\theta}}^{(like)}$, is measured through the mean absolute deviation as follows:

$$\| \hat{\boldsymbol{\theta}}_k^{(elicited)} - \hat{\boldsymbol{\theta}}^{(like)} \| = \frac{1}{2n} \sum_{j=1}^{2} \sum_{i=1}^{n} \left| \hat{\theta}_{j,k}^{(elicited)} - \hat{\theta}_{j,i}^{(like)} \right|.$$

A smaller deviation, $\left\|\hat{\boldsymbol{\theta}}_k^{(elicited)} - \hat{\boldsymbol{\theta}}^{(like)}\right\|$, indicates a higher level of agreement. Thus, physician weights are assigned inversely proportional to this deviation, ensuring that physicians whose elicited means are closer to the data-based estimates have higher weights:

$$w_k = \frac{\left\|\hat{\boldsymbol{\theta}}_k^{(elicited)} - \hat{\boldsymbol{\theta}}^{(like)}\right\|^{-1}}{\sum_{r=1}^{K}\left\|\hat{\boldsymbol{\theta}}_r^{(elicited)} - \hat{\boldsymbol{\theta}}^{(like)}\right\|^{-1}} .$$

When physician covariates are accessible, another approach to determine the physician weights involves utilizing these covariates. Assuming all physician covariates are positive and that a higher value of $X_{k,l}$ indicates a higher reliability for physician k (for instance, $X_{k,l}$ might represent years of experience), the weights can be established as follows:

$$w_k = \frac{1}{q}\sum_{l=1}^{q}\frac{X_{k,l}}{\sum_{r=1}^{K} X_{r,l}}.$$

In this formula, a greater ratio of $X_{k,l}/\sum_{r=1}^{K} X_{r,l}$ signifies a higher relative reliability of physician k's opinion based on the l^{th} covariate. Averaging over $l = 1, \cdots, q$ implies that each covariate is considered equally significant. Under this weighting scheme, especially when applied with Method 2, each physician's covariate information is utilized in a twofold manner: firstly, to derive the specific prior p_k for that physician, and secondly, to calculate the weight w_k. This dual use of covariate data aims to ensure that the more reliable or experienced physicians have a greater influence on the overall prior. An alternative, simpler approach is to assign equal weight to all physicians, with every w_k set to $1/K$, treating each physician's opinion as equally valid regardless of their specific covariates or experience.

4.3.5 Application to elicited data

Figure 4.4 shows the elicited histograms and the corresponding fitted beta distributions for the parameters θ_1 and θ_2, for a subset of three out of the seventeen physicians who contributed to the elicitation phase of the NEPHROMYCY trial. The process of eliciting these histograms involved physicians allocating a fixed number of "stickers" across different outcome probabilities, which then visually represented their beliefs regarding the treatment effects. In instances where physicians allocated fewer than 20 stickers for any histogram, adjustments were made to normalize these histograms, ensuring the total probability mass equaled 1 before proceeding with the beta distribution fitting.

The contour plots, in Figure 4.5, visualize the distribution of estimates for (μ, γ) derived from the beta distributions fitted to the elicited histograms of the 17 physicians participating in the study. These plots are organized to present the data for cyclophosphamide on the left and for MMF on the right, offering a comparative view of the physicians' beliefs regarding the two treatments. Additionally, histograms of the marginal distributions for the physician-specific estimates of μ (displayed at the top of the figure) and γ (shown on the right side) complement the contour plots. These histograms for μ_1 (associated with cyclophosphamide) and μ_2 (associated with MMF) indicate that, on average, the physicians perceived MMF as having a higher probability of response compared to cyclophosphamide. However, there is noted

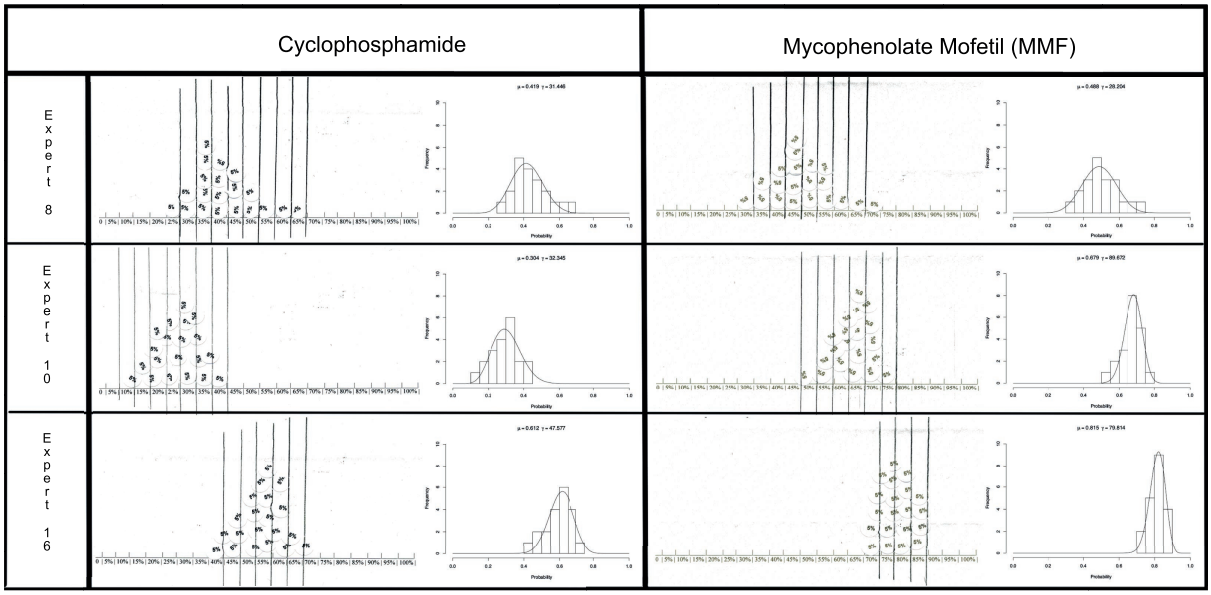


Figure 4.4 – Elicited histograms and fitted beta priors for θ_1 (left hand side - cyclophosphamide) and θ_2 (right hand side - MMF) for three of the 17 physicians who participated in the elicitation process in planning the NEPHROMYCY trial.

variability among physicians' opinions. Furthermore, the histograms for the precision parameters γ_1 and γ_2 exhibit a high level of dispersion but share remarkably similar shapes, with the majority of the distribution's mass concentrated between the values of 30 and 70.

Finally, the computed hyperparameters were

$$\boldsymbol{\Sigma} = \begin{bmatrix} 0.399 & -0.003 \\ -0.003 & 0.634 \end{bmatrix}$$

with $(\nu_{1,\mu}, \nu_{2,\mu}, \nu_{1,\gamma}, \nu_{2,\gamma})$ = (-0.708 , 0.237 , 3.387 , 3.395), $\boldsymbol{\beta}_\mu$ = (0.173, -0.049, -0.053), and $\boldsymbol{\beta}_\gamma$ = (-0.185, 0.239, 0.334). Figure 4.6 shows the two joint prior distributions for $(\bar{\theta}_1, \bar{\theta}_2)$ obtained using these values by Methods 1 and 2, using equal physician weights for the mixture. When comparing the priors resulting from Method 1 and Method 2, the smoother surface generated by Method 2, as opposed to the bimodal distribution produced by Method 1, suggests that Method 2 generated a more informative prior distribution.

4.3.6 Sensitivity to prior bias and informativeness

In conducting a sensitivity analysis of posterior inferences to the prior, both Method 1 and Method 2 define expert-specific location and precision parameters differently, adapting to the specifics of each method. For Method 1, the location parameters for each expert and treatment are denoted as $\xi_{j,k} = \upsilon_{j,k,\mu}$ and the precision parameters as $\chi_{j,k} = \upsilon_{j,k,\gamma}$. In contrast, Method 2 employs a regression-based approach, where the location parameters are defined as $\xi_{j,k} = \upsilon_{j,\mu} + \boldsymbol{\beta}_\mu \boldsymbol{X}_k$, and the precision parameters as $\chi_{j,k} = \upsilon_{j,\gamma} + \boldsymbol{\beta}_\gamma \boldsymbol{X}_k$, incorporating the influence of physician covariates $\boldsymbol{X}_k$. To assess the sensitivity of the posterior inferences to these priors, we

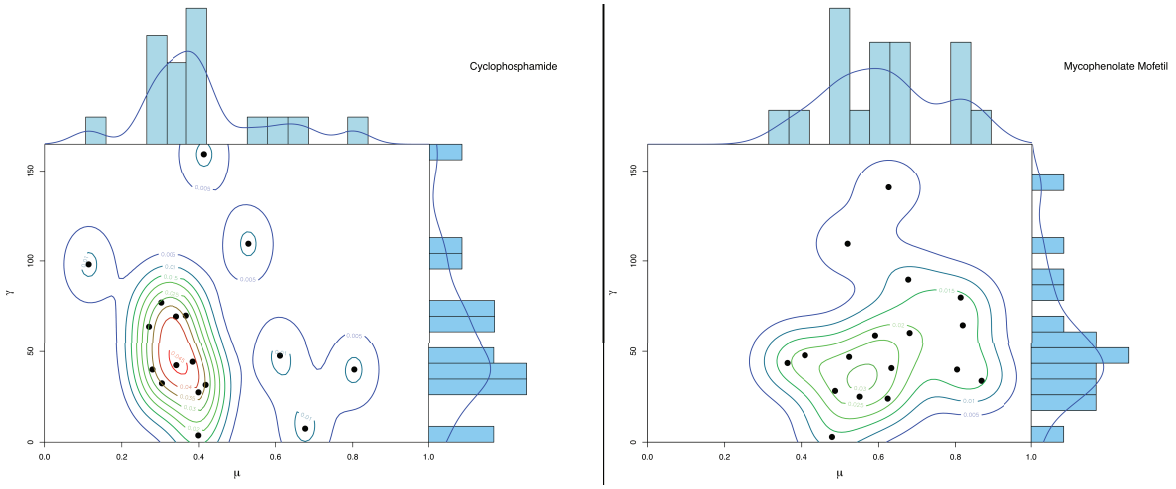


Figure 4.5 – Contour plots of estimated (μ, γ) for each expert in the domain $(0,1) \times R^+$ for the estimate prior response probabilities of Cyclophosphamide (left hand side) and MMF (right hand side). Marginal histograms are plotted on the top for μ and right-hand side for γ.

consider transformations on the vectors of location parameters $\boldsymbol{\xi}_j = (\xi_{j,1}, \ldots, \xi_{j,K})$ and precision parameters $\boldsymbol{\chi}_j = (\chi_{j,1}, \ldots, \chi_{j,K})$ for $j = 1, 2$. These transformations aim to adjust the prior location (bias) and informativeness.

For adjusting the prior location, a sensitivity parameter ϕ ranging from 0 to 1 is introduced. This parameter shifts the focus between the two sets of location parameters by replacing $\boldsymbol{\xi}_2$ with $(1-\phi)\boldsymbol{\xi}_1 + \phi\boldsymbol{\xi}_2$. The value of $\phi = 1$ retains the original bias (between treatments) of the unadjusted prior, whereas reducing ϕ towards 0 decreases the prior bias, with $\phi = 0$ eliminating the bias entirely ($\boldsymbol{\xi}_2 = \boldsymbol{\xi}_1$, making the prior bias 0).

For the prior precision, the introduction of a scale sensitivity parameter λ, which varies between $0 < \lambda \leq 1$, allows for an adjustment in the informativeness of the priors. By applying λ to $\boldsymbol{\chi}_j$ for both $j = 1$ and $j = 2$, the original prior precision is retained at $\lambda = 1$, signifying no change in informativeness. Conversely, as λ decreases towards 0, the priors for both treatment effects become progressively less informative, reflecting increased uncertainty in the prior beliefs about the treatment effects. This mechanism for altering the precision parameters facilitates a comprehensive sensitivity analysis, enabling the evaluation of how variations in the assumed precision of prior information impact the posterior inferences. By comparing posterior outcomes derived under these adjusted priors to those obtained from the unmodified priors ($\phi, \lambda = (1,1)$), researchers can discern the degree to which their conclusions are sensitive to changes in the precision of the priors.

Several posterior quantities can be of particular importance. These quantities offer insights into whether one treatment is superior to another by a predefined margin, δ_θ, or whether the treatments can be considered equivalent within a specified tolerance, ε. For example: (i) Posterior Probability of Superiority, to evaluate whether treatment 2 (θ_2) is superior to treatment 1 (θ_1) by at least δ_θ, one computes the posterior probability: $\pi^S_{1,2,\phi,\delta}(\delta) = P_{\phi,\lambda}(\theta_1 + \delta_\theta < \theta_2 | data)$. (ii) Symmetric Probability of Difference, where the interest might be in the probability that the

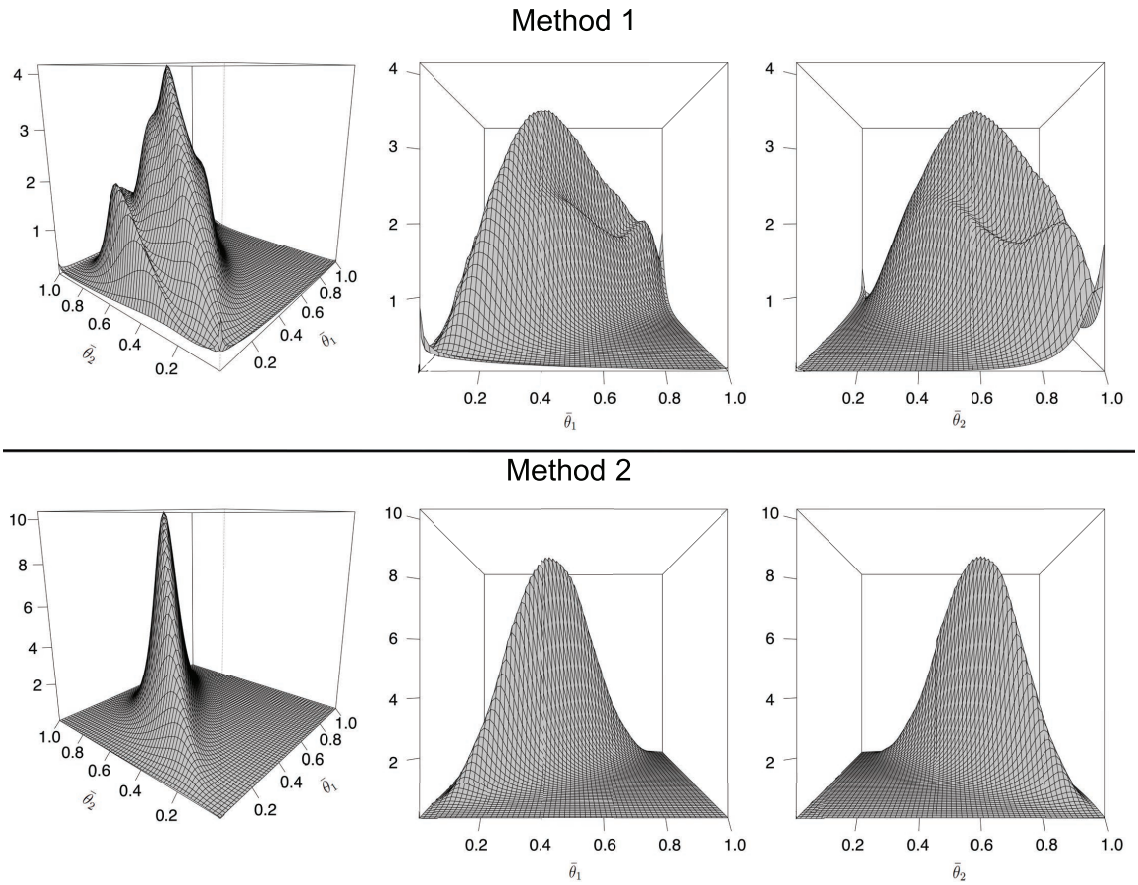


Figure 4.6 – 3-dimensional plots of prior distributions of $(\bar{\theta}_1, \bar{\theta}_2)$ using Method 1 (top row) and Method 2 (bottom row), with equal physician weights. For Method 2, the covariates $\boldsymbol{X}_k$ were the logarithm of the number of year as paediatrician, the logarithm of the average number of patients consulted per year, and a binary indicator of whether the physician had training in clinical trial methodology.

absolute difference between treatments exceeds δ_θ, regardless of which treatment is superior. This is given by: $P_{\phi,\lambda}(|\theta_1 - \theta_2| > \delta_\theta | data)$. (iii) Posterior Probability of Equivalence, when the goal is to establish the equivalence of an experimental treatment (treatment 2) to a standard treatment (treatment 1) within a margin of ε, the relevant posterior probability is: $\pi^E_{1,2,\phi,\delta}(\varepsilon)$, indexed by the prior transformation parameters (ϕ, λ). (iv) Lastly, the 95% posterior credible interval (CI) for $\theta_2 - \theta_1$, which we denote by $CI_{95,\phi,\lambda}(\theta_2 - \theta_1)$.

To demonstrate how a sensitivity analysis of prior-to-posterior transformations can be conducted, let us consider a hypothetical scenario inspired by the NEPHROMYCY trial data. In this scenario, a dataset is simulated with binary responses for two treatment groups, each consisting of 35 patients. The true effectiveness rates ($\theta_{1,true}$ and $\theta_{2,true}$) for both treatments are set at 40%. However, in the simulation, treatment 1 shows effectiveness in 14 out of 35 children (40%), and treatment 2 shows a slightly higher effectiveness in 16 out of 35 children (approximately 45.7%).

Table 4.2 – Prior-to-posterior sensitivity analyses performed on a 70-patient dataset with 14/35 responses in arm 1 and 16/35 responses in arm 2. The prior was constructed using Method 1 and equal physician weights, was transformed for each (ϕ, λ) pair, and the posterior quantities $\pi^E_{1,2,\phi,\lambda}(.05) = \Pr(\theta_1 - .05 < \theta_2 \mid data, \phi, \lambda)$, $\pi^S_{1,2,\phi,\lambda}(.15) = \Pr(\theta_1 + .15 < \theta_2 \mid data, \phi, \lambda)$ and $CI_{95,\phi,\lambda}(\theta_2 - \theta_1 \mid data)$ then were computed.

		$\lambda = 1$	$\lambda = 0.75$	$\lambda = 0.5$	$\lambda = 0.25$
$\phi = 1$	$\pi^E_{1,2,\phi,\delta}(.05)$	0.92	0.88	0.84	0.77
	$\pi^S_{1,2,\phi,\lambda}(.15)$	0.24	0.22	0.21	0.18
	$CI_{95,\phi,\lambda}(\theta_2 - \theta_1)$	(-0.09, 0.27)	(-0.13, 0.26)	(-0.17, 0.27)	(-0.20, 0.27)
$\phi = 0.5$	$\pi^E_{1,2,\phi,\lambda}(.05)$	0.77	0.76	0.75	0.75
	$\pi^S_{1,2,\phi,\delta}(.15)$	0.17	0.17	0.15	0.16
	$CI_{95,\phi,\lambda}(\theta_2 - \theta_1)$	(-0.18, 0.26)	(-0.2, 0.26)	(-0.19, 0.26)	(-0.19, 0.25)
$\phi = 0$	$\pi^E_{1,2,\phi,\lambda}(.05)$	0.77	0.75	0.76	0.74
	$\pi^S_{1,2,\phi,\lambda}(.15)$	0.16	0.15	0.15	0.15
	$CI_{95,\phi,\lambda}(\theta_2 - \theta_1)$	(-0.19, 0.26)	(-0.21, 0.26)	(-0.2, 0.26)	(-0.20, 0.25)

For the sensitivity analysis, a range of prior configurations is generated using Method 1 with equal weights for all physicians. This involves adjusting the priors across a matrix of (ϕ, λ) pairs, with ϕ taking values of 1, 0.5, and 0, and λ set at 1, 0.75, 0.50, and 0.25. These adjustments aim to explore how changes in the prior bias (ϕ) and informativeness (λ) affect the posterior outcomes. In Table 4.2, the posterior probabilities of .05-equivalence $\pi^E_{1,2,\phi,\delta}(.05) = \Pr(\theta_1 - .05 < \theta_2 | data, \phi, \delta)$ and .15-superiority $\pi^S_{1,2,\phi,\delta}(.15) = \Pr(\theta_1 + .15 < \theta_2 | data, \phi, \delta)$, and the posterior 95% credible interval $CI_{95,\phi,\lambda}(\theta_2 - \theta_1 \mid data)$, are reported in each cell.

With no adjustment in prior location ($\phi = 1$), the probability that treatment 2 is practically equivalent to treatment 1 within a 0.05 margin decreases to 0.77 when the prior precision is reduced to 25% ($\lambda = .25$), reflecting a prior effective sample size of about 17.5. For partial ($\phi = 0.5$) or complete ($\phi = 0$) shifts in prior location, the equivalence probability ($\pi^E_{1,2,\phi,\delta}(.05)$) varies slightly, remaining within a range of 0.74 to 0.77, indicating a degree of robustness to changes in prior bias and informativeness. The posterior probability of 0.15-superiority ($\pi^S_{1,2,\phi,\delta}(.15)$) for the original prior settings ($\phi = 1$) starts at 0.24 and decreases to 0.18 as λ reduces from 1 to 0.25, showing sensitivity to reductions in prior precision. Conversely, for shifted priors ($\phi = 0.5$ or 0), this probability remains relatively stable, between 0.15 and 0.17, regardless of λ, indicating insensitivity to changes in prior informativeness under these conditions. The upper limit of the posterior 95% credible interval for $\theta_2 - \theta_1$ stays consistent across all tested (ϕ, λ) values, suggesting that the potential maximum difference between treatments does not vary with adjustments in the priors. However, the lower limit is sensitive to changes in λ when $\phi = 1$, shifting from -0.09 to -0.20 as λ decreases, which implies increasing uncertainty about the minimum difference between treatments with less informative priors. For shifted priors ($\phi = 0.5$ or 0), the lower limit of the credible interval remains largely unaffected by variations in λ.

4.4 Discussion

In this chapter, I discussed two approaches for incorporating external information into prior distributions. The first approach involves leveraging data from previous clinical trials and evaluating potential conflicts between prior data and new data. According to our simulations, the degree of information borrowing varied, averaging between 0.3 and 0.5, depending on the method used, in situations where there was no data conflict. Although there may be a desire to borrow more extensively in such cases, it typically involves balancing the advantages against the potential risks. Notably, this is our first work to employ the Hellinger distance as a measure for assessing the compatibility between prior distributions and new data. This metric has subsequently been applied in further research, including Ollier et al. [2021] on the examination of the similarity between two completed trials within a bridging context, Calderazzo et al. [2023] on the investigations into studies resulted inconclusive by the Covid-19 pandemic, and Boulet et al. [2024], which is outlined in the following chapter. For instance, in Ollier et al. [2021], we coupled Eq. 4.3, which captures the entire dose-response curve distance, with a separate MTD distance metric that focuses on the disparity in point estimates. This dual approach allows us to derive more nuanced insights into dose behavior across two distinct populations (such as Caucasians and Asians, as outlined in our initial study) after conducting studies within each group. We implemented this methodology on real published datasets, and Figure 4.7 presents some of the visual outcomes. It is important to note that the synthetic datasets were built to exhibit: similarity in MTDs and dose-response curves (synthetic-1), similarity in MTDs but differences in dose-response curves (synthetic-2), and differences in both MTDs and dose-response curves (synthetic-3).

In the second approach, we have developed a method to create informative parametric mixture priors for treatment parameters, using histograms from expert physician inputs, to enhance randomized trials with limited sample sizes. Our approach, adaptable for binary, continuous, or time-to-event data, includes versions that consider physician covariates and offers three strategies for physician weighting in the mixture prior, resulting in six unique versions. Given the inherent debate over integrating expert opinion into Bayesian analysis priors, we have also proposed a clear strategy for constructing alternative priors through location adjustments and precision changes. This allows for a thorough sensitivity analysis, providing a solid foundation for informed decisions regarding the comparative effectiveness of treatments. It is worth mentioning that several adapted and/or simplified versions of our methodology are now implemented in two clinical trials: POOMA (NCT04554108), for primary endpoint analysis (this work initiated with my supervision of the internship of a master 2 level student), and TREOCAPA (NCT04459117), for safety analyses, though, in the latter, histograms were not utilized to further streamline the elicitation process. Two manuscripts detailing these applications are in the planning stages, as I aim to provide comprehensive guidelines to the broader community based on my experiences in both theoretical method development and practical application.

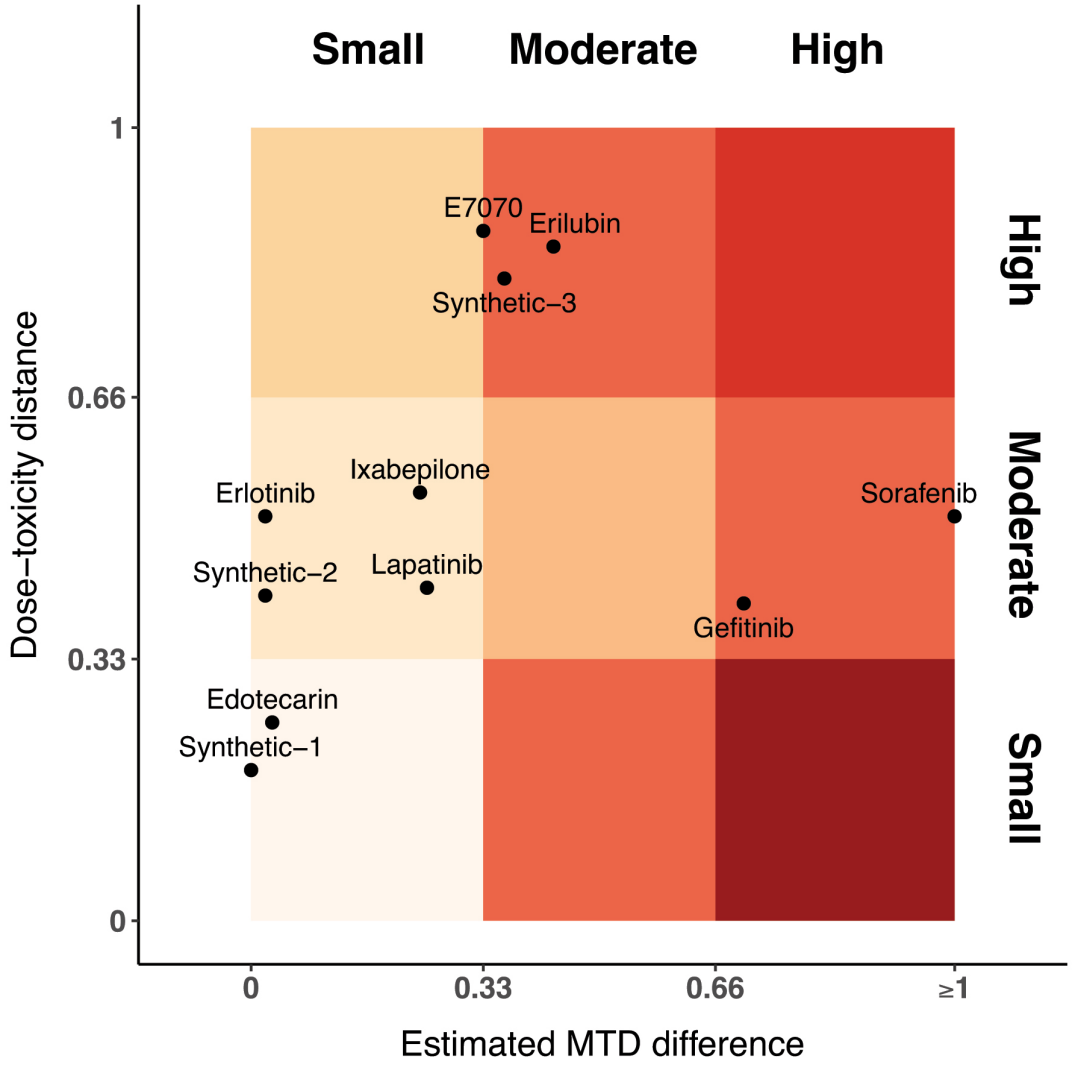


Figure 4.7 – Gradient plot representing the distance between dose–toxicity curves (y-axis), and between MTDs (x-axis). The intensity of the color varies along with the increasing distance value and coherence. Small dose–toxicity distance and high MTD distance is incoherent, as such it is plotted in a darker color.

Chapter 5

Meta-analysis approaches for early clinical phases

Contents

Unlike previous chapters where I concentrated on incorporating external or internal information into trial analyses, here I shift focus to leveraging data from completed studies to enhance the findings. Therefore, the two studies I will present here are related to the meta-analysis field. In the first work, I proposed a Bayesian mixed-effect model for meta-analysis of dose-finding trials [Ursino et al., 2021]. The second work, done by a postdoc student I have co-supervised, regards a framework to combine information of several preclinical trial before stepping to the first-in-man clinical trial [Boulet et al., 2024]. Following the structure of previous chapters, this section starts with a brief introduction to set the context for the two works discussed. Next, I will offer summaries of these studies, focusing on the methodologies applied and the results obtained. The chapter will end with a concise discussion.

5.1 Introduction

Several dose-finding studies are frequently carried out, featuring variations in how the drug is administered or the range of doses tested. Given that adverse effects may not necessarily correspond to the targeted condition, it could be beneficial to aggregate data from various studies. Table 5.1 shows an example on sorafenib (BAY 43-9006), which is a kinase inhibitor approved for the treatment of advanced renal cell carcinoma, hepatocellular carcinoma, and

Table 5.1 – The results of 14 studies on sorafenib monotherapy. For each dose considered in each trial, the numbers of patients experiencing DLT events, and the total numbers of exposed patients are given.

	Dose (mg)						
Study	100	200	300	400	600	800	1000
Clark et al. [2005]	0/3	0/3		1/4	1/6	3/3	
Awada et al. [2005]	0/4	0/3	1/5	1/10	7/12	1/3	
Moore et al. [2005]	0/3	1/6		0/8	3/7		
Strumberg et al. [2005]	1/5	1/6		0/15	4/14	2/7	
Minami et al. [2008]	0/3	1/12		0/6	1/6		
Miller et al. [2009]		8/34		6/20			
Nabors et al. [2011]		0/3		1/6	0/3	1/5	3/3
Chen et al. [2014]		0/3		1/16			
Jia et al. [2013]				3/4			
Borthakur et al. [2011]-1		0/3		0/15	2/8		
Borthakur et al. [2011]-2		0/3		1/7	2/6		
Crump et al. [2010]-1	0/4	1/6	0/6	1/6			
Crump et al. [2010]-2	0/3	1/6		0/3	2/6		
Furuse et al. [2008]		0/12		1/14			

radioactive iodine resistant advanced thyroid carcinoma. When integrating data from different trials, it is crucial to account for two main types of variability. First, there can be variations in the results due to factors such as differences in the study demographics or the ways in which toxicities are defined and measured. These differences can affect the estimated probabilities of toxicity. Second, the actual effects of the treatment, even when measured on a comparative scale, may differ from one study to another. Traditional meta-analysis methods tackle the first issue by grouping data according to the study. Random-effects meta-analyses address the second issue by allowing for variability in the treatment effects across studies. Estimating the relevant variance component and properly incorporating the uncertainty of this estimation into the inference of significant model parameters poses a challenge, particularly when the meta-analysis includes only a few studies [Friede et al., 2017]. Zohar et al. [2011] introduced a meta-analysis strategy specifically for Phase I clinical trials in oncology, aiming to aggregate data while acknowledging the sequential progression typical of these trials to more accurately determine the overall MTD. Nevertheless, this approach overlooked several critical aspects characteristic of phase I trials. Firstly, it combined data from trials with varying administration schedules, potentially leading to differing toxicity profiles. Secondly, it did not consider heterogeneity between trials, which could result in imprecise conclusions.

Meta-analysis approaches can also be valuable in aggregating information from the preclinical stage to the clinical phase. During preclinical research, such as *in vitro*, *in vivo*, and *in silico* experiments, the PK, PD, and toxicological profiles of a medication are assessed prior to initiating first-in-human (FIH) trials. Typically, these studies are examined separately, and the determination of the human dosage range does not fully utilize the insights obtained from all experiments. Integrating all preclinical findings via inferential methods can be notably beneficial

in establishing a more accurate initial dosage and dosage spectrum for humans. When studies are conducted in a sequential manner and the outcomes can be interconnected (through mathematical transformations, for instance), it becomes straightforward to incorporate the analysis within a Bayesian framework that updates posterior knowledge with each new set of data. For instance, in the realms of preclinical and clinical research, La Gamba et al. [2019] systematically incorporate findings from preclinical studies into a Bayesian PK/PD model, where the posterior distributions from one study serve as the prior distributions for the next. Furthermore, the Bayesian methodology has been broadly applied to leverage data from one group to enhance the analysis of another group. For example, Zheng and Hampson [2020] utilize preclinical data from animals to inform the design and prior distributions for a Phase I clinical trial through a Bayesian decision-theoretic framework, and Zheng et al. [2020] adopt a meta-analytic strategy for similar purposes.

5.2 Random-effects meta-analysis of Phase I

In Ursino et al. [2021], we introduce an innovative meta-analysis methodology for Phase I clinical trials in oncology. We have expanded upon the binomial-normal hierarchical model (BNHM), which is traditionally applied in meta-analyses of studies focusing on a single dose. The new method considers all doses tested in a trial for the meta-analysis, while also accounting for the variability between different trials.

As usual we start by setting the notation. Let k represent the index of the study, ranging from 1 to K, and i denote the index of the dose level, within the set $\mathcal{I} = \{1, \ldots, I\}$, where all doses d_i utilized across the K trials are arranged in ascending order. We introduce $\delta_{i,j}$ as a measure to determine the proximity or distance between different doses. This can be straightforwardly defined as the linear difference ($\delta_{i,j} = d_i - d_j$). However, in numerous instances, it may be more appropriate to consider the relative differences between dose levels on a logarithmic scale ($\delta_{i,j} = \log(d_i) - \log(d_j) = \log\left(\frac{d_i}{d_j}\right)$). Alternatively, considering unit increments for adjacent doses ($\delta_{i,j} = i - j$) might be another viable approach. For study k, the number of patients treated at dose i is denoted by n_{ik}, and X_{ik} represents the number of patients who experienced a DLT. We suggest the following model formulation:

$$\begin{aligned} X_{ik} &\sim \text{Binomial}(n_{ik}, p_{ik}) \\ \text{logit}(p_{ik}) &= \sum_{j \leq i} \mu_j + b_{ik} \end{aligned}$$

where p_{ik} is the toxicity probability of dose i in the kth study, modeled on the logit scale. If dose i is not utilized in study k, then it is assumed $n_{ik} = X_{ik} = 0$, and thus it does not affect the likelihood, following the convention that $0^0 = 1$. The model incorporates *fixed effects*, with μ_1 on the real line and μ_i on the positive real line for $i > 1$, consistent across all studies, ensuring progressively increasing mean toxicity probabilities with higher doses. *Random effects* are introduced to address the variability between studies, with study-specific vectors $\mathbf{b}_k$ following a normal distribution $N(\mathbf{0}, \Sigma)$, where $\mathbf{0}$ is a zero vector of dimension I and Σ denotes the variance-covariance matrix.

5.2.1 Gaussian process for the random effects

In our model for the random effects, we account for the position of dose d_i within the dose continuum without enforcing a monotonic relationship among the elements of $\mathbf{b}_k$. Instead, we utilize a relatively straightforward class of Gaussian processes. This model incorporates two notable scenarios: *independent* and *identical* residuals across all dose levels. To bridge the gap between these two scenarios, we implement a stationary *Ornstein-Uhlenbeck process (OUP)* to model the covariance, described as

$$\sigma^2_{i,j} = \sigma^2_m \exp\left(-\frac{|\delta_{i,j}|}{\ell}\right)$$

where σ^2_m is the marginal variance, and $\ell > 0$ acts as a smoothness parameter. This parameter controls the rate at which autocorrelation diminishes, making residuals less dependent on each other based on their distance $\delta_{i,j}$ [Uhlenbeck and Ornstein, 1930]. On smaller scales, relative to ℓ, the OUP mimics a Wiener process (or Brownian motion), aligning with the concept that knowledge about the residual at a specific dose diminishes as we move further away, with increments behaving in an (approximately) additive manner similar to the fixed effects model. For extreme values of ℓ, approaching 0 or infinity, the model predicts either independent or identical residuals across all doses, respectively. It is necessary to define prior distributions for both the marginal variance σ^2_m and the OUP distance parameter ℓ.

5.2.2 Gamma process for fixed effects prior distributions

The model construction, which defines the common effect through the addition of unknown increments, aligns it with the category of stochastic processes frequently utilized as nonparametric approaches to modeling unknown functions. This approach suggests that the priors for these increments could be derived from stochastic processes. A particularly suitable model class for this purpose is characterized by *infinitely divisible* probability distributions, ensuring that the summation of any two increments remains within the same distribution family, thereby maintaining model consistency [Steutel, 1979]. Given the requirement for strictly positive increments corresponding to increasing doses, the Gamma process emerges as a fitting choice for this scenario [Lawless and Crowder, 2004]. The Gamma distribution, defined by two parameters: the *shape* ($s > 0$) and the *scale* ($\theta > 0$), has a mean of $s\theta$ and a variance of $s\theta^2$. By designating the first dose (d_1) as the *reference dose*, we introduce the prior distributions within a Gamma process framework as follows:

$$\mu_1 \sim N(\mu^*, \sigma^*), \tag{5.1}$$

$$\mu_i \sim Gamma(s = \delta^*_{i,i-1}\kappa, \theta) \quad \text{for } i > 1. \tag{5.2}$$

This model allows δ^* to match the dose increments (δ) used in defining the random effects or to employ a different metric. The model further necessitates specifying the hyperparameters μ^*, σ^*, κ, and θ, with κ and θ influencing the model behavior on unit increments, that we consider via a re-parametrization in terms of a slope ($a = \kappa\theta$) and a coefficient of variation ($c = 1/\sqrt{\kappa}$). This re-parametrization reveals that lower c values suggest a nearly linear relationship in (logit) toxicity, while higher c values introduce the potential for non-linearity. In the scenario of pure

Table 5.2 – Settings and parameters in the 9 simulation scenarios. ℓ was chosen equal to 1 for all scenarios. OUP: Ornstein-Uhlenbeck process; CRM: continual reassessment method; 3+3: 3+3 algorithm design. a: $\mathbf{p}^\star$ =(0.15, 0.20, 0.33, 0,45, 0.55, 0.60, 0.65); b: $\mathbf{p}^\star$ =(0.05, 0.10, 0.15, 0.33, 0.60, 0.70, 0.75); c: $\mathbf{p}^\star$ =(0.05, 0.07, 0.11, 0.20, 0.33, 0.45, 0.50); d: $\mathbf{p}^\star$ =(0.04, 0.05, 0.07, 0.12, 0.20, 0.33, 0.45).

Scenario	Fixed effect true $\mathbf{p}^\star$	Random effect	Study designs
1	a)	OUP, $\sigma_m = 0.3$	CRM and 3+3
2	b)	OUP, $\sigma_m = 0.3$	CRM and 3+3
3	c)	OUP, $\sigma_m = 0.3$	CRM and 3+3
4	d)	OUP, $\sigma_m = 0.3$	CRM and 3+3
5	b)	OUP, $\sigma_m = 0.6$	CRM and 3+3
6	b)	$\Sigma = \left[\exp\left(-\frac{\lvert\delta_{i,j}\rvert}{\ell}\right)\sigma_i\sigma_j\right]$ and $\boldsymbol{\sigma} = (0.1, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6)$	CRM and 3+3
7	c)	$\Sigma = \left[\exp\left(-\frac{\delta_{i,j}^2}{2\ell^2}\right)\sigma_m^2\right]$, $\sigma_m = 0.3$	CRM and 3+3
8	c)	OUP, $\sigma_m = 0.3$	Only 3+3
9	c)	OUP, $\sigma_m = 0.3$	Only CRM

linearity, the model essentially becomes a logistic model, which, when dose increments are defined on a logarithmic scale, represents a specific instance of the *Emax* model.

5.2.3 MTD estimation

Various methods have been developed for determining the MTD; one widely accepted approach involves using posterior mean estimates of parameters to identify the MTD as the dose level with an estimated probability of DLT that most closely aligns with a predefined target, δ_T, within the range of [0,1]. In a meta-analysis framework, emphasis might be placed on the cumulative fixed effect. Therefore, we can express the probability of toxicity for dose i as:

$$\pi_i = \text{logit}^{-1}\left(\sum_{j=1}^{i}\mu_j\right)$$

where the inverse logit function is defined as $\text{logit}^{-1}(x) = (1+\exp(-x))^{-1}$. Based on this, the MTD is determined by finding the dose d_j for which the difference between the posterior expectation of π_i and the target δ_T is minimized:

$$\text{MTD} = d_j, \quad \text{where} \quad j = \arg\min_i |E[\pi_i|\mathbf{x}] - \delta_T|$$

Thus, the MTD corresponds to the dose with an estimated overall mean response that most closely matches the target. An alternative to using the posterior mean in this calculation involves applying the posterior median, offering a different perspective on estimating the MTD.

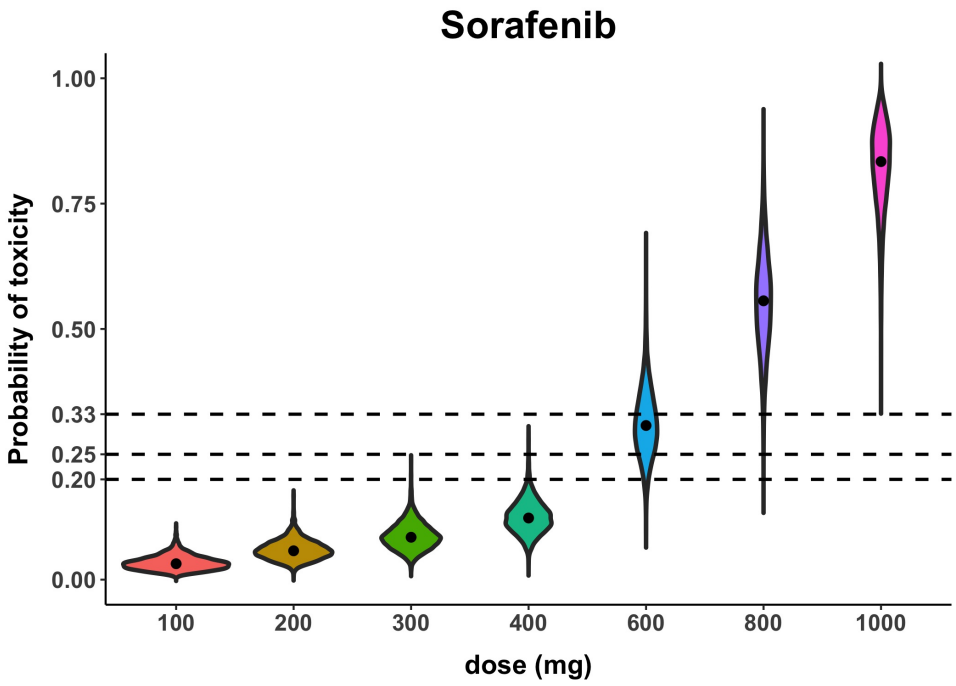


Figure 5.1 – Posterior distribution for probability of toxicity for each dose level in the sorafenib example.

5.2.4 Results

We conducted a comprehensive simulation study to assess the performance features of our proposed methodology (referred to as MADF henceforth), with a focus on comparing the accuracy of correct MTD selection to that of the ZKO [Zohar et al., 2011] method across various test scenarios. These scenarios, outlined in Table 5.2, include different MTD placements, heterogeneity frameworks, and trial designs, totaling nine distinct settings for evaluation. Table 5.3 presents the outcomes in terms of the frequency of selecting the MTD, highlighting the proportion of correct selections (PCS) in bold for the MADF method compared to the ZKO method across meta-analyses incorporating 10 studies each. MADF achieves a PCS ranging from 61% to 92%, outperforming ZKO, which has PCS rates between 50% and 72%. This disparity in PCS rates is anticipated as ZKO does not consider trial heterogeneity, often leading to a higher selection rate of overdoses compared to MADF. For instance, in Scenario 1, with the MTD identified at dose level 3, MADF rate of recommending overly toxic doses is 34%, compared to ZKO 41%. The PCS rates tend to decline with an increase in the σ value, as observed in Scenario 5, yet remain consistent even when there is a misestimation of random effects, as seen in Scenarios 6 and 7.

We applied the MADF method to the sorafenib example. Figure 5.1 illustrates the posterior distribution for the toxicity probability linked with each dose level in the study.

Table 5.3 – Proportion of dose selection using 10 studies in each meta-analysis. The proportion of correct MTD selection in each scenario is written in bold. MADF: proposed method; ZKO: Zohar et al. [2011] method; #patients: median (first quartile - third quartile) number of patients allocated to each dose.

	Dose levels						
	1	2	3	4	5	6	7
Scenario 1							
MADF	0.000	0.082	**0.612**	0.305	0.001	0.000	0.000
ZKO	0.022	0.190	**0.496**	0.253	0.034	0.002	0.003
#patients	31 (23, 41)	31 (23, 41)	54 (43, 65)	15 (9, 23)	6 (3, 12)	2 (0, 6)	0 (0, 3)
Scenario 2							
MADF	0.000	0.000	0.032	**0.920**	0.048	0.000	0.000
ZKO	0.000	0.002	0.052	**0.695**	0.233	0.013	0.005
#patients	22 (18, 26)	26 (20, 32)	29 (23, 37)	59 (50, 68)	14 (9, 20)	5 (0, 9)	0 (0, 3)
Scenario 3							
MADF	0.000	0.000	0.000	0.084	**0.834**	0.082	0.000
ZKO	0.000	0.000	0.002	0.075	**0.676**	0.216	0.031
#patients	22 (17, 26)	23 (19, 29)	26 (20, 33)	29 (22, 38)	45 (36, 54)	11.5 (6, 18)	6 (2, 12)
Scenario 4							
MADF	0.000	0.000	0.000	0.000	0.228	**0.758**	0.014
ZKO	0.000	0.000	0.000	0.001	0.131	**0.680**	0.188
#patients	43 (37, 51)	43 (37, 51)	24 (19, 31)	26 (20, 34)	26 (20, 33)	40 (32, 48)	11 (6, 18)
Scenario 5							
MADF	0.000	0.000	0.085	**0.781**	0.134	0.000	0.000
ZKO	0.004	0.037	0.162	**0.561**	0.215	0.017	0.004
#patients	24 (19, 31)	27 (20, 35)	28 (21, 37)	51 (41, 59)	13 (8, 20)	6 (2, 12)	0 (0, 6)
Scenario 6							
MADF	0.000	0.000	0.019	**0.882**	0.099	0.000	0.000
ZKO	0.000	0.000	0.022	**0.665**	0.287	0.015	0.011
#patients	21 (17, 26)	25 (20, 30)	30 (23, 37)	61 (53, 69)	14 (8, 20)	5 (0, 9)	0 (0, 4)
Scenario 7							
MADF	0.000	0.000	0.000	0.069	**0.830**	0.101	0.000
ZKO	0.000	0.000	0.002	0.075	**0.653**	0.245	0.025
#patients	22 (18, 26)	22.5 (18, 28)	27 (20, 33.25)	30 (23, 38)	45 (36, 54)	12 (6, 18)	6 (3, 12)
Scenario 8							
MADF	0.000	0.000	0.000	0.150	**0.773**	0.077	0.000
ZKO	0.000	0.000	0.002	0.078	**0.591**	0.295	0.034
#patients	24 (18, 27)	24 (18, 27)	24 (18, 27)	24 (18, 27)	30 (24, 36)	9 (3, 12)	3 (0, 6)
Scenario 9							
MADF	0.000	0.000	0.000	0.064	**0.837**	0.099	0.000
ZKO	0.000	0.000	0.001	0.076	**0.715**	0.194	0.014
#patients	20 (16, 25)	24 (18, 31)	30 (23, 39)	37 (27, 47)	60 (49, 71)	15 (9, 23)	10 (4, 16)

5.3 Bayesian framework for multi-source data integration

In this work [Boulet et al., 2024], where I co-supervised a postdoc, our goal is to introduce a Bayesian framework designed for the integration of data from multiple sources, which is adaptable and can be customized to meet the particular research question at hand. In details we focus on preclinical stage and we develop a framework that recognizes the similarities among preclinical studies, enabling the utilization of all pertinent data. Specifically, our strategy is to handle data from various sources (such as cell cultures, mice, etc.) more effectively than conventional methods, in order to enhance the prediction of critical human-related metrics like the Minimal Anticipated Biological Effect Level (MABEL), No Observed Adverse Effect Level (NOAEL), Minimal Effective Dose (MED), MTD, and others. We suppose that the development strategy encompasses K preclinical investigations, which are to be conducted or, at the very least, sequentially evaluated prior to the FIH trial. These studies focus on identical outcomes that can be interpreted through various methods, contingent on the specifics of each study. Our aim is to identify, through suitable extrapolation, transformation or linkage functions, the relevant quantities, such as the MTD, or an appropriate dose range for a FIH dose-ranging study. We introduce a Bayesian framework consisting of four steps (illustrated in Figure 5.2):

1. Initially, we estimate the parameters for each outcome.
2. Subsequently, we employ extrapolation using predetermined formulas to deduce the parameter distribution in humans for each criterion.
3. Next, we assess the consistency/compatibility of posterior distributions through a divergence-based metric.
4. Lastly, we integrate the chosen posterior distributions utilizing an expanded version of the Bayes formula.

5.3.1 First step: parameters estimation

For each of the K preclinical studies, identified by k ranging from 1 to K, the outcomes $\mathbf{y}_k$ recorded in the kth study are considered. These outcomes may be consistent across studies or, if differing, can contribute to determining the relevant doses, d_r, with r spanning from 1 to R, where R represents the total number of target doses (such as MABEL, NOAEL, MED, MTD, etc.). In every study k, Bayesian models, either fixed-effect or mixed-effect and potentially nonlinear, denoted as $\mathbf{f}_k(\mathbf{y}_k, \boldsymbol{\theta}_k)$, are applied to estimate the dose-outcome relationships, which include PK/PD measures for efficacy or toxicity. These models are fitted using Markov chain Monte Carlo (MCMC) techniques, with the introduction of weakly-informative prior distributions for the model parameters $\boldsymbol{\theta}_k$. The specific functions $\mathbf{f}_k(.)$ and the parameter vectors $\boldsymbol{\theta}_k$ might vary between studies, including differences in dimensionality or being tailored to each study. For instance, a study with a single outcome per subject might employ a fixed-effect model, whereas longitudinal studies could opt for mixed-effect models, maintaining the same structure for population effects but incorporating random effects for certain parameters.

For the first study, selection of prior distributions is guided by external pre-existing knowledge. In the absence of such knowledge, non-informative prior distributions are employed. For subsequent studies, should there be a link between any component of $\boldsymbol{\theta}_{k-1}$ and an element of $\boldsymbol{\theta}_k$,

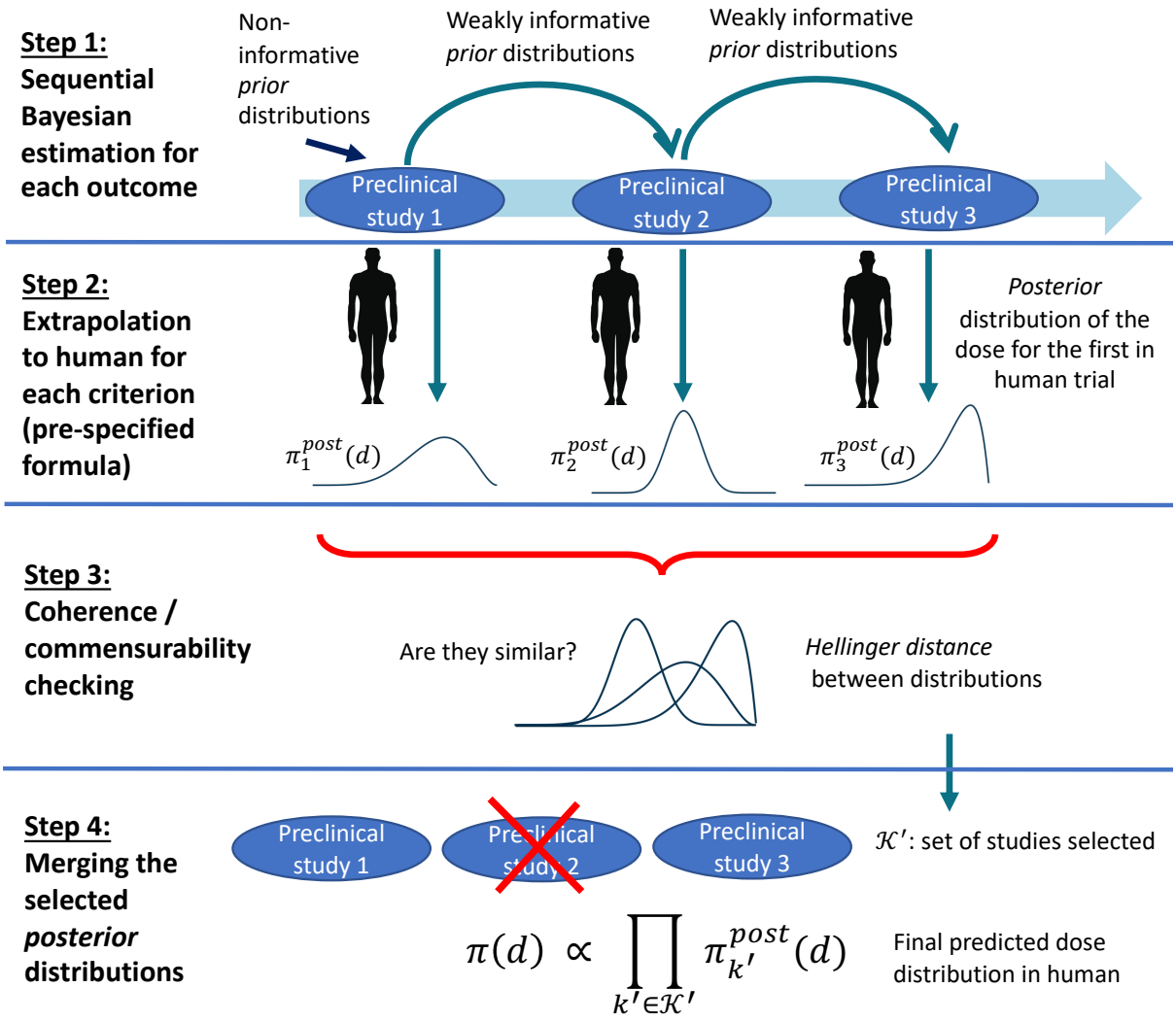


Figure 5.2 – Scheme of the four steps of the Bayesian framework.

prior distributions are derived from the estimated posterior means of the parameters from the preceding study, $k-1$. This process may include extrapolation between studies when needed, utilizing methodologies akin to those mentioned in subsequent steps, and is complemented by the use of suitably large standard deviations. Employing non-informative or weakly informative prior distributions at this stage is crucial to avoid redundant inclusion of the same data in the final stage of the framework. This approach ensures that the effective sample sizes (ESS) [Morita et al., 2008], as well as the influence of each preclinical study within the final analysis, remain comparable.

5.3.2 Second step: extrapolation to human

In the previous phase, the focus was on extrapolating data from one preclinical study to another via prior distributions. Now, our attention shifts to extrapolating preclinical findings

to human applications. For each study k, we extrapolate the preclinical results (the estimated parameters) onto human models using predefined transformation formulas $\mathbf{g}_k(.)$, resulting in $\boldsymbol{\theta}_k^{\text{human}} = \mathbf{g}_k(\boldsymbol{\theta}_k)$. Here, $\boldsymbol{\theta}_k^{\text{human}}$ denotes the vector of parameters extrapolated to humans from the kth study. These transformation formulas might also be employed in the preceding step for inter-study extrapolation for prior distributions.

In the Bayesian framework, this transformation is applied directly to the posterior parameter distributions to acquire the posterior distributions of the extrapolated parameters, in accordance with the random variable transformation theorem. Subsequently, the human-relevant doses, $d_r = d_{r,k}(\boldsymbol{\theta}_k^{\text{human}})$, where $d_{r,k}(.)$ links the extrapolated human parameter vector $\boldsymbol{\theta}_k^{\text{human}}$ from the kth study to the desired dose d_r, are calculated either analytically or through Monte Carlo simulations. The resulting posterior distributions of d_r, $\pi_k^{post}(d_r)$, are then obtained either through direct application of the random variable transformation theorem or approximated using Monte Carlo methods. It is important to note that this paper does not primarily focus on addressing the uncertainties associated with the fixed parameters used in extrapolation scaling, though this could be a topic for future research.

5.3.3 Third step: commensurability checking and posterior selection

At this stage, our objective is to evaluate the predicted human dose distributions, denoted as $D_{k,r} \sim \pi_k^{post}(d_r)$ for each dose r, across all K studies and identify those that appear most alike. In practical scenarios, it might be more suitable to initially choose the predicted human dose distribution(s) from the study (or studies) deemed most relevant, and subsequently, select additional distributions that closely match these. This process can be effortlessly tailored to accommodate such specific situations, although this work does not elaborate on this particular adaptation.

The direct comparison of posterior distributions is challenging due to the varying amounts of information they contain, influenced by the sample size of their respective studies. A method to enhance the comparability between two posterior distributions involves adjusting the likelihood of the study with a larger sample size, as suggested by Ollier et al. [2020]. This adjustment requires re-evaluating one of the datasets and deciding on a discount factor, a process that is relatively straightforward for fixed effect models but more complex for mixed effects models, especially in longitudinal studies. Drawing inspiration from the concept of ESS, we suggest transforming posterior distributions to approximate a Gaussian distribution, which are then normalized to match the highest variance observed. In our case study,

$$D_{k,r}^* = \frac{\max\limits_{h \in \mathcal{K}} S_{h,r}}{S_{k,r}} \log(D_{k,r}) + \left(1 - \frac{\max\limits_{h \in \mathcal{K}} S_{h,r}}{S_{k,r}}\right) M_{k,r} \sim \pi_k^{post}(d_r^*), \tag{5.3}$$

for $k \in \{1, ..., K\}$, where $M_{k,r} = E(\log D_{k,r})$ and $S_{k,r}^2 = Var(\log D_{k,r})$ are respectively the mean and the variance of $\log D_{k,r}$. This approach normalizes the distributions across studies while preserving their original means. This normalization is achieved through a logarithmic transformation of the dose to facilitate Gaussian assumption, though other transformations may be preferred in different contexts. According to Morita et al. [2008], the ESS for a normal distribution in a model with known variance is determined by the prior variance, allowing the ESS of our transformed distributions to be considered equal under a “Gaussian dose model".

After ensuring the posterior distributions are comparable, we assess the similarity of the adjusted distributions $\pi_k^{post}(d_r^*)$ across different studies, specifically between studies k_1 and k_2 where $k_1, k_2 \in \{1, ..., K\}$ and $k_1 \neq k_2$, using the Hellinger distance for each dose r. This distance is defined as $H_r(D_{k_1}, D_{k_2}) = \left[\frac{1}{2}\int\left(\sqrt{\pi_{k_1}^{post}(d_r^*)} - \sqrt{\pi_{k_2}^{post}(d_r^*)}\right)^2 \mathrm{d}d_r^*\right]^{\frac{1}{2}}$. The Hellinger distance is chosen for its properties: it ranges between 0 and 1, is symmetric, and simplifies the definition of consistency between study results. A Hellinger distance nearing 0 implies a high degree of consistency among the predicted human dose distributions across the studies, whereas a distance approaching 1 suggests a lack of agreement between them.

To determine which studies proceed to the subsequent phase, an algorithm incorporating a threshold derived from Hellinger distance measurements is necessary. For situations with a limited number of studies, developing specialized algorithms with straightforward decision criteria is feasible. In more intricate scenarios, clustering techniques may be employed. We recommend conducting simulations on chosen scenarios that are pertinent to optimize the algorithm and setting the threshold based on the precision of these results. For every scenario deemed suitable, we identify studies viewed as consistent and thus to be carried forward to the final step. This involves creating a binary indicator "true response" for each study comparison, assigned a value of 1 if the studies are theoretically similar in the scenario and 0 if not. Through simulation, we then calculate the Hellinger distances between studies. For each potential threshold value on this distance, we generate a "predicted response" indicator, which is 1 if the distance falls below the threshold and 0 otherwise. The selection of the threshold is guided by the relationship between accuracy and the Hellinger distance threshold.

Upon concluding this phase, studies that demonstrate consistency are chosen for further analysis. In cases where no studies form a "cluster" due to their similarity, only the findings from the study deemed most relevant are carried forward to the next step.

5.3.4 Fourth step: merging the selected posterior distributions

Let $\mathcal{K}$ represent the set of indices for the preclinical studies, with $\mathcal{K} = \{1, 2, \ldots, K\}$. We define $\mathcal{K}' \subset \mathcal{K}$ as the subset of studies chosen in the previous step. The extrapolated dose distributions $\pi_{k'}^{post}(d_r)$, for each $k' \in \mathcal{K}'$, are integrated using a modified version of Bayes' theorem to deduce the final predicted dose distribution d_r for humans as follows:

$$\pi(d_r) \propto \prod_{k' \in \mathcal{K}'} \pi_{k'}^{post}(d_r). \tag{5.4}$$

Eq. 5.4 is applicable when the selected distributions, denoted by $\mathcal{K}'$, share a common support or at least a portion of it. This condition prevents scenarios where the multiplicative term would be zero almost everywhere within the domain. Specifically, it addresses cases where one distribution is virtually zero in regions where others are nonzero. Additionally, the integrability of the second term in the equation is assured if the distributions in $\mathcal{K}'$ (or all but one) are bounded. While this boundedness is a sufficient condition for integrability, it is not mandatory. For instance, the multiplication of U-shaped (or J-shaped or inverse J-shaped) Beta distribution densities still results in a beta distribution. In the context of sequential Bayesian analyses, where the posterior distribution from one study is utilized as the prior for the subsequent study within a series

of K iterations of the same study design, the ultimate posterior distribution is represented as $\pi_K^{post}(\boldsymbol{\theta}) \propto \pi^{prior}(\boldsymbol{\theta}) \prod_{k=1}^{K} \mathcal{L}(\boldsymbol{\theta}|\text{data}_k)$. Assuming that $\pi^{prior}(\boldsymbol{\theta})$ is significantly non-informative, and considering $K-1$ similarly non-informative improper distributions adjacent to each likelihood function, it is possible to interpret each pair of product terms (the kth likelihood and its corresponding non-informative improper distribution) as a posterior distribution. Therefore, Eq. 5.4 can be viewed as a broadening of Bayes' theorem. Given that informative prior distributions would result in their information being accounted for multiple times (specifically, $\text{Card}(\mathcal{K}')$ times), it is recommended to employ non-informative prior distributions.

Eq. 5.4 gives the final dose distribution, incorporating all pertinent information for subsequent inference; namely, the expected value (or median) may serve as a point estimate, with credible intervals providing insight into the distribution variability. It is important to note that if, in the preceding step, only the most pertinent study was selected, Eq. 5.4 simplifies to that study posterior distribution. This approach utilizes the original dose posterior distributions, not those modified by Eq. 5.3, thus, it does not presume any of the predicted human dose distributions to follow a Gaussian model.

5.3.5 Results

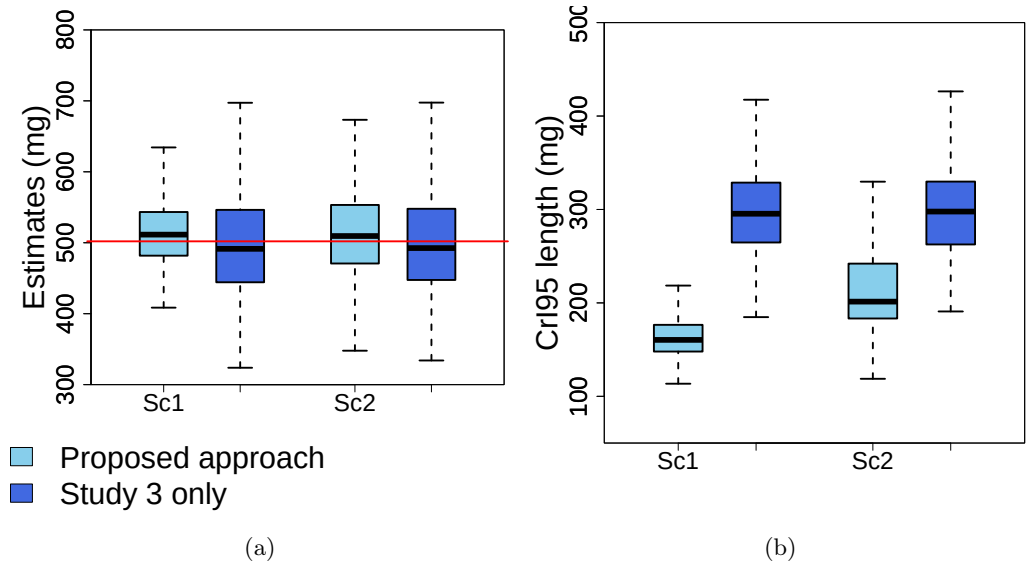
To demonstrate and assess our methodology, we employ the preclinical and clinical development of galunisertib (LY2157299 - the same used in the second chapter) as a case study, generating various simulation scenarios around it. Semi-mechanistic modeling was adopted to predict a safe dosing regimen for Phase I study from data in mouse, rat and dog [Bueno et al., 2008, Gueorguieva et al., 2014]. Allometric scaling, which integrated data from mouse, rat, and dog studies, was utilized to extrapolate PK parameters to humans. We focus on two main scenario: (1) where all species should give consistent results; (2) where the rat study results are inconsistent with the other two species. Data generation and simulations details are given in Boulet et al. [2024]. In this section, only results on MTD are shown.

In Figure 5.3(a), scenario 1 illustrates that using data from all animal species often leads to an accurate estimation of the MTD, with the Bayesian method estimating the MTD at 515 mg (with a standard deviation of 48 mg), closely aligning with the true value of 502 mg. In contrast, scenario 2 shows a slight overestimation of the MTD at 561 mg (with a standard deviation of 296 mg). Furthermore, the 95% credible interval (CrI95) length is significantly wider when solely dog data is utilized (the conventional method) for MTD estimation, as compared to the proposed method (referenced in Figure 5.3(b)). Specifically, the average CrI95 length using only dog data is approximately 300 mg (with a standard deviation of 47 mg), in stark contrast to the shorter lengths of 165 mg and 242 mg (with standard deviations of 26 mg and 153 mg, respectively) for scenarios 1 and 2 when employing the proposed method.

Moreover, Figure 5.4 shows the results of the framework applied to a single simulation run.

5.4 Discussion

In the first work, we introduced a novel approach for conducting random-effects meta-analysis of Phase I dose-finding trials, utilizing a Gaussian process for the random effects structure and a Gamma process for the prior distributions of fixed effects. This method enables more



(a) (b)

Figure 5.3 – Estimated MTD in humans (a) and the length of the 95% credibility interval (CrI95) (b) for scenarios 1 and 2 for the Bayesian approach and using the standard approach (that is, only study 3 - dog - data) over 500 replications, under the assumption that $\omega_{V,1} = 0.7$ for mouse. MTD: Maximum tolerated dose. This figure appears in color in the electronic version of this article.

information sharing between closer dose levels, reducing the correlation as the distance between doses increases. For instance, within a dose panel, the correlation between dose levels 3 and 4 is stronger than between levels 1 and 4, with the degree of correlation logically dependent on the distance rather than a fixed value. The Gamma process ensures adherence to the expected monotonic increase in toxicity. We advise against estimating the entire Gamma process due to the typically insufficient data for accurate parameter estimation in meta-analyses, despite the larger dataset compared to individual trials. Our focus was on modeling toxicities for the specific doses investigated in the trials. However, with ample data allowing for a reliable estimation of the Gamma process, the full model could potentially enable dose interpolation or extrapolation across a continuous range. It is worth mentioning that in a subsequent study, we introduced a simplified version of this meta-analysis by adopting a more straightforward two-stage approach [Röver et al., 2022]. Both methods were then applied in a work regarding meta-analyses of Phase I dose-finding studies of protein kinase inhibitors in oncology. This was part of an internship that I co-supervised, with the manuscript currently undergoing its first round of revisions. Through this applied work, we gained deeper insights into the strengths and limitations of each model. For instance, the two-stage approach necessitates estimable logistic regression for each trial and results in the exclusion of some trials where no DLT occurs, potentially leading to an overestimation of toxicities. Conversely, the method we describe here can result in a bivariate posterior probability of toxicity for certain doses, influenced by the trials substantial heterogeneity. However, overall, both approaches produced consistent results.

In the second study, we introduced a Bayesian framework designed for the integration of data from multiple sources, specifically tailored for extrapolating doses from preclinical to clinical phases. Particularly, steps 3 and 4 of our approach called for innovative methodologies. This

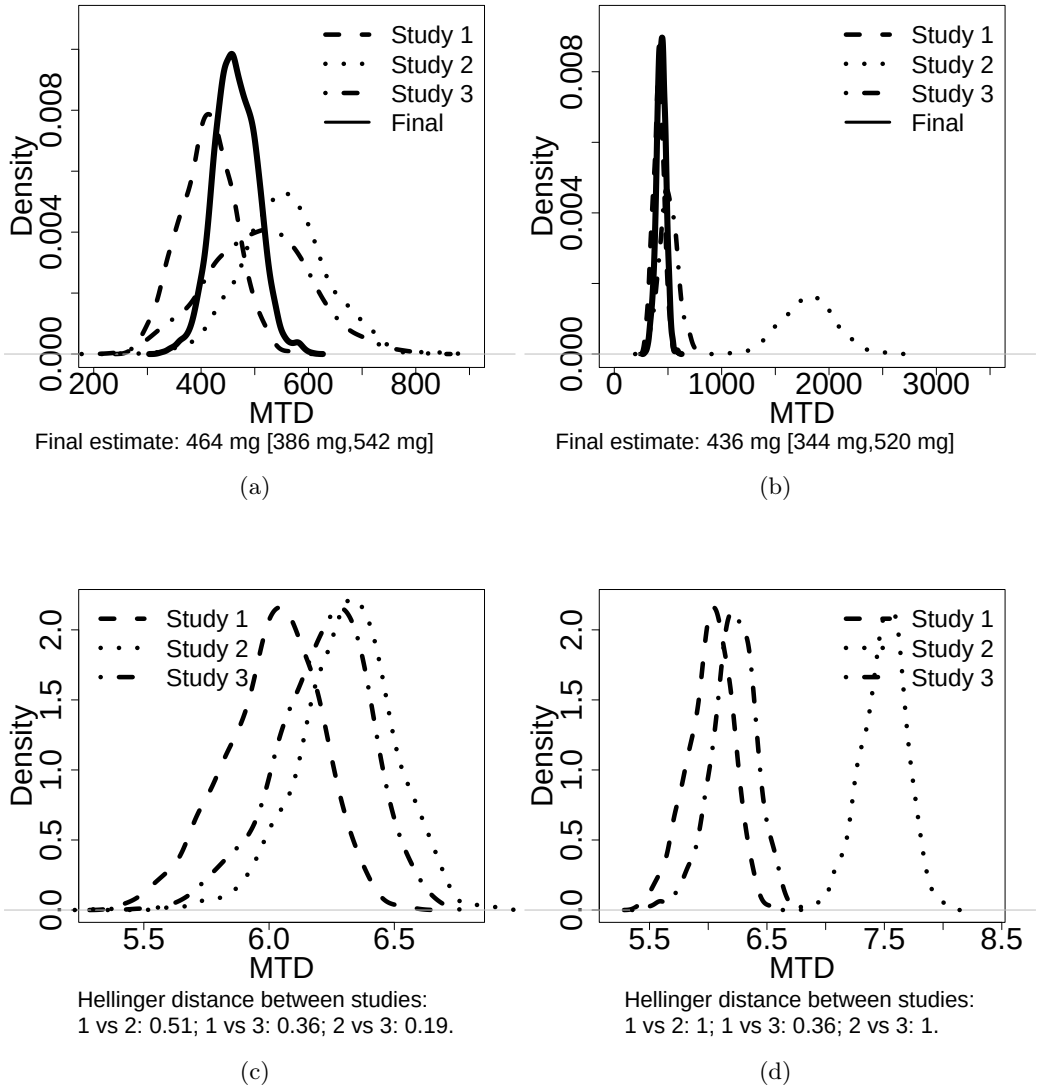


Figure 5.4 – Maximum tolerated dose (MTD) distributions for one simulated dataset in scenarios 1 and 2, for the Bayesian approach. Extrapolated preclinical studies (1 - mouse, 2 - rat, 3 - dog) to human MTD distributions (from Step 2) and final predicted MTD distribution (from Step 4) in the dataset from scenario 1 (a) and from scenario 2 (b); transformed extrapolated MTD distributions and Hellinger distances (from Step 3) in the dataset from scenario 1 (c) and from scenario 2 (d). The Hellinger distances equal to 1 are due to approximation in computation.

novel framework enhances the utilization of all available data compared to conventional methods, thereby diminishing uncertainty in predictions and potentially facilitating more efficient dose selection. A significant benefit of the Bayesian framework is its ability to yield the posterior distribution of extrapolated doses, irrespective of the distribution shape. This capability allows for the consideration of metrics beyond just the mean value or credible intervals, including distribution asymmetry, peak, or other summary statistics. Additionally, the framework offers

considerable flexibility; various submodels (such as linear, generalized, mixed-effect models) can be tailored to fit the specific outcomes of the study and may vary across different studies.

I view this work as a foundation for future enhancements and investigations. Indeed, when incorporating more than three studies, a clustering technique could prove more effective than an ad-hoc algorithm. This approach could identify main clusters of responses, which could then inform the design of FIH trials, including the selection of doses and the determination of prior distributions. Additionally, exploring metrics beyond the Hellinger distance could be beneficial for developing the clustering procedures.

Chapter 6

Perspectives

In this chapter, building on my previous research, I will show a few perspectives I intend to explore further in the upcoming period. I have already obtained funding through grants and collaborations to hire PhDs, postdocs, or engineers for these projects.

Firstly, I plan to continue exploring the use of PK/PD in dose-finding. Indeed, in the BEEP (Bayesian methods for Early Enriched Platform trials) project, of which I am principal investigator, one of the main axes is devoted to develop platform clinical trials for Phase I/II using PK/PD modeling. Since we are focusing on early phases of clinical trial, in this setting "platform" cannot be linked to classical randomized clinical trial. Nevertheless, as in the original definition, early platform phases will allow for flexibility, such as adding new arm or stopping treatments for futility (and/or safety in our case). For example, after sufficient data on the drug have been collected, it becomes possible to propose and incorporate a new arm with alternative schedules into the ongoing trial using the prediction of the efficacy and toxicity probability of untested doses/sequences via the PK/PD modeling.

Regarding other internal information in early clinical development, in collaboration with the ECSTRRA research team (affiliated to CRESS-UMR 1153), we will explore the enrichment designs in these early phases (starting in BEEP and then continuing via the project SMATCH, "Clinical study and trial designs for the evaluation of models and DMDs for their translation to patient's care", of the PEPR Santé Numerique program). These methods have rarely been applied in dose-finding studies, despite the critical importance of identifying an optimal dose for the subsequent development of a drug within a target population. We plan to work on a Bayesian framework for dose-finding that incorporates hierarchical modeling. This will enable consideration of population subgroups in connection with drug toxicity and effectiveness. The framework will include rules for ordering and stopping across these strata, facilitating dynamic dose adjustment and the exchange of information between different groups.

In SMATCH, I am also leading a task where I plan to extend the framework to integrate multi-source data. I will explore and adapt other metrics, probably still divergence-based, to serve as indicators of agreement between posterior distributions. Given that the informational content of each distribution is influenced by the sample size of its source, I will implement transformations to standardize all distributions for equitable comparison, but I will explore different standardization way, not only based on the ESS principle. We will address and quantify heterogeneity through modified clustering methodologies. Additionally, a comprehensive Bayesian solution will be introduced as an alternative to traditional meta-analysis, encompassing both non-parametric and semi-parametric methods. This work will be first done for univariate endpoints, as in the work of Boulet et al. [2024] regarding the MTD, to decide which information to carry on in priors for clinical trials. Then, this work will be adapted and extended to check the agreement between curves, for example the entire dose-response curve or a longitudinal path.

Indeed, in the DIGPHAT (Digital Pharmacological Twins) projects of the PEPR Santé Numerique program, I am workpackage leader and we wish to build pharmacological digital twin by linking and checking all existing models regarding a specific condition.

Regarding PK/PD and mechanistic modeling once more, I wish to account for uncertainties in the extrapolation process (preclinical to clinical or adults to children, for example), which can lead to inconsistencies not necessarily due to discrepancies between previous data and actual trial data. I will refine models for simulating patients in silico as part of the INVENTS EU project (Innovative designs, extrapolation, simulation methods and evidence tools for rare diseases addressing regulatory needs).

Another field I plan to explore in the medium term regards the evaluation of Artificial Intelligence (AI) algorithms in digital medical devices (DMDs) via clinical trials. This will be part of both the SMATCH project (as indicated by its full name) and the Meditwin project, which aims to create personalized virtual twins of organs, metabolism, and cancerous tumors for improved diagnosis and treatment. Indeed, agile AI algorithms in DMDs face regulatory challenges due to their constant updates. These devices, incorporating AI that requires frequent re-calibration, undergo analytical validation using health data sets but must secure clinical approval through expensive clinical trials. Therefore, exploring the potential of synthetic data (in silico) and internal extrapolation across software versions could turn out to be indispensable. Different models will be necessary depending on the DMD type and aims, whether for prevention, treatment administration, or other purposes.

In this final chapter, I have outlined several (though not all) projects I aim to address in the medium term, spanning the next 1 to 5 years. These projects remain connected to the field of clinical trials. This is not surprising, since I will be a co-leader of the third research axis of the HeKA team, focusing on “Data and model-driven designs for next generation clinical trials”.

Bibliography

J. A. Ajani, J. Faust, K. Ikeda, J. C. Yao, H. Anbe, K. L. Carr, M. Houghton, and P. Urrea. Phase I pharmacokinetic study of S-1 plus cisplatin in patients with advanced gastric carcinoma. *Journal of Clinical Oncology*, 23(28):6957–6965, 2005. (Cited on page 8.)

A. Awada, A. Hendlisz, T. Gil, S. Bartholomeus, M. Mano, D. De Valeriola, D. Strumberg, E. Brendel, C. G. Haase, B. Schwartz, et al. Phase I safety and pharmacokinetics of BAY 43-9006 administered for 21 days on / 7 days off in patients with advanced, refractory solid tumours. *British Journal of Cancer*, 92(10):1855, 2005. (Cited on page 64.)

N. Boissel, S. de Botton, X. G. Thomas, E. Rao, H. Bonnevaux, C. Rubin-Carrez, S. Guerif, E. Beys, A. Gosselin, A.-L. Bauchet, et al. An open-label, first-in-human, dose escalation study of a novel cd3-cd123 bispecific t-cell engager administered as a single agent by intravenous infusion in patients with relapsed or refractory acute myeloid leukemia, b-cell acute lymphoblastic leukemia, or high risk myelodysplastic syndrome., 2018. (Cited on page 13.)

B. Bornkamp, F. Bretz, A. Dmitrienko, G. Enas, B. Gaydos, C. H. Hsu, F. Konig, M. Krams, Q. Liu, B. Neuenschwander, T. Parke, J. Pinheiro, A. Roy, R. Sax, and F. Shen. Innovative approaches for designing and analyzing adaptive dose-ranging trials. *Journal of Biopharmaceutical Statistics*, 17(6):965–995, 2007. (Cited on page 19.)

G. Borthakur, H. Kantarjian, F. Ravandi, W. Zhang, M. Konopleva, J. J. Wright, S. Faderl, S. Verstovsek, S. Mathews, M. Andreeff, et al. Phase I study of sorafenib in patients with refractory or relapsed acute leukemias. *Haematologica*, 96(1):62–68, 2011. (Cited on page 64.)

S. Boulet, M. Ursino, R. Michelet, L. Aulin, C. Kloft, E. Comets, and S. Zohar. Bayesian framework for multi-source data integration-application to human extrapolation from preclinical studies. *Statistical Methods in Medical Research*, 33(4):574–588, 2024. (Cited on pages v, 6, 61, 63, 70, 74 and 79.)

W. Brannath, E. Zuber, M. Branson, F. Bretz, P. Gallo, M. Posch, and A. Racine-Poon. Confirmatory adaptive designs with Bayesian decision tools for a targeted therapy in oncology. *Statistics in Medicine*, 28(10):1445–1463, 2009. (Cited on page 22.)

T. M. Braun. Generalizing the TITE-CRM to adapt for early-and late-onset toxicities. *Statist. Med.*, 25(12):2071–2083, 2006. (Cited on page 29.)

T. M. Braun, Z. Yuan, and P. F. Thall. Determining a maximum-tolerated schedule of a cytotoxic agent. *Biometrics*, 61(2):335–343, 2005. (Cited on page 8.)

T. M. Braun, P. F. Thall, H. Nguyen, and M. De Lima. Simultaneously optimizing dose and schedule of a new cytotoxic agent. *Clinical Trials*, 4(2):113–124, 2007. (Cited on page 8.)

F. Bretz, H. Schmidli, F. König, A. Racine, and W. Maurer. Confirmatory seamless phase II/III clinical trials with hypotheses selection at interim: general concepts. *Biometrical Journal:*

Journal of Mathematical Methods in Biosciences, 48(4):623–634, 2006. (Cited on pages 6 and 22.)

F. Bretz, F. Koenig, W. Brannath, E. Glimm, and M. Posch. Adaptive designs for confirmatory clinical trials. *Statistics in medicine*, 28(8):1181–1217, 2009. (Cited on page 6.)

L. E. Broker, F. Y. de Vos, C. J. van Groeningen, B. C. Kuenen, H. E. Gall, M. H. Woo, M. Voi, J. A. Gietema, E. G. de Vries, and G. Giaccone. Phase I trial with BMS-275183, a novel oral taxane with promising antitumor activity. *Clinical Cancer Research*, 12(6):1760–1767, 2006. (Cited on page 8.)

L. Bueno, D. P. de Alwis, C. Pitou, J. Yingling, M. Lahn, S. Glatt, and I. F. Trocóniz. Semi-mechanistic modelling of the tumour growth inhibitory effects of LY2157299, a new type i receptor TGF-β kinase antagonist, in mice. *European Journal of Cancer*, 44(1):142–150, 2008. (Cited on pages 11 and 74.)

J. M. Bullock, T. Lin, and S. Bilic. Clinical pharmacology tools and evaluations to facilitate comprehensive dose finding in oncology: a continuous risk-benefit approach. *The Journal of Clinical Pharmacology*, 57:S105–S115, 2017. (Cited on page 8.)

S. Calderazzo, S. Tarima, C. Reid, N. Flournoy, T. Friede, N. Geller, J. L. Rosenberger, N. Stallard, M. Ursino, M. Vandemeulebroecke, et al. Coping with information loss and the use of auxiliary sources of data: A report from the niss ingram olkin forum series on unplanned clinical trial disruptions. *Statistics in Biopharmaceutical Research*, 16(2):141—-157, 2023. (Cited on page 61.)

K. Chaloner and F. S. Rhame. Quantifying and documenting prior beliefs in clinical trials. *Statistics in medicine*, 20(4):581–600, 2001. (Cited on page 46.)

K. Chaloner, T. Church, T. A. Louis, and J. P. Matts. Graphical elicitation of a prior distribution for a clinical trial. *The Statistician*, pages 341–353, 1993. (Cited on page 46.)

K. M. Chaloner and G. T. Duncan. Assessment of a beta prior distribution: Pm elicitation. *The Statistician*, pages 174–180, 1983. (Cited on page 46.)

A. G. Chapple and P. F. Thall. A hybrid phase I-II/III clinical trial design allowing dose re-optimization in phase III. *Biometrics*, 75(2):371–381, 2019. (Cited on page 22.)

X. Chen, C. Kamperschroer, G. Wong, and D. Xuan. A modeling framework to characterize cytokine release upon t-cell–engaging bispecific antibody treatment: methodology and opportunities. *Clinical and translational science*, 12(6):600–608, 2019. (Cited on pages 8 and 12.)

Y.-B. Chen, S. Li, A. A. Lane, C. Connolly, C. Del Rio, B. Valles, M. Curtis, K. Ballen, C. Cutler, B. R. Dey, et al. Phase I trial of maintenance sorafenib after allogeneic hematopoietic stem cell transplantation for fms-like tyrosine kinase 3 internal tandem duplication acute myeloid leukemia. *Biology of Blood and Marrow Transplantation*, 20(12):2042–2048, 2014. (Cited on page 64.)

Y. Cheung and R. Chappell. Sequential designs for phase I clinical trials with late-onset toxicities. *Biometrics*, 56(4):1177–1182, 2000. (Cited on pages 22 and 29.)

Y. K. Cheung. *Dose finding by the continual reassessment method*. CRC Press, 2011. (Cited on page 29.)

J. W. Clark, J. P. Eder, D. Ryan, C. Lathia, and H.-J. Lenz. Safety and pharmacokinetics of the dual action Raf kinase and vascular endothelial growth factor receptor inhibitor, BAY 43-9006, in patients with advanced, refractory solid tumors. *Clinical Cancer Research*, 11(15): 5472–5480, 2005. (Cited on page 64.)

R. T. Clemen and R. L. Winkler. Combining probability distributions from experts in risk analysis. *Risk analysis*, 19(2):187–203, 1999. (Cited on page 46.)

J. M. Collins, C. K. Grieshaber, and B. A. Chabner. Pharmacologically guided phase I clinical trials based upon preclinical drug development. *Journal of the National Cancer Institute*, 82 (16):1321–1326, 1990. (Cited on page 8.)

E. Comets and S. Zohar. A survey of the way pharmacokinetics are reported in published phase I clinical trials, with an emphasis on oncology. *Clinical Pharmacokinetics*, 48(6):387–395, 2009. (Cited on page 7.)

M. Crump, D. Hedley, S. Kamel-Reid, B. Leber, R. Wells, J. Brandwein, R. Buckstein, J. Kassis, M. Minden, J. Matthews, et al. A randomized phase I clinical and biologic study of two schedules of sorafenib in patients with myelodysplastic syndrome or acute myeloid leukemia: a NCIC (National Cancer Institute of Canada) clinical trials group study. *Leukemia & Lymphoma*, 51(2):252–260, 2010. (Cited on page 64.)

W. DuMouchel. A bayesian model and graphical elicitation procedure for multiple comparisons. *Bayesian statistics*, 3:127–145, 1988. (Cited on page 46.)

B. Duputel, N. Stallard, F. Montestruc, S. Zohar, and M. Ursino. Using dichotomized survival data to construct a prior distribution for a bayesian seamless phase ii/iii<? pag\break?> clinical trial. *Statistical Methods in Medical Research*, 32(5):963–977, 2023. (Cited on pages v, 6, 21 and 34.)

L. L. Fernandes, J. M. Taylor, and S. Murray. Adaptive phase I clinical trial design using Markov models for conditional probability of toxicity. *Journal of Biopharmaceutical Statistics*, 26(3): 475–498, 2016. (Cited on page 22.)

J. Fourie Zirkelbach, M. Shah, J. Vallejo, J. Cheng, A. Ayyoub, J. Liu, R. Hudson, R. Sridhara, G. Ison, L. Amiri-Kordestani, et al. Improving dose-optimization processes used in oncology drug development to minimize toxicity and maximize benefit to patients. *Journal of Clinical Oncology*, 40(30):3489–3500, 2022. (Cited on page 4.)

P. M. Fracasso, K. A. Blum, M. K. Ma, B. R. Tan, L. P. Wright, S. A. Goodner, C. L. Fears, W. Hou, M. A. Arquette, J. Picus, A. Denes, J. E. Mortimer, L. Ratner, S. P. Ivy, and H. L. McLeod. Phase I study of pegylated liposomal doxorubicin and the multidrug-resistance modulator, valspodar. *British Journal of Cancer*, 93(1):46–53, 2005. (Cited on page 8.)

T. Friede, N. Parsons, and N. Stallard. A conditional error function approach for subgroup selection in adaptive clinical trials. *Statistics in Medicine*, 31(30):4309–4320, 2012. (Cited on page 22.)

T. Friede, C. Röver, S. Wandel, and B. Neuenschwander. Meta-analysis of few small studies in orphan diseases. *Research Synthesis Methods*, 8(1):79–91, 2017. (Cited on page 64.)

T. Friede, N. Stallard, and N. Parsons. Adaptive seamless clinical trials using early outcomes for treatment or subgroup selection: Methods, simulation model and their implementation in R. *Biometrical Journal*, 62(5):1264–1283, 2020. (Cited on page 22.)

J. Furuse, H. Ishii, K. Nakachi, E. Suzuki, S. Shimizu, and K. Nakajima. Phase I study of sorafenib in Japanese patients with hepatocellular carcinoma. *Cancer Science*, 99(1):159–165, 2008. (Cited on page 64.)

J. Gabrielsson and D. Weiner. *Pharmacokinetic and pharmacodynamic data analysis: concepts and applications - Fourth Edition*. Sweden: Swedish Pharmaceutical Press, 2007. (Cited on pages 4 and 5.)

A. Gelman. Prior distributions for variance parameters in hierarchical models (comment on article by Browne and Draper). *Bayesian Analysis*, 1(3):515 – 534, 2006. (Cited on page 17.)

E. Gerard, S. Zohar, C. Lorenzato, M. Ursino, and M.-k. Riviere. Bayesian modeling of a bivariate toxicity outcome for early phase oncology trials evaluating dose regimens. *Statistics in Medicine*, 40(23):5096–5114, 2021. (Cited on page 20.)

E. Gerard, S. Zohar, H.-T. Thai, C. Lorenzato, M.-K. Riviere, and M. Ursino. Bayesian dose regimen assessment in early phase oncology incorporating pharmacokinetics and pharmacodynamics. *Biometrics*, 78(1):300–312, 2022. (Cited on pages v, 6, 7, 13 and 17.)

I. Gueorguieva, A. L. Cleverly, A. Stauber, N. S. Pillay, J. A. Rodon, C. P. Miles, J. M. Yingling, and M. M. Lahn. Defining a therapeutic window for the novel TGF-β inhibitor LY2157299 monohydrate based on a pharmacokinetic/pharmacodynamic model. *British Journal of Clinical Pharmacology*, 77(5):796–807, 2014. (Cited on pages 11 and 74.)

B. K. Günhan, S. Weber, and T. Friede. A bayesian time-to-event pharmacokinetic model for phase i dose-escalation trials with multiple schedules. *Statistics in Medicine*, 39(27):3986–4000, 2020. (Cited on pages 8 and 20.)

A. Hiance, S. Chevret, and V. Lévy. A practical approach for eliciting expert prior beliefs about cancer survival in phase iii randomized trial. *Journal of clinical epidemiology*, 62(4):431–437, 2009. (Cited on page 46.)

B. P. Hobbs, B. P. Carlin, S. J. Mandrekar, and D. J. Sargent. Hierarchical commensurate and power prior models for adaptive incorporation of historical information in clinical trials. *Biometrics*, 67(3):1047–1056, Sep 2011. (Cited on page 46.)

B. P. Hobbs, D. J. Sargent, and B. P. Carlin. Commensurate Priors for Incorporating Historical Information in Clinical Trials Using General and Generalized Linear Models. *Bayesian Anal*, 7(3):639–674, Aug 2012. (Cited on page 46.)

J. G. Ibrahim and M.-H. Chen. Power prior distributions for regression models. *Statist. Sci.*, 15 (1):46–60, 2000. ISSN 0883-4237. (Cited on pages 46 and 47.)

J. G. Ibrahim, M.-H. Chen, Y. Gwon, and F. Chen. The power prior: theory and applications. *Statistics in Medicine*, 34(28):3724–3749, 2015. (Cited on page 46.)

ICH. Ethnic Factors in the Acceptability of Foreign Clinical Data, 1998. URL https://www.ich.org/products/guidelines/efficacy/efficacy-single/article/ethnic-factors-in-the-acceptability-of-foreign-clinical-data.html. (Cited on page 47.)

Y. Ji, P. Liu, Y. Li, and B. Nebiyou Bekele. A modified toxicity probability interval method for dose-finding trials. *Clinical trials*, 7(6):653–663, 2010. (Cited on page 3.)

N. Jia, I. Liou, J. Halldorson, R. Carithers, J. Perkins, J. Reyes, M. Yeh, E. Stohr, S. Rao, and E. H. Lin. Phase I adjuvant trial of sorafenib in patients with hepatocellular carcinoma after orthotopic liver transplantation. *Anticancer Research*, 33(6):2797–2800, 2013. (Cited on page 64.)

S. R. Johnson, G. A. Tomlinson, G. A. Hawker, J. T. Granton, H. A. Grosbein, and B. M. Feldman. A valid and reliable belief elicitation method for bayesian priors. *Journal of clinical epidemiology*, 63(4):370–383, 2010. (Cited on pages 46 and 50.)

J. Kadane and L. J. Wolfson. Experiences in elicitation. *Journal of the Royal Statistical Society: Series D (The Statistician)*, 47(1):3–19, 1998. (Cited on page 46.)

P. K. Kimani, N. Stallard, and J. L. Hutton. Dose selection in seamless phase II/III clinical trials based on efficacy and safety. *Statistics in Medicine*, 28(6):917–936, 2009. (Cited on page 22.)

P. M. Kuhnert, T. G. Martin, and S. P. Griffiths. A guide to eliciting and using expert knowledge in bayesian ecological models. *Ecology letters*, 13(7):900–914, 2010. (Cited on page 46.)

F. La Gamba, T. Jacobs, H. Geys, T. Jaki, J. Serroyen, M. Ursino, A. Russu, and C. Faes. Bayesian sequential integration within a preclinical pharmacokinetic and pharmacodynamic modeling framework: Lessons learned. *Pharmaceutical Statistics*, 18(4):486–506, Apr. 2019. (Cited on page 65.)

J. Lawless and M. Crowder. Covariates and random effects in a gamma process model with application to degradation and failure. *Lifetime Data Analysis*, 10(3):213–227, 2004. (Cited on page 66.)

G. Lestini, C. Dumont, and F. Mentré. Influence of the size of cohorts in adaptive design for nonlinear mixed effects models: An evaluation by simulation for a pharmacokinetic and

pharmacodynamic model for a biomarker in oncology. *Pharmaceutical Research*, 32(10):3159–3169, 2015. (Cited on page 11.)

F. Liang, R. Paulo, G. Molina, M. A. Clyde, and J. O. Berger. Mixtures of g priors for Bayesian variable selection. *Journal of the American Statistical Association*, 103(481):410–423, 2008. (Cited on page 40.)

C. A. Liu and T. M. Braun. Parametric non-mixture cure models for schedule finding of therapeutic agents. *Journal of the Royal Statistical Society Series C: Applied Statistics*, 58(2): 225–236, 2009. (Cited on page 8.)

S. Liu and Y. Yuan. Bayesian optimal interval designs for phase i clinical trials. *Journal of the Royal Statistical Society: Series C: Applied Statistics*, pages 507–523, 2015. (Cited on page 3.)

J. Lyu, E. Curran, and Y. Ji. Bayesian adaptive design for finding the maximum tolerated sequence of doses in multicycle dose-finding clinical trials. *JCO Precision Oncology*, 2:1–19, 2018. (Cited on page 8.)

J. Maca, S. Bhattacharya, V. Dragalin, P. Gallo, and M. Krams. Adaptive seamless phase II/III designs—background, operational aspects, and examples. *Drug Information Journal*, 40(4): 463–473, 2006. (Cited on pages 5 and 22.)

S. Micallef, A. Sostelly, J. Zhu, P. G. Baverel, and F. Mercier. Exposure driven dose escalation design with overdose control: Concept and first real life experience in an oncology phase i trial. *Contemporary Clinical Trials Communications*, 26:100901, 2022. (Cited on page 20.)

A. A. Miller, D. J. Murry, K. Owzar, D. R. Hollis, E. B. Kennedy, G. Abou-Alfa, A. Desai, J. Hwang, M. A. Villalona-Calero, E. C. Dees, et al. Phase I and pharmacokinetic study of sorafenib in patients with hepatic or renal dysfunction: CALGB 60301. *Journal of Clinical Oncology*, 27(11):1800, 2009. (Cited on page 64.)

H. Minami, K. Kawada, H. Ebi, K. Kitagawa, Y.-i. Kim, K. Araki, H. Mukai, M. Tahara, H. Nakajima, and K. Nakajima. Phase I and pharmacokinetic study of sorafenib, an oral multikinase inhibitor, in Japanese patients with advanced refractory solid tumors. *Cancer Science*, 99(7):1492–1498, 2008. (Cited on page 64.)

M. Moatti, S. Zohar, T. Facon, P. Moreau, J.-Y. Mary, and S. Chevret. Modeling of experts' divergent prior beliefs for a sequential phase iii clinical trial. *Clinical Trials*, 10(4):505–514, 2013. (Cited on page 46.)

M. Moore, H. Hirte, L. Siu, A. Oza, S. Hotte, O. Petrenciuc, F. Cihon, C. Lathia, and B. Schwartz. Phase I study to determine the safety and pharmacokinetics of the novel Raf kinase and VEGFR inhibitor BAY 43-9006, administered for 28 days on / 7 days off in patients with advanced, refractory solid tumors. *Annals of Oncology*, 16(10):1688–1694, 2005. (Cited on page 64.)

S. Morita, P. F. Thall, and P. Müller. Determining the effective sample size of a parametric prior. *Biometrics*, 64(2):595–602, 2008. (Cited on pages 39, 46, 47, 71 and 72.)

S. Morita, P. F. Thall, and P. Muller. Evaluating the Impact of Prior Assumptions in Bayesian Biostatistics. *Stat Biosci*, 2(1):1–17, Jul 2010. (Cited on page 46.)

S. Morita, P. F. Thall, and P. Muller. Prior Effective Sample Size in Conditionally Independent Hierarchical Models. *Bayesian Anal*, 7(3), Sep 2012. (Cited on page 46.)

F. T. Musuamba, E. Manolis, N. Holford, S. A. Cheung, L. E. Friberg, K. Ogungbenro, M. Posch, J. Yates, S. Berry, N. Thomas, et al. Advanced methods for dose and regimen finding during drug development: summary of the ema/efpia workshop on dose finding (london 4–5 december 2014). *CPT: pharmacometrics & systems pharmacology*, 6(7):418–429, 2017. (Cited on page 8.)

L. Nabors, J. Supko, M. Rosenfeld, M. Chamberlain, S. Phuphanich, T. Batchelor, S. Desideri, X. Ye, J. Wright, S. Gujar, et al. Phase I trial of sorafenib in patients with recurrent or progressive malignant glioma. *Neuro-Oncology*, 13(12):1324–1330, 2011. (Cited on page 64.)

B. Neuenschwander, M. Branson, and T. Gsponer. Critical aspects of the Bayesian approach to phase I cancer trials. *Statistics in Medicine*, 27(13):2420–2439, 2008. (Cited on pages 3 and 23.)

B. Neuenschwander, M. Branson, and D. Spiegelhalter. A note on the power prior. *Statistics in Meddecine*, 28, Sep 2009. (Cited on page 46.)

B. Neuenschwander, S. Weber, H. Schmidli, and A. O'Hagan. Predictively consistent prior effective sample sizes. *Biometrics*, 76(2):578–587, 2020. (Cited on pages 39 and 46.)

A. O'Hagan. Eliciting expert beliefs in substantial practical applications. *Journal of the Royal Statistical Society: Series D (The Statistician)*, 47(1):21–35, 1998. (Cited on page 46.)

A. O'Hagan, C. E. Buck, A. Daneshkhah, J. R. Eiser, P. H. Garthwaite, D. J. Jenkinson, J. E. Oakley, and T. Rakow. *Uncertain judgements: eliciting experts' probabilities*. John Wiley & Sons, 2006. (Cited on page 46.)

A. Ollier, S. Morita, M. Ursino, and S. Zohar. An adaptive power prior for sequential clinical trials–application to bridging studies. *Statistical methods in medical research*, 29(8):2282–2294, 2020. (Cited on pages v, 6, 45, 47, 49 and 72.)

A. Ollier, S. Zohar, S. Morita, and M. Ursino. Estimating similarity of dose–response relationships in phase i clinical trials—case study in bridging data package. *International journal of environmental research and public health*, 18(4):1639, 2021. (Cited on page 61.)

J. O'Quigley, M. Pepe, and L. Fisher. Continual reassessment method: a practical design for phase 1 clinical trials in cancer. *Biometrics*, 46(1):33–48, 1990. (Cited on pages 3 and 19.)

J. O'Quigley, M. D. Hughes, T. Fenton, and L. Pei. Dynamic calibration of pharmacokinetic parameters in dose-finding studies. *Biostatistics*, 11(3):537—-545, 2010. (Cited on page 8.)

S. Patterson, S. Francis, M. Ireson, D. Webber, and J. Whitehead. A novel bayesian decision procedure for early-phase dose-finding studies. *Journal of Biopharmaceutical Statistics*, 9(4): 583–597, 1999. (Cited on pages 8 and 10.)

S. Piantadosi and G. Liu. Improved designs for dose escalation studies using pharmacokinetic measurements. *Statistics in Medicine*, 15(15):1605–1618, 1996. (Cited on pages 8 and 9.)

S. Postel-Vinay, L. Collette, X. Paoletti, E. Rizzo, C. Massard, D. Olmos, C. Fowst, B. Levy, P. Mancini, D. Lacombe, et al. Towards new methods for the determination of dose limiting toxicities and the assessment of the recommended dose for further studies of molecularly targeted agents – Dose-Limiting Toxicity and Toxicity Assessment Recommendation Group for Early Trials of Targeted therapies, an European Organisation for Research and Treatment of Cancer-led study. *European Journal of Cancer*, 50(12):2040–2049, 2014. (Cited on page 22.)

H. Quan, X. Luo, T. Zhou, and P.-L. Zhao. Seamless phase II/III/IIIb clinical trial designs with different endpoints for different phases. *Communications in Statistics-Theory and Methods*, 49(22):5436–5454, 2020. (Cited on page 22.)

M. Resche-Rigon, S. Zohar, and S. Chevret. Adaptive designs for dose-finding in non-cancer phase II trials: influence of early unexpected outcomes. *Clin. Trials*, 5(6):595–606, 2008. (Cited on page 32.)

M. Resche-Rigon, S. Zohar, and S. Chevret. Maximum-relevance weighted likelihood estimator: Application to the continual reassessment method. *Stat. Interface*, 3(2):177–184, 2010. (Cited on page 32.)

W. F. Rosenberger and L. M. Haines. Competing designs for phase i clinical trials: a review. *Statistics in Medicine*, 21(18):2757–2770, 2002. (Cited on page 19.)

C. Röver and T. Friede. Dynamically borrowing strength from another study through shrinkage estimation. *Statistical Methods in Medical Research*, 29(1):293–308, 2020. (Cited on page 46.)

C. Röver, M. Ursino, T. Friede, and S. Zohar. A straightforward meta-analysis approach for oncology phase i dose-finding studies. *Statistics in Medicine*, 41(20):3915–3940, 2022. (Cited on page 75.)

L. J. Savage. Elicitation of personal probabilities and expectations. *Journal of the American Statistical Association*, 66(336):783–801, 1971. (Cited on page 46.)

H. Schmidli, F. Bretz, and A. Racine-Poon. Bayesian predictive power for interim adaptation in seamless phase II/III trials where the endpoint is survival up to some specified timepoint. *Statistics in Medicine*, 26(27):4925–4938, 2007. (Cited on page 22.)

H. Schmidli, S. Gsteiger, S. Roychoudhury, A. O'Hagan, D. Spiegelhalter, and B. Neuenschwander. Robust meta-analytic-predictive priors in clinical trials with historical control information. *Biometrics*, 70(4):1023–1032, 2014. (Cited on page 46.)

M. Shah, A. Rahman, M. R. Theoret, and R. Pazdur. The drug-dosing conundrum in oncology-when less is more. *The New England journal of medicine*, 385(16):1445–1447, 2021. (Cited on page 4.)

A. Shimabukuro-Vornhagen, P. Gödel, M. Subklewe, H. J. Stemmler, H. A. Schlößer, M. Schlaak, M. Kochanek, B. Böll, and M. S. von Bergwelt-Baildon. Cytokine release syndrome. *Journal for immunotherapy of cancer*, 6(1):1–14, 2018. (Cited on page 12.)

D. J. Spiegelhalter, N. L. Harris, K. Bull, and R. C. Franklin. Empirical evaluation of prior beliefs about frequencies: methodology and a case study in congenital heart disease. *Journal of the American Statistical Association*, 89(426):435–443, 1994. (Cited on page 46.)

N. Stallard. Group-sequential methods for adaptive seamless phase II/III clinical trials. *Journal of Biopharmaceutical Statistics*, 21(4):787–801, 2011. (Cited on page 22.)

N. Stallard, C. U. Kunz, S. Todd, N. Parsons, and T. Friede. Flexible selection of a single treatment incorporating short-term endpoint information in a phase II/III clinical trial. *Statistics in Medicine*, 34(23):3104–3115, 2015. (Cited on page 22.)

F. W. Steutel. Infinite divisibilty in theory and practice. *Scandinavian Journal of Statistics*, 6 (2):57–64, 1979. (Cited on page 66.)

B. E. Storer. Design and analysis of phase i clinical trials. *Biometrics*, pages 925–937, 1989. (Cited on page 3.)

D. Strumberg, H. Richly, R. A. Hilger, N. Schleucher, S. Korfee, M. Tewes, M. Faghih, E. Brendel, D. Voliotis, C. G. Haase, et al. Phase I clinical and pharmacokinetic study of the novel Raf kinase and vascular endothelial growth factor receptor inhibitor BAY 43-9006 in patients with advanced refractory solid tumors. *Journal of Clinical Oncology*, 23(5):965–972, 2005. (Cited on page 64.)

X. Su, Y. Li, P. Müller, C.-W. Hsu, H. Pan, and K.-A. Do. A semi-mechanistic dose-finding design in oncology using pharmacokinetic/pharmacodynamic modeling. *Pharmaceutical statistics*, 21(6):1149–1166, 2022. (Cited on page 20.)

S.-B. Tan, Y.-F. A. Chung, B.-C. Tai, Y.-B. Cheung, and D. Machin. Elicitation of prior distributions for a phase iii randomized controlled trial of adjuvant therapy with surgery for hepatocellular carcinoma. *Controlled clinical trials*, 24(2):110–121, 2003. (Cited on page 46.)

D. T. Teachey, S. F. Lacey, P. A. Shaw, J. J. Melenhorst, S. L. Maude, N. Frey, E. Pequignot, V. E. Gonzalez, F. Chen, J. Finklestein, et al. Identification of predictive biomarkers for cytokine release syndrome after chimeric antigen receptor t-cell therapy for acute lymphoblastic leukemia. *Cancer discovery*, 6(6):664–679, 2016. (Cited on page 12.)

P. F. Thall, H. Q. Nguyen, S. Zohar, and P. Maton. Optimizing Sedative Dose in Preterm Infants Undergoing Treatment for Respiratory Distress Syndrome. *J. Am. Statist. Ass.*, 109 (507):931–943, Sep 2014. (Cited on page 22.)

P. F. Thall, M. Ursino, V. Baudouin, C. Alberti, and S. Zohar. Bayesian treatment comparison using parametric mixture priors computed from elicited histograms. *Statistical methods in medical research*, 28(2):404–418, 2019. (Cited on pages v, 6, 45, 53 and 54.)

G. E. Uhlenbeck and L. S. Ornstein. On the theory of Brownian motion. *Physical Review*, 36 (5):823–841, Sept. 1930. (Cited on page 66.)

M. Ursino, S. Zohar, F. Lentz, C. Alberti, T. Friede, N. Stallard, and E. Comets. Dose-finding methods for phase i clinical trials using pharmacokinetics in small populations. *Biometrical Journal*, 59(4):804–825, 2017. (Cited on pages v, 6, 7, 8 and 11.)

M. Ursino, Y. Yuan, C. Alberti, E. Comets, G. Favrais, T. Friede, F. Lentz, N. Stallard, and S. Zohar. A dose finding design for seizure reduction in neonates. *Journal of the Royal Statistical Society Series C: Applied Statistics*, 68(2):427–444, 2019. (Cited on pages v, 6 and 21.)

M. Ursino, C. Röver, S. Zohar, and T. Friede. Random-effects meta-analysis of phase i dose-finding studies using stochastic process priors. *Ann. Appl. Stat*, 15(1):174–193, 2021. (Cited on pages v, 6, 63 and 65.)

M. Ursino, L. Biard, and S. Chevret. Dice: A bayesian model for early dose finding in phase i trials with multiple treatment courses. *Biometrical Journal*, 64(8):1486–1497, 2022. (Cited on pages v, 6, 21 and 23.)

U.S. Food and Drug Administration (FDA). Adaptive designs for clinical trials of drugs and biologics guidance for industry. https://www.fda.gov/regulatory-information/search-fda-guidance-documents/adaptive-designs-clinical-trials-drugs-and-biologics-guidance-industry, 2018. (Cited on page 6.)

J. van Rosmalen, D. Dejardin, Y. van Norden, B. Löwenberg, and E. Lesaffre. Including historical data in the analysis of clinical trials: Is it worth the effort? *Statistical methods in medical research*, 27(10):3167–3182, 2018. (Cited on page 46.)

N. A. Wages, J. O'Quigley, and M. R. Conaway. Phase i design for completely or partially ordered treatment schedules. *Statistics in Medicine*, 33(4):569–579, 2014. (Cited on page 8.)

J. Whitehead and H. Brunier. Bayesian decision procedures for dose determining experiments. *Statistics in Medicine*, 14(9):885–893, 1995. (Cited on page 19.)

J. Whitehead, S. Patterson, D. Webber, S. Francis, and Y. Zhou. Easy-to-implement bayesian methods for dose-escalation studies in healthy volunteers. *Biostatistics*, 2(1):47–61, 2001. (Cited on page 10.)

J. Whitehead, Y. Zhou, L. Hampson, E. Ledent, and A. Pereira. A bayesian approach for dose-escalation in a phase i clinical trial incorporating pharmacodynamic endpoints. *Journal of Biopharmaceutical Statistics*, 17(6):1117–1129, 2007. (Cited on pages 8 and 10.)

M. Wiesenfarth and S. Calderazzo. Quantification of prior impact in terms of effective current sample size. *Biometrics*, 76(1):326–336, 2020. (Cited on pages 39 and 46.)

J. Zhang and T. M. Braun. A phase i bayesian adaptive design to simultaneously optimize dose and schedule assignments both between and within patients. *Journal of the American Statistical Association*, 108(503):892–901, 2013. (Cited on page 8.)

H. Zheng and L. V. Hampson. A Bayesian decision-theoretic approach to incorporate preclinical information into phase I oncology trials. *Biometrical Journal*, 62(6):1408–1427, Oct. 2020. (Cited on page 65.)

H. Zheng, L. V. Hampson, and S. Wandel. A robust Bayesian meta-analytic approach to incorporate animal data into phase I oncology trials. *Statistical methods in medical research*, 29 (1):94–110, 2020. (Cited on page 65.)

W. Zhong, J. S. Koopmeiners, and B. P. Carlin. A trivariate continual reassessment method for phase i/ii trials of toxicity, efficacy, and surrogate efficacy. *Statist. Med.*, 31(29):3885–3895, 2012. (Cited on page 22.)

S. Zohar, S. Katsahian, and J. O'Quigley. An approach to meta-analysis of dose-finding studies. *Statistics in Medicine*, 30(17):2109–2116, 2011. (Cited on pages 64, 68 and 69.)