



**HAL**  
open science

# Bayesian assessment and quantification of curves similarities : application to bridging studies

Adrien Ollier

► **To cite this version:**

Adrien Ollier. Bayesian assessment and quantification of curves similarities : application to bridging studies. Human health and pathology. Université Paris Cité, 2021. English. NNT : 2021UNIP5237 . tel-04557602

**HAL Id: tel-04557602**

**<https://theses.hal.science/tel-04557602>**

Submitted on 24 Apr 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Université de Paris

École doctorale Pierre Louis de santé publique  
ED393 Épidémiologie et Sciences de l'Information Biomédicale

INSERM UMRS 1138 Équipe 22  
"Sciences de l'Information au service de la médecine personnalisée"

## Bayesian assessment and quantification of curves similarities

*Application to bridging studies*

Par Adrien OLLIER

Thèse de doctorat de Biostatistiques

Dirigée par Sarah ZOHAR

Présentée et soutenue publiquement le 31 Mai 2021

Devant un jury composé de :

Xavier PAOLETTI	PU-PH à Université Versailles Saint Quentin / Paris Saclay Institut Curie (rapporteur)
Sylvie CHEVRET	PU-PH à Université de Paris, Hôpital Saint Louis APHP (Présidente du jury)
Lisa HAMPSON	Docteur, Novartis (examinatrice)
Sarah ZOHAR	Directrice de Recherche, Université de Paris, INSERM (Directrice de thèse)
Moreno URSINO	Docteur, Université de Paris (co-encadrant)



## Remerciements

Je tiens en premier lieu à remercier Xavier Paoletti et Nicolas Molinari d'avoir accepté de rapporter mon travail, de leurs commentaires qui ont permis d'en augmenter la qualité. Mes remerciements vont également à Lisa Hampson et Sylvie Chevret, pour avoir accepté d'être membre de mon jury.

Je tiens également à exprimer ma profonde reconnaissance à mes deux directeurs de thèse, Sarah Zohar et Moreno Ursino. Sarah, Moreno, je tiens à vous remercier de votre accompagnement, de votre soutien et de tous vos conseils, tant personnels que professionnels. Votre accompagnement a été très précieux pour moi, j'ai conscience de tout ce que je vous doit. Merci de votre bienveillance et surtout de votre patience.

Mes sincères remerciements à l'Institut National du Cancer d'avoir financé cette thèse.

Je suis également reconnaissant envers Anita Burgun d'avoir accepté de m'accueillir dans son équipe au Centre de Recherche des Cordeliers, et à mes collègues pour leur bonne humeur. J'ai une pensée particulière pour Sandrine, Anaïs, Marie et Samir, avec qui j'ai passé le plus clair de mon temps. Vous avez dû me supporter. Pas toujours facile, je sais bien. . . bravo à vous qui avez tenu bon ! Un remerciement particulier au professeur Ménard, qui nous a ouvert son bureau et son cœur, avec de savoureuses et intéressantes discussions.

Impossible pour moi de ne pas citer quelques personnes dans ces remerciements. Tout d'abord l'ensemble de ma famille. Mes parents et ma sœur en particulier, qui me supportent depuis maintenant 27 ans ! Merci de vos encouragements et de votre patience durant tout ce temps.

Une pensée pour mes amis de prépa, en particulier pour toi Mehdi, qui soutiendras prochainement ! Merci à vous d'être là depuis toutes ces années !  
Merci également au soutien des amis statisticiens, aussi rencontrés à l'ISUP: Ancelin,

Benjamin, Paul, Joffrey, Virginie, Amélie, Mélaine.

Il est enfin certaines personnes dans mon entourage que je ne saurais omettre. . . Des personnes rencontrées à l'ISUP. Que serait mon existence sans elles ? Par où ou par qui commencer.

Par toi Léa, rencontrée dès la première année, et pour qui le tryptique humour, Simpson et bière avec modération est également une religion. Merci d'avoir toujours été là, de ta bonne humeur. Par Toi Marion ensuite. Merci de ton réconfort, lorsque ça n'allait pas fort, camarade d'infortune ! Je ne doute pas qu'un jour l'Encyclopædia Universalis te citera comme une géographe d'exception. Merci de tes savoureuses anecdotes. Vient ensuite le tour de Max, dangereux pyromane lors de barbecues sur balcon, fervent défenseur du droit de répondre aux méchantes vieilles dans la rue, redoutable adversaire aux échecs : merci de ces moments incroyables ! Et enfin toi Kévin, merci de partager ta passion du cinéma, de l'Histoire et tant d'autres ! Merci et désolé. Désolé qu'à chaque fois que je fais une ânerie, la responsabilité t'en soit systématiquement imputée. . .

A vous quatre : merci de tous ces moments passés et à venir, de votre soutien, de votre amitié !

# Contents

<b>Scientific production</b>	<b>ix</b>
<b>Glossary</b>	<b>xi</b>
<b>List of Figures</b>	<b>xiii</b>
<b>List of Tables</b>	<b>xv</b>
<b>1 Context</b>	<b>1</b>
1.1 Clinical trials . . . . .	1
1.2 Bridging studies . . . . .	2
1.3 Dose-finding in oncology . . . . .	3
1.4 Thesis objectives . . . . .	5
<b>2 Incorporation and quantification of prior information in Bayesian statistics</b>	<b>9</b>
2.1 Introduction . . . . .	9
2.2 Prior distribution's choice . . . . .	10
2.2.1 Conjugate priors . . . . .	10
2.2.2 Maximum entropy priors . . . . .	11
2.2.3 Non-informative and weakly informative prior distributions . . . . .	12
2.3 Practical use of historical information in a Bayesian setting . . . . .	12
2.3.1 The power prior . . . . .	12
2.3.2 The commensurate priors . . . . .	15
2.3.3 The robust meta-analytic-predictive prior (MAP prior) . . . . .	16
2.4 Quantity of information of a prior distribution . . . . .	16
2.4.1 Variants of the effective sample size . . . . .	19
2.5 Discussion . . . . .	21

<b>3</b>	<b>Dose-finding and bridging studies: state of the art</b>	<b>23</b>
3.1	Dose-finding . . . . .	23
3.1.1	The continual reassessment method (CRM) . . . . .	25
3.1.2	Bayesian logistic regression model (BLRM) . . . . .	28
3.2	Bridging studies designs for dose-finding trials . . . . .	29
3.3	Similarity criteria between curves . . . . .	32
3.3.1	Degree of similarity between distributions: divergences . . . . .	33
3.3.2	Comparison between regression curves in clinical trials . . . . .	34
<b>4</b>	<b>An adaptive power prior for sequential clinical trials – Application to bridging studies</b>	<b>37</b>
<b>5</b>	<b>Estimating similarity of dose-response relationships in phase I clinical trials - case study in bridging data package</b>	<b>63</b>
	<b>Conclusion</b>	<b>81</b>
	<b>Discussion</b>	<b>85</b>
	<b>References</b>	<b>89</b>
	<b>Appendix A Bayesian statistics and decision theory</b>	<b>97</b>
	<b>Appendix B Dose-finding methods</b>	<b>99</b>
B.1	Algorithm-based methods . . . . .	99
B.1.1	The "3+3" algorithm . . . . .	99
B.1.2	The "up and down" method . . . . .	100
B.2	Model-based methods . . . . .	101
B.2.1	The EWOC method (Escalation With Overdose Control) . . . . .	101
B.2.2	Extensions . . . . .	103
B.2.3	Non-parametric approaches to dose-finding . . . . .	104
B.2.4	Dose-finding and pharmacokinetics . . . . .	105
B.2.5	Bivariate outcomes: efficacy and toxicity . . . . .	106
	<b>Appendix C Additional considerations about the ESS</b>	<b>109</b>
C.1	Alternative computational approach for the Effective Sample Size . . . . .	109
C.1.1	Bayesian central limit theorem . . . . .	109
C.1.2	Non-asymptotic view for the variance: the natural exponential family	110
C.1.3	Alternative variance approach . . . . .	110

---

C.2 ESS for the power prior . . . . .	113
<b>Summary in French</b>	<b>115</b>





# Scientific production

## Published articles

- Adrien Ollier, Satoshi Morita, Moreno Ursino, Sarah Zohar. An adaptive power prior for sequential clinical trials—Application to bridging studies. *Statistical methods in medical research*, 2020, vol. 29, no 8, p. 2282-2294, <https://doi.org/10.1177/0962280219886609>
- Adrien Ollier , Sarah Zohar, Satoshi Morita, Moreno 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*. 2021; 18(4):1639. <https://doi.org/10.3390/ijerph18041639>

## Oral communications

- Adaptive power prior for sequential clinical trials – applications to bridging studies; International Society for Biopharmaceutical Statistics (ISBS) 2019, Kyoto, Japan
- Bridging information across populations in early phase clinical trials in oncology, International Society for Clinical Biostatistics (ISCB) 2018, Melbourne, Australia
- Bridging information across populations in early phase clinical trials in oncology, Cancer Immunology and Immunopathology Department Day, 2017, Paris, France



# Glossary

- ADME: absorption, distribution, metabolism and excretion
- AUC: area under the curve
- BCRM: bridging continual reassessment method
- BLRM: Bayesian logistic regression model
- CIHM: conditionally independent hierarchical models
- CRM: continual reassessment method
- ECSS: effective current sample size
- ESS: effective sample size
- $ESS_{elir}$ : expected local-information-ratio effective sample size
- EWOC: escalation with overdose control
- ICH: international council for harmonisation of technical requirements for pharmaceuticals for human use
- LCPP: location commensurate power prior
- MAP: robust meta-analytic-predictive prior
- MPP: modified power prior
- MTD: maximum tolerated dose
- PK: pharmacokinetics
- PP: power prior
- TITE-CRM: time to event continual reassessment method



# List of Figures

- 1.1 Example of dose-toxicity relationship. The threshold of probability of toxicity, indicated by a red line, is 20%. The black curve represent the dose-toxicity relationship and each blue point represent the probability of toxicity associated to a dose level in the panel. The MTD is the third dose, since its probability of toxicity is equal to the target probability. . . . . 5
- 1.2 Example of dose-toxicity relationship where the MTD cannot be determined. The threshold of probability of toxicity is 20%, and all doses have a probability of toxicity higher than 20% . . . . . 6
  
- 3.1 An example of dose-toxicity relationship . . . . . 24
- 3.2 Examples of  $\psi$  functions for logistic and hyperbolic tangent models . . . . . 27
- 3.3 Example of three skeletons obtained with the BCRM method. The black curve represents the dose-toxicity relationship estimates based on the landmark trial. . . . . 32
  
- B.1 The "3+3" dose-finding algorithm . . . . . 100
- B.2 Figure inspired from article [45], about a generalisation of the "3+3" algorithm 101



# List of Tables

2.1	Examples of conjugate priors . . . . .	11
2.2	Examples of ESS for some usual models . . . . .	19
2.3	Quantities involved in CIHM' ESS. The prior of interest is the one for which the ESS should be calculated. Here, $\pi_{1,0}(\boldsymbol{\theta}/\tilde{\boldsymbol{\theta}})$ is an $\epsilon$ -information prior for $\pi_1(\boldsymbol{\theta}/\tilde{\boldsymbol{\theta}})$ and $f_1(\mathcal{Y}_M \tilde{\boldsymbol{\theta}})$ the marginal likelihood. . . . .	20
3.1	Examples of $f$ -divergences . . . . .	33





# Chapter 1

## Context

### 1.1 Clinical trials

Before the market approval, a drug has to pass a series of stages, preclinical and clinical ones. The purpose of clinical trials is to ensure the efficacy and safety of the patients to whom the drug to be marketed may be prescribed. Clinical trials are a very controlled practice: in the aftermath of the Second World War, the idea of carrying out experiments on human beings took on a very different dimension from what it had been before. Ethical issues took a decisive importance in the regulations that followed, notably on the consent of patients in full knowledge of the facts and of the risks incurred during a clinical trial. The french National Agency for the Safety of Medicines and Health Products (ANSM) defines a clinical trial as an: "organised and conducted biomedical research on humans with a view to develop biological or medical knowledge. The purpose of clinical trials on drugs is, depending on the case, to establish or verify certain pharmacokinetic (modalities of absorption, distribution, metabolism and excretion of the drug), pharmacodynamic (the mechanism of action of the drug) and therapeutic (efficacy and tolerance) data for a new drug or a new way of using a known treatment. The trial may be conducted on a healthy or sick volunteer". Clinical trials are conducted in several phases:

**Pre-clinical phase** The molecule is first tested *in vitro*. This permits to study its structure and potential effects on organs for example. Then the molecule is tested on animals *in vivo* (mouses for instance), to test its efficacy and its effects on living beings.

**Phase I** The drug (or treatment) is tested on humans being for the first time on a small group of healthy volunteers, except in oncology and pediatrics, where the drug is directly tested on patients. During this phase, the toxicity of the drug

is evaluated. This phase corresponds, among other things, to the research of the best dose to administer to the patients in subsequent trials. Such a dose could be the maximum tolerated dose as in oncology or a recommended dose according to some criteria. Moreover, this phase permits to collect efficacy, pharmacokinetic and pharmacodynamic data. Pharmacokinetics focuses on how the body treats the drug through four steps: administration, distribution, metabolism and excretion of the drug, while pharmacodynamic focuses on the drug's effects on the body.

**Phase II** The drug is tested on ill patients in order to confirm or evaluate its efficacy while toxicity is still recorded.

**Phase III** The drug is tested on large group of patients. This phase, also called confirmatory trial, is a randomised one. In order to establish its efficacy, its effect is compared either to a reference treatment or to a *placebo*. This phase is the last one that involves tests on humans, if the effectiveness is established, the drug can be marketed.

**Phase IV** This phase involves watching the drug once it is on the market, to control its related adverse events that would not have been detected during the previous phases, especially for long-term effects. It also permits to watch if there are potential commorbidities that might have gone undetected since the enrolled patients respond to very specific criteria.

Clinical trials are all carried out with general recommendations and in accordance with local legislations, such as the European medicines agency (EMA) for the European Union, to which each country adds its own laws (in France, the ANSM controls the conduct of clinical trials), or the food and drug administration (FDA) in the United States. These legislations are systematically dedicated to ensure the proper conduct of clinical trials, in order to guarantee patients' safety.

## 1.2 Bridging studies

The International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use (ICH) defines a bridging study as an additional study conducted in a new population to provide pharmacodynamic or clinical data on efficacy, safety, dosage and dose regimen in that new region that will possibly allow an extrapolation of clinical

data accrued in a previous population to the new one.

The term "population" is used here to distinguish two ethnic groups. It could also be applied to adult-to-child studies, but they are often called extrapolation studies. In this manuscript, we will focus on studies that make links between different ethnic groups.

These bridging studies are necessary in clinical research: it is observed that the effects of the same drug at the same dose does not have the same effects in different regions of the world between different ethnic groups, and not only for cancer. Actually, ethnic differences mechanically imply genetic and metabolic differences. We know, for example, that the proportion of alleles of a gene could vary from one ethnic group to another. A consequence of these differences is a different reaction to some molecules. A simple example is the ADH genes, whose allele proportions vary according to ethnicity, and plays a part in the metabolism of ethanol.

The case of the influence of ethnicity in the response to drug treatment has been highlighted in several ways, through differences in toxicity [39] or metabolism [93], and efficacy. Moreover, inferring toxicity ([39]) basing on pharmacokinetic data only is also noted. The issue of the transposition or validity of clinical data (e.g. efficacy and safety) from one ethnic group to another is addressed in the ICH E5 recommendations from both an ethical and a regulatory perspective [35]. The technical aspect of bridging studies is specifically addressed in specialised articles according to the pathology in question. Please refer to the book [47] for an overview of the issue.

In oncology clinical trials, significant differences in MTD between Caucasian and Japanese populations have already been remarked by Maeda et al. [50] and Ogura et al.[64]. In the context of our work, we highlight an essential aspect of these bridging studies: previous information is available. So, if a drug is tested *again* in a new population, there is a trial in an initial population, with data on toxicity, efficacy, pharmacokinetics, etc., and the data are available before starting a trial in the new population. The natural question is: how can this pre-existing information be best used *statistically*?

### 1.3 Dose-finding in oncology

Phase I are at up most important. On this phase, which must determine the maximum tolerated dose or recommended dose according to certain criteria, will depend the possible success of subsequent phases. Dose-finding studies are performed at this stage. The particular difficulty for this phase is the few number of patients involved: up to fifty patients in general, and no more than thirty most of the time. This small sample size

raises several statistical issues.

Phase I trials in oncology respond to very specific criteria. First, because of the high toxicity of anti-cancer drugs, it does not seem reasonable to test these drugs on healthy subjects. In addition, there is an ethical reason to test these drugs directly on subjects who are already ill: subjects included in an oncology clinical trial are those for whom existing drugs have been shown to be inefficient and their inclusion in a Phase I trial could save their life.

A fundamental assumption is often accepted in oncology for cytotoxic drugs: the higher is the dose given to the patient, the higher is the probability of efficacy. Nevertheless, the higher is the dose given, the higher is the probability of toxicity. The dilemma is therefore to find a dose high enough to increase the probability of efficacy but low enough to be considered safe. Dose-finding is the step which consists in finding the maximum tolerated dose (MTD), that is, the maximum dose that can be administered to patients according to a previously established probability of toxicity threshold. Doctors and clinicians define dose-limiting toxicities (DLT), then the MTD will be the dose whose probability of DLT equal to the established probability of toxicity threshold (e.g. 20%, see Figure 1.1). A dose higher than the MTD has a probability of toxicity higher than the threshold. This threshold is also called the "target toxicity", since we want to find the highest dose whose probability of toxicity equal to this threshold. To do so, a discrete panel of doses is used. In the case where none of the estimated probabilities of toxicity associated to each dose of the panel is equal to the target toxicity, the estimate of the MTD is the highest dose with probability of toxicity as close as possible to the threshold.

Ideally, most of the patients enrolled in the trial should be treated with a dose as close as possible to the MTD. There are several possible dose allocation methods, which can be random or sequential, but the sequential aspect of certain methods has the advantage of gradually administering the best possible dose. However, a rigorous methodology in the conduct of the clinical trial is not a guarantee of success. For instance, a comparative study of several dose finding methods [32] shows that the MTD is not always selected to a large extent, and moreover shows that the reference method ("3+3") gives very poor results. A study which compare the MTD in several dose-finding trial showed that this MTD is not always reached [50], or because sometimes the trial had to be stopped, or because the dose-response relationship is not monotonous, or for any other reason. Several explanations can be provided to explain this phenomenon. Since dose-finding trials involve few subjects, the high inter-subject variability in the response to the drug

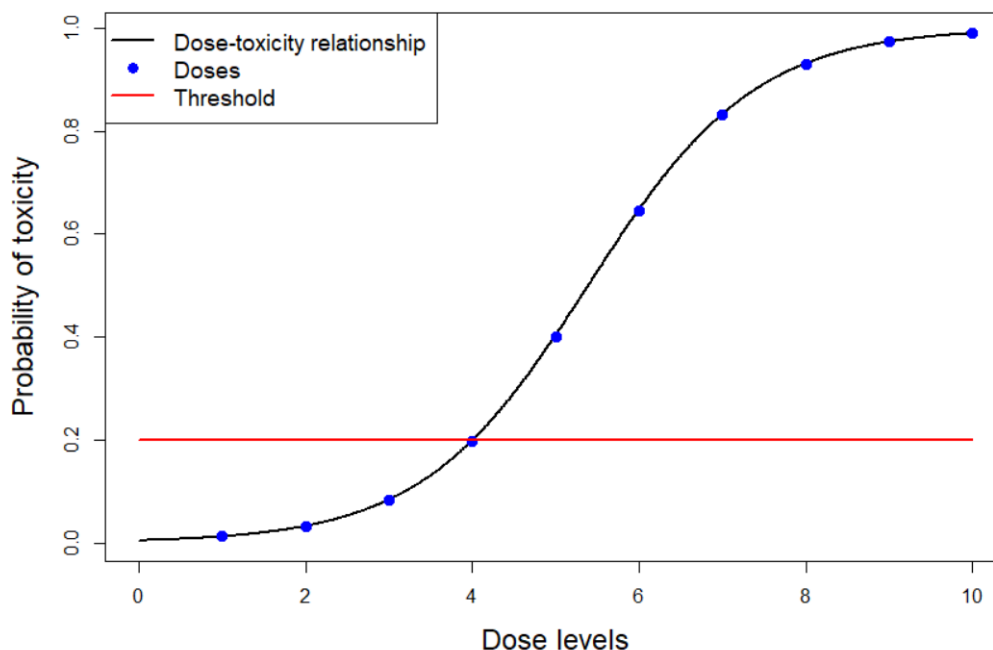


Fig. 1.1 Example of dose-toxicity relationship. The threshold of probability of toxicity, indicated by a red line, is 20%. The black curve represent the dose-toxicity relationship and each blue point represent the probability of toxicity associated to a dose level in the panel. The MTD is the third dose, since its probability of toxicity is equal to the target probability.

makes the estimate of the dose-response relationship difficult. Furthermore, it can be assumed that the dose panel (that is the set of doses tested on the patients) is not correctly selected, and this does not allow the identification of the MTD. For example, if the lowest dose of the panel has a probability of toxicity of thirty percent but the target probability is twenty percent, the MTD will not be found in such a dose panel (figure 1.2).

## 1.4 Thesis objectives

In bridging studies setting, before starting a clinical trial in a new population, it is important to notice that there is available information that can be used for the clinical trial. Taking into account historical data in adaptive trials showed benefits in general, as pointed by Harun et al.[28]. Moreover, in the context of clinical trials with small sample size, due to the high variability in dose-response relationship, it is desirable

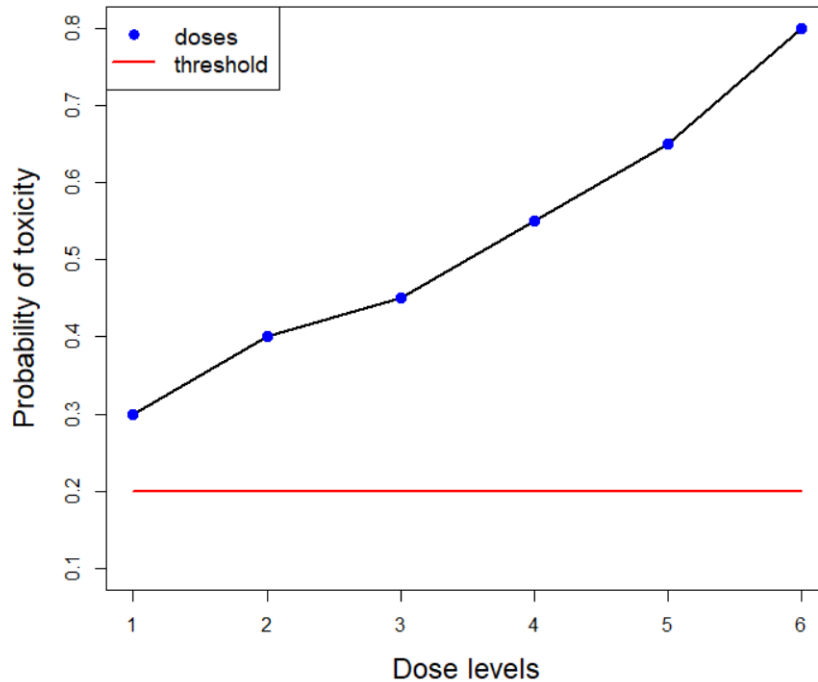


Fig. 1.2 Example of dose-toxicity relationship where the MTD cannot be determined. The threshold of probability of toxicity is 20%, and all doses have a probability of toxicity higher than 20%

to take previous available information into account (as we will explain later, chapter 2.1). Different methods can be proposed to incorporate this information through the prior distribution in Bayesian setting (these methods will be presented in the chapter 2). However, the possibility that the two datasets may be in conflict is not impossible (the so called "data-prior conflict" in Bayesian setting), and the amount of information in the prior distribution must be carefully weighted.

This raises the issue of an adapted methodology to take this information into account before, during and at the end of the new clinical trial. The Bayesian approach naturally appears to be consistent with incorporating historical datasets. So, in a prospective setting, we want to define an adapted Bayesian method to take into account some previous information, which permits to better estimate the dose-response relationship, to finally permit a better identification of the MTD. However, since the previous and current datasets can be different, it is highly desirable to be able to detect a data-prior conflict that could lead to biased estimation of the current dose-response relationship. The natural question is, then, which method can be used to incorporate the previous information? How to weight this information in the prior distribution? Is there a way to

calibrate this amount of information taking into account the similarity between previous and current data to avoid data-prior conflict?

Otherwise, bridging studies should help regulatory authorities to decide if an extrapolation from a population to another is possible. Therefore it can be interesting and important to retrospectively be able to quantify this difference according to different criteria.

The general issue is, in both prospective or retrospective setting, an objective quantification of two dose-response relationships similarity through a Bayesian approach.

However, an inherent difficulty of Bayesian Statistics is the choice of the prior distribution. This choice can be more or less subjective and strongly influences the estimate of the parameter of interest. For instance, let's consider the extreme case of a discrete distribution for which the prior distribution would be a Dirac in one point  $\theta_0$ . In such a case, whatever the data could be, the posterior distribution always gives the point  $\theta_0$ . Otherwise, a weakly informative prior distribution (with a very low variance for example) has a lower influence on the estimate. However, we must keep in mind an important property of Bayesian statistics that many prior distributions allow: in the continuous case, Bernstein Von Mises theorem ensures an asymptotic convergence of the posterior distribution to a normal distribution centered on the maximum likelihood estimator, under certain conditions. More generally, this theorem ensures that both the frequentist and Bayesian methods have *de facto* asymptotically similar results.

Our issue is that we do not work in an asymptotic setting but, on the contrary, in the context of small sample size: in such a context, the use of asymptotic results is excluded. Our objectives are thus, in the context of dose-finding clinical trial, to construct some Bayesian methods to properly weight and incorporate historical information for a current trial, and to define some similarity criteria between several datasets to assess the similarity between several dose-response curves for finished trials.





# Chapter 2

## Incorporation and quantification of prior information in Bayesian statistics

### 2.1 Introduction

Let  $x_1, \dots, x_n$  denote a sample from  $n$  independent and identically distributed random variables  $X_1, \dots, X_n$  from an assumed parametric density  $f(x|\theta)$ . To estimate the parameter  $\theta$ , frequentist statistics use the likelihood function  $\mathcal{L} = \prod_{i=1}^n f(x_i|\theta)$ . In Bayesian setting, the parameter  $\theta$  we try to estimate can be considered like a realisation of a random variable with density  $\pi_0$  with support  $\Theta$ , and that being given data realisations  $\mathbf{x} = (x_1, \dots, x_n)$  of  $X_1, \dots, X_n$ , the conditional density of  $\theta$  according to these realisations would be, by Bayes' theorem:

$$\pi(\theta|\mathbf{x}) = \frac{L(\theta|\mathbf{x})\pi_0(\theta)}{\int_{\Theta} L(\theta|\mathbf{x})\pi_0(\theta)d\theta}$$

$\pi(\theta|\mathbf{x})$  is called the posterior distribution of  $\theta$ . The Bayes' theorem can be applied sequentially: at each new observation, the posterior distribution is "updated", and instead of working on the unknown  $\theta$  density, we work on a conditional density based on a prior density which represents some knowledge about  $\theta$ . However, the choice of this prior distribution could influence the posterior distribution.

In the approach referring to the decision theory, we notice that the posterior distribution is the conditional density minimising the Bayesian risk associated with the  $\pi_0$  density, this point is detailed in the appendix [A](#).

Bayesian approach permits to take into account some available information via the prior distribution. There are many possible sources of information: a previous trial, expert opinions like presented by Boulet et al. [7], and also pre-clinical data ([98],[99]). In addition, information is also created during the trial itself, such as pharmacokinetic data, efficacy data, toxicity data, etc. For each of these situations, the Bayesian paradigm permits to incorporate external information. This particularity is precious in phase I clinical trials context, where due to the small sample size, frequentist statistics give imprecise results (for example in terms of confident intervals).

Our work is set in the context of bridging studies between different ethnic groups; therefore there will be previous available information. The question that arises in our case is: how to incorporate in practice the information from a previous trial into the prior distribution? Regarding prior distributions, there is a dilemma between the need for these distributions to be both "informative" but not to "overcome" the likelihood data. For example, a prior distribution with a high variance does not attenuate the data effect in the posterior distribution. However, it does not give much information about the parameter, except that with a very high variance, the prior distribution gives as small probability measure to an interval  $[a, b]$  as it does to  $[a + n, b + n]$ . On the opposite, a distribution with a very low variance will require more data to "correct" the posterior distribution if the prior distribution is badly calibrated (for example a normal prior distribution with a very low variance but a mean far from what the data one).

## 2.2 Prior distribution's choice

The choice of the prior distribution is a delicate question. It is important to understand that prior distributions are highly dependent on the parameter nature (see for more details Gelman et al. [24]). For example, for a variance parameter, truncated normal, inverse-gamma and half Cauchy distributions are usually chosen as prior distribution. We briefly present some usual prior distributions classification below.

### 2.2.1 Conjugate priors

**Definition:** A family  $\mathcal{F}$  of probability distributions on  $\Theta$  is said to be conjugate for a likelihood function  $f(x|\theta)$  if for every  $\pi \in \mathcal{F}$ , the posterior distribution  $\pi(\theta|x)$  is also in  $\mathcal{F}$ .

This approach has several advantages. Firstly, as indicated by Robert [74], conjugate

$f(x \theta)$	$\pi(\theta)$	$\pi(\theta x)$
$\mathcal{N}(\theta, \sigma^2)$	$\mathcal{N}(\mu, \tau^2)$	$\mathcal{N}(\frac{\sigma^2\mu + \tau^2x}{\sigma^2 + \tau^2}, \frac{\sigma^2\tau^2}{\sigma^2 + \tau^2})$
$\mathcal{N}(\mu, \frac{1}{\theta})$	$\mathcal{Gam}(\alpha, \beta)$	$\mathcal{Gam}(\alpha + 0.5, \beta + (\mu - x)^2/2)$
$\mathcal{Gam}(\nu, \theta)$	$\mathcal{Gam}(\alpha, \beta)$	$\mathcal{Gam}(\alpha + \nu, \beta + x)$
$\mathcal{Bin}(n, \theta)$	$\mathcal{Beta}(\alpha, \beta)$	$\mathcal{Beta}(\alpha + x, \beta + n - x)$
$\mathcal{Mult}(\theta_1, \dots, \theta_k)$	$\mathcal{Dir}(\alpha_1, \dots, \alpha_k)$	$\mathcal{Dir}(\alpha_1 + x_1, \dots, \alpha_k + x_k)$

Table 2.1 Examples of conjugate priors

priors have a structural coherence: since  $\pi(\theta|x)$  is computed from  $\pi(\theta)$  and  $x$ , it is natural that both distributions are in the same "space". Secondly, this approach gives an explicit formula of the posterior distribution, formula which is not obvious or explicit in general. Some examples of conjugate prior distributions with likelihood functions from the exponential family are given in table 2.1.

### 2.2.2 Maximum entropy priors

When some information about the prior distribution  $\pi$  are known or required as constraints (on the moments for example),

$$\mathbb{E}_\pi[g_k(\theta)] = \omega_k, \quad k = 1, \dots, K \quad (2.2.1)$$

a prior distribution can be chosen respecting those constraints and having a maximum entropy under those constraints: this prior is called maximum entropy prior, developed by Jaynes [38]. In the discrete case, the maximum entropy prior is

$$\pi(\theta_i) = \frac{\exp\left(\sum_{k=1}^K \lambda_k g_k(\theta_i)\right)}{\sum_i \exp\left(\sum_{k=1}^K \lambda_k g_k(\theta_i)\right)}$$

where the  $\lambda_k$  are Lagrange multipliers associated to equation 2.2.1. In the continuous case, the choice of a reference distribution of probability  $\pi_0$  is required and the entropy is substituted to the relative entropy, which is called the Kullback-Leibler's divergence between  $\pi$  and  $\pi_0$ . We therefore obtain the prior distribution:

$$\pi(\theta) = \frac{\exp\left(\sum_{k=1}^K \lambda_k g_k(\theta_i)\right) \pi_0(\theta)}{\int_{\Theta} \exp\left(\sum_{k=1}^K \lambda_k g_k(u)\right) \pi_0(u) du}$$

### 2.2.3 Non-informative and weakly informative prior distributions

The notion of non-informative prior distribution is often used but poorly defined. Although the aim of our work is to propose a method to incorporate information in the prior distribution, we can cite the non-informative prior distributions the Laplace prior, the Jeffreys prior, the matching priors and even the reference prior (see Robert [74]). However, there is no method to determine which one of the previous priors would be more informative than others, as explained by Robert [74]. In addition, several methods exist to quantify the "information" in a prior distribution, one of these methods will be precisely explained in chapter 2.4.

A weakly informative prior distribution is a prior with few information. They are used to stabilise the inference on the parameter, introducing some "scale" information about the parameter.

## 2.3 Practical use of historical information in a Bayesian setting

In this section we focus on prior distributions built to incorporate a likelihood into the prior distribution.

### 2.3.1 The power prior

When individual data are available, one possibility is to use the power prior, introduced by Ibrahim and Chen [33]. This approach offers various applications, notably in biostatistics. This prior distribution have already been used for logistic regressions ([11] for example). The idea is the following one: given an "historical" dataset  $D_0$  and a current dataset  $D$  with a common parameter  $\theta$ , and thus a common "model" since they have the same likelihood function, the power prior take into account the historical data incorporating the historical likelihood and weighting it through a coefficient  $\alpha_0 \in [0, 1]$ :

$$\pi^{PP}(\theta|D_0, \alpha_0) = \frac{L(\theta|D_0)^{\alpha_0} \pi_0(\theta)}{\int_{\Theta} L(\theta|D_0)^{\alpha_0} \pi_0(\theta) d\theta} \quad (2.3.1)$$

where  $L(\theta|D_0)$  is the historical likelihood and  $\pi_0$  an initial prior on  $\Theta$ , usually non-informative (see Duan et al. [19] for instance) or even improper (see Ibrahim et al. [34]). If  $\alpha_0 = 0$ , the prior distribution is only the initial prior  $\pi_0$  and the historical information

is not taken into account. On the contrary, if  $\alpha_0 = 1$ , the prior distribution 2.3.1 coincides with the posterior distribution of  $\theta$  given data  $D_0$ , and the historical data are fully taken into account in the prior distribution. As an illustration, we consider the case where the historical data  $D_0$  are independent identically distributed Gaussian observations  $y_{0i} \sim \mathcal{N}(\theta, \sigma^2)$ ,  $i = 1, \dots, n_0$ . The power prior is

$$\begin{aligned} \pi^{PP}(\theta|D_0, \alpha_0) &= \frac{L(\theta|D_0)^{\alpha_0} \pi_0(\theta)}{\int_{\Theta} L(\theta|D_0)^{\alpha_0} \pi_0(\theta) d\theta} \\ &\propto \frac{1}{(2\sigma\pi)^{\frac{\alpha_0 n_0}{2}}} \exp\left(-\frac{\alpha_0 \sum_{i=1}^{n_0} (y_{0i} - \theta)^2}{2\sigma^2}\right) \pi_0(\theta) \\ &\propto \exp\left(-\frac{1}{2} \frac{\sum_{i=1}^{n_0} (y_{0i} - \theta)^2}{\frac{\sigma^2}{\alpha_0}}\right) \pi_0(\theta), \text{ for } \alpha_0 \neq 0 \end{aligned}$$

In this case, we observe that the parameter  $\alpha_0$  directly influences the variance: the more the parameter is low, the more the variance is high and the less historical data are taken into account, and vice versa.

### 2.3.1.1 Variants

An issue raised by the power prior's approach is the choice of the parameter  $\alpha_0$ . The two possibilities are to choose a fixed  $\alpha_0$  or to estimate it. To estimate this parameter, a "fully Bayesian" method can be considered, choosing a prior distribution for  $\alpha_0$  and then estimate it through its posterior distribution. With this approach, the joint density of  $(\theta, \alpha_0)$  is

$$\pi^{PP}(\theta, \alpha_0|D_0) = \frac{L(\theta|D_0)^{\alpha_0} \pi_0(\theta) \pi(\alpha_0)}{\int_0^1 \int_{\Theta} L(\theta|D_0)^{\alpha_0} \pi_0(\theta) \pi(\alpha_0) d\theta d\alpha_0} \quad (2.3.2)$$

$$\propto L(\theta|D_0)^{\alpha_0} \pi_0(\theta) \pi(\alpha_0) \quad (2.3.3)$$

However, as explained by Neuenschwander et al. [59], this approach has several drawbacks. Firstly, it does not allow to consider the conditional distribution of  $\theta$  given  $\alpha_0$ . Secondly, it has a severe computational difficulty: the  $\alpha_0$  estimate from this method tends to be very low, or even equal to zero. Finally, it is in contradiction with the likelihood principle (as underlined by Duan et al. [19] for instance): multiplying the likelihood by a constant  $\delta$  modify the joint distribution, and so the posterior distribution:

$$\pi^{PP}(\theta, \alpha_0|D_0) = \frac{(\delta L(\theta|D_0))^{\alpha_0} \pi_0(\theta) \pi(\alpha_0)}{\int_0^1 \int_{\Theta} (\delta L(\theta|D_0))^{\alpha_0} \pi_0(\theta) \pi(\alpha_0) d\theta d\alpha_0} \propto \delta^{\alpha_0} L(\theta|D_0)^{\alpha_0} \pi_0(\theta) \pi(\alpha_0)$$

In this context and to overcome this issue, the *modified power prior* (MPP) was proposed.

### The modified power prior

A possibility to obtain a conditional distribution of  $\theta$  given  $\alpha_0$  is to "normalise" the power prior "part" and then to add a prior distribution on  $\alpha_0$ , this results in the modified power prior:

$$\pi^{MPP}(\theta, \alpha_0 | D_0) = \frac{L(\theta | D_0)^{\alpha_0} \pi_0(\theta)}{\int_{\Theta} L(\theta | D_0)^{\alpha_0} \pi_0(\theta) d\theta} \pi(\alpha_0) \quad (2.3.4)$$

$$= f(\theta | \alpha_0) \pi(\alpha_0) \quad (2.3.5)$$

In equation 2.3.4, we have a conditional distribution of  $\theta$  given  $\alpha_0$ :  $f(\cdot | \alpha_0)$ , as intended. In addition, the likelihood principle is respected in this version.

However, as pointed out by Hobbs et al. [29], this approach does not permit to directly parametrise the degree of similarity between both datasets  $D$  and  $D_0$ . Then, the resulting estimate of  $\alpha_0$  also tends to under evaluate the impact of historical data  $D_0$ , which should be offset by a too informative prior distribution for the parameter  $\alpha_0$  (as a beta distribution  $Beta(5, 1)$  for instance, giving strong weight to  $\alpha_0$ 's values near to 1) to compensate.

The difficulty is to determine the degree of similarity between  $D$  and  $D_0$  to properly calibrate the parameter  $\alpha_0$  as a function of such a degree, which is still not evaluated yet. In this context the *commensurate priors* were introduced by Hobbs et al. [29], and in particular the *location commensurate power prior* (LCPP).

The use of power prior was also investigated to control type I error in the setting of dynamic borrowing in clinical trials, notably by Nikolakopoulos et al. [63] through the concept of prior-data conflict calibrated power prior based on the prior predictive  $p$ -values.

The calibrated power prior proposed by Yuan et al. [96], introduced in the setting of biosimilar clinical trials, also demonstrated good performances to control the type I error. It is also possible to use the power prior for non-inferiority trials, as showed by Liu [46].

### 2.3.2 The commensurate priors

These prior distributions were first presented by Hobbs et al. [29], and are defined as follows:

$$\pi(\theta|D_0, \theta_0, \tau) \propto L(\theta_0|D_0)\pi(\theta|\theta_0, \tau)\pi_0(\theta)$$

with the conditions:

$$\begin{aligned} \pi(\theta|D_0, \theta_0, \tau) &\xrightarrow{\tau \rightarrow 0} \pi_0(\theta) \\ \pi(\theta|D_0, \theta_0, \tau) &\xrightarrow{\tau \rightarrow +\infty} L(\theta|D_0)\pi_0(\theta) \end{aligned}$$

where  $\tau$  being the prior distribution precision. Again,  $\tau$  estimation is challenging, as discussed in article [30], where the authors proposed a fully Bayesian and an empirical Bayesian approach for this estimation. Again, a fully Bayesian approach is proposed to estimate  $\tau$ .

Other variants are presented in a Gaussian setting, such as the *location commensurate prior* or *location scale commensurate prior*.

#### The location commensurate power prior

Up to now, we have considered that both  $D$  and  $D_0$  had in common an identical parameter  $\theta$  to estimate. Hobbs et al. [29] assumed henceforth that both datasets have the same structure, that is the same likelihood function, but have different parameters:  $\theta_0$  and  $\theta$  for  $D_0$  and  $D$ , respectively. In this way, the aim is to define a degree of similarity between  $D$  and  $D_0$  through a parameter  $\tau$  which will be incorporated into  $\alpha_0$ 's prior distribution. The example of a Gaussian distribution is considered:

$$\pi^{LCPP}(\theta, \alpha_0, \tau|D_0) \propto \frac{\int L(\theta_0|D_0)^{\alpha_0} N(\theta|\theta_0, 1/\tau) d\theta_0}{\int L(\theta_0|D_0)^{\alpha_0} d\theta_0} \text{Beta}(\alpha_0|g(\tau), 1) \pi(\tau) \quad (2.3.6)$$

$$\propto f(\theta|\alpha_0, \tau) \times \text{Beta}(\alpha_0|g(\tau), 1) \times \pi(\tau) \quad (2.3.7)$$

with  $N(\theta|\theta_0, 1/\tau)$  being the Gaussian distribution for  $\theta$ , with mean  $\theta_0$  and standard deviation  $1/\tau$ ,  $\text{Beta}(\alpha_0|g(\tau), 1)$  being the beta distribution for  $\alpha_0$  with parameters  $g(\tau)$  and 1 with  $g(\cdot)$  a positive monotonous function, close to zero if  $\tau$  is close to zero and large if  $\tau$  is large. Finally,  $f(\cdot|\alpha_0, \tau)$  is the conditional distribution of  $\theta$  given  $\alpha_0$  and  $\tau$ . In the formula 2.3.6, if  $\tau$  is low, which means that the similarity is low, the variance for  $\theta$  is high and vice versa. This hierarchical method indeed permits to guide the choice of  $\alpha_0$  by the criterion  $\tau$ .



### 2.3.3 The robust meta-analytic-predictive prior (MAP prior)

In randomised clinical trials where a control group and one or more treatment groups are considered, the incorporation of historical control trial information was proposed by Schmidli et al. [76]. We consider a new trial with  $X_*$  and parameter  $\phi_*$  for the treatment while the control data are  $Y_*$  with parameter  $\psi_*$ . Several historical control datasets are considered:  $Y_{\mathcal{H}} = (Y_1, \dots, Y_H)$  with their respective parameters  $\psi_{\mathcal{H}} = (\psi_1, \dots, \psi_H)$ . Each historical control data  $Y_h$  contains  $n_h$  patients. The robust MAP is based on the MAP prior, presented by Spiegelhalter et al. [80] and Neuenschwander et al. [60]. Considering  $n_*$  patients in the control group and  $F$  being a distribution, the model is:

$$\begin{aligned} Y_* &\sim F(\psi_*, n_*) \\ Y_h &\sim F(\psi_h, n_h) \\ \theta_*, \theta_1, \dots, \theta_H &\sim \mathcal{N}(\mu, \tau^2) \\ p(\mu, \tau) &= p(\mu)p(\tau) \end{aligned}$$

with  $\theta_i = f(\psi_i)$  being an exchangeable parameter,  $\tau$  the between-trial standard deviation and  $p(\mu)$  and  $p(\tau)$  the prior distributions for the hyper-parameters. With such a model, the MAP is  $p_H(\psi_*) = p(\psi_* | Y_1, \dots, Y_H)$ .

Schmidli et al. [76] focus on a two-steps approach: firstly, when designing the trial, a MAP prior  $p(\psi_* | Y_{\mathcal{H}})$  is constructed using historical trials. Secondly, at the end of the trial, both the MAP and current data  $Y_*$  are used with the Bayes's theorem to perform estimation:  $p(\psi_* | Y_*, Y_{\mathcal{H}}) \propto p(Y_* | \psi_*)p(\psi_* | Y_{\mathcal{H}})$ . There are two key points in the robust MAP construction: (i) the MAP is not analytically available, so the authors propose a process to estimate it through a mixture of conjugate distributions (ii) the MAP robustness is improved by adding a non-informative distribution into the mixture, which permits to deal with a data-prior possible conflict. The robust MAP also take into account the between-trial heterogeneity since the MAP is built through a hierarchical model that clearly considers this point, as explained before.

## 2.4 Quantity of information of a prior distribution

The question of the quantity of information in a prior distribution is also difficult, and several methods may be proposed. For instance: the variance, the entropy or even divergences can be considered. The variance and entropy are already well documented as method to quantify the quantity of information of a prior distribution. It were shown by

the fact that non-informative referred to a prior distribution with large variance and the existence of maximum entropy families.

However, other possibilities offer a way to quantify information of a prior distribution. Typically, a way to deal with information of prior distribution is to use information geometry: the quantity of information is in this context more like a comparison method between several distributions. As an example, the power prior theory can be approached through the Kullback-Leibler divergence (which belongs to f-divergences' family). This approach establishes a distance between two distributions, and as Ibrahim et al. [34] says, the point is that taking one of those distribution accepted as non-informative gives a point of comparison. More technical details about information geometry theory can be found in Amari's book [2]. However, the interpretation of such criterion is not easy, even for the most rudimentary, which is entropy or relative entropy (Kullback-Leibler divergence). The notion of *effective sample size*, presented just after, is an easy notion to interpret, for both statisticians and clinicians. The effective sample size notion here refers to the notion introduced by Morita et al. [53]

### The effective sample size (ESS)

The ESS is a tool which permits to judge to what extent a prior is informative or not. It is defined for a vector parameter  $\boldsymbol{\theta} = (\theta_1, \dots, \theta_d)$ . The construction is the following one. Given a prior  $p(\boldsymbol{\theta})$  and density function  $f(x/\boldsymbol{\theta})$ , we define a vague prior  $q_0(\boldsymbol{\theta})$  (that is a prior with large variance), which gives posterior  $q_m(\boldsymbol{\theta}/\mathbf{Y}_m)$  after  $m$  observations  $\mathbf{Y}_m = (Y_1, \dots, Y_m)$ . With a distance, we will define which  $m$  permits to  $q_m(\boldsymbol{\theta}/\mathbf{Y}_m)$  to be the closest to  $p(\boldsymbol{\theta})$ . If  $m^*$  minimises the distance, it means that  $p(\boldsymbol{\theta})$  is as informative as a vague prior after  $m^*$  observations: the higher  $m^*$  is, the more informative the prior  $p(\boldsymbol{\theta})$  is. This permits to have an easy interpretation of the "quantity of information": it is equivalent to the number of observations in a posterior distribution obtained from a prior distribution with a large variance. So there are two things to determine: the vague prior  $q_0$  and a distance to calculate the ESS. Morita et al.[53] defined  $q_0$  as an  *$\epsilon$ -information prior*, that is a prior distribution which satisfies the conditions:

$$\left\{ \begin{array}{l} \mathbb{E}_{q_0}[\boldsymbol{\theta}] = \mathbb{E}_p[\boldsymbol{\theta}] \\ \text{Corr}_{q_0}(\theta_j, \theta_{j'}) = \text{Corr}_p(\theta_j, \theta_{j'}), \quad j \neq j' \\ \text{Var}_{q_0}(\theta_j) \gg \text{Var}_p(\theta_j) \end{array} \right.$$

The two first conditions permit to keep the same mean for the two distributions  $q_0$  and  $p$  with the same correlation structure, while the third ensure that the  $q_0$  distribution is less

informative than  $p$  in term of variance. We denote  $\tilde{\theta}$  the vector of hyperparameter. With the likelihood function  $f_m(\mathbf{Y}_m/\theta)$  and an  $\epsilon$ -information prior  $q_0(\theta/\tilde{\theta}_0)$ , with notations of Morita et al. the posterior is

$$q_m(\theta/\tilde{\theta}_0, \mathbf{Y}_m) \propto q_0(\theta/\tilde{\theta}_0) f_m(\mathbf{Y}_m/\theta)$$

and the marginal distribution under  $p(\theta/\tilde{\theta})$  is

$$f_m(\mathbf{Y}_m/\tilde{\theta}) = \int f_m(\mathbf{Y}_m/\theta) p(\theta/\tilde{\theta}) d\theta$$

denoted by  $f_m(\mathbf{Y}_m)$  if  $\tilde{\theta}$  is fixed. Then we consider that either  $d = 1$  and  $p(\theta/\tilde{\theta})$  is an univariate parametric distribution, or  $d \geq 2$  and  $p(\theta/\tilde{\theta})$  is a multiparametric distribution. Then, we denote  $\bar{\theta} = \mathbb{E}_p[\theta]$ , and define

$$\begin{cases} D_{p,j}(\theta) = -\frac{\partial^2 \log\{p(\theta/\tilde{\theta})\}}{\partial \theta_j^2} \\ D_{q,j}(m, \theta, \mathbf{Y}_m) = -\frac{\partial^2 \log\{q_m(\theta/\tilde{\theta}, \mathbf{Y}_m)\}}{\partial \theta_j^2} \end{cases} \quad (2.4.1)$$

for  $j = 1 \dots d$ .

We also denote

$$\begin{cases} D_{p,+}(\theta) = \sum_{j=1}^d D_{p,j}(\theta) \\ D_{q,+}(m, \theta) = \sum_{j=1}^d \int D_{q,j}(m, \theta, \mathbf{Y}_m) f_m(\mathbf{Y}_m) d\mathbf{Y}_m \end{cases}$$

And we define the distance between  $p(\theta/\tilde{\theta})$  and  $q_m(\theta/\tilde{\theta}_0, \mathbf{Y}_m)$  as the absolute value of the two information matrices' trace's difference:

$$\delta(m, \bar{\theta}, p, q_0) = |D_{p,+}(\bar{\theta}) - D_{q,+}(m, \bar{\theta})|$$

And finally,

$$ESS[p(\theta/\tilde{\theta}), f_m(\mathbf{Y}_m)] = \arg \min_m \{\delta(m, \bar{\theta}, p, q_0)\}$$

Some examples are given in Table 2.2.

As explained by Morita et al. [53], this definition is in many points quite arbitrary. Several ways were investigated for the distance, and the previous one was the best for interpretations and computation of the resulting ESS. Several algorithms are given by Morita et al. [53] to compute the ESS.

Prior $p(\boldsymbol{\theta}/\tilde{\boldsymbol{\theta}})$	Likelihood $f_m(\mathbf{Y}_m/\boldsymbol{\theta})$	ESS
$\mathcal{N}(\tilde{\boldsymbol{\mu}}, \tilde{\boldsymbol{\sigma}}^2)$	$\mathcal{N}(\boldsymbol{\theta}, \boldsymbol{\sigma}^2)$	$\frac{\boldsymbol{\sigma}^2}{\tilde{\boldsymbol{\sigma}}^2}$
$\text{Be}(\tilde{\boldsymbol{\alpha}}, \tilde{\boldsymbol{\beta}})$	$\text{Bin}(n, \boldsymbol{\theta})$	$\tilde{\boldsymbol{\alpha}} + \tilde{\boldsymbol{\beta}}$
$\mathcal{G}(\tilde{\boldsymbol{\alpha}}, \tilde{\boldsymbol{\beta}})$	$\mathcal{E}(\boldsymbol{\theta})$	$\tilde{\boldsymbol{\alpha}}$
$\text{Inv}\chi^2(\tilde{\nu}, \tilde{\boldsymbol{\sigma}}^2)$	$\mathcal{N}(0, \boldsymbol{\sigma}^2)$	$\tilde{\nu}$

Table 2.2 Examples of ESS for some usual models

## Applications of the ESS

The ESS has several applications [53]. First of all, the ESS gives a new formalism to quantify the information incorporated into a prior distribution. In addition, is a powerful tool when designing sensitivity analysis checking several prior information level. Then, it can be used for prior adjustments: if a prior distribution's ESS is higher than the number of patients in a study, then the prior should be revised. This could be particularly interesting for adaptive designs to ensure that data dominate more than the prior distribution. Practical cases are investigated by Morita et al. [54].

### 2.4.1 Variants of the effective sample size

#### The ESS in *conditionally independent hierarchical models* (CIHM)

The ESS has been originally defined in the case of "one level Bayesian model", which means that hyper-parameter were fixed. If it is not the case, then we are in the context of hierarchical models, where hyper-prior must be considered, as explained by Morita et al. [55]. For CIHM, the construction of the ESS is extended by considering  $K$  subgroups, each having  $m$  observations. Then we denote  $\mathbf{Y}_k = (Y_{k,1}, \dots, Y_{k,m})$  for  $k \in \{1, \dots, K\}$  and  $\mathbf{Y}_M$  the  $m \times K$  matrix  $\mathbf{Y}_M = (\mathbf{Y}_1, \dots, \mathbf{Y}_K)$ . Then,  $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_K)$ .

In the case of CIHM, two prior distributions can be considered as the prior of interest for which an ESS should be calculated:

- the marginalised prior:

$$\pi_{12}(\boldsymbol{\theta}/\phi) = \int \pi_1(\boldsymbol{\theta}/\tilde{\boldsymbol{\theta}})\pi_2(\tilde{\boldsymbol{\theta}}/\phi)d\tilde{\boldsymbol{\theta}}$$

- the hyper-prior

$$\pi_2(\tilde{\boldsymbol{\theta}}/\phi)$$

A summary of all quantities used to calculate the ESS of both marginalised prior and the hyper-prior is given in table 2.3. Each of those prior distributions has its specificity

	Marginalised prior	Hyper-prior
Prior of interest	$\pi_{12}(\boldsymbol{\theta}/\boldsymbol{\phi}) = \int \pi_1(\boldsymbol{\theta}/\tilde{\boldsymbol{\theta}})\pi_2(\tilde{\boldsymbol{\theta}}/\boldsymbol{\phi})d\tilde{\boldsymbol{\theta}}$	$\pi_2(\tilde{\boldsymbol{\theta}}/\boldsymbol{\phi})$
$\epsilon$ -information prior	$\pi_{12,0}(\boldsymbol{\theta}/\boldsymbol{\phi}) = \int \pi_{1,0}(\boldsymbol{\theta}/\tilde{\boldsymbol{\theta}})\pi_2(\tilde{\boldsymbol{\theta}}/\boldsymbol{\phi})d\tilde{\boldsymbol{\theta}}$	$\pi_{2,0}(\tilde{\boldsymbol{\theta}}/\boldsymbol{\phi}_0)$
Considered likelihood	$f(\mathcal{Y}_M \boldsymbol{\theta}) = \prod_{k=1}^K f(\mathbf{Y}_k \boldsymbol{\theta}_k)$	$f_1(\mathcal{Y}_M \tilde{\boldsymbol{\theta}})$
Posterior distribution	$\pi_{12,M}(\boldsymbol{\theta}/\boldsymbol{\phi}, \mathcal{Y}_M) \propto \pi_{12,0}(\boldsymbol{\theta}/\boldsymbol{\phi})f(\mathcal{Y}_M/\boldsymbol{\theta})$	$\pi_{2,M}(\tilde{\boldsymbol{\theta}}/\boldsymbol{\phi}_0, \mathcal{Y}_M)$

Table 2.3 Quantities involved in CIHM' ESS. The prior of interest is the one for which the ESS should be calculated. Here,  $\pi_{1,0}(\boldsymbol{\theta}/\tilde{\boldsymbol{\theta}})$  is an  $\epsilon$ -information prior for  $\pi_1(\boldsymbol{\theta}/\tilde{\boldsymbol{\theta}})$  and  $f_1(\mathcal{Y}_M|\tilde{\boldsymbol{\theta}})$  the marginal likelihood.

and related methods, to define the appropriate  $\epsilon$ -information prior and then compute the corresponding ESS. In particular, the final distance doesn't consider the trace of the information matrix anymore but the determinant of the variance-covariance matrix. Several technical aspects and clinical uses are given by Morita et al. [55].

### The expected local-information-ratio ESS ( $ESS_{elir}$ )

An other definition of the ESS is given by Neuenschwander et al. [62]. The authors define a "predictive consistency criterion", which is "for a sample size  $N$ , the expected posterior ESS must be the sum of the prior ESS plus  $N$ ". After showing that other methods to define the ESS do not respect this criterion, the authors define the  $ESS_{elir}$  as

$$ESS_{elir} = \mathbb{E}_{\theta} \left[ \frac{i(p(\theta))}{i_F(\theta)} \right]$$

where

$$i(p(\theta)) = -\frac{d^2 \log p(\theta)}{d\theta^2}$$

$$i_F(\theta) = -\mathbb{E}_{Y_1|\theta} \left[ \frac{d^2 \log f(Y_1|\theta)}{d\theta^2} \right]$$

$i(p(\theta))$  being the information of the prior distribution and  $i_F(\theta)$  the observed Fisher information for one observation  $Y_1$  with likelihood  $f(Y_1|\theta)$ . The authors then show that the  $ESS_{elir}$  satisfies the consistency criterion, illustrating this point with the example of normal data using a student prior distribution. For one-parameter exponential families, ESS and  $ESS_{elir}$  are very close, however, they could significantly differ with some highly informative prior distributions. The  $ESS_{elir}$  is defined only for one dimensional parameter.

### The effective current sample size (ECSS)

The classic ESS is designed to assess the quantity of information of a prior distribution, in a sense as if the prior of interest contains historical data. However, it does not deal with the potential conflict between this prior of interest and the data with which it would be used through the Bayes' theorem. Basically, as explained by Wiesenfarth and Calderazzo [92], two normal priors with same variance have the same ESS, whatever are their respective means. However, their means influence the analysis with a potential bias. Presented by Wiesenfarth and Calderazzo [92], the ECSS is built to take into account a data-prior conflict. This ECSS quantifies the number of observations to be included to the current likelihood for the posterior inference to be equivalent of a reference prior. We denote  $\mathbf{y} = (y_1, \dots, y_k)$  the current hypothetical data with likelihood  $f_k(\mathbf{y}|\theta)$ ,  $\pi(\theta|\theta_\pi, \sigma_\pi^2)$  the prior of interest with mean  $\theta_\pi$  and variance  $\sigma_\pi^2$ , and  $\pi_b(\theta|\theta_b, \sigma_{\pi_b}^2)$  the baseline prior with mean  $\theta_b$  and variance  $\sigma_{\pi_b}^2$ . Those two priors lead respectively to posterior distributions  $\pi(\theta|\mathbf{y})$  and  $\pi_b(\theta|\mathbf{y})$ . The ECSS is then:

$$ECSS = \arg \min_m |D_{MSE}^{\theta_0}\{\pi(\theta|\mathbf{y}_{1:k-m})\} - D_{MSE}^{\theta_0}\{\pi_b(\theta|\mathbf{y}_{1:k})\}|$$

with  $D_{MSE}^{\theta_0}\{\pi(\theta|\mathbf{y})\} = \mathbb{E}_{\mathbf{y}|\theta_0}\{\mathbb{E}_\pi(\theta|\mathbf{y}) - \theta_0\}^2$  and  $\theta_0$  is either known or estimated by  $\widehat{\theta}_0 = \mathbb{E}_{\pi_b}(\theta|\mathbf{y})$

Consequently in some cases, this ECSS could even be negative, which means that the data-prior conflict is strong. Even if the ECSS evaluates the quantity of information of a prior distribution, its interpretation is thus different from the classic ESS one, since the question is not the quantity of information in comparison to a vague prior, but in a reference prior, with a particular interest on data-prior conflict.

## 2.5 Discussion

The the power prior approach and its variants were mainly investigated in the context of Gaussian distribution or generalised linear mixed models. The possibility of incorporating several historical data were also investigated by Ibrahim and Chen [33]. However, the more complex the method (e.g. commensurate power prior or the robust meta-analytic-predictive prior) the more challenging the parameters estimation. In the case of the MAP prior, several approximation stages can be considered to deal with this issue. Then, while MAP deals with the possible heterogeneity between populations, it was not taken into account in the power prior approach.

The different ESS notions presented here have each there own characteristics. The Morita

et al. ESS notion has the advantages of being intuitive, can be applied to multivariate prior distributions and is particularly adapted for exponential families, as pointed out by Neuenschwander et al. [62]. The  $ESS_{elir}$  notion permits to consider a consistency criterion, and can differ from Morita et al [53] for some distributions. It is adapted for univariate prior. Finally, the ECSS is designed to detect a potential data-prior conflict.

# Chapter 3

## Dose-finding and bridging studies: state of the art

### 3.1 Dose-finding

Our work is in the context of dose finding in oncology. In practice, with a limited sample size, in dose finding study we try to estimate the dose-toxicity relationship (see for example Figure 3.1). The dose-response relationship is what allows us to estimate the dose that will be the MTD. In phase I, the primary response outcome is toxicity. The main hypothesis in oncology is that the higher the dose, the more effective the drug is. The issue, however, is that a higher dose increases toxicity for patients.

More precisely, phase I oncology trials are designed to estimate the MTD linked to a pre-established toxicity threshold  $\nu$ . The  $\nu$  parameter, called target toxicity rate or target probability, represents a threshold of probability of toxicity: a dose with a probability of toxicity above this threshold is considered too dangerous for patients. In dose-finding clinical trials, the toxicity notion is defined through dose limiting toxicity (DLT). These DLTs are defined before the beginning of the clinical trial by doctors and clinicians, and often correspond to high toxicity grades (e.g. greater than 3 on a five-grade scale of adverse events: 1 being mild, 2 moderate, 3 severe, 4 life threatening, 5 fatal), with specific criteria. Once these DLTs have been identified, what we call the probability of toxicity is rigorously the probability of the occurrence of any DLT. This target toxicity rate is most often between 17% and 33%.

In practice, even if the dose could be seen as continuous, a panel of  $K$  doses  $d_1, \dots, d_K$  is decided. The MTD is then estimated as the dose which has the closest probability of toxicity near to the target toxicity rate  $\nu$  among the panel of  $K$  doses, keeping in mind that a higher dose involves a higher probability of toxicity.



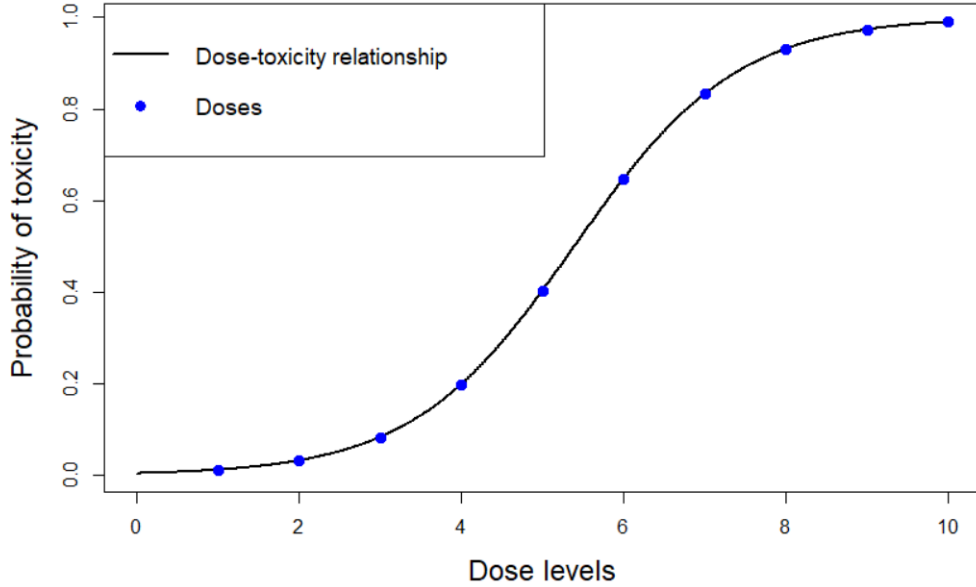


Fig. 3.1 An example of dose-toxicity relationship

We denote  $X_1, \dots, X_n$  the doses administered to the first  $n$  patients, and  $Y_1, \dots, Y_n$  their binary response: for dose  $X_i \in \{d_1, \dots, d_K\}$  given to the  $i^{\text{th}}$  patient,  $Y_i = 0$  if there were no observed DLT and  $Y_i = 1$  if there was an observed DLT. The probability of toxicity for the  $i^{\text{th}}$  patient with a given dose  $d_j$  is  $\mathbb{P}(Y_n = 1 | X_n = d_j)$ , with  $j \in \{1, \dots, K\}$ . In sequential design, for each new patient a new dose  $X_{n+1} \in \{d_1, \dots, d_K\}$  to administer must be found such that  $\mathbb{P}(Y_{n+1} = 1 | X_{n+1})$  be the closest as possible of the target toxicity rate  $\nu$ :  $X_{n+1} = \underset{d_j \in \{d_1, \dots, d_K\}}{\operatorname{argmin}} |\mathbb{P}(Y_{n+1} = 1 | d_j) - \nu|$ .

There are different dose-finding methods, with two main families of methods: non-parametric algorithmic methods and model-based methods. In the first family we find for example the "3+3" algorithm, which was and is still widely used, or "up and down" methods, and in the second family the famous *continual reassessment method* (CRM) and its variants, or the dose escalation method with toxicity control: EWOC (for Escalation With Overdose Control) and BLRM (Bayesian logistic regression model). A brief presentation of these different methods is given in Appendix B, while the continual reassessment method is presented just below. All the presented method are adaptive, which means with possible modifications of the design depending on interim analyses. Those adaptive designs are described more flexible, not only for dose-finding studies, as explain by Pallmann et al. [72].

### 3.1.1 The continual reassessment method (CRM)

The CRM was introduced by O'Quigley, Pepe and Fisher [69]. This method is widely used and is the subject of an extensive literature. In particular, Cheung [15] details the theoretical and practical aspects of CRM. Some extensions of the CRM are presented in Appendix B.

The CRM is a parametric dose-finding method. There exist a frequentist as well as a Bayesian version (the most used one).

The dose-toxicity model is

$$\mathbb{P}(Y_i = 1 | X_i = x_i) = \psi(x_i, \beta)$$

where  $\psi$  depends on the dose  $x_i$  given to the  $i^{\text{th}}$  patient and on the parameter  $\beta$  to estimate. For the  $n$  first observations, the binomial likelihood can be written as:

$$L_n(\beta) = \prod_{i=1}^n \{\psi(x_i, \beta)\}^{y_i} \{1 - \psi(x_i, \beta)\}^{1-y_i} \quad (3.1.1)$$

The general principle of CRM is, after each cohort of patients, to estimate the parameter  $\beta$ , which allows an estimate of the dose-toxicity relationship  $\psi(x_i, \hat{\beta})$ ,  $\hat{\beta}$  is the estimate of  $\beta$  and can be obtained either by the maximum likelihood, or often by the Bayesian approach:

$$\hat{\beta} = \mathbb{E}[Z] \text{ where } Z \sim \pi(\beta | (x_1, y_1), \dots, (x_n, y_n)) = \frac{L_n(\beta)\pi(\beta)}{\int L_n(\beta)\pi(\beta)d\beta}$$

where  $\pi(\beta)$  is the prior distribution of  $\beta$  and  $\pi(\beta | (x_1, y_1), \dots, (x_n, y_n))$  the posterior distribution of  $\beta$ .

We set the  $d_{n+1}$  dose that will be administered to the next patient (or the next cohort of patients) using the  $\hat{\beta}$  estimate, according to the rule:

$$\begin{aligned} d_{n+1} &= \operatorname{argmin}_{d_j \in \{d_1, \dots, d_K\}} |\mathbb{P}(Y_{n+1} = 1 | d_j) - \nu| \\ &= \operatorname{argmin}_{d_j \in \{d_1, \dots, d_K\}} |\psi(d_j, \hat{\beta}) - \nu| \end{aligned}$$

where  $\nu$  is the target toxicity rate.

At this stage, several things must be noticed.

- It is common to add a no-skipping rule: the dose to be given to the next cohort of patients is either already tested or immediately above the maximum tested dose.

In practice, if doses  $d_1$ ,  $d_2$  and  $d_3$  have been tested but the recommended dose is dose  $d_5$ , the dose  $d_4$  will actually be given to the next cohort of patients.

- The doses used in the model  $\mathbb{P}(Y_i = 1|X_i = x_i) = \psi(x_i, \beta)$  are not the real doses administered to the patients, but pseudo-doses (however, we keep the same notation  $x_i \in \{d_1, \dots, d_K\}$  for simplicity). In order to allow the model to be consistent and to work correctly, we choose pseudo-doses (as many as the real doses) which satisfy the equations  $p_{0,j} = \psi(d_j, \bar{\beta})$ ,  $j \in \{1, \dots, K\}$ , where  $\bar{\beta}$  is the mean of the prior distribution of  $\beta$ , and the  $p_{0,j}$  are initial toxicities elicited for each dose level. This set of initial toxicities is sometimes called "skeleton" of the CRM, or "working model" depending on the authors (as in [70]). The issue of the skeleton choice is discussed below.
- The choice of the  $\psi$  function is not unique. The most used functions are:

- The empiric function:

$$\psi(d_j, \beta) = \alpha_j^\beta, \alpha_j \in ]0, 1[, \beta > 0$$

with fixed  $\alpha_j$  for each dose before the beginning of the trial

- The one-parameter logistic function:

$$\psi(d_j, \beta) = \frac{\exp(a + \beta d_j)}{1 + \exp(a + \beta d_j)}$$

where  $a$  is a constant and the dose is in  $\mathbb{R}$ .

- The hyperbolic tangent:

$$\psi(d_j, \beta) = \left( \frac{\tanh(d_j) + 1}{2} \right)^\beta,$$

also for a dose of the skeleton in  $\mathbb{R}$ .

Some illustrations for logistic functions and hyperbolic tangent on a continuous standardised dose are given in figure 3.2.

The previous models have the advantage of having only one parameter to estimate.

### 3.1.1.1 CRM settings

The CRM is a widely used method and its theoretical and operating characteristics have been studied in detail in numerous publications. In the following we briefly discuss the

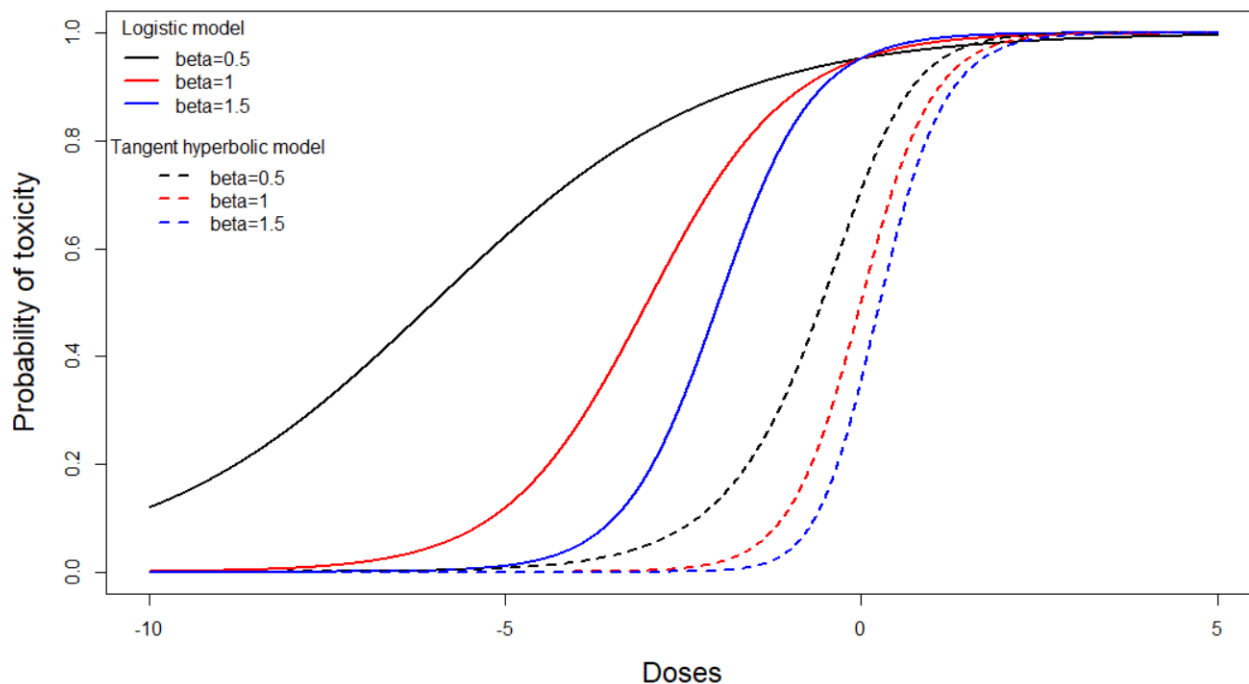


Fig. 3.2 Examples of  $\psi$  functions for logistic and hyperbolic tangent models

main practical setting of the CRM.

- **Doses panel choice**

The problem of the number of doses to be tested is twofold. One can wonder about the number of doses to be tested, but also about the "quantitative" difference between each dose. To determine the panel of doses to be tested, the Fibonacci sequence is often cited. However, the use of this sequence in the construction of the dose panel is rare, as shown in the study [75], while a modified Fibonacci sequence is preferred. For our part, we will highlight only one mechanical element of CRM: the increase in the number of doses in the panel does indeed allow a better precision in the estimation of MTD, but it also induces an increase in the number of patient cohorts, since the number of doses to be tested is higher. This element, underlined by Wheeler et al. [91], is interesting to consider in a clinical context involving a small number of patients.

- **Skeleton choice and robustness**

Clearly, the choice of initial toxicity guesses is central to the internal mechanism of

the CRM. It is therefore interesting and important to understand how the choice of this skeleton influences the results. This issue has been raised by O'Quigley and Zohar [70]. In this article, the authors tested different skeletons on several possible toxicity scenarios. The authors underlined the importance of having an "spaced-out" skeleton: the initial toxicities should not be too close to each other. Typically, with a target toxicity of 0.2, if the skeleton has the initial toxicities 0.19, 0.2 and 0.23 the results will be poor since the model will not be able to identify the MTD in good proportion. Several methods have been proposed for the choice of the skeleton, including the method based on indifference intervals (see [41]) and the Bayesian Model Averaging (see [95]).

- **Prior distribution variance**

If several approaches have been specifically proposed for the choice of the CRM skeleton, relatively few concern the variance of the prior distribution of the parameter to estimate. Considering the empiric model  $\psi(d, \beta) = d^{\exp(\beta)}$  and the one-parameter logistic model  $\psi(d, \beta) = \{1 + \exp(-a - \exp(\beta)d)\}^{-1}$  with a fixed  $a$ , Lee and Cheung [42] proposed in the case of a Gaussian prior distribution a choice of variance based on the hypothesis that each of the  $K$  doses would have the same probability as the others to be the MTD. This discrete uniform distribution leads the authors to suggest a variance equal to  $\frac{K^2-1}{12}$ , that is the variance of the discrete uniform distribution. The authors then propose empirical approaches to simultaneously choose the indifference intervals' window and the variance.

The variance of the prior distribution has in this Bayesian framework a very particular significance, its value influencing on the speed of convergence, in several possible ways: in the case of a low variance with a centred distribution around the "true" value of the parameter, the convergence is very fast, but the convergence will be slow if the variance is low and the distribution centred "far" from the value of the parameter of interest.

In a Bayesian setting with few observations and in which asymptotic convergence theorems can't be used, it is clear that the variance of the prior distribution must be rigorously calibrate, and if possible according to an objective criterion.

### 3.1.2 Bayesian logistic regression model (BLRM)

We have previously presented the general methodology of CRM, based on a logistic regression with one parameter and initial probabilities of toxicity: the skeleton. The BLRM presented by Neuenschwander et al.(2008) [58] consists in a two parameters

logistic regression, with continuous dose:

$$\text{logit}\{p_T(d)\} = \log(\alpha) + \beta \log(d/d^*) \quad (3.1.2)$$

where  $p_T(d)$  is the probability of toxicity at dose  $d$ , parameters  $\alpha, \beta > 0$  and  $d^*$  a reference dose.  $\alpha$  can be interpreted as the odds of toxicity at dose  $d^*$ . This model is described as more flexible and also more realistic to represent the dose-response curve. For the choice of the prior distribution, the authors suggested for instance to relate on prior information for each dose through  $M$  quantiles  $q_d = (q_d(p_1), \dots, q_d(p_M))$  where  $\mathbb{P}(p_T(d) \leq q_d(p_m)) = p_m$  for  $m = 1, \dots, M$ . This should be done for each dose, so we define  $q_{km} = q_{d_k}(p_m)$ ,  $k = 1, \dots, K$ ,  $m = 1, \dots, M$ . The prior choice consists of the choice of a prior which minimize  $\arg \max_{k,m} |q_{km} - q'_{km}|$  where  $q'_{km}$  is obtained with the prior distribution. The authors especially suggest a bivariate normal prior for the parameter  $\theta = (\log(\alpha), \log(\beta))$ , built as a product of univariate.

In the general setting of dose-escalation, this model was also used in a hierarchical way to incorporate knowledge from pre-clinical data by Zheng and Hampson [98].

## 3.2 Bridging studies designs for dose-finding trials

In the context of dose-finding trials, bridging studies were mainly discussed through the use of the CRM in the literature.

### CRM through population heterogeneity

Several approaches were proposed by Shu [79], with some asymptotic properties being studied by O'Quigley [67]. Although those method are not strictly speaking an bridging study in the sense that there is not one finished trial and one to be conducted, the heterogeneity of several populations is taken into account, but the two trials can conducted in parallel.

One method, call CRM shift model, consists in considering one of the two groups as a reference group, and to consider for the second one that the dose-toxicity relationship will be "shifted". It is assumed in this approach that the level of shifting is either known or if not a discrete prior can be considered. For the first group, the probability of toxicity for dose  $x_j$  is  $\mathbb{P}(Y_i = 1|X_i = x_i) = \psi(x_i; a) = \alpha_i^{\exp(a)}$  where  $Y_i$  is the binary outcome for  $i^{th}$  patient indicating toxicity and  $X_i$  the dose given to this patient among the panel of  $K$  doses  $\{d_1, \dots, d_K\}$ , and the recommended dose is supposed to be identified. Then, for

the second group, the dose-toxicity is

$$\psi(d_i; a) = \alpha_{\phi(i)}^{\exp(a)} \text{ for } i \in \{1, \dots, K\}$$

where

$$\phi(i) = i + zh(i)$$

$$\text{with } h(i) = L\mathbb{I}(1 \leq i + L \leq K) + K\mathbb{I}(i + L > K) + \mathbb{I}(i + L < 1) \text{ for } L \in \{-K, \dots, K\}$$

where  $z$  is an indicator variable for the group ( $z = 0$  for group 1 and  $z = 1$  for group 2),  $\mathbb{I}()$  is the indicator function and  $L$  the levels shifted. This method then deals with two trials, or one trial composed by two subgroups. In a way, this approach is twice adaptive, with an interdependence in the dose allocation rules. It is not designed to be particularly a Bayesian method (no specific prior for parameter  $a$ ).

Another possibility is the two parameter CRM model for two groups  $G_1$  and  $G_2$ :

$$\mathbb{P}(Y_i = 1 | X_i = x_i, G_1) = \psi_1(x_i, a)$$

$$\mathbb{P}(Y_i = 1 | X_i = x_i, G_2) = \psi_2(x_i, a, b)$$

with  $a$  is the parameter for the shared information and  $b$  is used to distinguish the two groups. The probability of toxicity for the dose  $d_i$  is then

$$\mathbb{P}(Y = 1 | X = d_i, z) = \psi(d_i, a, bz) = \alpha_i^{a+bz} \text{ for } i \in \{1, \dots, K\}$$

where  $z = 0$  for the first group and  $z = 1$  for the second group. The estimated probability of toxicity for group 1 and 2 for a dose  $d_i$  are then  $\psi(d_i, \hat{a})$  and  $\psi(d_i, \hat{a}, \hat{b})$  respectively. Further details about those bridging methods depending on the setting (two trials in parallel or one finished for instance) were studied by O'Quigley and Iasonos [68]

## The Bridging continual reassessment method (BCRM)

An other method is proposed by Liu et al.(2015) [48]. This one, also based on the use of the CRM, directly defines or sets the skeleton of the CRM. The previous finished trial is considered to be the landmark one. The idea is to take for each dose of the new trial  $d_1, \dots, d_K$  the estimate probability of toxicity  $\hat{p}_1, \dots, \hat{p}_K$  based on the dose-toxicity relationship estimate obtained from the previous trial. Let's consider three skeleton:

- **Skeleton 1:**  $p_i = \hat{p}_i$

- **Skeleton 2:**

$$p_i = \hat{p}_{i+1} \text{ for } i \in \{1, \dots, K - 1\}$$

$$p_K = \frac{\hat{p}_K + 1}{2}$$

- **Skeleton 3:**

$$p_i = \hat{p}_{i-1} \text{ for } i \in \{2, \dots, K\}$$

$$p_1 = \frac{\hat{p}_1}{2}$$

Each skeleton is associated to a model  $M_k$ , for  $k = 1, 2, 3$ . The skeletons correspond to the three cases where the new population has the same dose-toxicity relationship than the landmark one, or is more sensitive or less sensitive for skeletons 2 and 3 respectively. Skeleton 1 corresponds to the skeleton for which the toxicity probabilities are estimated from a previous trial. Skeleton 2 to the same skeleton as before but with a shift of one rank lower in the toxicity probabilities associated with the doses: the toxicity probability estimated for dose  $i$  in the previous trial is now associated with dose  $i - 1$ , this skeleton assumes a higher sensitivity to the drug in the population. Finally, skeleton 3 is the same idea as before but the shift is now of one rank higher, assuming a lower sensitivity to the drug. An illustration is given in figure 3.3.

Then, the probability of toxicity associated to each dose is calculated through the Bayesian model averaging method (the principle of the Bayesian model averaging is reminded in the appendix of [65] or is detailed in [31]). The results show better performance than the classic CRM in most scenarios. However, it does not take into account the possible opposition between the two data sets, and the method is strongly dependant on the skeleton calculated from the landmark trial.

## CRM through historical-to-current parameter

Takeda and Morita (2018) [85] proposed to consider an historical-to-current parameter space. The authors proposed to retrospectively analyse an historical dataset  $D_H$  using a two parameter logistic model

$$\pi(x_i, \alpha, \beta) = \mathbb{P}(Y_i = 1 | x_i, \alpha, \beta) = \frac{\exp(\alpha + \beta x_i)}{1 + \exp(\alpha + \beta x_i)}$$



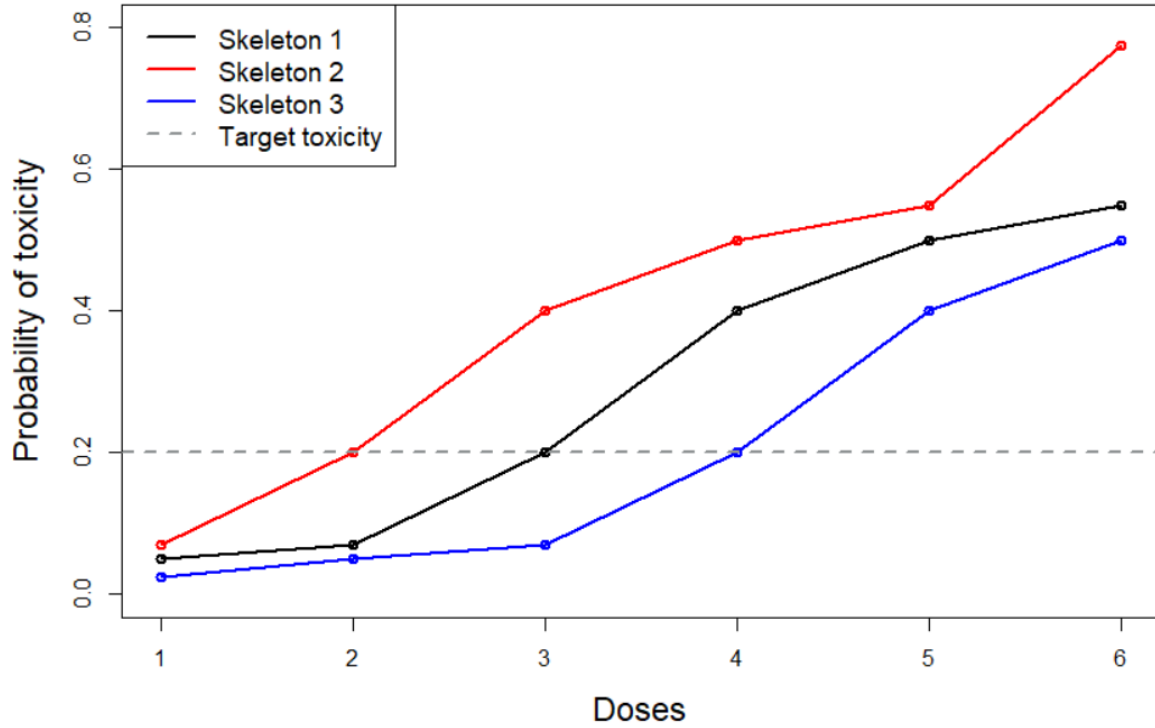


Fig. 3.3 Example of three skeletons obtained with the BCRM method. The black curve represents the dose-toxicity relationship estimates based on the landmark trial.

where  $x_i$  is the standardized dose given to patient  $i$ :  $x_i = \log(d_i) / \frac{1}{K} \sum_{j=1}^K \log(d_j)$  with  $d_1, \dots, d_K$  the doses administered. Then, considering  $\tilde{\mu}_{\alpha,H}$  and  $\tilde{\sigma}_{\alpha,H}^2$  the posterior mean and variance of the parameter  $\alpha$ , the authors assume that the current  $\alpha$  parameter  $\alpha_C \sim \mathcal{N}(\tilde{\mu}_{\alpha,H}, \frac{1}{\omega_\alpha} \tilde{\sigma}_{\alpha,H}^2)$  with  $\omega_\alpha$  being the historical-to-current parameter, included between 0 and 1. The authors investigated several priors for the historical-to-current parameter and calibrated the two prior distributions for  $\alpha_C$  and  $\beta_C$  for to have an effective sample size fixed at 2. The authors' simulation study showed that their method had comparable results in terms of MTD recommendation.

### 3.3 Similarity criteria between curves

In this section, we explore existing criteria or methods to evaluate the similarity between several curves. In the context of bridging studies, it is often necessary to compare several dose-response relationships to then assess if an extrapolation is possible or not [35]. Therefore, it is necessary to have some appropriate tools to be able to compare several

Name	Expression of $D_f(p  q)$	Expression $f(t)$
Kullback-Leibler	$\int p(x) \log \left( \frac{p(x)}{q(x)} \right) dx$	$t \log(t)$
Squared Hellinger	$\int \left( \sqrt{p(x)} - \sqrt{q(x)} \right)^2 dx$	$(\sqrt{t} - 1)^2$
Total variation	$\int  p(x) - q(x)  dx$	$\frac{1}{2} t - 1 $
Pearson $\chi_P^2$	$\int \frac{(q(x) - p(x))^2}{p(x)} dx$	$(t - 1)^2$
$\alpha$ -divergences	$\frac{4}{1-\alpha^2} \left( 1 - \int p(x)^{\frac{1-\alpha}{2}} q(x)^{\frac{1+\alpha}{2}} dx \right)$	$\frac{4}{1-\alpha^2} (1 - t^{\frac{1+\alpha}{2}})$ , $\alpha \neq \pm 1$

Table 3.1 Examples of  $f$ -divergences

dose-response relationships. We decided to deal with the question of curves comparison in two parts. In the first one the *divergences* approach is considered to compare distributions. The second part deals with dose-response curves distances, directly developed in a clinical context.

### 3.3.1 Degree of similarity between distributions: divergences

Several criteria are available to compare two distributions. Here we focus on the *divergences* approach, and particularly in the setting of continuous density functions.

#### $f$ -divergences

The concept of  $f$ -divergences was introduced independently by Csiszàr [17], Ali and Silvey [1] and Morimoto [52]. Let's consider  $p$  and  $q$  two density functions and let  $f$  be a convex function satisfying  $f(1) = 0$ . The  $f$ -divergence  $D_f(p||q)$  between  $p$  and  $q$  is

$$D_f(p||q) = \int q(x) f \left( \frac{p(x)}{q(x)} \right) dx$$

Those divergences have the some specific properties, developed below, as being always positive, being jointly convex and having some invariance properties (see [2] for example), which can be useful, for instance, for image reconstruction (see [40]). The most known  $f$ -divergences are given in table 3.1.

Several links were established between the several divergences. For instance, the know Neyman divergence is the reverse of the Pearson one  $\chi_P^2(p, q) = \chi_N^2(q, p)$ , and the  $\alpha$ -divergence with  $\alpha = 0$  is proportional to the Hellinger divergence. The Kullback-

Leibler divergence is perhaps one of the most used  $f$ -divergence. This is due to its interpretation as a *relative entropy* between the two density functions. As explained before, the  $f$ -divergences has many interesting properties:

- They are positive:  $D_f(p||q) \geq 0$
- They are jointly convex: for  $\lambda \in [0, 1]$  and  $p_1, p_2, q_1, q_2$  some density functions,  $D_f(\lambda p_1 + (1 - \lambda)p_2 || \lambda q_1 + (1 - \lambda)q_2) \leq \lambda D_f(p_1 || q_1) + (1 - \lambda)D_f(p_2 || q_2)$ .
- They satisfy some invariance properties, for example:
  - Denoting  $\tilde{f}(t) = f(t) + at + b$ , then  $D_{\tilde{f}}(p||q) = D_f(p||q)$
  - Denoting  $\hat{f}(t) = tf(\frac{1}{t})$ , then  $D_{\hat{f}}(p||q) = D_f(q||p)$  ( $\hat{f}$  is called the conjugate of  $f$ , see [71]).
- Two conjugate functions  $f$  and  $\hat{f}$  satisfy  $D_f(p||q) = D_{\hat{f}}(p||q)$  of all  $p$  and  $q$  if and only if  $f(t) - \hat{f}(t) = c(t - 1)$ ,  $c$  being a constant.

Other  $f$ -divergences' properties can be found in [71], [2], [17].

### Bregman divergences

Bregman divergences, introduced by Bregman [9], are adapted to very various areas. They can be defined for matrix, vectors, functions, density functions. There are particularly used for machine learning and classification problems. Generally,  $f$  being a strictly convex and continuously differentiable function, define on a convex set, and  $x, y$  being two points (vectors, matrix, functions...) and  $\langle \cdot, \cdot \rangle$  the inner product, the Bregman divergence between  $x$  and  $y$  is

$$D_f^B(x, y) = f(x) - f(y) - \langle \nabla f(y), x - y \rangle$$

This general definition can be adapted depending on the context. For our setting [5], the Bregman divergences between two density functions  $p$  and  $q$  are:

$$D_f^B(p, q) = \int \{f(p(x)) - f(q(x)) - (p(x) - q(x))f'(q(x))\}dx$$

### 3.3.2 Comparison between regression curves in clinical trials

Several ways were proposed to compare regression curves, most of the time based on hypothesis testing. In the case where the two regression curves have some common

parameters, Möllenhoff et al. [51] proposed the approach below. We consider

$$Y_{l,j,i} = m_l(d_{l,j}, \boldsymbol{\beta}_l) + \eta_{l,j,i}$$

where  $Y_{l,j,i}$  is the response of the  $i$ th subject at the  $j$ th dose under the  $l$ th dose-response model, with  $l = 1, 2$ ,  $j = \{1, \dots, k_l\}$ ,  $i \in \{1, \dots, n_{l,j}\}$ .  $k_l$  is the number of doses in group  $l$ . The  $\eta_{l,j,i}$  terms are independent and identically distributed. The two regression models  $m_1$  and  $m_2$  can be different, the the vector parameters  $\boldsymbol{\beta}_1$  and  $\boldsymbol{\beta}_2$  too. Considering the approach presented in [49] and [26], and defining  $d_\infty(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2) = \max_{d \in \mathcal{D}} |m_1(d, \boldsymbol{\beta}_1) - m_2(d, \boldsymbol{\beta}_2)|$  with  $\mathcal{D}$  the region containing all doses and  $\epsilon$  being a pre-specified constant, often considered as a relevant threshold, the authors proposed a test for the hypothesis

$$H_0 : d_\infty(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2) \geq \epsilon \text{ vs } H_1 : d_\infty(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2) < \epsilon \quad (3.3.1)$$

The test is then performed through a bootstrap procedure. Afterwards, considering that both  $\boldsymbol{\beta}_1 = (\beta_{1,1}, \dots, \beta_{1,p_1})$  and  $\boldsymbol{\beta}_2 = (\beta_{2,1}, \dots, \beta_{2,p_2})$  have a common part  $(\beta_{l,1}, \dots, \beta_{l,p'})$  for  $l = 1, 2$ , the authors proposed to test the equivalence of model parameters through the hypothesis

$$K_0 : \max_{i=1, \dots, p'} |\beta_{i,1} - \beta_{2,i}| \geq \delta \text{ vs } K_1 : \max_{i=1, \dots, p'} |\beta_{i,1} - \beta_{2,i}| < \delta$$

A similar approach was proposed by Bretz et al. [10]. The authors first constructed their own method without bootstrap for testing hypothesis 3.3.1, using confidence intervals for  $m_1(d, \boldsymbol{\beta}_1) - m_2(d, \boldsymbol{\beta}_2)$  obtained with the delta-method. Then, the authors considered the problem of similarity between two target doses from two dose-response curves from two non-overlapping subgroups. The target dose for them is the *minimum effective dose*: the MED, that is the smallest dose showing a clinical relevant benefit over placebo. The authors then propose an asymptotic test to assess the similarity between two MED calculated through two different models. In particular, if  $\widehat{MED}_i$  denotes the estimated MED of group  $i$ ,  $i = 1, 2$ , the authors established that  $\widehat{MED}_1 - \widehat{MED}_2 - (MED_1 - MED_2)$  follows a normal distribution with mean zero and a variance  $\tau^2$  calculated with the delta method. With confidence intervals calculation, the authors show that rejecting hypothesis  $H'' : |MED_1 - MED_2| \geq \eta$  if  $|\widehat{MED}_1 - \widehat{MED}_2| < c$  gives an asymptotic test with level  $\alpha$ , if  $c$  is such as  $\alpha = \Phi\left(\frac{c-\eta}{\tau}\right) - \Phi\left(\frac{-c-\eta}{\tau}\right)$ .



# Chapter 4

## An adaptive power prior for sequential clinical trials – Application to bridging studies

In this chapter, we present article [65]: Adrien Ollier, Satoshi Morita, Moreno Ursino, Sarah Zohar. An adaptive power prior for sequential clinical trials–Application to bridging studies. *Statistical methods in medical research*, 2020, vol. 29, no 8, p. 2282-2294, <https://doi.org/10.1177/0962280219886609>.

### **Context**

In early phase clinical trials, it is common to have previous available information from clinical trials conducted in another population for example. This is the case for bridging studies in particular. It is highly desirable to share this previous information in the setting of sequential dose-finding trials.

### **Objective**

We want to propose a Bayesian method which permits to adaptively set prior distribution to better estimate the dose-response relationship. Such a method must take into account the weight given to historical data and the potential data-prior conflict.

**Method:** The power prior approach was used to incorporate historical data. The weight of the prior distribution was considered through its variance. The variance was calibrated by selecting the power in the power prior's likelihood corresponding to a desirable ESS, fixed by advance. This desirable prior distribution's ESS depends on the number of patients in the current trial: it must be low if there are few patients, and

could increase if there are more patients. It is also the maximum ESS we authorise the prior distribution to have, and it could be decreased. Then, to assess the potential data-prior conflict, we built a similarity criterion between both the historical and the current datasets. This similarity criterion is based on the Hellinger's distance, and permits to decrease the maximum authorised ESS fixed before.

### **Results**

Through an extensive simulation study, we explored several ways to fixed the maximum desirable ESS and the similarity criterion setting and compared it to the classic CRM. We noted, as expected, that a high ESS permits to increase the MTD selection if both historical and current datasets are similar. However, it also induces a bias if they are not similar, even taking the difference into account through the similarity criterion. Then, a low ESS does not permit to increase the MTD selection. We finally suggested a setting that permits to increase the selection of the MTD when the two datasets are similar and has similar results to the classic CRM when they are not.

### **Conclusion**

Even if we suggested a setting that permits to increase the selection of the MTD, a sensitivity analysis is required before the starting of the trial. More precisely, the number of patients in each cohort, the desired ESS, and the similarity criterion must be considered. The built similarity criterion has an easy interpretation since it is between 0 and 1. However, it also depends on the model, since it is based on the likelihood function. In other words, the quality of this criterion also depends on the quality of the matching between the model and the data.

# An adaptive power prior for sequential clinical trials – Application to bridging studies

Adrien Ollier<sup>1</sup>, Satoshi Morita<sup>2</sup>, Moreno Ursino<sup>1,3,\*</sup>  and Sarah Zohar<sup>1,\*</sup>

Statistical Methods in Medical Research  
0(0) 1–13

© The Author(s) 2019



Article reuse guidelines:

sagepub.com/journals-permissions

DOI: 10.1177/0962280219886609

journals.sagepub.com/home/smm



## Abstract

During drug evaluation trials, information from clinical trials previously conducted on another population, indications or schedules may be available. In these cases, it might be desirable to share information by efficiently using the available resources. In this work, we developed an adaptive power prior with a commensurability parameter for using historical or external information. It allows, at each stage, full borrowing when the data are not in conflict, no borrowing when the data are in conflict or “tuned” borrowing when the data are in between. We propose to apply our adaptive power prior method to bridging studies between Caucasians and Asians, and we focus on the sequential adaptive allocation design, although other design settings can be used. We weight the prior information in two steps: the effective sample size approach is used to set the maximum desirable amount of information to be shared from historical data at each step of the trial; then, in a sort of Empirical Bayes approach, a commensurability parameter is chosen using a measure of distribution distance. This approach avoids elicitation and computational issues regarding the usual Empirical Bayes approach. We propose several versions of our method, and we conducted an extensive simulation study evaluating the robustness and sensitivity to prior choices.

## Keywords

Bayesian, bridging studies, power priors, dose-finding, phase I, early phase

## 1 Introduction

Bayesian inference is increasingly used in clinical trial planning, implementation and analysis. It has the advantage of using external information (historical data from previous clinical data, electronic health records, the medical literature, expert opinion, etc.) into the statistical framework. This property allows the reduction of sample size while increasing the statistical power.<sup>1–3</sup> However, incorporating external sources of data into the prior distribution needs to be done carefully, as these data can either be in conflict with or empower the resulting posterior. In these cases, specific tools for prior distribution calibration are required.

Two efficient approaches have been proposed in recent years to calibrate and tune prior information: the effective sample size (ESS) method and the power prior approach. The ESS method allows us to interpret the calibrated parametric prior in terms of the number of hypothetical patients who were used to develop the prior

<sup>1</sup>Centre de Recherche des Cordeliers, INSERM, Sorbonne Université, USPC, Université de Paris, Paris, France

<sup>2</sup>Department of Biomedical Statistics and Bioinformatics, Kyoto University Graduate School of Medicine, Kyoto, Japan

<sup>3</sup>F-CRIN PARTNERS platform, AP-HP, Université de Paris, Paris, France

\*These co-last authors contributed equally to this article.

### Corresponding author:

Moreno Ursino, Centre de Recherche des Cordeliers, INSERM, Sorbonne Université, USPC, Université de Paris, 15 rue de l’Ecole de Médecine, 75006 Paris, France.

Email: [moreno.ursino@inserm.fr](mailto:moreno.ursino@inserm.fr)



distribution. The ESS can be viewed as “how informative” a prior distribution is and can quantify it.<sup>4-6</sup> In addition, the power prior was first proposed when incorporating historical data into the analysis of clinical trials.<sup>7-9</sup> In this context, one weights the amount of information that will be used in the posterior computation. For this aim, a weight is introduced as a power parameter and is defined between 0 and 1 (where 0 = non-informative prior and 1 = full borrowing prior). However, this parameter requires subjective elicitation which can lead to prior misspecification if it is not done correctly.<sup>10</sup> Recently, a class of commensurate priors has been proposed, that can be viewed as an attempt to quantify the degree of similarity between the informative prior distribution and the likelihood. Indeed, if the two distributions are superposed, there would be a full borrowing. Therefore, the amount of “borrowing” depends, in this case, on the amount of commensurability between the prior and the likelihood.<sup>1,11</sup> To do that, the authors define a commensurability parameter that is estimated once the trial data (that are resumed in the likelihood) are observed. Other efforts have also been made to set the power prior to control the type I error in the case of data-prior conflict,<sup>12,13</sup> or when historical data are simulated from medical device and their uncertainty is taken into account.<sup>14</sup> Regarding commensurability, a meta-analytic approach was also proposed by Schmidli et al.<sup>15</sup> with the robust meta-analytic predictive (MAP) prior.

However, when designing a sequential adaptive trial, interim analyses are performed regarding dose modification, safety or efficacy estimation depending on the trial design and the clinical trial phase (early or confirmatory). Owing to the sequential aspect of the process, (1) one cannot wait the end of the trial to estimate the commensurability or power prior parameter, (2) one cannot necessarily decisively elicitate the power parameter, as it either has a large weight that is too informative during the initial interim analysis because the trial did not yet reach its full sample size or a small weight that renders it non-informative at the time of the final analysis, and (3) one can propose a small value of the power parameter but all the advantages of information borrowing are lost.

Early phase dose-finding clinical trials aiming at estimating the maximum tolerated dose (MTD) are sequential. In model-based methods, the dose administered to the next cohort of patients depends on all the doses given and the associated toxicities observed so far. Recently, methods were proposed that allow using external information into the dose-finding design, where either external data were used for choosing the skeleton (working model) of the design or for calibrating the prior of the dose-toxicity relationship parameter(s).<sup>16,17</sup> Liu et al. proposed using a Bayesian model averaging (BMA) dose-finding method in which the estimated probabilities of toxicity at the end of the previous trial are used to build three different skeletons, that will be averaged during the present trial.<sup>18</sup> Takeda and Morita defined a “historical-to-current” (H-C) parameter representing the degree of borrowing based on a retrospective analysis of previous trial data.<sup>19</sup> Finally, Petit et al. used external information to calibrate the dose range and the working model.<sup>20</sup>

## 1.1 Motivating case study

Bridging studies are designed to bridge the gap of clinical data, such as efficacy, comorbidities, safety and dosing regimens between two populations. According to the International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use E5 (ICH-E5) guidelines, a bridging study of a medicine can be defined as an additional study executed in a new population, for example, in another ethnic group, to “build a bridge” to the new clinical data on safety, efficacy, and dose response.<sup>21</sup> These studies are important for both pharmacodynamic and pharmacokinetic reasons. Indeed, ethnic diversity in drug response for some drugs with respect to safety, efficacy and the resulting similarities or differences in recommended doses have been well described.<sup>22</sup> Some of these differential responses may be related to the pharmacogenomics of a particular drug.<sup>23</sup> In some situations, it might be desirable to share information between populations to look for a more efficient use of resources but in other cases, it should not be done if the populations are very different from each other. For instance, the MTD of lapatinib, a tyrosine kinase inhibitor used in breast cancer, was estimated to be 1800 mg in Japanese patients but was higher in US and European patients. By contrast, the MTD of sunitinib, a multi-targeted receptor tyrosine kinase inhibitor, which is given for renal cell carcinoma, is equal to 50 mg and is similar in all countries. Accordingly, it will be pertinent to achieve “full borrowing” when the populations are similar, “no borrowing” when they are different or “tuned borrowing” when some, but not all, information can be borrowed.<sup>24</sup>

The aim of our work is to propose a new method which, using the information coming from an existing fixed historical trial, helps the design and dose-allocation of a new prospective clinical trial. In this paper, we propose an adaptive power prior (APP) approach, based on a criterion constructed using the power prior, the ESS and the Hellinger distance. Using the advantages of each approach, the ESS is used for checking the maximum desired amount of information, the power prior is used for adding historical data, and the Hellinger distance is used for

tuning the final borrowing. A criterion measures the similarity between distributions and decides how much information should be used in the clinical trial. We used the example of a phase I dose-finding study, because sequential and adaptive dose allocation need to be performed often during the trial.

In Section 2, the proposed method is described in detail along with several variants of the application to phase I bridging studies. Then, the simulation study is shown in Section 3, followed by the results in Section 4. The article ends with a discussion on strengths, weaknesses, and future improvements of the method.

## 2 Methods

Let  $\theta$  be the parameter or the vector of parameters of interest. For simplicity, we will write all notations as if we are in one dimension, but everything can be easily generalised to vectors and matrices. Let  $D_0$  denote the historical data,  $D_0 = \{y_j\}_{n_0}$ ,  $n_0$  the sample size of  $D_0$ , and  $L(\theta|D_0)$  the likelihood function of  $\theta$  given the historical data. In a similar manner, define  $D$  as the current data,  $D = \{y_i\}_n$ ,  $n$  as the sample size of  $D$  and  $L(\theta|D)$  as the likelihood function of  $\theta$  given the current data. We propose the following adaptive power prior  $\pi^{APP}$

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

where  $\pi_0$  represents a non-informative prior distribution for  $\theta$  and the original power prior parameter, introduced in Ibrahim and Chen,<sup>8</sup> is split into two parts,  $0 \leq \alpha_0 \leq 1$  and  $0 \leq \gamma \leq 1$ . Since the two new parameters,  $\alpha_0$  and  $\gamma$  called the ‘‘quantity of information’’ and ‘‘commensurability’’ parameters, respectively, have two separate and specific interpretations, we propose a two-steps approach to set their values.

### 2.1 Quantity of information parameter value

In the first step,  $\alpha_0$  is chosen to add an upper limit on the quantity of information that is desirable to borrow.  $\gamma$  is temporarily set to 0, and the ESS of equation (1) is computed as  $s = ESS[\pi^{APP}(\theta|\gamma = 0)] = \alpha_0 ESS[L(\theta|D_0)] + ESS[\pi_0(\theta)]$ , where  $L(\theta|D_0)$  should be seen as a distribution.<sup>4</sup> If a very non-informative prior is chosen for  $\pi_0$ , for example an improper uniform distribution when possible, the second term of the summation,  $ESS[\pi_0(\theta)] = s_0$ , tends towards zero. Moreover, since  $L(\theta|D_0)$  includes  $n_0$  observations, it is straightforward to think that the ESS can be approximately equal to  $n_0$  (a proof of the convergence is given in the Supplementary Material). The resulting ESS can be written as  $s \approx \alpha_0 n_0 + s_0$ , and it is linearly dependent on the parameter  $\alpha_0$ . Therefore, after setting a desirable ESS  $s^*$  for the upcoming analysis,  $\alpha_0$  can be chosen to invert the previous equation, that is  $\alpha_0 = (s^* - s_0)/n_0$ .  $\alpha_0$  can be viewed as an upper limit on the quantity of information borrowed, because  $ESS[\pi^{APP}(\theta|\gamma = 0)] > ESS[\pi^{APP}(\theta|\gamma > 0)]$ . A desirable  $s^*$  depends on the application and on the sample size  $n$  of the actual data. Except in rare cases, it is usually accepted that  $s^* < n$  to avoid the situation where the prior distributions overcome the actual data.

### 2.2 Commensurability parameter value

In the second step, we set the commensurability parameter  $\gamma$ . This parameter takes care of the possible conflict between the historical and current datasets. When the two datasets are very different, a non-informative prior should be preferable; when they are similar, a complete borrowing is preferred. We suggest linking  $\gamma$  to a measure of distance between the two datasets, that is, between  $D_0$  and  $D$ . This distance should be a positive number between 0 and 1, and it should tend towards the maximum value when the two datasets are very different, and to zero when  $D_0$  and  $D$  are close. Since in some applications, such as dose finding, the data can come from a non-homogeneous population, the approach of Pan and Yuan<sup>25</sup> cannot be directly used. For instance, in dose-finding trial data, each dose of the panel will produce a different population outcome. To overcome this issue, we propose, in a sort of Empirical Bayes method, to use the Hellinger distance with the normalised likelihoods, as follows

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

where  $d^2$  refers to the square of the Hellinger distance. Each likelihood has a normalisation constant to ensure that it can be viewed as a probability distribution. Moreover,  $L(\theta|D_0)$  is raised to a factor  $n/n_0$  to allow it to be comparable to  $L(\theta|D)$ . Since  $n_0 \geq n$ , we expect that the information included in the historical likelihood is more accurate, or in other words, has less variance than the actual data. Therefore, it is not directly comparable to a likelihood with fewer data points. We propose the addition of this factor to avoid this inconvenience. For example, in the Bernoulli case, e.g.  $y_i$  follows a Bernoulli distribution, we have that  $L(\theta|D_0) = p^{\sum_{n_0} y_i} (1-p)^{n_0 - \sum_{n_0} y_i}$ , and, when we add the exponent, we obtain  $L(\theta|D_0)^{n/n_0} = p^{n\bar{y}_{n_0}} (1-p)^{n - n\bar{y}_{n_0}}$ , where  $\bar{y}_{n_0} = (\sum_{n_0} y_i)/n_0$ . If we rewrite  $\sum_{n_0} y_i$  as  $n_0 \bar{y}_{n_0}$  in the likelihood without the exponent, we can easily see that the mean of the data is preserved, and the historical and the actual likelihood can now be compared with regard to their variability. Another example is given by the Gaussian case, where  $y_i \sim N(\mu, \sigma^2)$ . In this situation,  $L(\theta|D_0) = (2\pi\sigma^2)^{-n_0/2} \exp\left(-(\sigma^2)^{-1} \sum_{n_0} (y_i - \mu)^2\right)$  and  $L(\theta|D_0)^{n/n_0} = (2\pi\sigma^2)^{-n/2} \exp\left(-(\sigma^2)^{-1} n (\sum_{n_0} (y_i - \mu)^2)/n_0\right)$ . Again, writing  $\sum_{n_0} (y_i - \mu)^2$  as  $n_0 (\sum_{n_0} (y_i - \mu)^2)/n_0$ , we can see that the quantity  $(\sum_{n_0} (y_i - \mu)^2)/n_0$  is preserved but its ‘‘intensity’’ is reduced to  $n$ . The same reasoning can be done for the variance parameter. To be noted, equation (2) assumes that  $n_0 \geq n$ , but it can be easily generalised as follows

$$d^2(D_0, D) = \frac{1}{2} \int \left( \sqrt{\frac{L(\theta|D)^{\min(1, \frac{n_0}{n})}}{\int L(\theta|D)^{\min(1, \frac{n_0}{n})} d\theta}} - \sqrt{\frac{L(\theta|D_0)^{\min(1, \frac{n_0}{n_0})}}{\int L(\theta|D_0)^{\min(1, \frac{n_0}{n_0})} d\theta}} \right)^2 d\theta \quad (3)$$

where each time we downgrade the more accurate likelihood.

Finally, we set  $\gamma = \sqrt{d^2(D_0, D)}$ , considering, of course, the only real root (this reasoning will be applied for now on for all other  $c$ -roots). Any other power of  $d$  can be used to define the parameter  $\gamma$ , since  $\gamma = d^c(D_0, D) \in [0, 1] \forall c \in R^+$ . Values greater than 1 will reduce the computed distance and will lead to more borrowing, while values less than 1 will lead to a more conservative approach and will increase the computed distance.

### 2.3 Application in phase I bridging studies

The adaptive power prior distribution naturally fits the sequential nature of the phase I bridging study. It is only necessary to set the maximum ESS that a prior can have at each stage of the trial. One can set the maximum ESS as a vector where each number is related to a trial cohort, or set it as a function with the number of patients already accrued as an independent variable. In this article, as an example, we apply the APP on phase I bridging studies where the continual reassessment method (CRM) was used as the design. In particular, we chose the logistic model, that is

$$\text{logit}(p_i) = a + \exp(\beta) \tilde{d}_i$$

where  $p_i$  refers to the probability of toxicity at dose  $i$ ,  $a = 3$  is a constant parameter,  $\tilde{d}_i$  is the ‘‘effective’’ dose, which is defined as the prior estimate of the probability of toxicity associated with dose level  $i$  and  $\beta$  is the parameter of interest (to be estimated).<sup>26</sup> Usually, a normal distribution with a mean equal to 0 and variance equal to 1.34 is used as prior distribution for  $\beta$ .<sup>27</sup> In our setting, we used the same distribution as  $\pi_0(\beta)$ , that is as the non-informative distribution, to build the  $\pi^{APP}(\beta)$ , and we propose several possible versions of it.

The classical CRM, that is when  $\gamma = 1$  or directly  $\alpha = 0$  for all cohorts to ensure that no historical data is shared, will be called P\_NI from now on. P\_ESS refers to the model where only the ESS part is specified and constant, and in parentheses, the amount of information desired will be specified. AP\_L denotes the model with  $\gamma = \sqrt{d^2(D_0, D)}$  and a linear ESS that depends on the number of patients,  $s^*(n) = n$ . Other forms, including the rounded sigmoid form, gave almost the same results; therefore, we opted for the simpler version. The version called AP\_S is the same as before but with  $\gamma = \sqrt[4]{d^2(D_0, D)}$ , that is, to have the square root of the distance equal to  $\gamma$ . To make the prior more robust, we checked the performance of the method using a mixture prior, AP\_MIX( $\omega$ ), built as follows  $\pi(\theta|D) = \omega \pi^{APP}(\beta) + (1 - \omega) \pi_0(\beta)$ , where  $\omega$  represents the mixture weight parameter. Then, following the example of Liu et al.,<sup>18</sup> we also propose using the Bayesian model averaging technique, where the two models involved are AP\_L, as  $M_1$ , and AP\_NI, as  $M_2$ .<sup>28</sup> Details on the BMA methods, called AP\_BMA, are given in the

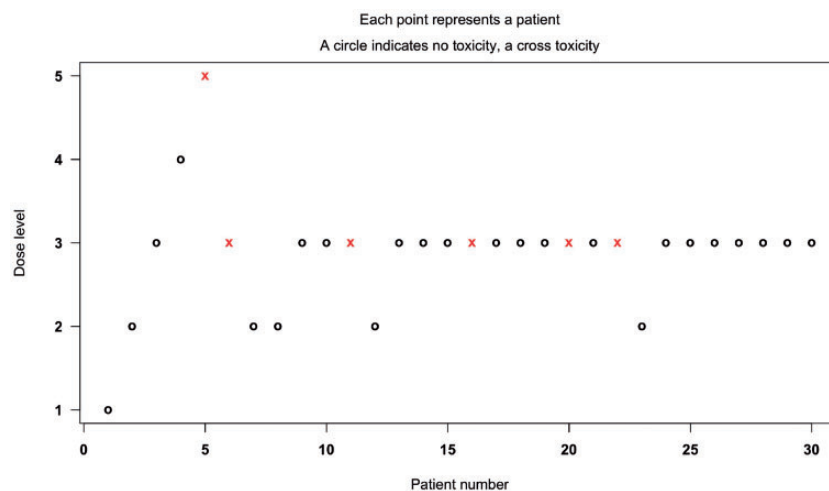
Supplementary Material. We have also reshaped the idea of Occam's windows,<sup>29</sup> frequently used in BMA, to set a threshold on  $\alpha = \alpha_0(1 - \gamma)$  or on  $\gamma$ . In this way,  $\alpha = \alpha I(\alpha > \tau_\alpha)$ , where  $I$  is the indicator function. In other words,  $\alpha$  is set to zero if its value is less than the pre-specified threshold. In the same manner, we can define  $\gamma = \gamma I(\gamma < \tau_\gamma) + I(\gamma \geq \tau_\gamma)$ .

### 3 Simulation studies

To evaluate the performance of the proposed method, we carried out an extensive simulation study, in which 1000 independent phase I trials per scenario were simulated. The aim is to evaluate the performance of the dose-finding methods in different scenarios, when the generating probabilities of the prospective trial are similar to the ones estimated by  $D_0$  and when they are different. In any case,  $D_0$  is considered fixed, since we are not interested in their generated probabilities: the best guesses are, of course, the estimated ones.

First, to compare the dose-finding methods, we simulated subject responses, which follow Bernoulli distributions with parameters defined depending on the scenario, to all doses for each trial. Then, each simulated dataset was stored, and when running a trial, regardless of the method applied, subject responses were read from this stored dataset. In this way, when two methods coincide when proposing the dose allocation to the next patient, the results on the simulated patient are the same.

Regarding the design, each trial had a maximum sample size of 30 patients, six dose levels, a cohort size of one patient. The no-skipping rule was applied; that is, a higher dose was not proposed if all previous ones were not already given. For simplicity, stopping rules were not applied, except for scenario 6. Six scenarios were studied with a target toxicity of 20% and the trial plotted in Figure 1 was assumed as historical data  $D_0$ . In particular,  $D_0$  followed a CRM design with six doses, a cohort of one and a maximum sample size of 30, as the main simulation trials. There were five dose-limiting toxicities (DLTs) at dose level 3 and one at dose level 5. The final posterior probability of toxicity was estimated to be (0.052, 0.073, 0.205, 0.406, 0.506, 0.555), and the MTD was declared to be dose level 3. Scenarios (that is, the choice of the probability of toxicity at each dose level) were generated to have the MTD be in several positions in the dose panel. This attribute allowed us to study cases where the probabilities of toxicity per dose were higher than the historical one, a scenario where these probabilities coincide in the two trials and in others where the probabilities of toxicity per dose in the actual trial were lower than in the historical one. As a skeleton, we adopted probabilities (0.05, 0.07, 0.2, 0.4, 0.5, 0.55) that reflect the probabilities of the historical data. In order words, we checked the performance of borrowing information when we already had a very good setting.



**Figure 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 six doses,  $d_i, i = 1, \dots, 6$ , 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.

After preliminary simulations (not shown here), we decided to use all methods that require a distance computation after the 10th patient enrolled in the trial in order to gather some information before starting to compute the commensurability parameter. For comparison, we ran the classic CRM (denoted by P\_NI) and the CRM with power prior where the power  $\alpha$ , which is denoted by AP\_EB, was chosen using the Empirical Bayes method.<sup>30</sup> In the latter case,  $\alpha$  is chosen to maximise the marginal likelihood

$$m(D, D_0, \alpha) = \frac{\int L(\theta|D)L(\theta|D_0)^\alpha \pi_0(\theta) d\theta}{\int L(\theta|D_0)^\alpha \pi_0(\theta) d\theta}$$

As a third competitor model, we run the bridging CRM (denoted by BCRM) of Liu et al.<sup>18</sup> We utilised the code that was provided online by the authors. However, using  $D_0$ , the principal skeleton estimated gave two doses at the same prior probability; therefore, we added 0.01 to the estimation of the higher dose in the couple to distinguish between them and then ran the BCRM.

Computations were carried out in R (version 3.5.0),<sup>31</sup> running under macOS High Sierra 10.13.6. rstan package (2.17.3)<sup>32</sup> was used for Bayesian inference, the ks package (1.11.3)<sup>33</sup> was adopted to approximate the distribution-like likelihoods in equation (2), and Monte Carlo sampling was used to approximate the final integral. R scripts will be available at the GitHub of the corresponding author.

## 4 Results

Among the models compared, we added the AP\_SOC1 that refers to the AP\_S with an Occam's window with  $\tau_\alpha = 0.2$ , and AP\_SOC2, which is the same version of AP\_SOC1 with the ESS term replaced with  $s^*(n) = \min(n, 20)$ . In the latter case, we reduced the possible amount of ESS to 20 instead of 30, as in the other models.

Table 1 summarized the characteristics of each method compared through simulation study. Table 2 shows the main results: the percentage of correct selection (PCS) and the number of DLTs, each for six different scenarios.

In scenario 1, where the MTD was placed at dose level 5, that is two ranks higher than the MTD of the historical data, the best PCS (54%) was obtained by P\_NI, followed by AP\_SOC1 (52%) and AP\_SOC2 (52%). These three methods share the same percentage of dose allocation (shown in Supplementary Material) and median number of DLTs. The PCS of the other methods is smaller by 10–22% (in absolute change), except for the full borrowing method P\_ESS(30). In the second scenario, the MTD was set one position higher with respect to the historical data. P\_NI, P\_ESS(10) and AP\_SOC2 achieved a similar PCS, approximately 60%, while P\_NI had a higher median number of DLTs than the other two methods. The performance of the rest of the methods, except of P\_ESS(30), was between 44% and 54%. Scenario 3 represents the situation where both the historical data and the actual data came from the same population. All methods increase the PCS with respect to P\_NI, except for the BCRM, which, by contrast, experiences a lower number of DLTs. The maximum PCS (95%) was reached by P\_ESS(30), followed by AP\_EB (90%). The other methods ranged between 80% and 87%, and the dose allocation table shows that more patients were treated at the MTD. In scenario 4, where the MTD was one rank less than the historical data, P\_ESS(10) had the best result, 73%, followed by the other methods whose PCSs ranged between 58% and 68%. The lowest performance was achieved by AP\_EB with a PCS of 54% and a higher median number of DLTs. In scenario 5, the MTD was set to two ranks lower than the historical data, and all methods, except for P\_ESS(30), had an high PCS, ranging from 76% to 90%. P\_ESS(10) and P\_ESS(30) tended to assign more patients to toxic doses. Finally, scenario 6 represents the situation where all doses are toxic. We added the same stopping rule to all methods, that it the trial is stopped if the posterior probability that the probability of toxicity of the first dose is higher than the pre-specified toxicity threshold is higher than 0.9. P\_ESS methods stopped and/or selected no dose with a lower percentage than the rest of the method, whose PCS ranges from 81% to 88%.

Sensitivity analyses regarding the ESS shape, mixture priors with different weights, BMA and Occam's window types are given in the Supplementary Material.

Then, we focused on the amount of borrowing, defined as the total power used in the power prior, that is,  $\alpha = \alpha_0(1 - \gamma)$ , at the end of trial for three methods: AP\_SOC2, AP\_S and AP\_MIX(05). The results are shown in Figure 2, where the scenarios are resumed in the x-axis through the difference of the probability of toxicity at dose

**Table 1.** Methods notation summary.

Method	$s^*(n)$ (ESS)	$\gamma$	Description
P_NI	$s^*(n) = 0$	–	Classic CRM, without historical information
P_ESS(10)	$s^*(n) = 10$	$\gamma = 0$	Constant ESS
AP_L	$s^*(n) = n$	$\gamma = d(D_0, D)$	Linear ESS, commensurability criterion equal to the modified Hellinger distance
AP_S	$s^*(n) = n$	$\gamma = \sqrt{d(D_0, D)}$	Linear ESS, commensurability criterion equal to the square root of the modified Hellinger distance
AP_Mix( $\omega$ )	$s^*(n) = n$	$\gamma = d(D_0, D)$	Mixture prior with $\pi(\theta D) = \omega\pi^{APP}(\theta) + (1 - \omega)\pi_0(\theta)$ , with $\pi^{APP}$ denoting the prior distribution obtained with AP_L
AP_SOC1	$s^*(n) = n$	$\gamma = \sqrt{d(D_0, D)}$	Linear ESS, commensurability criterion equal to the square root of the modified Hellinger distance, $\alpha = \alpha_{\mathbb{I}_{x>0.2}}$
AP_SOC2	$s^*(n) = \min(20, n)$	$\gamma = \sqrt{d(D_0, D)}$	Threshold (equals to 20) on the linear ESS, commensurability criterion equal to the square root of the modified Hellinger distance, $\alpha = \alpha_{\mathbb{I}_{x>0.2}}$
AP_EB	–	–	Empirical Bayes, <sup>30</sup> $\alpha$ is chosen to maximise the marginal likelihood $m(D, D_0, \alpha) = \frac{\int L(\theta D)L(\theta D_0)^\alpha \pi_0(\theta) d\theta}{\int L(\theta D_0)^\alpha \pi_0(\theta) d\theta}$
BCRM	–	–	Bridging CRM <sup>18</sup>

CRM: continual reassessment method; ESS: effective sample size.

3 with respect to the historical data (assumed to be 0.20). The median and the first and third quartiles of the final  $\alpha$  are represented on the y-axis. Between the three methods, AP\_MIX(05) is the one that borrows more information in scenario 3 (denoted by SC3 in the figure), with a median greater than 0.7. However, AP\_MIX(05) also adds more bias in the other scenarios, since its median is always greater than 0.1. Going through AP\_S to AP\_SOC2, the amount of borrowing is decreased in all scenarios, but borrowing information is still permitted in scenario 3 (median greater than 0.3) and does not add too much bias in the other scenarios.

We also examined at the convergence properties of  $\alpha$  for AP\_SOC2. We increased the sample size of each trial up to 100 and we used equation (3) to compute the distance to be used in  $\gamma$ . The results are plotted. Figure 3 shows the median value of  $\alpha$  along with the first and the third quartiles for scenarios 1, 2 and 3. In scenario 3, where full borrowing is desired, the median has an increasing trend, while the range tends to become narrower as the sample size increases. After 70 patients,  $\alpha$  starts to be greater than 0.4. In scenario 1, the median and the first and third quartiles coincide and are all equal to zero. In scenario 2, the median goes up to 0.20 when the sample size is approximately 20 patients, but then it decreases quickly to zero. Additionally, the third quartile decreases to zero after 40 patients. Scenarios 4 and 5 give similar results as scenarios 2 and 1, respectively; therefore, we did not plot them.

Finally, in Figure 4, we showed the dose-allocation/toxicity of the same trial, simulated under scenario 3, using the non-informative prior (left side) and the AP\_SOC2 (right side). Both dose-allocation schemes coincide until the 14th patient. Then, the AP\_SOC2 remains at the MTD level, while P\_NI is more conservative and decreases the dose level.

## 5 Discussion

The aim of our work was to propose a modification of the power prior by incorporating external fixed data into a sequential adaptive design. Indeed, during sequential clinical trials, several adaptations are possible, including dose or regimen modification which are usually performed in early phase dose-finding studies. As noted in the Introduction, using the data from another population, indication or schedule can empower the trial results when

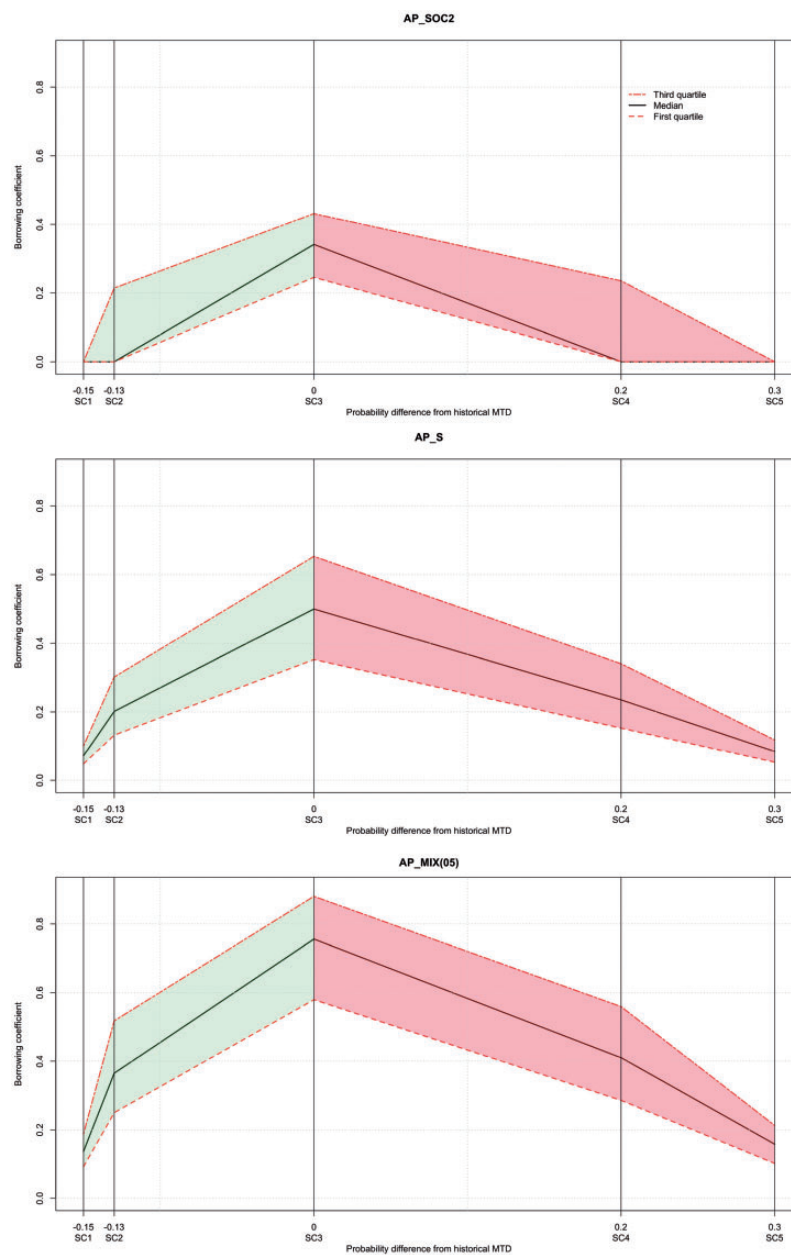
**Table 2.** Results for each method and each scenario in terms of the percentage of dose selection at the end of the trial, the PCS in bold and median number of DLTs, along with the first and the third quartiles. At the beginning of each scenario section, the true probabilities used for the scenario simulation are displayed. All methods were provided with stopping rules for scenario 6.

Method	% dose selection						DLTs Median (25q, 75q)	% dose selection						DLTs Median (25q, 75q)	
	1	2	3	4	5	6		0	1	2	3	4	5		6
Scenario 1								Scenario 2							
$p_{tox}$	0.001	0.01	0.05	0.07	<b>0.2</b>	0.4		0.01	0.05	0.07	<b>0.2</b>	0.4	0.5		
P_NI	0	0	2	28	<b>54</b>	16	5 (5, 6)	0	1	25	<b>61</b>	12	1	6 (5, 7)	
P_ESS(10)	0	0	10	51	<b>32</b>	7	2 (2, 3)	0	0	36	<b>61</b>	3	0	3 (3, 4)	
P_ESS(30)	0	0	68	32	<b>0</b>	0	1 (1, 2)	0	0	84	<b>16</b>	0	0	2 (1, 3)	
AP_L	0	0	8	37	<b>39</b>	16	4 (4, 5)	0	0	45	<b>48</b>	6	1	5 (4, 6)	
AP_S	0	0	6	32	<b>47</b>	15	5 (4, 5)	0	0	37	<b>54</b>	8	0	5 (4, 6)	
AP_MIX(0.5)	0	0	6	34	<b>46</b>	14	5 (4, 5)	0	0	39	<b>53</b>	7	1	5 (4, 6)	
AP_SOC1	0	0	5	26	<b>52</b>	17	5 (5, 6)	0	0	38	<b>50</b>	11	1	5 (4, 6)	
AP_SOC2	0	0	3	28	<b>52</b>	17	5 (5, 6)	0	0	30	<b>58</b>	11	1	5 (5, 6)	
AP_EB	0	0	9	34	<b>43</b>	14	5 (4, 5)	0	0	46	<b>47</b>	6	1	5 (4, 6)	
BCRM	0	0	16	45	<b>38</b>	1	3 (2, 4)	0	1	51	<b>44</b>	3	0	4 (3, 5)	
Scenario 3								Scenario 4							
$p_{tox}$	0.05	0.07	<b>0.2</b>	0.4	0.5	0.55		0.07	<b>0.2</b>	0.4	0.5	0.55	0.65		
P_NI	1	18	<b>70</b>	11	0	0	6 (5, 7)	17	<b>68</b>	15	0	0	0	6 (6, 8)	
P_ESS(10)	0	12	<b>84</b>	4	0	0	6 (5, 7)	10	<b>73</b>	17	0	0	0	8 (7, 8)	
P_ESS(30)	0	4	<b>95</b>	0	0	0	6 (4, 7)	1	<b>64</b>	35	0	0	0	9 (8, 10)	
AP_L	0	9	<b>87</b>	4	0	0	6 (5, 7)	11	<b>61</b>	28	0	0	0	7 (6, 8)	
AP_S	0	11	<b>84</b>	5	0	0	6 (5, 7)	15	<b>63</b>	22	0	0	0	7 (6, 8)	
AP_MIX(0.5)	0	10	<b>85</b>	4	0	0	6 (5, 7)	14	<b>62</b>	24	0	0	0	7 (6, 8)	
AP_SOC1	1	10	<b>85</b>	5	0	0	6 (5, 7)	18	<b>58</b>	24	0	0	0	7 (6, 8)	
AP_SOC2	1	11	<b>80</b>	8	0	0	6 (5, 7)	18	<b>62</b>	19	0	0	0	7 (6, 8)	
AP_EB	0	6	<b>90</b>	4	0	0	6 (5, 7)	12	<b>54</b>	34	0	0	0	8 (7, 9)	
BCRM	1	22	<b>73</b>	4	0	0	5 (4, 6)	21	<b>68</b>	10	0	0	0	6 (5, 7)	
Scenario 5								Scenario 6							
$p_{tox}$	<b>0.2</b>	0.4	0.5	0.55	0.65	0.7		0.35	0.45	0.5	0.6	0.7	0.8		
P_NI	<b>86</b>	14	1	0	0	0	8 (7, 9)	<b>88</b>	10	2	0	0	0	9 (8, 10)	
P_ESS(10)	<b>76</b>	23	1	0	0	0	9 (8, 10)	<b>54</b>	39	7	0	0	0	10 (9, 11)	
P_ESS(30)	<b>40</b>	57	3	0	0	0	12 (11, 13)	<b>7</b>	54	37	2	0	0	13 (12, 14)	
AP_L	<b>77</b>	22	1	0	0	0	8 (7, 9)	<b>86</b>	10	3	1	0	0	9 (8, 10.25)	
AP_S	<b>79</b>	20	1	0	0	0	8 (7, 9)	<b>86</b>	10	2	1	0	0	9 (8, 10)	
AP_MIX(0.5)	<b>79</b>	20	1	0	0	0	8 (7, 9)	<b>86</b>	10	3	1	0	0	9 (8, 10)	
AP_SOC1	<b>86</b>	13	1	0	0	0	8 (7, 9)	<b>88</b>	10	1	1	0	0	9 (8, 10)	
AP_SOC2	<b>86</b>	13	1	0	0	0	8 (7, 9)	<b>88</b>	10	2	1	0	0	9 (8, 10)	
AP_EB	<b>76</b>	21	3	0	0	0	8 (7, 9)	<b>85</b>	11	3	2	0	0	9 (8, 11)	
BCRM	<b>90</b>	10	0	0	0	0	7 (6, 9)	<b>81</b>	18	1	0	0	0	9 (8, 9)	

DLT: dose-limiting toxicity; PCS: percentage of correct selection.

the data are not in conflict. Our method has been shown to be able to detect data conflict during sequential adaptive trials and to add it when there is a strong reliable belief in its usefulness. To note, our method is tailored to check if the prospective trial is similar to a fixed historical dataset, not if both datasets come from the same population, that is, if both trials have the same generating probabilities.

We proposed several ways of using  $\pi^{APP}$  and the amount of historical information that could be used in the analysis of clinical trials. In our setting, the best method, on average, was AP\_SOC2, but in other settings, another choice could be better. Therefore, a sensitivity analysis should always be performed before the study onset. The Occam's window threshold on  $\alpha$  (the total borrowing) was set to 0.2, that is, if the total borrowing is less than 0.2, we set it to zero. Therefore, a great variability in SC2 is shown, since all numbers below 0.2 are set to zero by the method. Otherwise, we have an increasing alpha and then, after approximately 20-22 patients, a slower decrease.



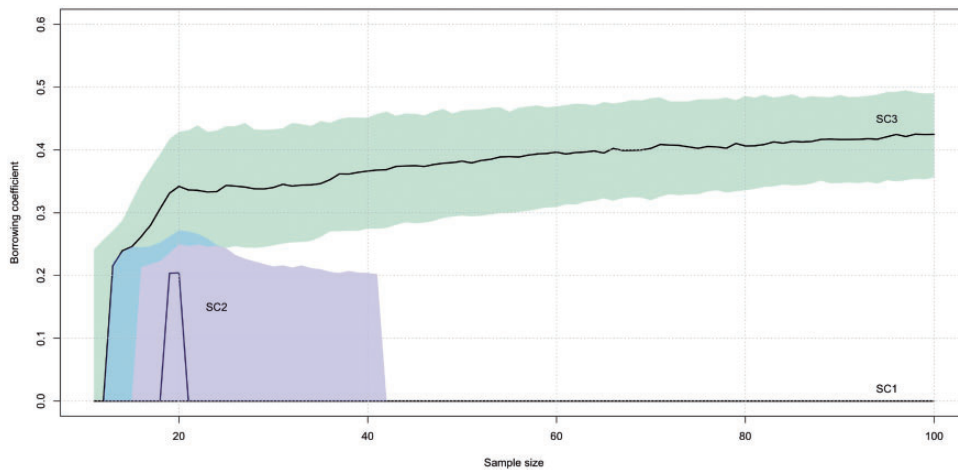
**Figure 2.** Evolution of the amount of borrowing value, that is,  $\alpha = \alpha_0(1 - \gamma)$ , at the end of the trial, for models AP\_SOC2, AP\_S and AP\_MIX(05) in all five scenarios. On the x-axis, scenarios (SC) are represented by their difference in the probability of toxicity at dose 3 with respect to the historical data (assumed to be 0.20). On the right, more toxic scenarios than SC3 are plotted.

The effect of the Occam's window is a strong cut and we “feel” a greater variability than we expect in SC2. In our point of view, large decrease in PCS, in scenarios where the actual population is different from the previous one, should have more weight in the decision of which method to use. As usual, there is a trade-off between benefit and risk.

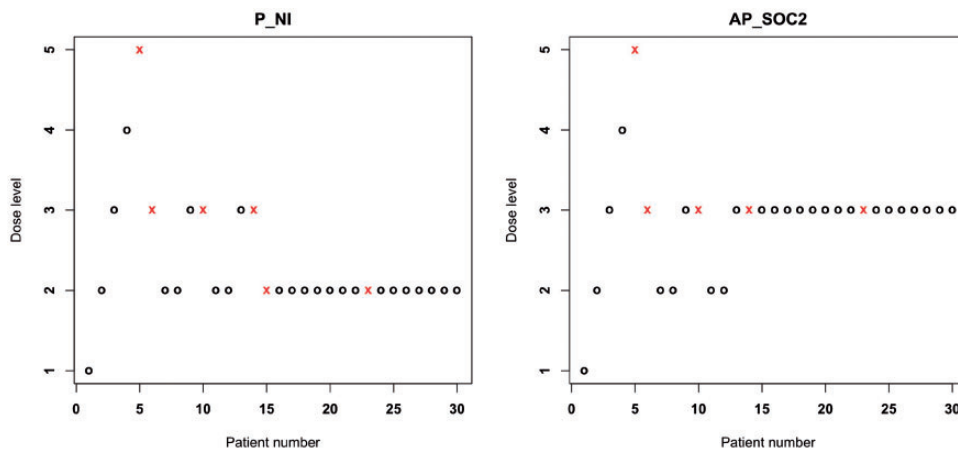
From our simulations, the amount of borrowing value was, on average, between 0.3 and 0.5 according to the method when the data were not in conflict. As the sample size of the clinical study was fixed at 30, which represents the additional information of 10 to 15 patients, the overall trial accuracy increased.

We compared our propositions to the Empirical Bayes approach and to the BCRM. It has been shown that the EB approach works best when there are no data conflict. However, in situations where there is major conflict between the data, it fails to detect that no borrowing is necessary. A way of improving its performance would be to incorporate the EB approach in our  $\gamma$  parameter, such as  $\gamma = \text{argmax}(m(D, D_0, \alpha))$ . However, it could





**Figure 3.** Evolution of the  $\alpha$  value in model AP\_SOC2 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).



**Figure 4.** 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.

experience computational issues in optimisation when moving in multidimensional parameter space. Even if equation (2) involves three possibly multidimensional integrals, where the dimension depends on the length of the vector  $\theta$ , it can be easily approximated using Monte Carlo approaches. First, to reduce computational time, we suggest an MCMC approach, where the two pseudo distributions,  $L(\theta|D)/\int L(\theta|D)d\theta$  and  $L(\theta|D_0)^{n/n_0}/\int L(\theta|D_0)^{n/n_0}d\theta$ , can be obtained as Bayesian posterior distributions, for example, by setting an improper uniform prior on  $\theta$  and forcing the likelihoods to be  $L(\theta|D)$  and  $L(\theta|D_0)^{n/n_0}$ , respectively. Then, the final integral of the distance in equation (2) can be computed using the Monte Carlo approach and the density kernel estimation from the previous sampling results. To the best of our knowledge, this approach works well in up to six dimensions and is robust. Regarding the BCRM, its aim is to select a proper skeleton for the prospective trial and not to increase the information used in the analysis. Therefore, the comparison is not fair in our setting, since it is made to maintain a constant PCS and to not increase in the case of no data conflict between populations. In our case, we experienced some issue in computing the skeleton. The final proposed skeleton had two couples of doses with the same probability of toxicity. Changing the skeleton and/or adding our method as one of the models inside the BMA procedure in the BCRM could increase the performance.

Our method is based on several parametrisation choices. First, we used the ESS to tune the quantity of information used from the historical data. In this case, we fixed a threshold that reflects the maximum amount

of hypothetical patients who are incorporated into the prior with respect to the number of patients who will be included in the present trial. In our case, for the AP\_SOC2 method, we simulated a phase I clinical trial with 30 patients and we fixed the maximum ESS threshold to 20. Even if our method is built to add information only when necessary, increasing the threshold can decrease the PCS when data are in conflict, as shown above. However, this threshold is already high, as it already allows the incorporation of 67% (20/30) additional information.

Second, the value of  $\gamma$  is defined between zero and one. We chose to parametrise it using the Hellinger distance  $d(\cdot)$  or its square root. As proposed in the methods section,  $\gamma$  can also be parametrised as  $d^c(D_0, D) \in [0, 1] \forall c \in \mathbb{R}^+$ . If  $c$  is higher than 1, it will reduce the computed distance and will lead to more borrowing, while a value less than 1 will lead to a more conservative approach with an increase in the computed distance. Trial statisticians can decide to use other possibilities for  $\gamma$ , but an extensive sensitivity analysis is required.

Third, the method is based on the power prior approach and, therefore, on some notion of exchangeability.<sup>34</sup> When the two datasets can be seen as a realisation from the same process, we would assume full exchangeability; that is, the two datasets would have the same parameter, but not otherwise. Many models, such as hierarchical models and the power prior, are based on the assumption of exchangeability. As pointed out by Psioda et al.,<sup>35</sup> we prefer to see historical data that give the prior as non-random. In this point of view, the prior simply reflects the previous knowledge, which can be near the truth or not with respect to the new prospective trial. Therefore, even though exchangeability is an intrinsic part of the design, whether the patients in the two trials are actually exchangeable is not important when evaluating the impact of borrowing the prior information on the PCS.

Our approach is not limited to early phase dose-finding clinical trials. It can be generalised to any adaptive sequential method and trial phase. It can adapt the amount of the historical information that could be used in any interim analysis. Of course, some choices need to be fixed before the beginning of the trial, such as the ESS, the Occam's window threshold and the parametrisation of  $\gamma$ . The generalisation of our method is straightforward because it does not depend on the number of interim analyses or on the type of outcomes. Then, it should be noticed that the term "fixed historical dataset" does not mean necessarily a unique dataset; this can be extended either for the averaging of several datasets or for a meta-analysis of several phase I datasets in order to take into account the inter- and intra-trials' variability.<sup>36</sup>

Being able to incorporate external data into small sample trials is an advantage when the data are not in conflict. Proposing a full borrowing when not appropriate will conduct to a wrong choice of the MTD in the majority of cases. Full borrowing should not be considered, unless there is a strong evidence of similarity. Our method was shown to be able to detect conflict and avoid it. This method will help investigators and clinical trial statisticians use historical data without the fear of adding bias to the results when the data are in conflict. We believe that such approaches should be further developed in the future.

### Authors' note

Codes which are a part of the supplemental material will be available at the GitHub of the corresponding author.

### Acknowledgements

The authors would like to thank Dr Emmanuelle Comets for her precious advice and the two anonymous reviewers whose comments/suggestions improved this article.

### Declaration of conflicting interests

The author(s) declare no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

### Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: The research of Adrien Ollier and Moreno Ursino was funded by the Institut National Du Cancer, grant numbers INCa\_11324 and INCa\_9539, respectively.

### ORCID iD

Moreno Ursino  <https://orcid.org/0000-0002-5709-4322>

## Supplemental material

Supplemental material for this article is available online.

## References

1. Hobbs BP, Carlin BP, Mandrekar SJ et al. Hierarchical commensurate and power prior models for adaptive incorporation of historical information in clinical trials. *Biometrics* 2011; **67**: 1047–1056.
2. Röver C and Friede T. Dynamically borrowing strength from another study through shrinkage estimation. *Stat Meth Med Res* 2019; 10.1177/0962280219833079.
3. van Rosmalen J, Dejardin D, van Norden Y et al. Including historical data in the analysis of clinical trials: is it worth the effort? *Stat Meth Med Res* 2018; **27**: 3167–3182.
4. Morita S, Thall PF and Muller P. Determining the effective sample size of a parametric prior. *Biometrics* 2008; **64**: 595–602.
5. Morita S, Thall PF and Muller P. Evaluating the impact of prior assumptions in Bayesian biostatistics. *Stat Biosci* 2010; **2**: 1–17.
6. Morita S, Thall PF and Muller P. Prior effective sample size in conditionally independent hierarchical models. *Bayesian Anal* 2012; **7**: 591–614.
7. Pocock SJ. The combination of randomized and historical controls in clinical trials. *J Chronic Dis* 1976; **29**: 175–188.
8. Ibrahim JG and Chen MH. Power prior distributions for regression models. *Stat Sci* 2000; **15**: 46–60.
9. Neuenschwander B, Branson M and Spiegelhalter D. A note on the power prior. *Stat Med* 2009; **28**: 3562–3566.
10. Ibrahim JG, Chen MH, Gwon Y et al. The power prior: theory and applications. *Stat Med* 2015; **34**: 3724–3749.
11. Hobbs BP, Sargent DJ and Carlin BP. Commensurate priors for incorporating historical information in clinical trials using general and generalized linear models. *Bayesian Anal* 2012; **7**: 639–674.
12. Nikolakopoulos S, van der Tweel I and Roes KC. Dynamic borrowing through empirical power priors that control type I error. *Biometrics* 2018; **74**: 874–880.
13. Liu GF. A dynamic power prior for borrowing historical data in noninferiority trials with binary endpoint. *Pharm Stat* 2018; **17**: 61–73.
14. Haddad T, Himes A, Thompson L, et al. Incorporation of stochastic engineering models as prior information in Bayesian medical device trials. *J Biopharm Stat* 2017; **27**: 1089–1103.
15. Schmidli H, Gsteiger S, Roychoudhury S et al. Robust meta-analytic-predictive priors in clinical trials with historical control information. *Biometrics* 2014; **70**: 1023–1032.
16. O’Quigley J and Iasonos A. Bridging solutions in dose finding problems. *Stat Biopharm Res* 2014; **6**: 185–197.
17. Zohar S, Baldi I, Forni G et al. Planning a Bayesian early-phase phase I/II study for human vaccines in HER2 carcinomas. *Pharm Stat* 2011; **10**: 218–226.
18. Liu S, Pan H, Xia J et al. Bridging continual reassessment method for phase I clinical trials in different ethnic populations. *Stat Med* 2015; **34**: 1681–1694.
19. Takeda K and Morita S. Incorporating historical data in Bayesian phase I trial design: the Caucasian-to-Asian toxicity tolerability problem. *Ther Innov Regul Sci* 2015; **49**: 93–99.
20. Petit C, Samson A, Morita S et al. Unified approach for extrapolation and bridging of adult information in early-phase dose-finding paediatric studies. *Stat Meth Med Res* 2018; **27**: 1860–1877.
21. ICH. Ethnic factors in the acceptability of foreign clinical data, [www.ich.org/products/guidelines/efficacy/efficacy-single/article/ethnic-factors-in-the-acceptability-of-foreign-clinical-data.html](http://www.ich.org/products/guidelines/efficacy/efficacy-single/article/ethnic-factors-in-the-acceptability-of-foreign-clinical-data.html) (1998, accessed 5 March 2019).
22. Maeda H and Kurokawa T. Differences in maximum tolerated doses and approval doses of molecularly targeted oncology drug between Japan and Western countries. *Invest New Drugs* 2014; **32**: 661–669.
23. Yasuda SU, Zhang L and Huang SM. The role of ethnicity in variability in response to drugs: focus on clinical pharmacology studies. *Clin Pharmacol Ther* 2008; **84**: 417–423.
24. Reigner B, Watanabe T, Schüller J et al. Pharmacokinetics of capecitabine (Xeloda) in Japanese and Caucasian patients with breast cancer. *Cancer Chemother Pharmacol* 2003; **52**: 193–201.
25. Pan H, Yuan Y and Xia J. A calibrated power prior approach to borrow information from historical data with application to biosimilar clinical trials. *J R Stat Soc: Series C (Applied Statistics)* 2017; **66**: 979–996.
26. Zohar S, Resche-Rigon M and Chevret S. Using the continual reassessment method to estimate the minimum effective dose in phase II dose-finding studies: a case study. *Clin Trials* 2013; **10**: 414–421.
27. Cheung YK. *Dose finding by the continual reassessment method*. New York: Chapman and Hall/CRC, 2011.

28. Hoeting JA, Madigan D, Raftery AE et al. Bayesian model averaging: a tutorial (with comments by M. Clyde, David Draper and E. I. George, and a rejoinder by the authors. *Statist Sci* 1999; **14**: 382–417.
29. Yin G and Yuan Y. Bayesian model averaging continual reassessment method in phase I clinical trials. *J Am Stat Assoc* 2009; **104**: 954–968.
30. Gravestock, Isaac and Held, Leonhard and COMBACTE-Net consortium. Adaptive power priors with empirical Bayes for clinical trials. *Pharm Stat* 2017; **16**: 349–360.
31. R Core Team. *R: a language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing, 2013.
32. Stan Development Team. RStan: the R interface to Stan, <http://mc-stan.org/>. R package version 2.17.3 (2018, accessed 5 March 2019).
33. Duong T. *ks: Kernel Smoothing*, <https://CRAN.R-project.org/package=ks>. R package version 1.11.3. (2018, accessed 5 March 2019).
34. Bernardo JM. The concept of exchangeability and its applications. *Far East J Math Sci* 1996; **4**: 111–122.
35. Psioda MA, Soukup M and Ibrahim JG. A practical Bayesian adaptive design incorporating data from historical controls. *Stat Med* 2018; **37**: 4054–4070.
36. Zohar S, Katsahian S and O’Quigley J. An approach to meta-analysis of dose-finding studies. *Stat Med* 2011; **30**: 2109–2116.

---

# Supplementary material of the paper entitled “Adaptive power prior for sequential clinical trials - application to bringing studies in early phase trials”

Journal Title

XX(X):1–11

© The Author(s) 2016

Reprints and permission:

sagepub.co.uk/journalsPermissions.nav

DOI: 10.1177/ToBeAssigned

www.sagepub.com/

SAGE

Adrien Ollier<sup>1</sup>, Satoshi Morita<sup>2</sup>, Moreno Ursino<sup>1,\*</sup> and Sarah Zohar<sup>1,\*</sup>

## 1 Full tables

Table 1 represents the full table results for scenario 1-5, while Table 2 show the full results in scenario 6, where we added the same stopping rules to all methods,  $P(p_1 > \tau) > 0.9$ , that is the trial is stopped if the posterior probability of toxicity that the first dose exceeds the toxicity target is higher than 0.9.

---

<sup>1</sup>Centre de Recherche des Cordeliers, INSERM, Sorbonne Université, USPC, Université Paris Descartes, Université Paris Diderot, F-75006 Paris, France.

<sup>2</sup>Department of Biomedical Statistics and Bioinformatics, Kyoto University Graduate School of Medicine, Kyoto, Japan.

\*Co-last authors: Moreno Ursino and Sarah Zohar made equal contributions.

### Corresponding author:

Centre de Recherche des Cordeliers, INSERM, Sorbonne Université, USPC, Université Paris Descartes, Université Paris Diderot, 15 rue de l'Ecole de Médecine, 75006, Paris, France.

Email: moreno.ursino@inserm.fr

Method	% dose selection						% dose allocation						DLTs median (25q, 75q)
	1	2	3	4	5	6	1	2	3	4	5	6	
Scenario 1													
Prob of toxicity	0.001	0.01	0.05	0.07	<b>0.2</b>	0.4							
P_NI	0	0	0.02	0.277	<b>0.541</b>	0.162	0.037	0.042	0.097	0.281	0.331	0.212	5 (5, 6)
P_ESS(10)	0	0	0.096	0.511	<b>0.324</b>	0.069	0.033	0.037	0.379	0.423	0.118	0.01	2 (2, 3)
P_ESS(30)	0	0	0.68	0.32	<b>0</b>	0	0.033	0.033	0.857	0.077	0	0	1 (1, 2)
AP_L	0	0	0.081	0.368	<b>0.386</b>	0.165	0.037	0.041	0.158	0.351	0.248	0.164	4 (4, 5)
AP_S	0	0	0.056	0.325	<b>0.469</b>	0.15	0.037	0.042	0.136	0.332	0.276	0.177	5 (4, 5)
AP_MIX(0.5)	0	0	0.062	0.338	<b>0.461</b>	0.139	0.037	0.042	0.141	0.336	0.268	0.176	5 (4, 5)
AP_SOC1	0	0	0.048	0.262	<b>0.517</b>	0.173	0.037	0.042	0.115	0.265	0.33	0.211	5 (5, 6)
AP_SOC2	0	0	0.033	0.278	<b>0.515</b>	0.174	0.037	0.042	0.113	0.268	0.33	0.211	5 (5, 6)
AP_EB	0	0	0.09	0.34	<b>0.433</b>	0.137	0.037	0.041	0.176	0.321	0.258	0.166	5 (4, 5)
BCRM	0	0	0.161	0.446	<b>0.383</b>	0.01	0.037	0.046	0.185	0.382	0.33	0.02	3 (2, 4)
Scenario 2													
Prob of toxicity	0.01	0.05	0.07	<b>0.2</b>	0.4	0.5							
P_NI	0	0.006	0.253	<b>0.614</b>	0.119	0.008	0.047	0.066	0.278	0.42	0.128	0.06	6 (5, 7)
P_ESS(10)	0	0.005	0.355	<b>0.611</b>	0.028	0.001	0.034	0.05	0.54	0.364	0.011	0	3 (3, 4)
P_ESS(30)	0	0	0.845	<b>0.155</b>	0	0	0.033	0.034	0.897	0.036	0	0	2 (1, 3)
AP_L	0	0.004	0.454	<b>0.48</b>	0.055	0.007	0.047	0.057	0.431	0.339	0.078	0.048	5 (4, 6)
AP_S	0	0.004	0.374	<b>0.537</b>	0.08	0.005	0.047	0.059	0.371	0.379	0.091	0.052	5 (4, 6)
AP_MIX(0.5)	0	0.004	0.386	<b>0.532</b>	0.072	0.006	0.047	0.058	0.384	0.371	0.088	0.052	5 (4, 6)
AP_SOC1	0	0.003	0.385	<b>0.497</b>	0.108	0.007	0.047	0.059	0.361	0.348	0.125	0.06	5 (4, 6)
AP_SOC2	0	0.004	0.3	<b>0.58</b>	0.109	0.007	0.047	0.059	0.347	0.362	0.125	0.06	5 (5, 6)
AP_EB	0	0.003	0.455	<b>0.472</b>	0.064	0.006	0.047	0.054	0.459	0.304	0.086	0.05	5 (4, 6)
BCRM	0	0.009	0.514	<b>0.444</b>	0.033	0	0.049	0.075	0.427	0.366	0.082	0.001	4 (3, 5)
Scenario 3													
Prob of toxicity	0.05	0.07	<b>0.2</b>	0.4	0.5	0.55							
P_NI	0.006	0.184	<b>0.7</b>	0.107	0.003	0	0.085	0.215	0.496	0.151	0.032	0.021	6 (5, 7)
P_ESS(10)	0	0.12	<b>0.837</b>	0.043	0	0	0.036	0.156	0.757	0.052	0	0	6 (5, 7)
P_ESS(30)	0	0.045	<b>0.954</b>	0.001	0	0	0.033	0.059	0.907	0	0	0	6 (4, 7)
AP_L	0.001	0.087	<b>0.872</b>	0.04	0	0	0.076	0.154	0.632	0.097	0.024	0.018	6 (5, 7)
AP_S	0.004	0.108	<b>0.836</b>	0.049	0.002	0.001	0.078	0.168	0.602	0.107	0.027	0.019	6 (5, 7)
AP_MIX(0.5)	0.004	0.097	<b>0.853</b>	0.044	0.001	0.001	0.077	0.163	0.611	0.103	0.026	0.019	6 (5, 7)
AP_SOC1	0.007	0.095	<b>0.847</b>	0.048	0.003	0	0.085	0.159	0.588	0.115	0.032	0.02	6 (5, 7)
AP_SOC2	0.007	0.11	<b>0.801</b>	0.079	0.003	0	0.085	0.162	0.582	0.119	0.032	0.02	6 (5, 7)
AP_EB	0.003	0.057	<b>0.904</b>	0.035	0.001	0	0.076	0.119	0.675	0.086	0.025	0.018	6 (5, 7)
BCRM	0.01	0.217	<b>0.729</b>	0.043	0	0.001	0.1	0.244	0.538	0.096	0.021	0.001	5 (4, 6)
Scenario 4													
Prob of toxicity	0.07	<b>0.2</b>	0.4	0.5	0.55	0.65							
P_NI	0.166	<b>0.681</b>	0.151	0.002	0	0	0.263	0.473	0.209	0.037	0.012	0.006	6 (6, 8)
P_ESS(10)	0.099	<b>0.732</b>	0.169	0	0	0	0.115	0.561	0.321	0.003	0	0	8 (7, 8)
P_ESS(30)	0.008	<b>0.645</b>	0.347	0	0	0	0.036	0.378	0.586	0	0	0	9 (8, 10)
AP_L	0.107	<b>0.609</b>	0.284	0	0	0	0.196	0.451	0.309	0.03	0.01	0.005	7 (6, 8)
AP_S	0.151	<b>0.632</b>	0.216	0.001	0	0	0.22	0.457	0.275	0.033	0.01	0.005	7 (6, 8)
AP_MIX(0.5)	0.14	<b>0.618</b>	0.241	0.001	0	0	0.213	0.456	0.284	0.032	0.01	0.005	7 (6, 8)
AP_SOC1	0.182	<b>0.578</b>	0.237	0.003	0	0	0.268	0.408	0.273	0.034	0.011	0.006	7 (6, 8)
AP_SOC2	0.179	<b>0.624</b>	0.193	0.004	0	0	0.268	0.414	0.266	0.034	0.011	0.006	7 (6, 8)
AP_EB	0.117	<b>0.54</b>	0.343	0	0	0	0.202	0.37	0.383	0.029	0.01	0.005	8 (7, 9)
BCRM	0.206	<b>0.685</b>	0.105	0.004	0	0	0.306	0.473	0.177	0.037	0.007	0	6 (5, 7)
Scenario 5													
Prob of toxicity	<b>0.2</b>	0.4	0.5	0.55	0.65	0.7							
P_NI	<b>0.859</b>	0.135	0.006	0	0	0	0.725	0.195	0.058	0.015	0.005	0.002	8 (7, 9)
P_ESS(10)	<b>0.764</b>	0.229	0.007	0	0	0	0.499	0.384	0.116	0.001	0	0	9 (8, 10)
P_ESS(30)	<b>0.399</b>	0.568	0.033	0	0	0	0.152	0.543	0.306	0	0	0	12 (11, 13)
AP_L	<b>0.767</b>	0.219	0.014	0	0	0	0.639	0.265	0.076	0.013	0.004	0.002	8 (7, 9)
AP_S	<b>0.791</b>	0.196	0.013	0	0	0	0.671	0.239	0.07	0.014	0.004	0.002	8 (7, 9)
AP_MIX(0.5)	<b>0.789</b>	0.198	0.013	0	0	0	0.661	0.249	0.069	0.014	0.004	0.002	8 (7, 9)
AP_SOC1	<b>0.858</b>	0.128	0.014	0	0	0	0.722	0.187	0.069	0.015	0.005	0.002	8 (7, 9)
AP_SOC2	<b>0.857</b>	0.129	0.014	0	0	0	0.722	0.187	0.069	0.015	0.005	0.002	8 (7, 9)
AP_EB	<b>0.764</b>	0.208	0.028	0	0	0	0.634	0.243	0.102	0.013	0.004	0.002	8 (7, 9)
BCRM	<b>0.899</b>	0.099	0.002	0	0	0	0.775	0.158	0.049	0.016	0.002	0	7 (6, 9)

**Table 1.** Results for each method and each scenario (1-5) in terms of the percentage of dose selection at the end of the trial, the percentage of dose allocation and median number of DLTs, along with the first and the third quartiles. At the beginning of each scenario section, the true probabilities used for the scenario simulation are displayed.

Method	% dose selection							% dose allocation						DLTs median (25q, 75q)
	0	1	2	3	4	5	6	1	2	3	4	5	6	
Scenario 6														
Prob of toxicity		0.3	0.4	0.5	0.6	0.7	0.8							
P_NI	<b>0.877</b>	0.104	0.016	0.003	0	0	0	0.629	0.213	0.114	0.032	0.009	0.003	9 (8, 10)
P_ESS(10)	<b>0.538</b>	0.389	0.069	0.004	0	0	0	0.57	0.319	0.11	0.001	0	0	10 (9, 11)
P_ESS(30)	<b>0.071</b>	0.535	0.37	0.024	0	0	0	0.235	0.504	0.261	0	0	0	13 (12, 14)
AP_L	<b>0.861</b>	0.1	0.032	0.007	0	0	0	0.568	0.258	0.135	0.027	0.009	0.003	9 (8, 10.25)
AP_S	<b>0.864</b>	0.104	0.025	0.007	0	0	0	0.586	0.245	0.129	0.029	0.009	0.003	9 (8, 10)
AP_MIX(0.5)	<b>0.865</b>	0.101	0.027	0.007	0	0	0	0.588	0.249	0.123	0.028	0.009	0.003	9 (8, 10)
AP_SOC1	<b>0.878</b>	0.101	0.012	0.009	0	0	0	0.624	0.201	0.131	0.031	0.009	0.003	9 (8, 10)
AP_SOC2	<b>0.878</b>	0.1	0.015	0.007	0	0	0	0.624	0.202	0.131	0.031	0.009	0.003	9 (8, 10)
AP_EB	<b>0.852</b>	0.107	0.026	0.015	0	0	0	0.559	0.237	0.165	0.028	0.009	0.003	9 (8, 11)
BCRM	<b>0.806</b>	0.185	0.009	0	0	0	0	0.768	0.14	0.068	0.022	0.003	0	9 (8, 9)

**Table 2.** Comparison of the proposed methods in a toxic scenario.

## 2 ESS sensitivity analysis

We compare, when  $\gamma = 0$ , the linear ESS, that is  $s(n) = n$ , with respect a sigmoid ESS, computed as

$$s(n) = \min \left( 30, \left[ \frac{31}{1 + \exp(-0.25(n - 15))} \right] \right),$$

where  $[x]$  refers to the integer part of  $x$ . Results, shown in Table 3, suggest not big difference between them in percentage of correct selection.

Method	% dose selection						% dose allocation						DLTs median (25q, 75q)	
	1	2	3	4	5	6	1	2	3	4	5	6		
Scenario 1														
Prob of toxicity	0.001	0.01	0.05	0.07	<b>0.2</b>	0.4								
ESS_LIN	0	0	0.299	0.701	<b>0</b>	0	0.034	0.04	0.324	0.602	0	0	2 (1, 3)	
ESS_SIG	0	0	0.312	0.688	<b>0</b>	0	0.035	0.04	0.327	0.488	0.11	0	2 (1, 3)	
Scenario 2														
Prob of toxicity	0.01	0.05	0.07	<b>0.2</b>	0.4	0.5								
ESS_LIN	0	0.001	0.802	<b>0.197</b>	0	0	0.038	0.051	0.651	0.26	0	0	3 (2, 4)	
ESS_SIG	0	0.002	0.782	<b>0.216</b>	0	0	0.041	0.054	0.661	0.197	0.048	0	3 (2, 4)	
Scenario 3														
Prob of toxicity	0.05	0.07	<b>0.2</b>	0.4	0.5	0.55								
ESS_LIN	0	0.032	<b>0.966</b>	0.002	0	0	0.047	0.122	0.77	0.061	0	0	6 (5, 7)	
ESS_SIG	0	0.035	<b>0.96</b>	0.005	0	0	0.056	0.126	0.744	0.058	0.017	0	6 (5, 7)	
Scenario 4														
Prob of toxicity	0.07	<b>0.2</b>	0.4	0.5	0.55	0.65								
ESS_LIN	0.003	<b>0.599</b>	0.398	0	0	0	0.079	0.5	0.395	0.025	0	0	8 (7, 9)	
ESS_SIG	0.004	<b>0.594</b>	0.402	0	0	0	0.097	0.462	0.406	0.028	0.007	0	8 (7, 9)	
Scenario 5														
Prob of toxicity	<b>0.2</b>	0.4	0.5	0.55	0.65	0.7								
ESS_LIN	<b>0.256</b>	0.717	0.027	0	0	0	0.327	0.552	0.11	0.011	0	0	10 (9, 11)	
ESS_SIG	<b>0.296</b>	0.672	0.032	0	0	0	0.314	0.565	0.105	0.013	0.002	0	11 (9, 11)	

**Table 3.** Comparison of two ESS shapes: linear (ESS\_LIN) vs sigmoidal (ESS\_SIG).



### 3 Mixture priors and Bayesian Model Averaging

We performed a sensitivity analysis for the mixture prior, AP\_MIX( $\omega$ ) choosing several  $\omega$ . Then, for the Bayesian Model Averaging technique (BMA), the two models involved are AP\_L, as  $M_1$ , and the AP\_NI, as  $M_2$ . We consider the posterior distribution  $p(\theta|D) = P(\theta|M_1, D)P(M_1|D) + P(\theta|M_2, D)P(M_2|D)$  as a mixture of posterior distribution with  $P(\theta|M_1, D)$  the posterior with *adaptive power prior* and  $P(\theta|M_2, D)$  the posterior with vague prior for the parameter of interest.  $M_1$  and  $M_2$  are called model 1 and 2. The posterior probability of each model is  $P(M_i, D) = \frac{P(D|M_i)P(M_i)}{\sum_{i=1}^2 P(D|M_i)P(M_i)}$  with  $P(D|M_i) = \int P(D/\theta_i, M_i)P(\theta_i|M_i)d\theta_i$  the integrated likelihood of model  $M_i$ ,  $\theta_i$  the parameter of model  $M_i$ ,  $P(\theta_i|M_i)$  the prior of  $\theta_i$  for  $M_i$ ,  $P(D/\theta_i, M_i)$  the likelihood and  $P(M_i)$  the prior probability that  $M_i$  is the good model, set as 0.5.

For all models, we set ESS linear, that is  $s(n) = n$ , and  $\gamma = \sqrt{d^2(D_0, D)}$ , except for AP\_BMANG where  $\gamma = 0$ . Results are shown in Fig 4.

Method	% dose selection						% dose allocation						DLTs median (25q, 75q)
	1	2	3	4	5	6	1	2	3	4	5	6	
Scenario 1													
Prob of toxicity	0.001	0.01	0.05	0.07	<b>0.2</b>	0.4							
AP_MIX(0.3)	0	0	0.055	0.32	<b>0.467</b>	0.158	0.037	0.042	0.129	0.325	0.282	0.185	5 (4, 6)
AP_MIX(0.8)	0	0	0.078	0.34	<b>0.424</b>	0.158	0.037	0.042	0.153	0.346	0.254	0.168	5 (4, 5)
AP_MIX(0.9)	0	0	0.08	0.358	<b>0.397</b>	0.165	0.037	0.042	0.156	0.35	0.25	0.165	5 (4, 5)
AP_BMA	0	0	0.063	0.337	<b>0.457</b>	0.143	0.037	0.042	0.14	0.336	0.27	0.175	5 (4, 5)
AP_BMANG	0	0	0.117	0.343	<b>0.367</b>	0.173	0.037	0.041	0.172	0.363	0.222	0.164	4 (4, 5)
Scenario 2													
Prob of toxicity	0.01	0.05	0.07	<b>0.2</b>	0.4	0.5							
AP_MIX(0.3)	0	0.003	0.354	<b>0.564</b>	0.069	0.01	0.047	0.06	0.354	0.388	0.097	0.054	5 (4, 6)
AP_MIX(0.8)	0	0.004	0.431	<b>0.491</b>	0.061	0.013	0.047	0.058	0.414	0.349	0.083	0.05	5 (4, 6)
AP_MIX(0.9)	0	0.004	0.441	<b>0.494</b>	0.05	0.011	0.047	0.057	0.422	0.347	0.079	0.049	5 (4, 6)
AP_BMA	0	0.004	0.402	<b>0.514</b>	0.075	0.005	0.047	0.059	0.384	0.37	0.088	0.052	5 (4, 6)
AP_BMANG	0	0.004	0.55	<b>0.391</b>	0.046	0.009	0.047	0.057	0.464	0.315	0.069	0.048	5 (4, 5)
Scenario 3													
Prob of toxicity	0.05	0.07	<b>0.2</b>	0.4	0.5	0.55							
AP_MIX(0.3)	0.003	0.11	<b>0.836</b>	0.049	0.002	0	0.078	0.174	0.589	0.111	0.027	0.019	6 (5, 7)
AP_MIX(0.8)	0.001	0.094	<b>0.873</b>	0.031	0.001	0	0.076	0.156	0.626	0.099	0.025	0.018	6 (5, 7)
AP_MIX(0.9)	0	0.084	<b>0.874</b>	0.042	0	0	0.075	0.156	0.629	0.098	0.024	0.018	6 (5, 7)
AP_BMA	0.003	0.091	<b>0.858</b>	0.045	0.002	0.001	0.077	0.163	0.611	0.104	0.026	0.019	6 (5, 7)
AP_BMANG	0	0.068	<b>0.907</b>	0.025	0	0	0.074	0.148	0.647	0.089	0.024	0.018	6 (5, 7)
Scenario 4													
Prob of toxicity	0.07	<b>0.2</b>	0.4	0.5	0.55	0.65							
AP_MIX(0.3)	0.154	<b>0.624</b>	0.221	0.001	0	0	0.225	0.459	0.27	0.031	0.01	0.005	7 (6, 8)
AP_MIX(0.8)	0.122	<b>0.597</b>	0.281	0	0	0	0.198	0.452	0.305	0.03	0.01	0.005	7 (6, 8)
AP_MIX(0.9)	0.109	<b>0.617</b>	0.274	0	0	0	0.197	0.454	0.305	0.03	0.009	0.005	7 (6, 8)
AP_BMA	0.14	<b>0.619</b>	0.24	0.001	0	0	0.213	0.455	0.285	0.032	0.01	0.005	7 (6, 8)
AP_BMANG	0.074	<b>0.595</b>	0.331	0	0	0	0.176	0.455	0.324	0.03	0.009	0.005	8 (7, 9)
Scenario 5													
Prob of toxicity	<b>0.2</b>	0.4	0.5	0.55	0.65	0.7							
AP_MIX(0.3)	<b>0.782</b>	0.204	0.014	0	0	0	0.681	0.23	0.068	0.014	0.005	0.002	8 (7, 9)
AP_MIX(0.8)	<b>0.783</b>	0.202	0.015	0	0	0	0.645	0.261	0.074	0.013	0.004	0.002	8 (7, 9)
AP_MIX(0.9)	<b>0.784</b>	0.198	0.018	0	0	0	0.644	0.262	0.074	0.013	0.004	0.002	8 (7, 9)
AP_BMA	<b>0.779</b>	0.206	0.015	0	0	0	0.659	0.25	0.071	0.014	0.004	0.002	8 (7, 9)
AP_BMANG	<b>0.703</b>	0.278	0.019	0	0	0	0.599	0.303	0.078	0.013	0.004	0.002	8 (8, 10)

**Table 4.** Performance comparison of mixture priors and BMA.

### 4 Occam window sensitivity analysis

We re-shaped the idea of the Occam windows to set a threshold on  $\alpha = \alpha_0(1 - \gamma)$  or on  $\gamma$ . In this way,  $\alpha = \alpha I(\alpha > \tau_\alpha)$ , where  $I$  is the indicator function. In other words,  $\alpha$  is set to zero if its value is lower than the pre-specified threshold. This model is denoted as OC\_A( $\tau_\alpha$ ). In the same manner, we can define  $\gamma = \gamma I(\gamma < \tau_\gamma) + I(\gamma \geq \tau_\gamma)$  and the model is called OC\_G( $\tau_\gamma$ ). For all models, we set ESS linear, that is  $s(n) = n$ , and  $\gamma = \sqrt{d^2(D_0, D)}$ . Results are shown in Fig 5.

Method	% dose selection						% dose allocation						DLTs median (25q, 75q)
	1	2	3	4	5	6	1	2	3	4	5	6	
Scenario 1													
Prob of toxicity	0.001	0.01	0.05	0.07	<b>0.2</b>	0.4							
OC_A(0.2)	0	0	0.082	0.267	<b>0.489</b>	0.162	0.037	0.041	0.161	0.238	0.314	0.209	5 (4, 6)
OC_A(0.3)	0	0	0.078	0.227	<b>0.526</b>	0.169	0.037	0.042	0.131	0.251	0.328	0.211	5 (5, 6)
OC_A(0.4)	0	0	0.054	0.245	<b>0.529</b>	0.172	0.037	0.042	0.105	0.272	0.332	0.211	5 (5, 6)
OC_G(0.6)	0	0	0.081	0.254	<b>0.506</b>	0.159	0.037	0.041	0.159	0.24	0.311	0.211	5 (4, 6)
OC_G(0.7)	0	0	0.086	0.258	<b>0.49</b>	0.166	0.037	0.041	0.16	0.278	0.268	0.216	5 (4, 6)
OC_G(0.8)	0	0	0.079	0.209	<b>0.53</b>	0.182	0.037	0.041	0.158	0.338	0.217	0.209	5 (4, 6)
Scenario 2													
Prob of toxicity	0.01	0.05	0.07	<b>0.2</b>	0.4	0.5							
OC_A(0.2)	0	0.003	0.465	<b>0.428</b>	0.096	0.008	0.047	0.057	0.436	0.282	0.118	0.06	5 (4, 6)
OC_A(0.3)	0	0.003	0.499	<b>0.378</b>	0.114	0.006	0.047	0.058	0.41	0.301	0.124	0.06	5 (4, 6)
OC_A(0.4)	0	0.002	0.482	<b>0.398</b>	0.11	0.008	0.047	0.062	0.338	0.365	0.127	0.06	5 (5, 6)
OC_G(0.6)	0	0.004	0.459	<b>0.426</b>	0.105	0.006	0.047	0.056	0.436	0.28	0.12	0.06	5 (4, 6)
OC_G(0.7)	0	0.004	0.471	<b>0.404</b>	0.111	0.01	0.047	0.056	0.435	0.298	0.102	0.061	5 (4, 6)
OC_G(0.8)	0	0.004	0.455	<b>0.409</b>	0.117	0.015	0.047	0.057	0.431	0.328	0.079	0.058	5 (4, 6)
Scenario 3													
Prob of toxicity	0.05	0.07	<b>0.2</b>	0.4	0.5	0.55							
OC_A(0.2)	0	0.085	<b>0.883</b>	0.03	0.002	0	0.079	0.149	0.636	0.085	0.031	0.02	6 (5, 7)
OC_A(0.3)	0.006	0.068	<b>0.892</b>	0.029	0.005	0	0.085	0.145	0.617	0.101	0.032	0.02	6 (5, 7)
OC_A(0.4)	0.008	0.076	<b>0.876</b>	0.037	0.003	0	0.085	0.166	0.568	0.128	0.032	0.02	6 (5, 7)
OC_G(0.6)	0.007	0.082	<b>0.869</b>	0.036	0.005	0.001	0.082	0.147	0.633	0.087	0.031	0.021	6 (5, 7)
OC_G(0.7)	0.006	0.083	<b>0.871</b>	0.036	0.004	0	0.077	0.152	0.632	0.092	0.027	0.02	6 (5, 7)
OC_G(0.8)	0.001	0.087	<b>0.871</b>	0.041	0	0	0.076	0.154	0.632	0.096	0.024	0.018	6 (5, 7)
Scenario 4													
Prob of toxicity	0.07	<b>0.2</b>	0.4	0.5	0.55	0.65							
OC_A(0.2)	0.107	<b>0.612</b>	0.279	0.002	0	0	0.213	0.432	0.309	0.029	0.011	0.006	7 (6, 8)
OC_A(0.3)	0.168	<b>0.536</b>	0.294	0.002	0	0	0.268	0.378	0.305	0.032	0.011	0.006	7 (6, 8)
OC_A(0.4)	0.182	<b>0.51</b>	0.306	0.002	0	0	0.265	0.406	0.276	0.035	0.011	0.006	7 (6, 8)
OC_G(0.6)	0.195	<b>0.523</b>	0.281	0.001	0	0	0.261	0.384	0.309	0.029	0.011	0.006	7 (6, 8)
OC_G(0.7)	0.202	<b>0.515</b>	0.283	0	0	0	0.228	0.418	0.309	0.029	0.01	0.006	7 (6, 8)
OC_G(0.8)	0.12	<b>0.596</b>	0.284	0	0	0	0.196	0.45	0.309	0.03	0.01	0.005	7 (6, 8)
Scenario 5													
Prob of toxicity	<b>0.2</b>	0.4	0.5	0.55	0.65	0.7							
OC_A(0.2)	<b>0.765</b>	0.22	0.015	0	0	0	0.654	0.249	0.077	0.013	0.005	0.002	8 (7, 9)
OC_A(0.3)	<b>0.855</b>	0.125	0.02	0	0	0	0.723	0.182	0.073	0.015	0.005	0.002	8 (7, 9)
OC_A(0.4)	<b>0.851</b>	0.127	0.022	0	0	0	0.721	0.19	0.067	0.015	0.005	0.002	8 (7, 9)
OC_G(0.6)	<b>0.861</b>	0.125	0.014	0	0	0	0.709	0.194	0.077	0.013	0.004	0.002	8 (7, 9)
OC_G(0.7)	<b>0.888</b>	0.098	0.014	0	0	0	0.675	0.229	0.076	0.013	0.004	0.002	8 (7, 9)
OC_G(0.8)	<b>0.779</b>	0.207	0.014	0	0	0	0.639	0.265	0.076	0.013	0.004	0.002	8 (7, 9)

**Table 5.** Performance comparison of methods with Occam window threshold.

## 5 Results in another setting

We performed an other simulation study where, in each scenario, 1000 trial were simulated with 4 different dose levels, 18 patients and 0.3 as target toxicity. The  $D_0$  was chosen as a trial with 1 patient allocated to the first dose level, 2 patients at the second dose level, 9 and 6 patients allocated to the third and fourth dose level, respectively. Three DLTs were seen at both dose level 3 and 4. The final CRM analysis gave the following estimation of probability of toxicity: 0.06, 0.16, 0.32, and 0.47.

For AP\_SOC2 we set  $s^*(n) = \min(12, n)$ . Result trends, Table 6, are similar to the ones showed in the main simulation analysis.

Method	% dose selection				% dose allocation				DLTs median (25q, 75q)
	1	2	3	4	1	2	3	4	
Scenario 1b									
Prob of toxicity	0.05	0.15	<b>0.30</b>	0.45					
P_NI	0.011	0.229	<b>0.546</b>	0.214	0.125	0.266	0.345	0.264	5 (4, 6)
P.ESS(10)	0	0.159	<b>0.709</b>	0.132	0.057	0.21	0.638	0.096	5 (4, 6)
P.ESS(18)	0	0.099	<b>0.833</b>	0.068	0.056	0.143	0.769	0.033	5 (4, 6)
AP_L	0.003	0.141	<b>0.738</b>	0.118	0.117	0.228	0.438	0.218	5 (4, 6)
AP_S	0.006	0.168	<b>0.677</b>	0.149	0.119	0.24	0.412	0.23	5 (4, 6)
AP_MIX(0.5)	0.007	0.159	<b>0.692</b>	0.142	0.119	0.239	0.413	0.229	5 (4, 6)
AP_SOC1	0.011	0.163	<b>0.677</b>	0.149	0.124	0.235	0.412	0.23	5 (4, 6)
AP_SOC2	0.011	0.179	<b>0.642</b>	0.168	0.125	0.237	0.405	0.233	5 (4, 6)
AP_EB	0.002	0.1	<b>0.821</b>	0.077	0.117	0.205	0.481	0.197	5 (4, 6)
BCRM	0.014	0.238	<b>0.536</b>	0.212	0.115	0.274	0.389	0.222	5 (4, 6)
Scenario 2b									
Prob of toxicity	0.15	<b>0.30</b>	0.45	0.60					
P_NI	0.226	<b>0.543</b>	0.206	0.025	0.339	0.389	0.188	0.084	5 (5, 6)
P.ESS(10)	0.049	<b>0.608</b>	0.331	0.012	0.092	0.499	0.393	0.017	6 (5, 7)
P.ESS(18)	0.008	<b>0.539</b>	0.45	0.003	0.059	0.408	0.529	0.004	7 (6, 8)
AP_L	0.111	<b>0.521</b>	0.359	0.009	0.288	0.388	0.252	0.073	6 (5, 7)
AP_S	0.149	<b>0.53</b>	0.308	0.013	0.303	0.388	0.233	0.076	6 (5, 7)
AP_MIX(0.5)	0.143	<b>0.534</b>	0.31	0.013	0.299	0.393	0.232	0.076	6 (5, 7)
AP_SOC1	0.215	<b>0.463</b>	0.309	0.013	0.337	0.355	0.233	0.076	6 (5, 6)
AP_SOC2	0.229	<b>0.478</b>	0.277	0.016	0.34	0.357	0.227	0.076	6 (5, 6)
AP_EB	0.118	<b>0.451</b>	0.43	0.001	0.289	0.349	0.293	0.069	6 (5, 7)
BCRM	0.203	<b>0.554</b>	0.221	0.022	0.308	0.41	0.212	0.07	5 (4, 6)
Scenario 3b									
P Prob of toxicity	<b>0.30</b>	0.45	0.60	0.70					
P_NI	<b>0.728</b>	0.242	0.029	0.001	0.668	0.234	0.074	0.024	7 (6, 8)
P.ESS(10)	<b>0.401</b>	0.549	0.048	0.002	0.277	0.565	0.155	0.004	8 (7, 9)
P.ESS(18)	<b>0.167</b>	0.75	0.082	0.001	0.113	0.628	0.258	0	8 (7, 9)
AP_L	<b>0.584</b>	0.354	0.062	0	0.609	0.276	0.093	0.023	7 (6, 8)
AP_S	<b>0.654</b>	0.303	0.042	0.001	0.633	0.258	0.086	0.023	7 (6, 8)
AP_MIX(0.5)	<b>0.637</b>	0.32	0.043	0	0.626	0.264	0.086	0.023	7 (6, 8)
AP_SOC1	<b>0.713</b>	0.245	0.041	0.001	0.666	0.225	0.086	0.023	7 (6, 8)
AP_SOC2	<b>0.729</b>	0.233	0.037	0.001	0.668	0.225	0.085	0.023	7 (6, 8)
AP_EB	<b>0.605</b>	0.302	0.092	0.001	0.612	0.253	0.112	0.023	7 (6, 8)
BCRM	<b>0.694</b>	0.272	0.032	0.002	0.635	0.265	0.081	0.019	7 (6, 8)

**Table 6.** Results for each method and each scenario (1-5) in terms of the percentage of dose selection at the end of the trial, the percentage of dose allocation and median number of DLTs, along with the first and the third quartiles. At the beginning of each scenario section, the true probabilities used for the scenario simulation are displayed.

## 6 ESS of a likelihood

We detail the computation of the ESS for a likelihood function, and we show its asymptotic value. We consider  $\mathbf{X}_0 = (X_{1,0}, \dots, X_{n_0,0})$  a vector of  $n_0$  i.i.d. observations from  $f(X_{i,0}|\boldsymbol{\theta})$  with  $\boldsymbol{\theta} = (\theta_1, \dots, \theta_q)$ , and  $L(\boldsymbol{\theta}|\mathbf{X}_0) = \prod_{i=1}^{n_0} f(X_{i,0}|\boldsymbol{\theta})$  the corresponding likelihood.

### Theorem:

For a given likelihood  $L(\boldsymbol{\theta}|\mathbf{X}_0) = \prod_{i=1}^{n_0} f(X_{i,0}|\boldsymbol{\theta})$ ,  $\mathbf{X} = (X_1, \dots, X_n)$  a vector of  $n$  generic i.i.d. observations which comes from the same distribution than  $X_{1,0}, \dots, X_{n_0,0}$ , a vague prior  $q_0(\boldsymbol{\theta})$ , with the following hypothesis:

- (i)  $\mathbb{E} \left[ \left| \frac{\partial^2 \log(f(X_i|\boldsymbol{\theta}))}{\partial \theta_j^2} \right| \right] < +\infty$  for all  $j$
- (ii)  $\frac{\partial^2 \log(q_0(\boldsymbol{\theta}))}{\partial \theta_j^2} \rightarrow 0$  for all  $j$

then,  $ESS[L(\boldsymbol{\theta}|\mathbf{X}_0)] \xrightarrow[n_0 \rightarrow \infty]{} n_0$ .

### Proof:

We consider the likelihood as a general function of theta to calculate its ESS. Considering a vague prior  $q_0(\boldsymbol{\theta})$ , we denote  $q_m(\boldsymbol{\theta}|\mathbf{X}) = q_0(\boldsymbol{\theta}) \prod_{i=1}^n f(X_i|\boldsymbol{\theta})$  the posterior distribution of  $\boldsymbol{\theta}$  after  $n$  observations.  $f_n(\mathbf{X})$  is the marginal density of  $\mathbf{X}$  With Morita's notations,

$$\begin{aligned} D_{q,+}(\boldsymbol{\theta}, n) &= \sum_{j=1}^q \int \left\{ -\frac{\partial^2 \log(q_0(\boldsymbol{\theta}))}{\partial \theta_j^2} - \sum_{i=1}^n \frac{\partial^2 \log(f(X_i|\boldsymbol{\theta}))}{\partial \theta_j^2} \right\} f_n(\mathbf{X}) d\mathbf{X} \\ &= -\sum_{j=1}^q \frac{\partial^2 \log(q_0(\boldsymbol{\theta}))}{\partial \theta_j^2} - n \sum_{j=1}^q \mathbb{E} \left[ \frac{\partial^2 \log(f(X_i|\boldsymbol{\theta}))}{\partial \theta_j^2} \right] \end{aligned}$$

Then, considering that in the particular case of the ESS of a likelihood, the prior of interest is the likelihood with observation,  $p(\boldsymbol{\theta}) = L(\boldsymbol{\theta}|\mathbf{X}_0)$ . So still with Morita's notations:

$$D_{p,+}(\boldsymbol{\theta}) = -\sum_{j=1}^q \sum_{i=1}^{n_0} \frac{\partial^2 \log(f(X_{i,0}|\boldsymbol{\theta}))}{\partial \theta_j^2}$$

Reminding that

$$\begin{aligned} \delta(n, q_0) &= |D_{p,+}(\bar{\boldsymbol{\theta}}) - D_{q,+}(\bar{\boldsymbol{\theta}}, n)| \\ ESS[L(\boldsymbol{\theta}|\mathbf{X}_0)] &= \arg \min_n \delta(n, q_0) \end{aligned}$$

Therefore,

$$\delta(n, q_0) = \left| - \sum_{j=1}^q \sum_{i=1}^{n_0} \frac{\partial^2 \log(f(X_{i,0}|\boldsymbol{\theta}))}{\partial \theta_j^2} \Big|_{\boldsymbol{\theta}=\bar{\boldsymbol{\theta}}} + \sum_{j=1}^q \frac{\partial^2 \log(q_0(\boldsymbol{\theta}))}{\partial \theta_j^2} \Big|_{\boldsymbol{\theta}=\bar{\boldsymbol{\theta}}} + n \sum_{j=1}^q \mathbb{E} \left[ \frac{\partial^2 \log(f(X_i|\boldsymbol{\theta}))}{\partial \theta_j^2} \right] \Big|_{\boldsymbol{\theta}=\bar{\boldsymbol{\theta}}} \right|$$

Considering hypothesis (ii), minimising  $\delta(n, q_0)$  is equivalent to minimise

$$\Delta(n, n_0) = - \sum_{j=1}^q \sum_{i=1}^{n_0} \frac{\partial^2 \log(f(X_{i,0}|\boldsymbol{\theta}))}{\partial \theta_j^2} \Big|_{\boldsymbol{\theta}=\bar{\boldsymbol{\theta}}} + n \sum_{j=1}^q \mathbb{E} \left[ \frac{\partial^2 \log(f(X_i|\boldsymbol{\theta}))}{\partial \theta_j^2} \right] \Big|_{\boldsymbol{\theta}=\bar{\boldsymbol{\theta}}}$$

And with hypothesis (i) and Law of Large Number, for  $\boldsymbol{\theta} = \bar{\boldsymbol{\theta}}$ ,

$$\Delta(n, n_0) \xrightarrow[n_0 \rightarrow \infty]{} (n - n_0) \sum_{j=1}^q \mathbb{E} \left[ \frac{\partial^2 \log(f(X_i|\boldsymbol{\theta}))}{\partial \theta_j^2} \right] \Big|_{\boldsymbol{\theta}=\bar{\boldsymbol{\theta}}}$$

With  $\Delta(n, n_0) = 0$  if  $n = n_0$ , so  $\arg \min_n \delta(n, q_0) = n_0$ . ■

## 7 Copyright statement

Please be aware that the use of this  $\text{\LaTeX} 2_{\epsilon}$  class file is governed by the following conditions.

### 7.1 Copyright

Copyright © 2016 SAGE Publications Ltd, 1 Oliver's Yard, 55 City Road, London, EC1Y 1SP, UK. All rights reserved.

## Chapter 5

# Estimating similarity of dose-response relationships in phase I clinical trials - case study in bridging data package

In this chapter, we present article [66]: Adrien Ollier, Sarah Zohar, Satoshi Morita, Moreno 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*.2021; 18(4):1639. <https://doi.org/10.3390/ijerph18041639>.

### Context

It is not uncommon for a drug to be tested several times, on different populations. In particular, bridging studies are additional studies conducted into a new population to provide clinical data from an historical trial. The ICH-E5 [35] considers that clinical data can be extrapolated from a population to another if the two dose-response curves are similar. However, the similarity between two dose-response curve is not precisely defined.

### Objective

Our objective is to propose several statistical similarity criteria to evaluate the similarity of two dose-response curves obtained in two Phase I clinical trials, and more generally the similarity between the two populations response to the drug. This should help the regional authorities to decide if the extrapolation is possible in term of dose-toxicity and MTD.



**Method:** We extended the existing criterion developed in Ollier et al.[65]. This criterion was first proposed for an adaptive context. The proposed extensions permit to evaluate the similarity in terms of parameter posterior distributions and MTD posterior distributions. The new criteria do not assess the same kind of similarity and have not the same numerical particularities (some are bounded, some are not). To evaluate the performance of the proposed criteria, synthetic datasets were built as example, and six drugs with published data were also considered. Illustrative plots are proposed to guide the interpretation of the results.

### **Results**

The synthetic example considered permitted to illustrate the expected differences between several populations. Indeed, two populations could have a same MTD but different dose-response relationship, or both can be different, etc. The different crieteria appear to correctly assess the expected kind of similarity or dissimilarity.

### **Conclusion**

The proposed criteria appeared to be in adequacy with the expected results, and demonstrated that one single criterion is not enough to assess the similarity between two datasets or two dose-response curves. However, several criteria are based on the MTD, and so depends on the definition of the MTD at the beginning of the trials.



Article

# Estimating Similarity of Dose–Response Relationships in Phase I Clinical Trials—Case Study in Bridging Data Package

Adrien Ollier <sup>1,\*</sup>, Sarah Zohar <sup>1,\*</sup> , Satoshi Morita <sup>2</sup> and Moreno Ursino <sup>1,3</sup>

<sup>1</sup> INSERM, Centre de Recherche des Cordeliers, Sorbonne Université, USPC, Université de Paris, F-75006 Paris, France; adrien.ollier@inserm.fr (A.O.); moreno.ursino@inserm.fr (M.U.)

<sup>2</sup> Department of Biomedical Statistics and Bioinformatics, Kyoto University Graduate School of Medicine, Kyoto 606-8501, Japan; smorita@kuhp.kyoto-u.ac.jp

<sup>3</sup> F-CRIN PARTNERS Platform, Assistance Publique-Hôpitaux de Paris, Université de Paris, F-75010 Paris, France

\* Correspondence: sarah.zohar@inserm.fr

**Abstract:** Bridging studies are designed to fill the gap between two populations in terms of clinical trial data, such as toxicity, efficacy, comorbidities and doses. According to ICH-E5 guidelines, clinical data can be extrapolated from one region to another if dose–response curves are similar between two populations. For instance, in Japan, Phase I clinical trials are often repeated due to this physiological/metabolic paradigm: the maximum tolerated dose (MTD) for Japanese patients is assumed to be lower than that for Caucasian patients, but not necessarily for all molecules. Therefore, proposing a statistical tool evaluating the similarity between two populations dose–response curves is of most interest. The aim of our work is to propose several indicators to evaluate the distance and the similarity of dose–toxicity curves and MTD distributions at the end of some of the Phase I trials, conducted on two populations or regions. For this purpose, we extended and adapted the commensurability criterion, initially proposed by Ollier et al. (2019), in the setting of completed phase I clinical trials. We evaluated their performance using three synthetic sets, built as examples, and six case studies found in the literature. Visualization plots and guidelines on the way to interpret the results are proposed.

**Keywords:** bridging studies; distribution distance; oncology; phase I; dose-finding; dose–response; bayesian inference



**Citation:** Ollier, A.; Zohar, S.; Morita, S.; Ursino, M. Estimating Similarity of Dose–Response Relationships in Phase I Clinical Trials—Case Study in Bridging Data Package. *Int. J. Environ. Res. Public Health* **2021**, *18*, 1639. <https://doi.org/10.3390/ijerph18041639>

Academic Editor: Paul B. Tchounwou  
Received: 14 December 2020  
Accepted: 5 February 2021  
Published: 9 February 2021

**Publisher’s Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Bridging studies are designed to fill the gap between two populations in terms of clinical trial data, such as toxicity, efficacy, comorbidities and doses. A bridging data package consists of selected data from the Clinical Data Package of the population in the new region, including pharmacokinetic, any pharmacodynamic, dose–toxicity or dose–efficacy data, and if appropriate, a bridging study to extrapolate the foreign dose–response data to the new region [1].

According to the International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use E5 (ICH-E5) guidelines, data can be extrapolated from one region to another if “a bridging study [...] indicates that a different dose in the new region results in a safety and efficacy profile that is not substantially different from the one derived from the original region; it will often be possible to extrapolate the foreign data to the new region, with an appropriate dose adjustment, if this can be adequately justified (e.g., by pharmacokinetic and/or pharmacodynamic data)” [1]. This is the reason why proposing a statistical tool evaluating the similarity between two foreign dose–response curves is of great interest. If this is proven, then, other clinical trials data can be used and extrapolated for the new region.

In Japan, the Pharmaceuticals and Medical Devices Agency (PMDA) recommends the re-evaluation of a drug if there are insufficient data from Japanese patients [2]. Indeed, Phase I clinical trials in oncology, which aim to estimate the maximum tolerated dose (MTD), are often repeated. Ogura et al. [3] pointed out that MTD differences between populations could be due to the different distribution of genetic polymorphisms in enzymes involved in drug metabolism or of biomarker incidences in different populations. In particular, in Japan, Phase I trials are repeated based on a physiological/metabolic paradigm: MTDs for Japanese patients are often lower than the ones of for Caucasian patients [4]. Based on this assumption, Maeda and Kurokawa [5] have performed an intensive study comparing the MTD of 21 molecularly targeted cancer drugs in Japanese versus Caucasian populations. They found out that this assumption does not hold well: in their study, the MTD was lower for Japanese patients in only two cases, there were no differences between the two populations with 10 drugs and MTD was incommensurable as the evaluated dose range acted different with nine drugs. Moreover, Mizugaki et al. [6] have analyzed data of single-agent Phase I trials at the National Cancer Center Hospital between 1995 and 2012, comparing the dose-limiting toxicity (DLT) profiles and MTDs of Japanese trials with the trials from Caucasian populations.

Recently, methods for bridging dose-finding design have been proposed where previous population data were used to either calibrate the prior distribution of the Bayesian model parameter(s) or to choose the “working model” of the design for prospective trials [7]. Liu et al. [8] proposed using a Bayesian model to average the dose-finding method where the previous trial data were used to build three different skeletons which would then be averaged during the study. Moreover, Takeda and Morita recently defined an “historical-to-current” parameter that could describe the degree of borrowing from one population to the other [9]. Ollier et al. [10] proposed a bridging method where a borrowing parameter was estimated sequentially in a response adaptive design which quantifies the amount of reasonable borrowing according to the similarity between the two populations’ estimates. Usually, the proposed methods focus on one parameter, strictly related to the MTD and not on the full dose–toxicity response curve. All these methods were proposed with the purpose of using the foreign data to plan and conduct the future Phase I trial in the new region. Indeed, at this stage, the idea is to use the foreign data to calibrate model-based priors to be used in the new region trial. However, in most cases, the trial in the new region will not be planned this way, but rather by using the MTD information from the foreign region only, if available. The sophisticated statistical approach will not be used.

Another option is to compare the two dose–response curves estimated from each region and to evaluate how similar they are. In this case, the overall purpose is different from before; if the curves prove to be similar (under the uncertainty estimation), the new purpose will be to extrapolate other trial data—such as that of Phase II—to the new region and to avoid further repetition of clinical investigations. For dose–response curves, Bretz et al. [11] introduced an asymptotic test to evaluate the difference of the *minimum efficient dose* among several groups of subjects, according to a threshold. However, this method was built for later clinical phases and presents weaknesses when applied to a small sample size. By contrast, Bayesian methods could mitigate the issue of estimation based on a small sample size setting, since they do not rely on asymptotic approximations and prior distributions can be used to ensure more stability in computation. Thereafter, the degree of similarity could be considered directly at the posterior distributions level. Therefore, methods proposing to estimate the similarity between dose–toxicity curves should be proposed when there is the need to evaluate if the safety data can be extrapolated or not.

The aim of our work is to propose some Bayesian indicators that evaluate the distance and the similarity of (1) dose–toxicity curves, taking into account the variability, (2) the MTD posterior distributions, by extending and adapting the commensurability criterion initially proposed by Ollier et al. [10]. These indicators were applied to several Phase I trials presented in Maeda and Kurokawa [5] and Mizugaki et al. [6], evaluating the similarity

between Western dose–toxicity data to Eastern ones. The proposed tools should be used by trial stakeholders in order to decide if other trials data could be extrapolated from the new region, and, if so, to avoid the repetition of multiple clinical trials. In the next section, the original commensurability parameter is summarized along with the proposed extensions and the dose–toxicity model used. The case studies are described in Section 3, while Section 4 details the computational settings. The results are given in Section 5, followed by a Discussion section.

## 2. Methods

In this section, we briefly recall the Bayesian commensurability measure used in Ollier et al. [10], which was originally adopted into a power prior setting [12]; we then propose extensions and modifications to this measure to be applied at the end of the study. We also introduce the Bayesian dose–toxicity model, which will be used for retrospective data analyses.

Let  $\mathbf{D}_c$  denote the Caucasian data,  $\mathbf{D}_c = \{(y_j, x_j)\}_{n_c}$ ,  $n_c$  the sample size of  $\mathbf{D}_c$ , and  $y_j$  the binary outcome of the  $j$ -th patient which received dose  $x_j$ . In a similar way, we can define  $\mathbf{D}_a$ , the Japanese data and associated parameters. Let us also set a model for the probability of toxicity vs dose;  $p_T(x) = f(x, \beta)$ , where  $f(\cdot)$  denotes a convenient monotonous link function parametrized by  $\beta$ . The likelihood function for each population can be written as  $L(\beta|\mathbf{D}_m) = \prod_{j=1}^{n_m} f(x, \beta)^{y_j} (1 - f(x, \beta))^{1-y_j}$ , for  $m = c, a$ .

### 2.1. Commensurability Distances

Ollier et al. [10] suggested to consider the likelihood function as a distribution, divided by a normalization constant. This type of normalized likelihood can also be seen as the resulting Bayesian posterior distribution when constant (probably improper) priors are used for the analysis. Then, the authors defined a measure of “commensurability” between the two data-sets through a distance  $d(\mathbf{D}_c, \mathbf{D}_a)$ , the Hellinger one, in the parameters space via the following relation

$$d^2(\mathbf{D}_c, \mathbf{D}_a) = \frac{1}{2} \int \left( \sqrt{\frac{L(\beta|\mathbf{D}_c)^{\min(1, \frac{n_a}{n_c})}}{\int L(\beta|\mathbf{D}_c)^{\min(1, \frac{n_a}{n_c})} d\beta}} - \sqrt{\frac{L(\beta|\mathbf{D}_a)^{\min(1, \frac{n_c}{n_a})}}{\int L(\beta|\mathbf{D}_a)^{\min(1, \frac{n_c}{n_a})} d\beta}} \right)^2 d\beta. \quad (1)$$

The commensurability measure, denoted by  $\gamma$ , is then defined as  $\gamma = d^q(\mathbf{D}_c, \mathbf{D}_a)$ , with  $q \in \mathbb{R}^+$ . Values of  $q$  higher than 1 will reduce the computed distance, while values lower than 1 will lead to a more conservative method, increasing the computed distance. In case of sequential trials, the authors proved that, when coupled with the power prior approach, a conservative value of  $\gamma$  leads to a better result in terms of operating characteristics, as a percentage of the right MTD selection. However, at the end of the trial, we are interested in comparing the achieved results, without any discount in the resulting distance. Therefore, in this paper, we will focus on the original Hellinger distance, which is  $q = 1$ . This computed distance is a positive number between 0 and 1, it tends towards the maximum value when the two datasets are quite different, and towards zero when they are close to each other. Each likelihood is divided by a normalization constant in order to ensure that it can be viewed as a probability distribution. The variance of the likelihood density depends on the sample size of the trial. To make the two likelihoods comparable in terms of precision (variance), if  $n_c > n_a$ ,  $L(\beta|\mathbf{D}_c)$  is raised to a power of less than 1, otherwise,  $L(\beta|\mathbf{D}_a)$  is raised to a power of less than 1. Following this method, the variance of likelihood density of the trial with more patients is increased to almost fit the one of the trial with fewer patients. Practical examples are given in Ollier et al. [10].

A straightforward modification of the distance in Equation (1) was performed by changing the underlying flat prior into a proper one. The posterior distribution obtained with the weighted likelihood is then used in the Hellinger formula. Thus, denoted by

$\pi_{post,c}(\beta|\mathbf{D}_c) \propto L(\beta|\mathbf{D}_c)^{\min(1, \frac{n_a}{n_c})} \pi_{prior}(\beta)$  and by  $\pi_{post,a}(\beta|\mathbf{D}_a) \propto L(\beta|\mathbf{D}_a)^{\min(1, \frac{n_c}{n_a})} \pi_{prior}(\beta)$  the posterior distribution of  $\beta$  given  $\mathbf{D}_c$  and  $\mathbf{D}_a$ , respectively, we have

$$d_{mod}^2(\mathbf{D}_c, \mathbf{D}_a) = \frac{1}{2} \int \left( \sqrt{\pi_{post,c}(\beta|\mathbf{D}_c)} - \sqrt{\pi_{post,a}(\beta|\mathbf{D}_a)} \right)^2 d\beta. \tag{2}$$

This modification will ensure more stability in computation when the likelihoods involve more than one parameter. When flat/constant priors are used for  $\pi_{prior}(\beta)$ , Equation (2) is equivalent to Equation (1). Even if, theoretically, two different priors can be chosen for the two trials, we suggest using a single one for the sake of comparability.

Both previous distances work at the parameter level. They check if the whole dose–toxicity curve is similar or not. Using a single parameter model for the dose–toxicity relationship, as a one parameter logistic model used in the continual reassessment method (CRM) [13], is also equivalent to check the MTD distance. However, in models with more parameters, such as the Bayesian Logistic Regression Model (BLRM) [14] where we have two parameters, intercept and slope, we check if the bivariate distribution of  $\beta$  is the same. Since the distance is difficult to interpret in case of the multidimensional parameters space, we propose a summary distance using the resulting posterior MTD distribution. In our setting, the MTD,  $x^*$ , is estimated as the dose linked to a pre-specified toxicity target  $\tau$ , that is,  $x^* = f^{-1}(\tau|\beta)$ , where  $f^{-1}(\cdot)$  is the inverse function of  $f(\cdot)$ . The posterior MTD distribution,  $\pi_{MTD,m}(x^*|\mathbf{D}_m)$ , is obtained evaluating  $x^*$  through the posterior distribution of the parameter,  $\pi_{post,m}(\beta|\mathbf{D}_m)$ , for  $m = c, a$ . Therefore, we can define

$$d_{MTD}^2(\mathbf{D}_c, \mathbf{D}_a) = \frac{1}{2} \int \left( \sqrt{\pi_{MTD,c}(x^*|\mathbf{D}_c)} - \sqrt{\pi_{MTD,a}(x^*|\mathbf{D}_a)} \right)^2 dx^*. \tag{3}$$

Note that this distance always involves a one dimensional integral.

Previous distances focused on understanding the similarity of the whole dose–toxicity curve between two populations. However, even with different slopes and intercepts, two populations can still have the same MTD. Those differences should generally indicate a difference in responsiveness to a drug and it is important to know when MTDs are similar but not the underlying curves. Therefore, we propose to couple the distances, previously described, with a measure denoting the difference in MTD point estimations. We can build this measure as a percentage using the median of the posterior MTD distributions, such as

$$d_{p1}(\mathbf{D}_c, \mathbf{D}_a) = \left( \frac{\text{med}_c}{\text{med}_a} \right)^{1-2I(\text{med}_c < \text{med}_a)} - 1, \tag{4}$$

where  $I(\cdot)$  is the indicator function, which assumes the value 1 if the statement in parentheses is true and zero otherwise, and  $\text{med}_i$  with  $i = c, a$ , is the median of the posterior MTD distribution of Caucasians and Japanese, respectively. This formulation was chosen for its easy interpretation, indeed, we check how much the highest MTD differs in percentage in respect to the lowest one. For this reason, the formula implies the exponent  $1 - 2I(\text{med}_c < \text{med}_a)$ , which allows us to always have the highest estimate at the numerator, and the  $-1$  term. Similarly to the three previous measures, Equation (4) tends to zero when the two MTDs are very similar. However, this measure does not have an upper bound. We propose the use of the median since it is less impacted by outliers than the mean. The *maximum a posteriori* is another possible candidate, that is

$$d_{p2}(\mathbf{D}_c, \mathbf{D}_a) = \left( \frac{\tilde{x}_c^*}{\tilde{x}_a^*} \right)^{1-2I(\tilde{x}_c^* < \tilde{x}_a^*)} - 1, \tag{5}$$

where

$$\tilde{x}_i^* = \arg \max_{x^*} \pi_{MTD,i}(x^*|\mathbf{D}_i).$$

To summarize, the first three measures  $d$ ,  $d_{mod}$ , and  $d_{MTD}$  are bounded between 0 and 1. Even if they are not built as percentages, their interpretation could be strictly linked to the percentage. Otherwise, the last two measures  $d_{p1}$  and  $d_{p2}$  have a ratio-like measure, lower bounded at 0. In practice, they give the information on the number of times the maximum MTD is higher than the lowest one.

## 2.2. Dose–Toxicity Model

In this section, we describe the model selected for the link function  $f(\cdot)$ . Instead of the CRM, originally used in Ollier et al. [10], which is better suited to prospective trials than retrospective analyses (retrospective CRM requires special techniques), we opted for a more flexible BLRM model, with two parameters, the intercept  $\beta_0$  and the (logarithm of the) slope  $\beta_1$  [14]. The dose–toxicity relationship is represented by

$$\text{logit}\{p_T(x)\} = \beta_0 + \exp(\beta_1) \log\left(\frac{x}{x_r}\right)$$

where  $\beta \in \mathbb{R}^2$ ,  $x_r$  denotes a reference dose and  $\exp(\beta_1)$  assures a positive final slope in the model. In this case,  $f^{-1}(\cdot)$  is equal to the logit function and the BLRM formulation is similar to the one of Zheng and Hampson [15]. To close the Bayesian model, we suggest a bivariate normal distribution as prior for  $(\beta_0, \beta_1)$ .

Following the described model, the final MTD is estimated as  $x^* = x_r \exp\left(\frac{\text{logit}(\tau) - \beta_0}{\exp(\beta_1)}\right)$ . In order to minimize the overdispersion generated by this formula, we compared the distribution of the log ratio of the MTD and the reference dose,  $x^{**} = \log(x^*/x_r)$  (instead of the real MTD). Therefore, we have also changed Equations (4) and (5), accordingly, to the new formulation ( $x^{**}$ ) in order to preserve the original distance meaning, that is  $d_{p1}(\mathbf{D}_c, \mathbf{D}_a) = \exp|\text{med}_c - \text{med}_a| - 1$  and  $d_{p2}(\mathbf{D}_c, \mathbf{D}_a) = \exp|\tilde{x}_c^* - \tilde{x}_a^*| - 1$ .

Finally, in a previous sensitivity analysis (not shown), even when comparing the distribution of the log ratio of the MTD and the reference dose, we faced instability in computation due to the issue of outliers. We have found that truncating the posterior distribution of  $x^{**}$  between the 10 and 90 percentiles gives a good compromise between preserving trial information and computation stability.

## 3. Case Studies

To show the results and the interpretation of the proposed measures, we first introduce four different synthetic datasets (1 for Caucasian and 3 for Japanese), to check the results when two datasets are similar or not. We fixed the Caucasian dataset first: setting  $\tau$  equal to 0.3, the MTD at dose 600 mg/day. The same setting was used for the Japanese synthetic-1 set. Moreover, the two datasets were generated to have the same dose–toxicity shape. Japanese synthetic-2 set shares the same MTD with the Caucasian set, but has a different dose–toxicity shape: the Japanese dose–toxicity is steeper at the MTD than the Caucasian one. The Japanese synthetic-3 set has a different dose–toxicity curve and MTD (200 mg/day). The data are summarized in Table 1.

Then, we applied our methods to eight examples found in the literature. Our research started by looking at the drugs presented in Maeda and Kurokawa [5] and Mizugaki et al. [6]. We selected only drugs for which both Caucasian and Japanese trial data were available. We then extracted the number of toxicities and the number of allocated patients to the administered doses in each trial. All those data are shown in Table 2, each time with the reference article. The MTD declared at the end of the trial is shown in a box. As we can see from Table 2, Caucasians and Japanese trials were not usually used with the same set of doses.

**Table 1.** Number of dose-limiting toxicity and total number of patients accrued at each dose for 1 Caucasian trial and 3 Japanese synthetic trials. In the first column, the trial population is specified. A dash (-) means that the dose was not tested in the specified population. A box denotes the dose that has been defined as maximum tolerated dose (MTD).

Example (mg/day)	Doses					
	100	200	400	500	600	800
Caucasian (DLTs/nb pt)	0/3	0/3	0/6	-	3/9	2/3
Japanese						
Synthetic-1 (DLTs/nb pt)	-	-	-	1/10	2/8	2/2
Synthetic-2 (DLTs/nb pt)	-	-	0/3	0/9	4/12	3/3
Synthetic-3 (DLTs/nb pt)	0/3	1/6	3/3	-	-	-

**Table 2.** Value of dose-limiting toxicity and total number of patients accrued at each dose for all trials analysed in this manuscript. In the first column, the trial population is specified. A dash (-) means that the dose was not tested in the specified population. A box denotes the dose that has been defined as MTD, if the MTD was reached in the trial. For Sorafenib, the doses were given twice daily (bid).

Investigated Drug	Doses								
	0.25	0.5	0.7	1.0	1.4	2	2.8	4	
<b>Erlubun (mg/m<sup>2</sup>)</b>									
Caucasian [16] (DLTs/nb pt)	0/1	0/4	-	0/3	-	1/7	2/3	3/3	
Japanese [17] (DLTs/nb pt)	-	-	0/3	0/3	2/6	3/3	-	-	
<b>Lapatinib (mg/day)</b>	<b>500</b>	<b>650</b>	<b>900</b>	<b>1000</b>	<b>1200</b>	<b>1600</b>	<b>1800</b>		
Caucasian [18] (DLTs/nb pt)	0/13	1/15	0/11	1/3	1/12	1/13	-	-	
Japanese [19] (DLTs/nb pt)	-	-	0/6	-	0/6	1/6	1/6	-	
<b>Sorafenib (mg bid)</b>	<b>100</b>	<b>200</b>	<b>400</b>	<b>600</b>					
Caucasian [20] (DLTs/nb pt)	0/3	1/6	0/8	3/7					
Japanese [21] (DLTs/nb pt)	0/3	1/12	0/6	1/6					
<b>Ixabepilone (mg/m<sup>2</sup>)</b>	<b>7.4</b>	<b>15</b>	<b>30</b>	<b>40</b>	<b>50</b>	<b>57</b>	<b>65</b>		
Caucasian [22] (DLTs/nb pt)	0/3	0/3	0/3	-	3/22	3/3	2/3		
Japanese [23] (DLTs/nb pt)	-	0/3	0/3	1/6	2/2	-	-		
<b>Edotecarin (mg/m<sup>2</sup>)</b>	<b>6</b>	<b>8</b>	<b>11</b>	<b>13</b>	<b>15</b>				
Caucasian [24] (DLTs/nb pt)	0/3	0/3	0/6	1/9	4/9				
Japanese [25] (DLTs/nb pt)	-	0/3	1/6	1/9	2/6				
<b>E7070 (mg/m<sup>2</sup>)</b>	<b>50</b>	<b>100</b>	<b>200</b>	<b>400</b>	<b>600</b>	<b>700</b>	<b>800</b>	<b>900</b>	<b>1000</b>
Caucasian [26] (DLTs/nb pt)	0/4	0/3	0/3	0/3	0/4	2/7	2/4	-	3/3
Japanese [27] (DLTs/nb pt)	-	-	-	0/3	0/3	0/6	1/6	2/3	-

#### 4. Settings

We chose  $\tau$ , the target toxicity probability, to be used to define the MTD, which equals 0.3 for the three synthetic set examples, while it equals 0.25 for the real case studies. Most of real case studies followed an algorithm base allocation; therefore, it seemed more natural to have a threshold lower than 0.3, which is more frequently used when model based designs are adopted in oncology.

A non-informative bivariate prior distribution, commonly used in this setting, was chosen for the BLRM model as follows:

$$\begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \text{logit}(0.1) \\ \log 1 \end{pmatrix}, \begin{bmatrix} 4 & 0 \\ 0 & 4 \end{bmatrix}\right).$$

The hyperprior parameters of the bivariate prior were chosen after a preliminary sensitivity analysis (not shown) in order to ensure computational stability. In detail, this prior choice suggests a mean prior probability of toxicity at the reference dose,  $x_r$ , of 0.1 and that the slope has the prior median centered at zero. Therefore,  $x_r$  was chosen in the first half of the total dose panel for each example. In detail, 400 mg/day was set for the three synthetic examples, 1 mg/m<sup>2</sup> for Erlubun, 900 mg/day for Lapatinib, 200 mg/day for Sorafenib, 30 mg/m<sup>2</sup> for Ixabepilone, 8 mg/m<sup>2</sup> for Edotecarin and 700 mg/m<sup>2</sup> for E7070.

All distances were computed with  $q = 1$ , which is why we focus on the square root of Equation (1)–(3) and on the original value for Equation (4) and (5). The reference doses selected are reported along with the results in Table 3. All computations were performed in R, version 3.5.2. Monte Carlo approximations were adopted for all integrals involved, and uniform prior distribution on compact supports was set to approximate weighted likelihoods (as posterior distributions) in Equation (4). Details can be found in R scripts in the Supplementary Materials.

**Table 3.** Results in terms of  $d$ ,  $d_{mod}$ ,  $d_{MTD}$ ,  $d_{p1}$  and  $d_{p2}$  for the synthetic examples and the real case studies.  $x_r$  denotes the reference dose selected for the Bayesian Logistic Regression Model (BLRM).

Drug	$d$	$d_{mod}$	$d_{MTD}$	$d_{p1}$	$d_{p2}$
Synthetic-1	0.23	0.18	0.19	0	0
Synthetic-2	0.53	0.37	0.41	0.02	0.02
Synthetic-3	0.91	0.83	1.00	1.50	1.27
Erlubun	0.92	0.83	0.91	0.47	0.43
Lapatinib	0.58	0.39	0.50	7.29	0.35
Sorafenib	0.45	0.43	0.57	10.07	0.75
Ixabepilone	0.77	0.56	0.62	0.34	0.26
Edotecarin	0.38	0.24	0.32	0.32	0.04
E7070	0.63	0.63	0.88	0.59	0.23

### 5. Results

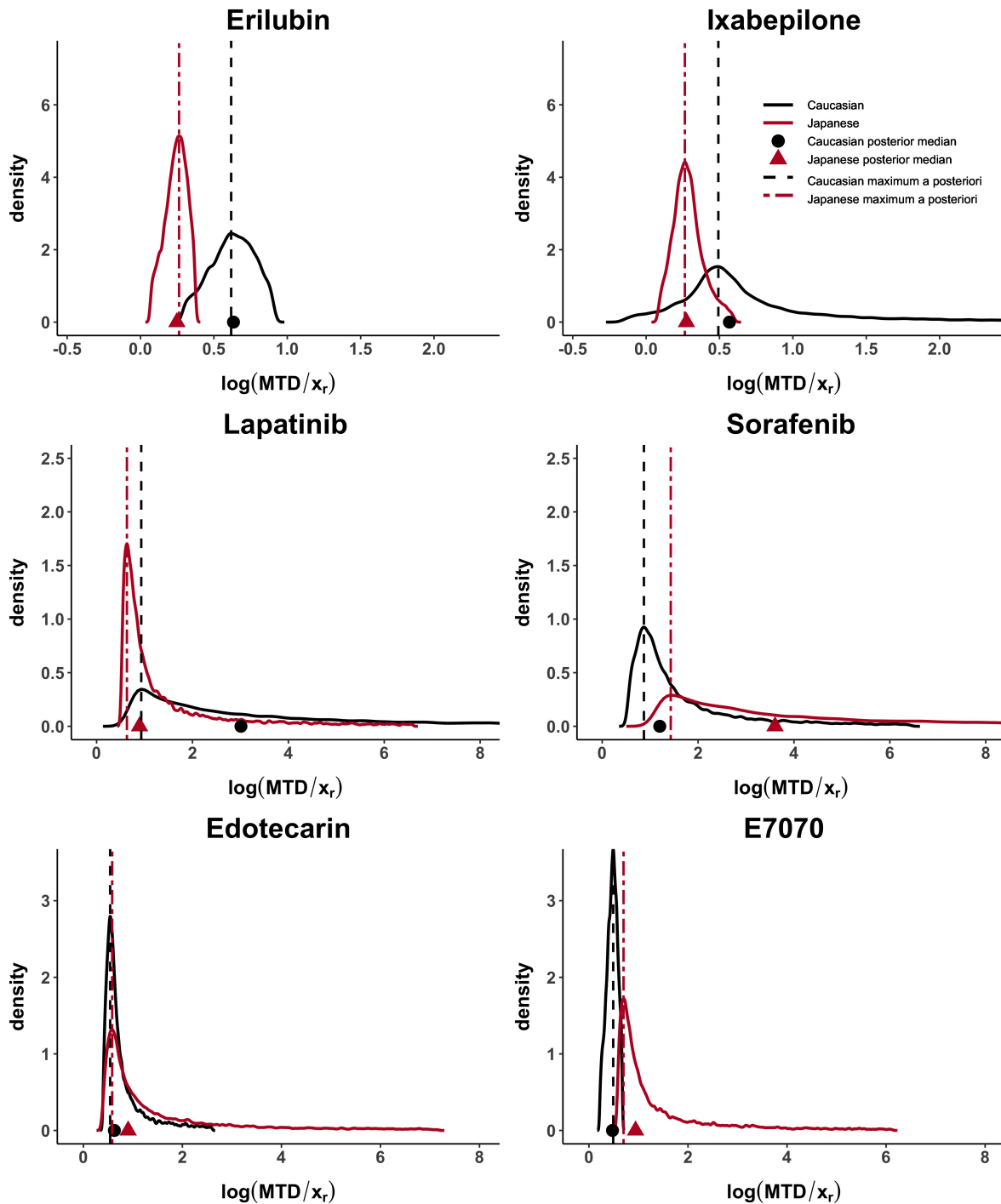
The computed distances under all the proposed methods are shown in Table 3. When the MTD and the dose–toxicity curves are similar, like in synthetic-1 data,  $d$ ,  $d_{mod}$ ,  $d_{MTD}$  are lower than 0.23 and  $d_{p1} = d_{p2} = 0$ . When only the MTDs are similar (synthetic-2 data) but not the dose–toxicity curves,  $d_{p1} = d_{p2} = 0.02$  but  $d$ ,  $d_{mod}$ ,  $d_{MTD}$  are higher than 0.37. Finally, when both curves and MTDs (synthetic-3 data) differ  $d_{p1} = 1.50$ ,  $d_{p2} = 1.27$  and  $d$ ,  $d_{mod}$ ,  $d_{MTD}$  are higher than 0.83.

Taking these cases’ studies as reference, we then analyse the data from published papers with Caucasian and Japanese datasets. Erlubun has the highest values of  $d$ ,  $d_{mod}$  and  $d_{MTD}$ , greater than 0.80, which suggests differences between the dose–toxicity curves. It is also shown in Figure 1. Its values of  $d_{p1}$  and  $d_{p2}$  are around 0.45. Ixabepilone and E7070 have quite large  $d$ ,  $d_{mod}$  and  $d_{MTD}$ , greater than 0.56 and they also have similar results in term of  $d_{p2}$ . The value of  $d_{p1}$  is different in these two examples and reflects the presence of unbalanced heavy tails in the E7070 case. The heavy tail concern is observed, in at least one population, in all examples except for Erlubun. The results obtained in Table 3 show that  $d_{p1}$  is directly impacted by this phenomenon. For example, Lapatinib and Sorafenib have a very high value of  $d_{p1}$ , greater than 7.29, whereas the maximum a posteriori,  $d_{p1}$ , has more stable and usual results. Edotecarin has close values of  $d$ ,  $d_{mod}$  and  $d_{MTD}$ , around 0.3, representing similar dose–toxicity curves.

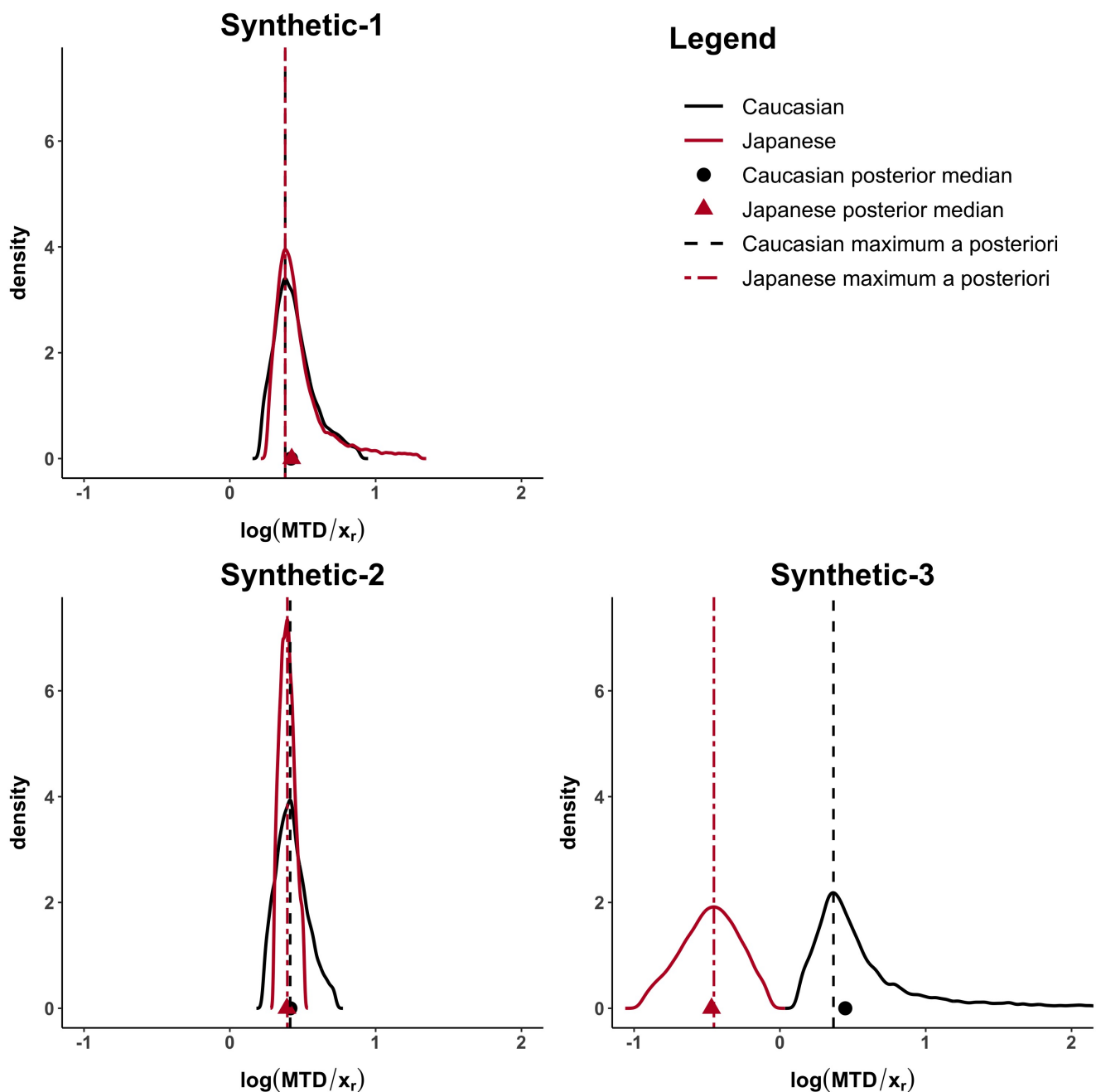
Figure 2 and Figure A1, in the Appendix A, show how the Caucasian posterior distribution is different in the three synthetic examples even if it comes from the same



Caucasian dataset. This behaviour is due to the variance adjustment given by  $\min\left(1, \frac{n_a}{n_c}\right)$ . In general, the posterior peak is preserved and the variance increases when the exponent is less than 1 (as in the synthetic-3 example).

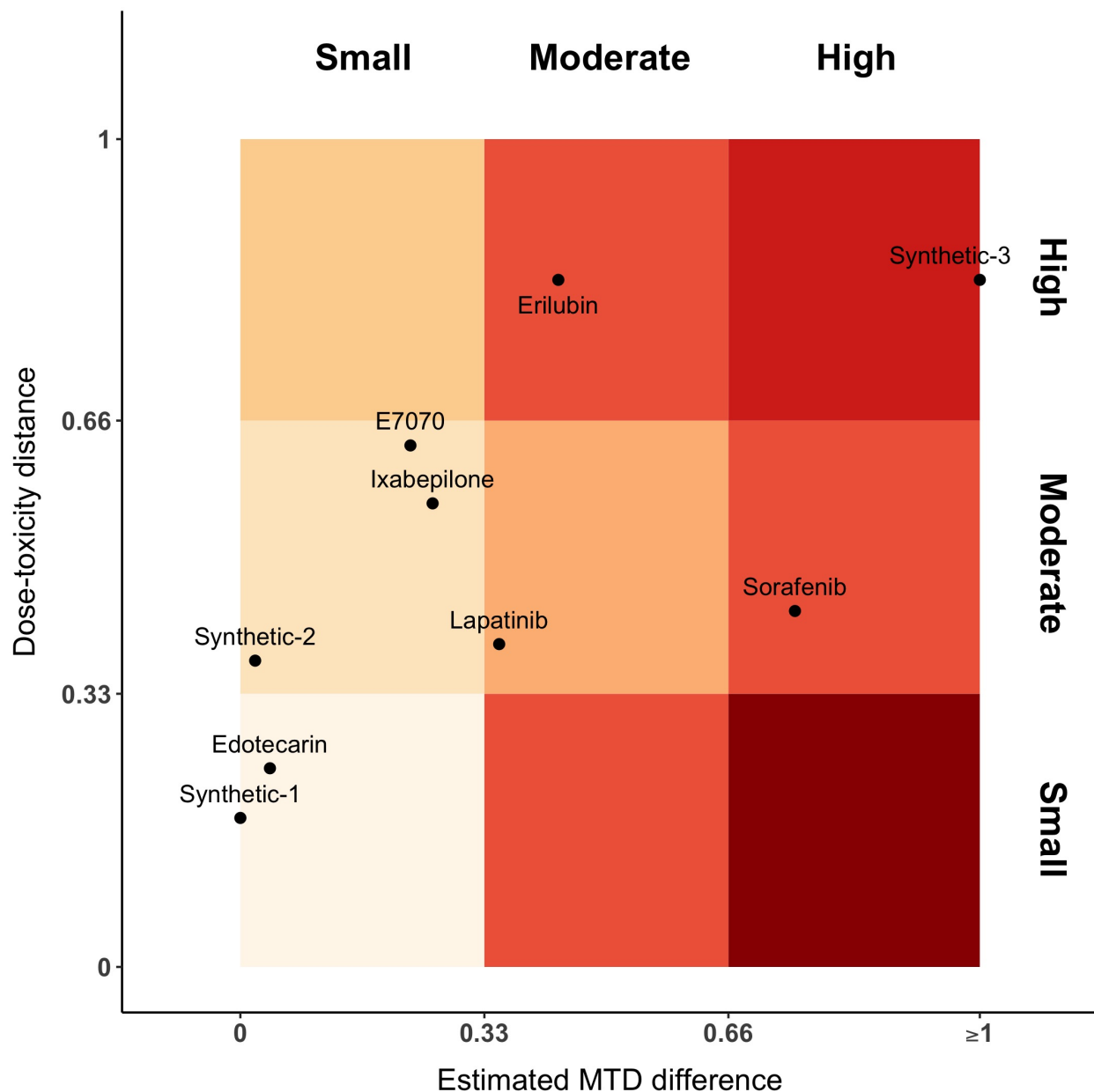


**Figure 1.** MTD posterior distributions for Erilubin, Ixabepilone, Lapatinib, Sorafenib, Edotecarin and E7070 case studies. Posterior medians are represented by a circle for Caucasian and a triangle for Japanese, while *maximum a posteriori* is represented by a dashed line for Caucasian and a two-dash line for Japanese.



**Figure 2.** MTD posterior distributions for the Synthetic-1, Synthetic-2 and Synthetic-3 examples. Posterior medians are represented by a circle for Caucasian and a triangle for Japanese, while *maximum a posteriori* by a dashed line for Caucasian and a two-dash line for Japanese.

Figure 3 represents the distance between dose–toxicity curves,  $d_{mod}$ , and maximum of the posterior MTD distribution,  $d_{p2}$ . For the sake of interpretability, we have equally divided the axes into three parts, each one denoting a small, moderate or high distance, respectively. In this plot, Sorafenib has moderate distances between curves and high difference between MTDs. This is the opposite for Erlotinib, where there is a moderate difference between MTD and a large distance between curves. When MTDs are similar or close (first column of the gradient), Edotecarin has similar dose–toxicity curves, while the distance between curves of Ixabepilone and E7070 is moderate. Lapatinib shows a moderate distance of both dose–toxicity curve and estimated MTDs.



**Figure 3.** Gradient plot representing the distance between dose–toxicity curves,  $d_{mod}$  ( $y$ -axis), and maximum of the posterior MTD distribution,  $d_{p2}$  ( $x$ -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.

## 6. Discussion

The aim of our work was to propose several Bayesian indicators to support further decisions when using a bridging data package [1]. Bayesian methods permit the definition of a similarity degree based on posterior distribution, which do not rely on asymptotic approximations and can be used also in small sample size settings. Specifically, we proposed Bayesian indicators which evaluate the distance and the similarity of dose–toxicity curves and MTD. When evaluating a drug among different populations, assessing the dose–response curves similarity is of most importance, since, if it is proved, other clinical trial data can be used, as well as extrapolation from one population to the other. Maeda and Kurokawa [5] pointed out the difficulty of defining a commensurability measure for different populations.

We presented and studied five criteria, where three of them,  $d$ ,  $d_{mod}$  and  $d_{MTD}$ , measure the similarity between dose–toxicity curves, and two of them,  $d_{p1}$  and  $d_{p2}$ , measure

the distance between the median and the *maximum a posteriori* of the MTD posterior distributions. The first three measures are bounded between 0 and 1 and their interpretation could be linked to a proportion. The second ones,  $d_{p1}$  and  $d_{p2}$  have a ratio-like value with a lower bound at 0. In practice, they represent a relative risk measure.

Our approach allows for the identification and discussion of similarities and differences between dose–toxicity curves and MTDs. However, as small samples were used in these studies, estimation of the entire dose–toxicity curve, when only part of the doses in the panel were evaluated, is complex and leads to an estimation with high variability. This is reflected in the values of  $d$ ,  $d_{mod}$  and  $d_{MTD}$ , which in our real case studies were above 0.2. When high differences between  $d$  and  $d_{mod}$  are observed, this is probably due to computational difficulties in Equation (1), especially in computing the weighted likelihood without a stabilization term. In general,  $d_{mod}$  is lower than  $d_{MTD}$ . This could be expected for two reasons: (i)  $d_{MTD}$  introduces, via the transformation, more variability (increased in the density estimation step); (ii)  $d_{MTD}$  is computed after truncating the posterior induced distribution of the MTD. Moreover, we showed that  $d_{p2}$ , based on the maximum a posteriori, is more stable than  $d_{p1}$ , which is based on the median, in the presence of unbalanced heavy tails. Therefore,  $d_{p2}$  could be suggested as a more reliable measure in this setting. We have attempted the analysis while varying the variance matrix of the bivariate normal prior distribution and  $d_{p1}$  was less stable (results not shown).

The MTD definition can vary according to the trial and to the population. Therefore, even if the same MTD is claimed in both Caucasian and Japanese populations, our analysis can identify differences. For instance, in the Japanese trial of Sorafenib, 400 mg/day is defined in the clinical trial as the MTD, but at the closest higher dose level, 600 mg/day, only one patient experienced toxicities (16.7%). Otherwise, in the Caucasian trial, three patients out of seven experienced toxicity at 600 mg/day (42.6%). Even if the two trials find the same MTD, the toxicity probability associated with each one differs. That is the reason why our results showed otherwise. Indeed, in the published clinical trials, there is a discrepancy between the method section defining the MTD and the real given MTD at the end of the trial. Our methods are based on data only and allow for evaluation of the actual similarity.

We decided to present the plot of the posterior densities (of the parameters and of the MTD) as it shows the super-position (or not) of the information. Plotting directly one-dimensional dose–response curves could, instead, be misleading and give hazardous interpretation.

A first limitation of our work is that we used published data, where the reporting can be sometimes incomplete in terms of DLTs and doses. For instance, in the paper of Burris et al. [18], we had to re-compose the DLT table and the dose-allocation sequence. Therefore, some interpretation discrepancy can be found in our Table 2. The issue of poor reporting in cancer trials was already raised by Zohar et al. [28] and Comets and Zohar [29]. As a second limitation, we did not provide fixed cut-offs for each criterion. In our opinion, the choice of the cut-offs depends on the application and on the quantity of information in the two trials. The more information we have, the more stringent cut-offs can be considered. Figure 3 only represents a proposition on the way to display the results.

The criteria proposed in this manuscript may be extended to be used in other settings. For example, when several trials are available, a meta-analysis of the dose–toxicity curves or of the MTDs can be considered [30–32]. In this case, pairwise distances can be previously estimated, in an empirical Bayes approach, and then be used to model the heterogeneity parameter(s) or to set prior distribution(s). Other extensions, which do not involve necessarily Phase I studies, could be considered: (i) in adults–children extrapolation; (ii) when we are interested to jointly evaluate efficacy and toxicity [33]; (iii) when comparing outcomes (efficacy or toxicity) of the same drug in different indications; (iv) when dealing with similarities in subgroups; (v) in comparing historical control data with respect to the actual trial in randomized Phase III trials.

Being able to quantify distance and bridging between two populations at the end of early Phase I trials can be useful to better characterize the dose–toxicity relationship and

differences. In case of small or acceptable differences, the extrapolation process can be considered, as suggested in the ICH-E5.

**Supplementary Materials:** The following are available at <https://www.mdpi.com/1660-4601/18/4/1639/s1>.

**Author Contributions:** Conceptualization, A.O., M.U. and S.Z.; methodology, A.O. and M.U.; validation, S.Z. and S.M.; writing—original draft preparation, A.O.; writing—review and editing, M.U., S.Z. and S.M. All authors have read and agreed to the published version of the manuscript.

**Funding:** The research of Adrien Ollier and Moreno Ursino was funded by the Institut National Du Cancer, grant numbers INCa\_11324 and INCa\_9539, respectively.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** R scripts are given as Supplemental Materials.

**Conflicts of Interest:** The authors declare no conflict of interest.

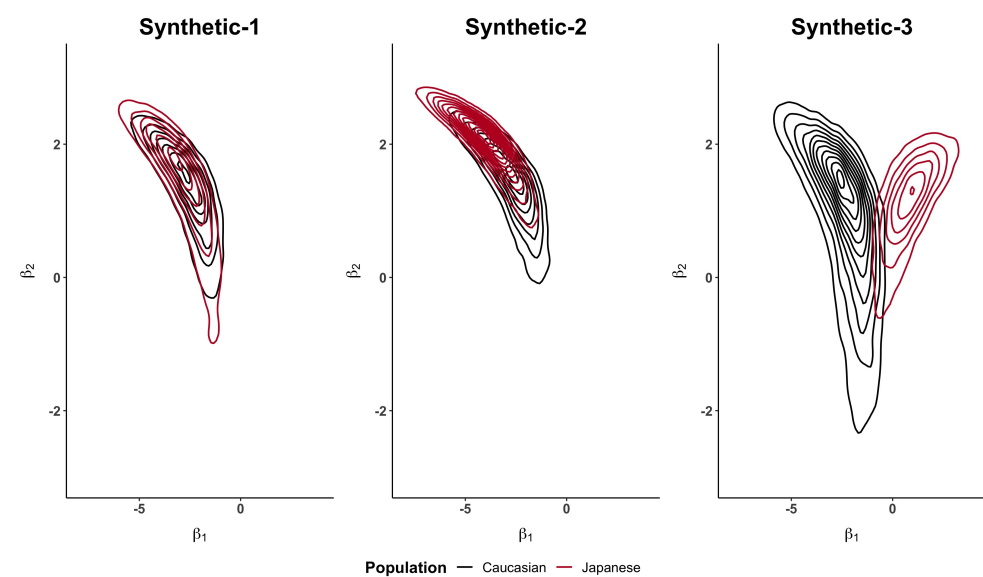
## Abbreviations

The following abbreviations are used in this manuscript:

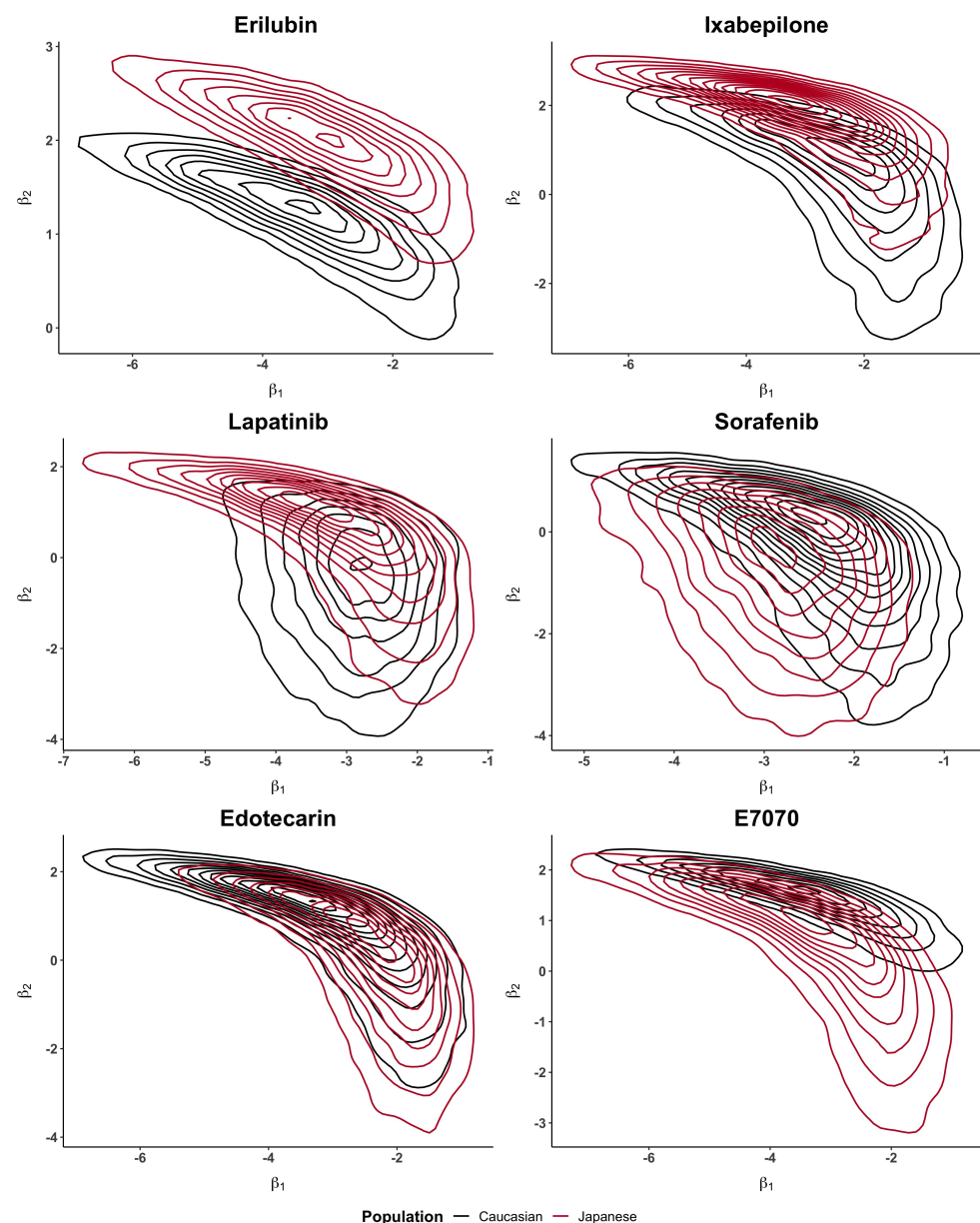
bid	<i>bis in die</i> : twice a day
BLRM	Bayesian Logistic Regression Model
CRM	Continual reassessment method
DLT	dose-limiting toxicity
ICH	International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use
MTD	maximum tolerated dose
PMDA	Pharmaceuticals and Medical Devices Agency

## Appendix A. Bivariate Posterior Plots

Figures A1 and A2 show the bivariate posterior distributions of  $\beta_0$  and  $\beta_1$  when using  $d_{mod}$ .



**Figure A1.** Bivariate posterior distributions of  $\beta_0$  and  $\beta_1$  when using  $d_{mod}$  for the three synthetic examples.



**Figure A2.** Bivariate posterior distributions of  $\beta_0$  and  $\beta_1$  when using  $d_{mod}$  for the real case studies shown in Table 2.

## References

1. ICH E5 (R1). *Ethnic Factors in the Acceptability of Foreign Clinical Data E5 (R1)*; International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use: Geneva, Switzerland, 1998.
2. Pharmaceuticals and Medical Devices Agency. Basic Principles for Conducting Phase I Trials in the Japanese Population Prior to Global Clinical Trials. 2015. Available online: <https://www.pmda.go.jp/files/000157777.pdf> (accessed on 26 September 2019).
3. Ogura, T.; Morita, S.; Yonemori, K.; Nonaka, T.; Urano, T. Exploring Ethnic Differences in Toxicity in Early-Phase Clinical Trials for Oncology Drugs. *Ther. Innov. Regul. Sci.* **2014**, *48*, 644–650. [[CrossRef](#)] [[PubMed](#)]
4. Malinowski, H.J.; Westelinck, A.; Sato, J.; Ong, T. Same drug, different dosing: Differences in dosing for drugs approved in the United States, Europe, and Japan. *J. Clin. Pharmacol.* **2008**, *48*, 900–908. [[CrossRef](#)]
5. Maeda, H.; Kurokawa, T. Differences in maximum tolerated doses and approval doses of molecularly targeted oncology drug between Japan and Western countries. *Investig. New Drugs* **2014**, *32*, 661–669. [[CrossRef](#)]
6. Mizugaki, H.; Yamamoto, N.; Fujiwara, Y.; Nokihara, H.; Yamada, Y.; Tamura, T. Current status of single-agent phase I trials in Japan: toward globalization. *J. Clin. Oncol.* **2015**, *33*, 2051–2061. [[CrossRef](#)]
7. O’Quigley, J.; Iasonos, A. Bridging Solutions in Dose Finding Problems. *Stat. Biopharm. Res.* **2014**, *6*, 185–197. [[CrossRef](#)] [[PubMed](#)]

8. Liu, S.; Pan, H.; Xia, J.; Huang, Q.; Yuan, Y. Bridging continual reassessment method for phase I clinical trials in different ethnic populations. *Stat. Med.* **2015**, *34*, 1681–1694. [[CrossRef](#)] [[PubMed](#)]
9. Takeda, K.; Morita, S. Incorporating Historical Data in Bayesian Phase I Trial Design: The Caucasian-to-Asian Toxicity Tolerability Problem. *Ther. Innov. Regul. Sci.* **2015**, *49*, 93–99. [[CrossRef](#)] [[PubMed](#)]
10. Ollier, A.; Morita, S.; Ursino, M.; Zohar, S. An adaptive power prior for sequential clinical trials—Application to bridging studies. *Stat. Methods Med. Res.* **2020**, *29*, 2282–2294. [[CrossRef](#)]
11. Bretz, F.; Möllenhoff, K.; Dette, H.; Liu, W.; Trampisch, M. Assessing the similarity of dose response and target doses in two non-overlapping subgroups. *Stat. Med.* **2016**. [[CrossRef](#)] [[PubMed](#)]
12. Ibrahim, J.G.; Chen, M.H.; Gwon, Y.; Chen, F. The power prior: Theory and applications. *Stat. Med.* **2015**, *34*, 3724–3749. [[CrossRef](#)]
13. O’Quigley, J.; Pepe, M.; Fisher, L. Continual reassessment method: A practical design for phase 1 clinical trials in cancer. *Biometrics* **1990**, *33*–48. [[CrossRef](#)]
14. Neuenschwander, B.; Branson, M.; Gsponer, T. Critical aspects of the Bayesian approach to phase I cancer trials. *Stat. Med.* **2008**, *27*, 2420–2439. [[CrossRef](#)]
15. Zheng, H.; Hampson, L.V. A Bayesian decision-theoretic approach to incorporate preclinical information into phase I oncology trials. *Biom. J.* **2020**, *62*, 1408–1427. [[CrossRef](#)] [[PubMed](#)]
16. Tan, A.R.; Rubin, E.H.; Walton, D.C.; Shuster, D.E.; Wong, Y.N.; Fang, F.; Ashworth, S.; Rosen, L.S. Phase I study of eribulin mesylate administered once every 21 days in patients with advanced solid tumors. *Clin. Cancer Res. Off. J. Am. Assoc. Cancer Res.* **2009**, *15*, 4213–4219. [[CrossRef](#)]
17. Mukohara, T.; Nagai, S.; Mukai, H.; Namiki, M.; Minami, H. Eribulin mesylate in patients with refractory cancers: A Phase I study. *Investig. New Drugs* **2011**, *30*, 1926–1933. [[CrossRef](#)]
18. Burris, H.A.; Hurwitz, H.I.; Dees, E.C.; Dowlati, A.; Blackwell, K.L.; O’Neil, B.; Marcom, P.K.; Ellis, M.J.; Overmoyer, B.; Jones, S.F.; et al. Phase I Safety, Pharmacokinetics, and Clinical Activity Study of Lapatinib (GW572016), a Reversible Dual Inhibitor of Epidermal Growth Factor Receptor Tyrosine Kinases, in Heavily Pretreated Patients With Metastatic Carcinomas. *J. Clin. Oncol. Off. J. Am. Soc. Clin. Oncol.* **2005**, *23*, 5305–5313. [[CrossRef](#)]
19. Nakagawa, K.; Minami, H.; Kanezaki, M.; Mukaiyama, A.; Minamide, Y.; Uejima, H.; Kurata, T.; Nogami, T.; Kawada, K.; Mukai, H.; et al. Phase I Dose-escalation and Pharmacokinetic Trial of Lapatinib (GW572016), a Selective Oral Dual Inhibitor of ErbB-1 and -2 Tyrosine Kinases, in Japanese Patients with Solid Tumors. *Jpn. J. Clin. Oncol.* **2009**, *39*, 116–123. [[CrossRef](#)]
20. Moore, M.; Hirte, H.; Siu, L.; Oza, A.; Hotte, S.; Petrenciuc, O.; Cihon, F.; Lathia, C.; Schwartz, B. 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. *Ann. Oncol.* **2005**, *16*, 1688–1694. [[CrossRef](#)] [[PubMed](#)]
21. Minami, H.; Kawada, K.; Ebi, H.; Kitagawa, K.; Kim, Y.i.; Araki, K.; Mukai, H.; Tahara, M.; Nakajima, H.; Nakajima, K. Phase I and pharmacokinetic study of sorafenib, an oral multikinase inhibitor, in Japanese patients with advanced refractory solid tumors. *Cancer Sci.* **2008**, *99*, 1492–1498. [[CrossRef](#)]
22. Aghajanian, C.; Burris, H.A.; Jones, S.; Spriggs, D.R.; Cohen, M.B.; Peck, R.; Sabbatini, P.; Hensley, M.L.; Greco, F.A.; Dupont, J.; et al. Phase I Study of the Novel Epothilone Analog Ixabepilone (BMS-247550) in Patients with Advanced Solid Tumors and Lymphomas. *J. Clin. Oncol.* **2007**, *25*, 1082–1088. [[CrossRef](#)] [[PubMed](#)]
23. Shimizu, T.; Yamamoto, N.; Yamada, Y.; Fujisaka, Y.; Yamada, K.; Fujiwara, Y.; Takayama, K.; Tokudome, T.; Klimovsky, J.; Tamura, T. Phase I clinical and pharmacokinetic study of 3-weekly, 3-h infusion of ixabepilone (BMS-247550), an epothilone B analog, in Japanese patients with refractory solid tumors. *Cancer Chemother. Pharmacol.* **2008**, *61*, 751–758. [[CrossRef](#)] [[PubMed](#)]
24. Hurwitz, H.I.; Cohen, R.B.; McGovren, J.P.; Hirawat, S.; Petros, W.P.; Natsumeda, Y.; Yoshinari, T. A phase I study of the safety and pharmacokinetics of edotecarin (J-107088), a novel topoisomerase I inhibitor, in patients with advanced solid tumors. *Cancer Chemother. Pharmacol.* **2007**, *59*, 139–147. [[CrossRef](#)] [[PubMed](#)]
25. Yamada, Y.; Tamura, T.; Yamamoto, N.; Shimoyama, T.; Ueda, Y.; Murakami, H.; Kusaba, H.; Kamiya, Y.; Saka, H.; Tanigawara, Y.; et al. Phase I and pharmacokinetic study of edotecarin, a novel topoisomerase I inhibitor, administered once every 3 weeks in patients with solid tumors. *Cancer Chemother. Pharmacol.* **2006**, *58*, 173–182. [[CrossRef](#)] [[PubMed](#)]
26. Raymond, E.; ten Bokkel Huinink, W.; Taïeb, J.; Beijnen, J.; Faivre, S.; Wanders, J.; Ravic, M.; Fumoleau, P.; Armand, J.; Schellens, J. Phase I and Pharmacokinetic Study of E7070, a Novel Chloroindolyl Sulfonamide Cell-Cycle Inhibitor, Administered as a One-Hour Infusion Every Three Weeks in Patients with Advanced Cancer. *J. Clin. Oncol.* **2002**, *20*, 3508–3521. [[CrossRef](#)] [[PubMed](#)]
27. Yamada, Y.; Yamamoto, N.; Shimoyama, T.; Horiike, A.; Fujisaka, Y.; Takayama, K.; Sakamoto, T.; Nishioka, Y.; Yasuda, S.; Tamura, T. Phase I pharmacokinetic and pharmacogenomic study of E7070 administered once every 21 days. *Cancer Sci.* **2005**, *96*, 721–728. [[CrossRef](#)]
28. Zohar, S.; Lian, Q.; Levy, V.; Cheung, K.; Ivanova, A.; Chevret, S. Quality assessment of phase I dose-finding cancer trials: Proposal of a checklist. *Clin. Trials* **2008**, *5*, 478–485. [[CrossRef](#)]
29. Comets, E.; Zohar, S. A survey of the way pharmacokinetics are reported in published phase I clinical trials, with an emphasis on oncology. *Clin. Pharmacokinet.* **2009**, *48*, 387–395. [[CrossRef](#)]
30. Zohar, S.; Katsahian, S.; O’Quigley, J. An approach to meta-analysis of dose-finding studies. *Stat. Med.* **2011**, *30*, 2109–2116. [[CrossRef](#)]

- 
31. Ursino, M.; Röver, C.; Zohar, S.; Friede, T. Random-effects meta-analysis of phase I dose-finding studies using stochastic process priors. *arXiv* **2019**, arXiv:1908.06733.
  32. Röver, C.; Friede, T. Dynamically borrowing strength from another study through shrinkage estimation. *Stat. Methods Med. Res.* **2019**, *29*, 293–308. [[CrossRef](#)]
  33. Thall, P.F.; Cook, J.D. Dose-finding based on efficacy–toxicity trade-offs. *Biometrics* **2004**, *60*, 684–693. [[CrossRef](#)] [[PubMed](#)]





# Conclusion

In oncology, phase I clinical trials aim is to evaluate the toxicity of a new drug and to find the MTD. Sequential designs are often used for such trials. It is common for a drug to be tested several times, for example in different regions, since two populations will not necessary have the same response. This could be due to ethnic diversity in metabolism and genetics. Bridging studies are recommended to bridge the gap between different populations, in terms of dose-response relationship, toxicity etc. In this setting, our work focused on proposing similarity criteria between two trials, to understand when sharing information between them is pertinent or not.

Indeed, when a drug has already been tested, it is advisable to take into account the data collected during the previous clinical trial for planning and analysing a new clinical trial, notably bridging studies. Bayesian statistics is a suitable approach to take into account these historical data, through the notion of prior distributions. However, two main questions arise: (i) the possible contrast between this prior distribution and future data; (ii) the amount of information that this prior should contain.

In the first part of our work, we developed a statistical method to incorporate historical information from an historical to a current trial, in the setting of dose-finding trials. An appropriate borrowing of historical information was supposed to allow a better MTD selection. In a Bayesian setting, the power prior distribution allows to incorporate an historical likelihood raised by a power  $\alpha \in [0, 1]$  as a prior distribution. This power  $\alpha$  permits to calibrate the amount of historical information into the prior distribution, through the ESS approach. Then, to detect a possible data-prior conflict, we developed a similarity criterion. This similarity criterion based on the Hellinger's distance compares the historical and current likelihood functions, taking into account the difference of number of patients into the two likelihoods. Comparing two normalised likelihood function permits to compare two dose-response relationships.

Thereby, our method allows both a control of the quantity of information in a prior distribution through the ESS notion taking into account historical data according to the

size of the current dataset, but also the possibility of a data-prior conflict. These two aspects had been studied separately in the scientific literature, we adapted and used them simultaneously. Our simulation study showed that borrowing historical information is highly desirable. When both historical and current datasets were similar, the MTD selection was strongly improved, while it was comparable when both datasets were different.

The simulation study also showed that the sensitivity analysis, performed at planning stage, should be done carefully. Indeed, several points must be considered. First of all, the number of patients in the current trial: our simulations were done with thirty patients, but the results could change with the number of patients enrolled into the trial. Second, the distance was evaluated at each new cohort of one patient. Cohort with more than one patient can be considered. Finally, we should remind that the similarity criterion was built to compare two likelihoods. It means that the statistical model chosen for the likelihood function induces differences for the computed distance criterion.

The second part of the presented work focuses more specifically on similarity criteria for finished clinical trials. Although bridging studies should permit to decide if an extrapolation between two populations is possible [35], the use of statistical criteria to assess such a possibility were rarely specifically developed.

We proposed several extensions of the similarity criterion proposed in the first part. Indeed, two dose-response curves can be different in several ways. In the first part of our work, the likelihood functions were compared, it permitted to compare two dose-response relationships. However, it is also possible for two populations to have the same MTD but different dose-relationships. It was therefore necessary to propose other criteria to compare two datasets. The proposed criteria were evaluated through three synthetic datasets with same MTD and similar dose-toxicity curves, same MTD but different dose-toxicity curves or all different. The criteria appeared to be able to assess the several kind of similarity/dissimilarity. The criteria were then evaluated with published datasets from phase I clinical trials. We also provide illustrative plots to help for the interpretation of the results. The provided plots illustrated the fact that only one similarity criterion is not enough to assess the similarity of two dose-response curves, but at least two are necessary.

In a context where clinical trials for the same drug are repeated, it is necessary to have objective tools to calibrate the sharing of information from an old clinical trial to a new one in order to avoid possible data-prior conflict. Furthermore, such criteria have

proven to be effective in judging the similarity between two trials from two different regions.



# Discussion

The aim of my PhD research work was to propose a framework to incorporate external data either when planning new clinical trials in a different population or when comparing distributions of parameter of interest in two distinct populations. To do so, we proposed a first tool for quantifying the maximum amount of information to be incorporated using Bayesian inference, in particular, the ESS driven prior distribution calibration. Second, we proposed a modelling approach to evaluate the commensurability, that is, the similarity between two datasets or two populations. These tools allow to quantify the prior amount of information and the data-prior conflict before deciding whenever to use external data or not.

In this project, we evaluated the proposed framework in the setting of phase I dose-finding clinical trials in oncology. However, our modelling approach can be used at any evaluation phase either in interventional or observational studies. The issue of "if" and "how" to use external information is currently in discussion in the methodological field ([61],[77]). Especially, recently, with unplanned clinical trial disruptions due to the Covid 19 pandemic. Indeed, many clinical trials have been stopped and could not restart for different reasons. Nevertheless, for some of them, using external data sources could allow to cope with information loss and to avoid research waste, more critical, to keep from including patients in inconclusive trials. Our methods could be used in this setting when, first, evaluating the similarity between data-sets and if it is acceptable, second, in calibrating how much information should be used (avoiding overpowering the current trial data-set).

Yet, there are some limits to the proposed methods. One of them is associated with the computational difficulties and approximation that needed to be done in order to cope with it. These approximations choices have influenced the estimation of the commensurability criterion; however, we have considered it be minor. Some limits could be related to the CRM design itself. For instance if the model is misspecified, the interpretation of

the results could be more difficult. However, since the Hellinger distance is applied to normalised likelihoods, the commensurability criterion should work even in this situation. The two datasets are fitted with the same misspecified model and their "distance" in results should be preserved. For instance, when comparing two exponential datasets but fitted with gaussian likelihoods, the distance estimation was increasing but still able to detect when the two datasets were similar or not. In the case of dose-finding design with logistic regression models, it could be interesting to systematically assess the consistency of the model with the data [88]. Our method based on adaptive power prior applied to phase I dose-finding trials focused on the CRM parameter of interest. This parameter is then used to estimate the dose-toxicity curve and the MTD. Thus, the information of the dose-response curve is given by the posterior distribution of the unique parameter of the model, and the Hellinger distance applied to normalised likelihoods (of one parameter) can be considered equivalent to the Hellinger distance between the posterior distributions of the MTD. However, the MTD itself could be a parameter of interest, like in the EWOC approach (with two parameters). As shown in the second work presented in this manuscript, two populations can have the same MTD but a different dose-response curve. A model with only one parameter is not able to catch both quantities. This is why investigating the proposed framework with other dose-finding designs might be desirable and will be a future work. For example, a sequential dose-finding method will be investigated in the setting of the EWOC design, where we will aim at calibrating the prior distributions considered for the joint density for the probability of toxicity of the first dose and the MTD.

In this work, we considered that the prospective trial follows the same design of the historical one, that is, the CRM in our case. We can imagine to extend this method when the historical trial was done under a different design. For example, algorithm designs are still widespread for phase I dose-finding trial. Therefore, a retrospective analysis of historical data, for instance with the retrospective CRM to first assess the probability of toxicity of each dose of the historical trial, could be necessary if the CRM is considered as design for the prospective trial. In this situation, the method should be generalised. Indeed, the retrospective CRM does not provide historical data as results, but a parameter estimation. In this situation, the power prior approach cannot be used as it was formulated and a new prior distribution should be defined, starting from the retrospective CRM results. Then, its variance can be calibrate using our framework, that is, ESS and commensurability tools. More generally, this question can be raised into other statistical settings, when the design could influence the data analysis.

An extension to several historical datasets could and will be investigated extending the multiple power priors model proposed by Ibrahim and Chen [33]. Indeed, we can consider to add the commensurability parameter for each historical trial. However, the computational time can be very long (as it is also the case for the meta-analytic prior): while it is not a main issue in oncology setting, where the time window between the patient screening and the his/her trial inclusion is usually large, but it can be difficult in other setting, as in paediatrics, notably in newborn setting, where the new dose level should be estimated rapidly in an emergency situation. Moreover, these historical trials might have been performed with several designs, a retrospective analysis could be required and a mixed framework (power prior and the extensions described in the previous paragraph) would be necessary.

Future works regard the exploration of two key points of clinical trials: the sample size and the dose panel choice. If the two datasets (historical and prospective one) seems to be similar, the number of patients in the ongoing trial to obtain a MTD estimation associated to a pre-specified precision threshold can be reduced with respect to the total planned sample size. Otherwise, the total sample size will be used when historical and new data populations seems to be in conflict. Moreover, the historical data can be used not only to define prior distribution but also to chose the dose panel. An example of planning a dose-finding trial in paediatrics using historical adult data have been proposed by [73].

A main part of our methods depend on the likelihood, which implies the weakness of a model could possibly have repercussions on the method. This is why an extension to non-parametric approaches could be interesting (see Appendix B). In case of non-parametric density estimation, our framework can be easily adapted since it is based distribution-kind object: indeed the normalised likelihood used in the commensurability part, can be seen as a posterior distribution generated by an uniform prior distribution (degenerate if necessary).





# References

- [1] Syed Mumtaz Ali and Samuel D Silvey. A general class of coefficients of divergence of one distribution from another. *Journal of the Royal Statistical Society: Series B (Methodological)*, 28(1):131–142, 1966.
- [2] Shun-ichi Amari. *Information geometry and its applications*, volume 194. Springer, 2016.
- [3] Revathi Ananthakrishnan, Stephanie Green, Daniel Li, and Michael LaValley. 2d (2 dimensional) teqr design for determining the optimal dose for safety and efficacy. *Contemporary Clinical Trials Communications*, 16:100461, 2019.
- [4] James Babb, Andre Rogatko, and Shelemyahu Zacks. Cancer phase i clinical trials: efficient dose escalation with overdose control. *Statistics in Medicine*, 17(10):1103–1120, 1998.
- [5] Michèle Basseville. Divergence measures for statistical data processing—an annotated bibliography. *Signal Processing*, 93(4):621–633, 2013.
- [6] Paola Bortot and Alessandra Giovagnoli. Up-and-down experiments of first and second order. *Journal of Statistical Planning and Inference*, 134:236–253, 09 2005.
- [7] Sandrine Boulet, Moreno Ursino, Peter Thall, Bruno Landi, Céline Lepère, Simon Pernot, Anita Burgun, Julien Taieb, Aziz Zaanani, Sarah Zohar, et al. Integration of elicited expert information via a power prior in bayesian variable selection: Application to colon cancer data. *Statistical methods in medical research*, 29(2):541–567, 2020.
- [8] Thomas M Braun. The bivariate continual reassessment method: extending the crm to phase i trials of two competing outcomes. *Controlled clinical trials*, 23(3):240–256, 2002.
- [9] Lev M Bregman. The relaxation method of finding the common point of convex sets and its application to the solution of problems in convex programming. *USSR computational mathematics and mathematical physics*, 7(3):200–217, 1967.
- [10] Frank Bretz, Kathrin Möllenhoff, Holger Dette, Wei Liu, and Matthias Trampisch. Assessing the similarity of dose response and target doses in two non-overlapping subgroups. *Statistics in medicine*, 37(5):722–738, 2018.
- [11] M.-H. Chen, J. G. Ibrahim, and C. Yiannoutsos. Prior elicitation, variable selection and bayesian computation for logistic regression models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 61(1):223–242, 1999.

- 
- [12] Ying Kuen Cheung. Dose finding with delayed binary outcomes in cancer trials. *Statistical Methods for Dose Finding Experiments*, pages 225–242, 2006.
- [13] Ying Kuen Cheung and Rick Chappell. Sequential designs for phase i clinical trials with late-onset toxicities. *Biometrics*, 56(4):1177–1182, 2000.
- [14] Ying Kuen Cheung and Rick Chappell. Sequential designs for phase i clinical trials with late-onset toxicities. *Biometrics*, 56(4):1177–1182, 2000.
- [15] Ying Kuen Ken Cheung. *Dose Finding by the Continual Reassessment Method*. Chapman and Hall/CRC, 2011.
- [16] Emmanuelle Comets and Sarah 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.
- [17] Imre Csiszár. Information-type measures of difference of probability distributions and indirect observation. *studia scientiarum Mathematicarum Hungarica*, 2:229–318, 1967.
- [18] Kristen M Cunanan and Joseph S Koopmeiners. Efficacy/toxicity dose-finding using hierarchical modeling for multiple populations. *Contemporary clinical trials*, 71:162–172, 2018.
- [19] Yuyan Duan, Keying Ye, and Eric P Smith. Evaluating water quality using power priors to incorporate historical information. *Environmetrics: The Official Journal of the International Environmetrics Society*, 17(1):95–106, 2006.
- [20] Stephen D. Durham and Nancy Flournoy. Up-and-down designs i: Stationary treatment distributions. *Lecture Notes-Monograph Series*, 25:139–157, 1995.
- [21] Jeff Evans, Emma Dean, L Rhoda Molife, Juanita Lopez, Malcolm Ranson, Fatima El-Khouly, Ishtiaq Zubairi, Claudio Savulsky, Larisa Reyderman, Yan Jia, Lorna Sweeting, Alastair Greystoke, Jorge Barriuso, and Rebecca Kristeleit. Phase 1 dose-finding and pharmacokinetic study of eribulin-liposomal formulation in patients with solid tumours. *British Journal of Cancer*, 120:1, 01 2019.
- [22] Mauro Gasparini and Jeffrey Eisele. A curve-free method for phase i clinical trials. *Biometrics*, 56(2):609–615, 2000.
- [23] Andrew Gelman, John B Carlin, Hal S Stern, David B Dunson, Aki Vehtari, and Donald B Rubin. *Bayesian data analysis*. CRC press, 2013.
- [24] Andrew Gelman, Daniel Simpson, and Michael Betancourt. The prior can often only be understood in the context of the likelihood. *Entropy*, 19(10):555, 2017.
- [25] Emma Gerard, Sarah Zohar, Hoai-Thu Thai, Christelle Lorenzato, Marie-Karelle Riviere, and Moreno Ursino. Bayesian dose-regimen assessment in early phase oncology incorporating pharmacokinetics and pharmacodynamics. *arXiv preprint arXiv:2011.11480*, 2020.

- [26] S Gsteiger, F Bretz, and W Liu. Simultaneous confidence bands for nonlinear regression models with application to population pharmacokinetic analyses. *Journal of Biopharmaceutical Statistics*, 21(4):708–725, 2011.
- [27] Burak Kürsüd Günhan, Sebastian Weber, and Tim 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.
- [28] Nusrat Harun, Chunyan Liu, and Mi-Ok Kim. Critical appraisal of bayesian dynamic borrowing from an imperfectly commensurate historical control. *Pharmaceutical statistics*, 19(5):613–625, 2020.
- [29] Brian P. Hobbs, Bradley P. Carlin, Sumithra J. Mandrekar, and Daniel J. Sargent. Hierarchical commensurate and power prior models for adaptive incorporation of historical information in clinical trials. *Biometrics*, 67(3):1047–1056, 2011.
- [30] Brian P Hobbs, Daniel J Sargent, and Bradley P Carlin. Commensurate priors for incorporating historical information in clinical trials using general and generalized linear models. *Bayesian analysis (Online)*, 7(3):639, 2012.
- [31] JA Hoeting, D Madigan, AE Raftery, and CT Volinsky. Bayesian model averaging: a tutorial (with comments by m. clyde, david draper and ei george, and a rejoinder by the authors. *stat sci* 14 (4): 382–417. *Referred to on*, page 2, 1999.
- [32] Alexia Iasonos, Andrew S Wilton, Elyn R Riedel, Venkatraman E Seshan, and David R Spriggs. A comprehensive comparison of the continual reassessment method to the standard 3 + 3 dose escalation scheme in phase i dose-finding studies. *Clinical Trials*, 5(5):465–477, 2008. PMID: 18827039.
- [33] Joseph G. Ibrahim and Ming-Hui Chen. Power prior distributions for regression models. *Statistical Science*, 15(1):46–60, 2000.
- [34] Joseph G Ibrahim, Ming-Hui Chen, Yeongjin Gwon, and Fang Chen. The power prior: theory and applications. *Statistics in medicine*, 34(28):3724–3749, 2015.
- [35] ICH. Ethnic Factors in the Acceptability of Foreign Clinical Data, 1998.
- [36] Anastasia Ivanova. Escalation, group and a + b designs for dose-finding trials. *Statistics in Medicine*, 25(21):3668–3678, 2006.
- [37] Anastasia Ivanova. Escalation, group and a + b designs for dose-finding trials. *Statistics in Medicine*, 25(21):3668–3678, 2006.
- [38] ET Jaynes. Marginalization and prior probabilities in bayesian analysis in econometrics and statistics, 1980.
- [39] Hirotsugu Kenmotsu and Yusuke Tanigawara. Pharmacokinetics, dynamics and toxicity of docetaxel: Why the japanese dose differs from the western dose. *Cancer Science*, 106(5):497–504, 2015.
- [40] Henri Lantéri, Céline Theys, and Claude Aime. Scale invariant divergences for signal and image reconstruction. In *2015 23rd European Signal Processing Conference (EUSIPCO)*, pages 2162–2166. IEEE, 2015.

- 
- [41] Shing M Lee and Ying Kuen Cheung. Model calibration in the continual reassessment method. *Clinical Trials*, 6(3):227–238, 2009. PMID: 19528132.
- [42] Shing M. Lee and Ying Kuen Cheung. Calibration of prior variance in the bayesian continual reassessment method. *Statistics in Medicine*, 30(17):2081–2089, 2011.
- [43] Denis Heng-Yan Leung and You-Gan Wang. Isotonic designs for phase i trials. *Controlled clinical trials*, 22(2):126–138, 2001.
- [44] Daniel H Li, James B Whitmore, Wentian Guo, and Yuan Ji. Toxicity and efficacy probability interval design for phase i adoptive cell therapy dose-finding clinical trials. *Clinical Cancer Research*, 23(1):13–20, 2017.
- [45] Yong Lin and Weichung J. Shih. Statistical properties of the traditional algorithm-based designs for phase I cancer clinical trials. *Biostatistics*, 2(2):203–215, 06 2001.
- [46] G Frank Liu. A dynamic power prior for borrowing historical data in noninferiority trials with binary endpoint. *Pharmaceutical statistics*, 17(1):61–73, 2018.
- [47] Jen-pei Liu, Shein-Chung Chow, and Chin-Fu Hsiao. *Design and analysis of bridging studies*. CRC Press, 2012.
- [48] Suyu Liu, Haitao Pan, Jielai Xia, Qin Huang, and Ying Yuan. Bridging continual reassessment method for phase i clinical trials in different ethnic populations. *Statistics in medicine*, 34(10):1681–1694, 2015.
- [49] W Liu, F Bretz, AJ Hayter, and HP Wynn. Assessing nonsuperiority, noninferiority, or equivalence when comparing two regression models over a restricted covariate region. *Biometrics*, 65(4):1279–1287, 2009.
- [50] Hideki Maeda and Tatsuo Kurokawa. Differences in maximum tolerated doses and approval doses of molecularly targeted oncology drug between japan and western countries. *Investigational New Drugs*, 32(4):661–669, Aug 2014.
- [51] Kathrin Möllenhoff, Frank Bretz, and Holger Dette. Equivalence of regression curves sharing common parameters. *Biometrics*, 2019.
- [52] Tetsuzo Morimoto. Markov processes and the h-theorem. *Journal of the Physical Society of Japan*, 18(3):328–331, 1963.
- [53] Satoshi Morita, Peter F Thall, and Peter Müller. Determining the effective sample size of a parametric prior. *Biometrics*, 64(2):595–602, 2008.
- [54] Satoshi Morita, Peter F Thall, and Peter Müller. Evaluating the impact of prior assumptions in bayesian biostatistics. *Statistics in biosciences*, 2(1):1–17, 2010.
- [55] Satoshi Morita, Peter F Thall, and Peter Müller. Prior effective sample size in conditionally independent hierarchical models. *Bayesian Analysis (Online)*, 7(3), 2012.
- [56] Carl N Morris et al. Natural exponential families with quadratic variance functions: statistical theory. *The Annals of Statistics*, 11(2):515–529, 1983.

- 
- [57] B Nebiyu Bekele and Yu Shen. A bayesian approach to jointly modeling toxicity and biomarker expression in a phase i/ii dose-finding trial. *Biometrics*, 61(2):343–354, 2005.
- [58] Beat Neuenschwander, Michael Branson, and Thomas Gsponer. Critical aspects of the bayesian approach to phase i cancer trials. *Statistics in medicine*, 27(13):2420–2439, 2008.
- [59] Beat Neuenschwander, Michael Branson, and David J Spiegelhalter. A note on the power prior. *Statistics in medicine*, 28(28):3562–3566, 2009.
- [60] Beat Neuenschwander, Gorana Capkun-Niggli, Michael Branson, and David J Spiegelhalter. Summarizing historical information on controls in clinical trials. *Clinical Trials*, 7(1):5–18, 2010.
- [61] Beat Neuenschwander, Satrajit Roychoudhury, and Heinz Schmidli. On the use of co-data in clinical trials. *Statistics in Biopharmaceutical Research*, 8(3):345–354, 2016.
- [62] Beat Neuenschwander, Sebastian Weber, Heinz Schmidli, and Anthony O’Hagan. Predictively consistent prior effective sample sizes. *arXiv preprint arXiv:1907.04185*, 2019.
- [63] Stavros Nikolakopoulos, Ingeborg van der Tweel, and Kit CB Roes. Dynamic borrowing through empirical power priors that control type i error. *Biometrics*, 74(3):874–880, 2018.
- [64] Takashi Ogura, Satoshi Morita, Kan Yonemori, Takahiro Nonaka, and Tsutomu Urano. Exploring ethnic differences in toxicity in early-phase clinical trials for oncology drugs. *Therapeutic innovation & regulatory science*, 48(5):644–650, 2014.
- [65] Adrien Ollier, Satoshi Morita, Moreno Ursino, and Sarah Zohar. An adaptive power prior for sequential clinical trials—application to bridging studies. *Statistical methods in medical research*, 2019.
- [66] Adrien Ollier, Sarah Zohar, Satoshi Morita, and Moreno 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.
- [67] John O’Quigley. Efficiency of bridging between related dose finding studies. *Comptes rendus mathématique*, 351(9-10):401–404, 2013.
- [68] John O’Quigley and Alexia Iasonos. Bridging solutions in dose-finding problems. *Statistics in biopharmaceutical research*, 6(2):185–197, 2014.
- [69] John O’Quigley, Margaret Pepe, and Lloyd Fisher. Continual reassessment method: a practical design for phase 1 clinical trials in cancer. *Biometrics*, 46(1):33–48, 1990.
- [70] John O’Quigley and Sarah Zohar. Retrospective robustness of the continual reassessment method. *Journal of biopharmaceutical statistics*, 20:1013–25, 09 2010.

- [71] Ferdinand Österreicher. Csiszár's f-divergences-basic properties. *RGMA Res. Rep. Coll*, 2002.
- [72] Philip Pallmann, Alun W Bedding, Babak Choodari-Oskooei, Munyaradzi Dimairo, Laura Flight, Lisa V Hampson, Jane Holmes, Adrian P Mander, Matthew R Sydes, Sofia S Villar, et al. Adaptive designs in clinical trials: why use them, and how to run and report them. *BMC medicine*, 16(1):1–15, 2018.
- [73] Caroline Petit, Adeline Samson, Satoshi Morita, Moreno Ursino, Jérémie Guedj, Vincent Jullien, Emmanuelle Comets, and Sarah Zohar. Unified approach for extrapolation and bridging of adult information in early-phase dose-finding paediatric studies. *Statistical methods in medical research*, 27(6):1860–1877, 2018.
- [74] Christian Robert. *The Bayesian choice: from decision-theoretic foundations to computational implementation*. Springer Science & Business Media, 2007.
- [75] André Rogatko, David Schoeneck, William Jonas, Mourad Tighiouart, Fadlo R. Khuri, and Alan Porter. Translation of innovative designs into phase i trials. *Journal of Clinical Oncology*, 25(31):4982–4986, 2007. PMID: 17971597.
- [76] Heinz Schmidli, Sandro Gsteiger, Satrajit Roychoudhury, Anthony O'Hagan, David Spiegelhalter, and Beat Neuenschwander. Robust meta-analytic-predictive priors in clinical trials with historical control information. *Biometrics*, 70(4):1023–1032, 2014.
- [77] Heinz Schmidli, Dieter A Häring, Marius Thomas, Adrian Cassidy, Sebastian Weber, and Frank Bretz. Beyond randomized clinical trials: Use of external controls. *Clinical Pharmacology & Therapeutics*, 107(4):806–816, 2020.
- [78] V Seegers, S Chevret, and M Resche-Rigon. Dose-finding design driven by efficacy in onco-hematology phase i/ii trials. *Statistics in medicine*, 30(13):1574–1583, 2011.
- [79] Jianfen Shu. *The CRM (Continual Reassessment Method) Designs in the Presence of Population Heterogeneity*. University of Virginia, 2012.
- [80] David J Spiegelhalter, Keith R Abrams, and Jonathan P Myles. *Bayesian approaches to clinical trials and health-care evaluation*, volume 13. John Wiley & Sons, 2004.
- [81] Barry E. Storer. Design and analysis of phase i clinical trials. *Biometrics*, 45(3):925–937, 1989.
- [82] Oleksandr Sverdlov, Weng Wong, and Yevgen Ryznik. Adaptive clinical trial designs for phase i cancer studies. *Statistics Surveys [electronic only]*, 8, 06 2014.
- [83] Ami Takahashi and Taiji Suzuki. Bayesian optimization design for dose-finding based on toxicity and efficacy outcomes in phase i/ii clinical trials. *Pharmaceutical Statistics*, 2020.
- [84] Kentaro Takeda, Kanji Komatsu, and Satoshi Morita. Bayesian dose-finding phase i trial design incorporating pharmacokinetic assessment in the field of oncology. *Pharmaceutical Statistics*, 17(6):725–733, 2018.

- [85] Kentaro Takeda and Satoshi Morita. Bayesian dose-finding phase i trial design incorporating historical data from a preceding trial. *Pharmaceutical Statistics*, 17(4):372–382, 2018.
- [86] Niansheng Tang, Songjian Wang, and Gen Ye. A nonparametric bayesian continual reassessment method in single-agent dose-finding studies. *BMC medical research methodology*, 18(1):172, 2018.
- [87] Peter F Thall and John D Cook. Dose-finding based on efficacy–toxicity trade-offs. *Biometrics*, 60(3):684–693, 2004.
- [88] Tue Tjur. Coefficients of determination in logistic regression models—a new proposal: The coefficient of discrimination. *The American Statistician*, 63(4):366–372, 2009.
- [89] Moreno Ursino, Sarah Zohar, Frederike Lentz, Corinne Alberti, Tim Friede, Nigel Stallard, and Emmanuelle Comets. Dose-finding methods for phase i clinical trials using pharmacokinetics in small populations. *Biometrical Journal*, 59(4):804–825, 2017.
- [90] Nolan A Wages, John O’Quigley, and Mark R Conaway. Phase i design for completely or partially ordered treatment schedules. *Statistics in medicine*, 33(4):569–579, 2014.
- [91] Graham M Wheeler, Adrian P Mander, Alun Bedding, Kristian Brock, Victoria Cornelius, Andrew P Grieve, Thomas Jaki, Sharon B Love, Christopher J Weir, Christina Yap, et al. How to design a dose-finding study using the continual reassessment method. *BMC medical research methodology*, 19(1):1–15, 2019.
- [92] Manuel Wiesenfarth and Silvia Calderazzo. Quantification of prior impact in terms of effective current sample size. *Biometrics*, 76(1):326–336, 2020.
- [93] SU Yasuda, L Zhang, and S-M Huang. The role of ethnicity in variability in response to drugs: Focus on clinical pharmacology studies. *Clinical Pharmacology & Therapeutics*, 84(3):417–423, 2008.
- [94] Guosheng Yin, Yisheng Li, and Yuan Ji. Bayesian dose-finding in phase i/ii clinical trials using toxicity and efficacy odds ratios. *Biometrics*, 62(3):777–787, 2006.
- [95] Guosheng Yin and Ying Yuan. Bayesian model averaging continual reassessment method in phase i clinical trials. *Journal of the American Statistical Association*, 104(487):954–968, 2009.
- [96] Ying Yuan, Jielai Xia, et al. A calibrated power prior approach to borrow information from historical data with application to biosimilar clinical trials. *Journal of the Royal Statistical Society. Series C, Applied statistics*, 66(5):979, 2017.
- [97] Yong Zang, J Jack Lee, and Ying Yuan. Adaptive designs for identifying optimal biological dose for molecularly targeted agents. *Clinical Trials*, 11(3):319–327, 2014.
- [98] Haiyan Zheng and Lisa V Hampson. A bayesian decision-theoretic approach to incorporate preclinical information into phase i oncology trials. *Biometrical Journal*, 62(6):1408–1427, 2020.



- [99] Haiyan Zheng, Lisa V Hampson, and Thomas Jaki. Bridging across patient subgroups in phase i oncology trials that incorporate animal data. *Statistical Methods in Medical Research*, page 0962280220986580, 2021.

# Appendix A

## Bayesian statistics and decision theory

In the setting of decision theory of Wald, we consider the triplet  $(E, \mathcal{B}, \mathcal{P})$  where  $\mathcal{P}$  is a set of probability measures on the measure space  $(E, \mathcal{B})$ , a measure space  $(D, \mathcal{D})$  called decision space, a set  $\mathcal{R}$  of  $\mathcal{B} - \mathcal{D}$  measurable functions called decisions functions.

By this way, we define a loss function  $U : (\Theta \times D, \mathcal{T} \otimes \mathcal{D}) \rightarrow (\mathbb{R}_+, \mathcal{B}_{\mathbb{R}_+})$  to which we associate a risk function  $R$ :

$$R(\theta, d) = \mathbb{E}[U(\theta, d(X))], \theta \in \Theta, d \in \mathcal{R}$$

where  $X \sim f(\cdot|\theta)$ .

In a Bayesian setting,  $\Theta$  has a  $\sigma$ -algebra  $\mathcal{T}$  and we denote  $\pi$  a probability measure on  $(\Theta, \mathcal{T})$ . Finally, we define  $r$  the Bayesian risk function associated to  $R$

$$r(\tau, d) = \int_{\Theta} R(\theta, d) d\pi(\theta)$$

Then, rewriting  $r$  as

$$\begin{aligned} r(\tau, d) &= \int R(\theta, d) d\pi(\theta) = \int U(\theta, d(x)) f(x|\theta) dx d\pi(\theta) \\ &= \int \left\{ \left[ \int U(\theta, d(x)) g(x, \theta) d\pi(\theta) \right] \left[ \int f(x|\theta) d\pi(\theta) \right] \right\} dx \end{aligned}$$

where  $g(x, \theta) = f(x|\theta) [\int f(x|\theta) d\pi(\theta)]^{-1}$ . The objective is to minimise  $\int U(\theta, d(x)) t(x, \theta) d\tau(\theta)$  for  $d$ . Now,  $g$  is the conditional distribution of  $\theta$  given  $X = x$ :

$$d(x) = \int \theta g(x, \theta) d\pi(\theta)$$

where the posterior distribution  $g$  naturally appears, and we can interpret that with a Bayesian risk and a quadratic loss function, the best way to estimate the parameter  $\theta$  is to consider the expected value of  $\theta$  with its posterior distribution.

# Appendix B

## Dose-finding methods

### B.1 Algorithm-based methods

#### B.1.1 The "3+3" algorithm

This method is one of the most widely used in dose-finding trials, due to the simplicity of its application. The idea is to systematically start from the lowest dose in the  $\{d_1, \dots, d_K\}$  panel, and by including cohorts (that is sets of patients) of three patients by three patients, select the next dose under some conditions. The algorithm works as follows: we include three patients at the minimum dose. If no toxicity is observed, three additional patients are included at the next higher dose. If more than two of three patients have toxicities, the MTD is considered to have been exceeded, and the trial is stopped. If one out of three patients have toxicities, three additional patients are included at the same dose. In the latter case, if toxicity is observed in one out of three patients, then the MTD is the tested dose. If one out of six patients has toxicity, then include three patients at the higher dose. If more than three out of six patients have toxicity, then the dose is the dose previously tested. And so on. This algorithm is illustrated in Figure [B.1](#).

This method has the significant advantage of being easy to implement in practice. It should also be noted that it allows a simple interpretation of the dose identified as the MTD. However, this algorithm allows only one target toxicity rate: thirty-three percent [\[82\]](#). Moreover, the final estimate is only done with six patients: such an estimate gives very wide confidence intervals. In addition, there is a statistical curiosity associated with this method: if intuition dictates that the target toxicity rate is 33%, Ivanova [\[36\]](#) showed that such an algorithm targets a toxicity rate between 16% and 27%. Finally, if the final estimate is only based on the last six patients, it is due to the algorithm itself: it is a memoryless design. In fact, at each step, there are at most six patients used for

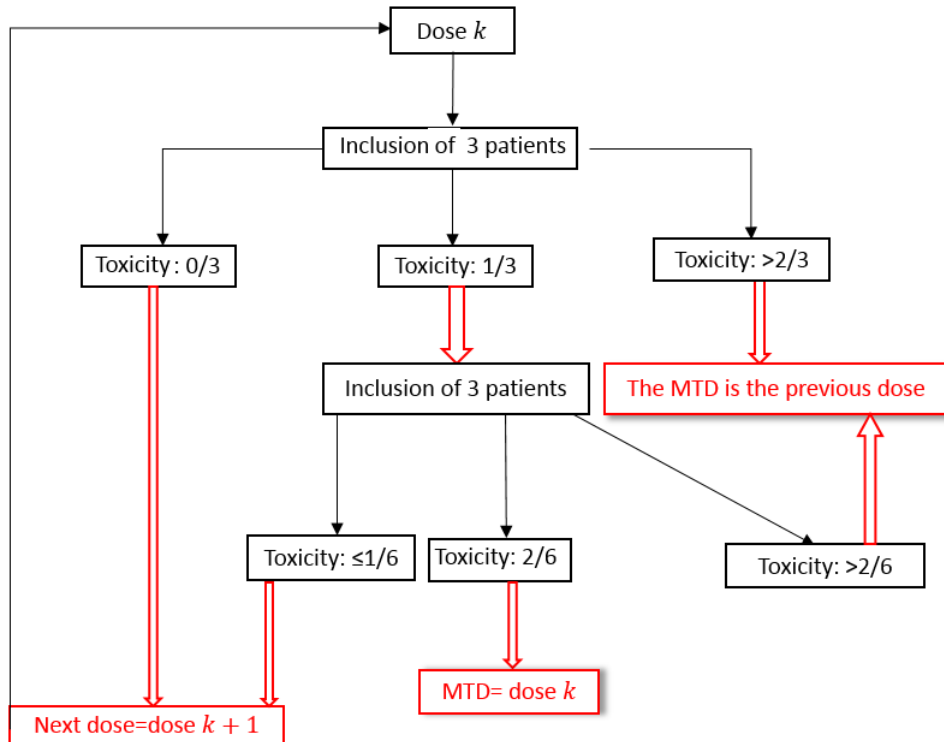


Fig. B.1 The "3+3" dose-finding algorithm

the estimate, and the information brought by patients previously treated at the same (or another) dose is no longer taken into account.

### B.1.1.1 Generalisation: "A+B" algorithms

Lin and Shih [45] proposed a generalisation of the "3+3" algorithm, more flexible in terms of cohort size (see Figure B.2). In addition, Ivanova [37] gave a method to calibrate the algorithm according to the target toxicity rate  $\nu$ . Note that "A+B" methods are sometimes called "up and down" methods (for example by Storer [81]), to be distinguished from the one we're going to present next.

### B.1.2 The "up and down" method

A presentation of this method, sometimes called "random walk rule" is given by Durham et al. [20]. The idea is to start from a target toxicity rate  $\nu \in ]0,0.5]$  and then to define  $b = \frac{\nu}{1-\nu}$ . For the  $i^{\text{th}}$  patient, we denote  $X_i$  and  $Y_i$  the dose assigned to the patient and

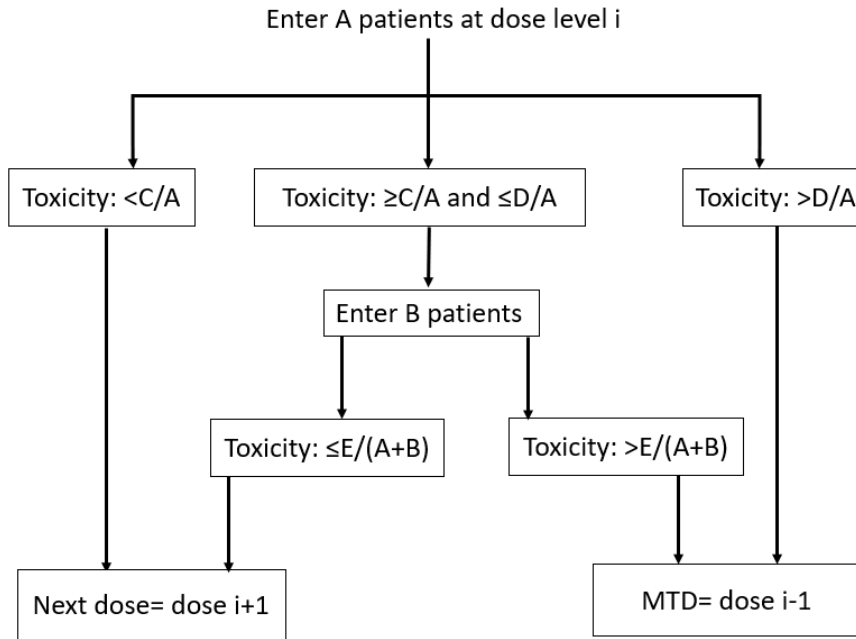


Fig. B.2 Figure inspired from article [45], about a generalisation of the "3+3" algorithm

his (binary) response. If  $Y_i = 1$ , the  $i + 1^{th}$  patient will have the dose immediately below that given to the  $i^{th}$  patient. If  $Y_i = 0$ , the  $i + 1^{th}$  patient will have the immediately higher dose with probability  $b$ , or the same dose as the  $i^{th}$  with probability  $1 - b$ . In the special case where  $(X_i, Y_i) = (d_K, 0)$  (so the dose is the higher of the panel) the next patient has the dose  $d_K$ . In this method, it is common to choose the number of patients before starting the trial [82]. The asymptotic properties of this method have been investigated by Bortot and Giovagnoli [6] in a Markovian setting.

## B.2 Model-based methods

### B.2.1 The EWOC method (Escalation With Overdose Control)

This method, presented by Babb et al. [4], is a parametric and Bayesian one. Its aim is to treat patients with a dose as close as possible to MTD, but avoiding exposing these patients to a dose whose toxicity would be above the threshold  $\nu$ .

The method is based on the following parametric model:

$$\mathbb{P}(Y_i = 1 | X_i = x_i) = F(b_0 + b_1 x_i)$$

where  $x_i \in \{d_1, \dots, d_K\}$ ,  $F$  is a cumulative distribution function and  $b_0$  and  $b_1$  are the parameters to be estimated. Based on the assumption that the probability of toxicity is an increasing function of dose,  $b_1 > 0$  is required. It immediately follows that the dose  $d_\nu$  corresponding to the toxicity threshold  $\nu$  is

$$d_\nu = \frac{F^{-1}(\nu) - b_0}{b_1}$$

However, it is common and convenient to place ourselves in the special case where  $F(z) = \frac{1}{1+\exp(-z)}$  is the logistic function. In this case and by defining  $\rho$  the probability of toxicity of  $d_1$  we obtain

$$b_0 = \frac{d_1 \text{logit}(\nu) - d_\nu \text{logit}(\rho)}{d_1 - d_\nu}$$

$$b_1 = \frac{\text{logit}(\rho) - \text{logit}(\nu)}{d_1 - d_\nu}$$

with  $\text{logit}(x) = \log(x/(1-x))$ . The EWOC method aim is to control the risk of administering a dose higher than the MTD to the  $i+1^{\text{th}}$  patient. To control this risk, we define

$$r = \mathbb{P}(d_\nu \leq x_{i+1} | x_1, \dots, x_i, y_1, \dots, y_i)$$

with  $(x_1, \dots, x_i)$  and  $(y_1, \dots, y_i)$  the doses and toxicities of the  $i$  first patients. The model, in the case where  $F$  is the logistic function, can then be written as a function  $f$  of the dose,  $d_\nu$  and  $\rho$ :  $F(b_0 + b_1 x_i) = f(x_i, \rho, d_\nu)$ . The likelihood of the model is  $L(\rho, d_\nu) = \prod_{j=1}^i \{f(x_j, \rho, d_\nu)\}^{y_j} \{1 - f(x_j, \rho, d_\nu)\}^{1-y_j}$ , and defining  $g(\rho, d_\nu)$  a joint prior distribution for  $(\rho, d_\nu)$ , we get the posterior distribution  $g(\rho, d_\nu | x_1, \dots, x_i, y_1, \dots, y_i) \propto L(\rho, d_\nu)g(\rho, d_\nu)$ . Note that the support of  $g$  can be as  $[0, \rho_0] \times [d_1, d_K]$ , where  $\rho_0$  would be chosen by a clinician.

The parameter of interest here is  $d_\nu$  and its posterior distribution is

$$g(d_\nu | x_1, \dots, x_i, y_1, \dots, y_i) = \int_0^{\rho_0} g(\rho, d_\nu | x_1, \dots, x_i, y_1, \dots, y_i) d\rho$$

and

$$r = \int_{d_1}^{x_{n+1}} g(d_\nu | x_1, \dots, x_i, y_1, \dots, y_i) d\{d_\nu\}$$

with  $r$  defined previous, "small enough" for the risk of overdose being low.

Considering that the doses' space is not continuous but discrete, the  $i+1^{\text{th}}$  patient's

dose is

$$\tilde{x}_{i+1} = \max\{d_j \in \{d_1, \dots, d_K\} \text{ avec: } d_j - X_{i+1} \leq T_1 \text{ et } \mathbb{P}(d_j < X_{i+1}) - r \leq T_2\}$$

where  $T_1$  and  $T_2$  are tolerance thresholds. Simulation studies [4] have shown that this method gives comparable results to those of the CRM, but with a lower proportion of overdosing, which is not surprising: avoiding overdosing is at the heart of the method.

## B.2.2 Extensions

### Time to event CRM (TITE-CRM)

One difficulty with dose-finding designs and particularly with an adaptive design like the CRM is the DLTs evaluations, which are done in long-term. The protocol is indeed defined to consider only the DLTs that appear after few weeks, for example three or four after the beginning of the treatment. The problem is therefore on one hand to evaluate and take into account the DLTs after such a delay, and on the other hand the duration of a trial for which several weeks of observations per patient are necessary to evaluate the toxicity of a dose and to include another patient. To overcome this problem, a method was proposed by Cheung and Chappell [13]. If one assumes that a patient is followed up during a time  $T$ , denoting  $t_i$  the time to toxicity for patient  $i$ , in the traditional CRM  $\mathbb{P}(t_i \leq T|d) = \psi(d, \beta)$ , as noted by Cheung [15], while the TITE-CRM considers that the time to toxicity is perhaps not observed, which leads Cheung and Chappell [13] to weight the  $\psi$  function:

$$\begin{aligned} \mathbb{P}(t_i \leq t) &= \mathbb{P}(t_i \leq t | t_i \leq T) \mathbb{P}(t_i \leq T) \\ &= w(t, T) \psi(x_i, \beta) \end{aligned}$$

with  $x_i$  the dose assigned to patient  $i$ . Therefore, the basic idea is to replace the likelihood function 3.1.1 by a weighted one, by changing the model  $\psi(d, \beta)$  in  $\tilde{\psi}(d, \beta, w)$ , with  $\tilde{\psi}$  also monotonous in  $d$ ,  $\beta$  and  $w$ , and such as  $\tilde{\psi}(d, \beta, 0) = 0$  and  $\tilde{\psi}(d, \beta, 1) = \psi(d, \beta)$ .

The resulting likelihood for the first  $n$  patients is

$$\tilde{L}_n(\beta) = \prod_{i=1}^n \{\tilde{\psi}(x_i, \beta, w_{i,n})\}^{Y_i} \{1 - \tilde{\psi}(x_i, \beta, w_{i,n})\}^{1-Y_i}$$

The weight  $w_{i,n}$  is the one associated to the  $i^{\text{th}}$  patient. The choice of the weight function is indeed a central question, and several answers can be considered. The author proposes



a weight function  $w(t, T) = \frac{t}{T}$ , with  $t$  the time during which the patient is being followed up. An easy generalisation is given by Cheung [15], which is  $w(t, T) = \min(\frac{t}{T}, 1)$ . This choice correspond to consider that the weight does not depends neither on the function  $\psi$  nor on the dose  $x_i$  given to patient  $i$ . It is possible to consider that the weight depends on the dose [12], or the weight can even be adaptive [14].

### B.2.3 Non-parametric approaches to dose-finding

The CRM provides an important ethical and practical framework for dose-finding studies. However, the choice of a parametric approach can be discussed ([22],[15]), as it involves complications (use of pseudo-doses, choice of skeleton or other, see Gasparini and Eisele [22]). In order to free oneself from the constraints induced by the choice of a model, several authors proposed some non-parametric approaches. The idea is basically not to estimate a parameter of a model to estimate the probability of toxicity associated to a dose, but to directly consider a probability distribution associated to each probability of toxicity: each probability of toxicity become a parameter of interest. There are many ways to address this idea.

For instance, Gasparini and Eisele [22] consider a joint prior distribution  $\pi(p_1, \dots, p_K)$  for the vector of probability of toxicity associated to each dose  $d_1, \dots, d_K$ . The authors recommend this prior to respect some conditions, as to be centred around clinicians prior guesses, to be disperse to allow dose-escalation and that each marginal distribution of  $p_i$  to be unimodal. The authors then propose a reparametrisation of the model:

$$\begin{aligned}\theta_1 &= 1 - p_1 \\ \theta_2 &= \frac{1 - p_2}{1 - p_1} \\ &\dots \\ \theta_K &= \frac{1 - p_K}{1 - p_{K-1}}\end{aligned}$$

with each  $\theta_i$  being independent beta distribution  $Beta(a_i, b_i)$ ,  $a_i, b_i$  being fixed for  $i \in \{1, \dots, K\}$ . With this model, the distribution of each  $p_i$  is then called a product-of-beta prior (PBP) by the authors since  $p_i = 1 - \prod_{j=1}^K \theta_j$ . The properties of this seems to be as good as the standard CRM, and it avoids over-dosing issue in dose-finding.

An other possibility, more complex, is given by Tang et al. [86]. Denoting  $n_k$  and  $N_k$  the number of observed toxicity at dose  $d_k$  and the number of patients treated at dose  $d_k$

respectively, the authors propose the following model:

$$\begin{aligned} n_k | p_k &\sim \text{Bin}(N_k, p_k), \quad k \in \{1, \dots, K\} \\ p_k &= F(d_k) \text{ with } F | \eta, \alpha \sim \text{DP}(\alpha F_0(x | \eta)) \end{aligned}$$

with  $F_0(\cdot)$  a function used as starting point to construct the non-parametric distribution of  $F(d_k)$ , which associate a probability of toxicity to each dose:  $F(d_k) = \mathbb{P}(Y = 1 | d_k)$ , and  $\text{DP}$  denotes a Dirichlet process. This hierarchical model is then extended in the particular case where  $F_0$  is based on the cumulative probability function of a Gaussian variable. Although the results shown with this model seem remarkable, it should be noticed that the simulations were performed with the basis of sixty patients, which is not usual in dose-finding studies. Then, the performances of such a model are not assured to be better than some usual methods with fewer sample sizes.

Finally, an isotonic approach was investigated by Leung and Wang [43].

## B.2.4 Dose-finding and pharmacokinetics

The purpose of the pharmacokinetics is the study of the evolution and transformation of a drug in the body or in an organ. This evolution of the drug in the body or organ goes through four processes: absorption (A), distribution (D), metabolism (M) and excretion (E), summarised by the ADME concept.

During Phase I trials it is common to conduct simultaneously dose-finding and pharmacokinetics study. The interests of pharmacokinetics are multiple, including understanding how a patient's body will respond to a molecule, or establishing the presence or absence of a link between toxicity and pharmacokinetics values (e.g. in [21]). Comets and Zohar [16] showed that most phase I studies doing both dose-finding and pharmacokinetics study do not bring these two aspects together. However, it is possible to consider dose-finding methods that closely combine PK data with the dose-finding process, as shown by Ursino et al. [89]. The authors explain that a basic idea could be to link the dose to the area under the curve (AUC, which is a pharmacokinetics important value: the area under the curve which represents the concentration of the drug), and then to link this AUC to the probability of toxicity. Several methods were compared in Ursino et al. [89], and the authors concluded that even if the percentage of right MTD selection was not improved, the dose-toxicity relationship curve estimation was more precise.

An other method was proposed by Takeda et al. [84]. It consists in a CRM but instead of considering the dose-toxicity relationship to consider the standardized AUC-toxicity

relationship, the MTD selection is generally higher with this method than with the classic CRM.

Otherwise, Günhan et al.[27] proposed a time to event approach through a pharmacokinetic model for phase I dose escalation with multiple schedules: the TITE-PK (time-to-event pharmacokinetic model), which showed comparable or better performances than a reference model for dose-schedule finding design, the POCRM (partial order CRM, introduced by Wages et al.[90]).

Gerard et al.[25] proposed the DRtox dose regimen model, with better performances than other dose regimen models in term of MTD-regimen.

### B.2.5 Bivariate outcomes: efficacy and toxicity

We have explained the assumption that increasing a dose implies an increase in efficacy and toxicity. Then, ethics imposes a control of toxicity. However, for certain drugs, it seems that the efficacy is not monotonous with the dose but become constant at some level. In this setting, there are methods that allow a simultaneous evaluation of toxicity and efficacy. Here is a quick overview of what exists to take into account both toxicity and efficacy.

In the context of adoptive cell therapy (ACT), Li et al. [44] proposed a method which consists in proposing a toxicity and efficacy probability interval (TEPI) design. It involves dose escalation, de-escalation or staying at the same dose considering both probability of efficacy and probability of toxicity of the drug. This method permits to not increase the dose if it is efficient but too toxic and to increase the dose if the toxicity is reasonable. The compromise is performed through a utility function. Ananthakrishnan et al. [3] proposed a frequentist version of the method proposed by Li et al. [44].

The bivariate CRM (bCRM) method, proposed by Braun [8], takes into account the efficacy of the drug (which corresponds to a low progression rate of a disease) through a CRM-based model. In this model the probability of efficacy by dose is given by a logistic regression. The MTD finally corresponds to the dose which minimises a distance  $d_i^n = \sqrt{\sum_{k=1}^2 w_k (p_{k,i}^n - p_k^*)^2}$  where  $p_k^*$ ,  $k = 0, 1$  is a rate of toxicity and efficacy, respectively, and  $p_{k,i}$  the estimate probability of toxicity and efficacy. A CRM approach to take into account efficacy is also considered by Seegers et al. [78], where the authors use the CRM empiric function (called *power model*), and a correlation parameter between efficacy and toxicity is considered. The empiric model and a nonparametric one are investigated by Cunanan and Koopmeiners [18] through a hierarchical approach for dose-finding with multiple populations.

Thall and Cook [87] proposed a method that can be applied to different dose-finding

models (bivariate or trivariate), consisting of searching for a compromise between probability of toxicity and probability of efficacy in contours within a two-dimensional domain. Yin et al. [94] proposed to adopt the toxicity-efficacy trade-off investigate by Thall and Cook [87]. However, instead of considering a direct dose-response model, efficacy and toxicity are modelled jointly through odd-ratios.

A score approach was given by Bekele and Shen [57], where the authors also introduced a correlation between toxicity and efficacy, which is evaluated *via* a biomarker expression. A comparison of several designs, taking into account both toxicity and efficacy, can be found in [97], in the context of muscularly targeted agent use. The authors conclude that in their setting, a non-parametric (isotonic) regression or a semi-parametric (locally logistic model) have both good operating characteristics to find a optimal biological dose, which is the lowest dose with highest efficacy rate while safe for patients.

A utility function approach was proposed by Takahashi and Suzuki [83] in a phase I/II setting, where this utility function is a function to be optimised to obtain the optimal dose, defined as the dose that provides sufficient efficacy under an acceptable toxicity rate. The dose-efficacy and dose-toxicity functions are estimated through Gaussian prior process.



# Appendix C

## Additional considerations about the ESS

### C.1 Alternative computational approach for the Effective Sample Size

As Morita et al. [53] explained, the definition of the ESS is in many points quite arbitrary. Several ways were investigated, and the previous one was the best for interpretations of the resulting ESS, given "*known*" results.

#### C.1.1 Bayesian central limit theorem

A key point is that Morita's approach is based on the "*normal approximations of the prior  $p(\boldsymbol{\theta})$  and the posterior  $q_m(\boldsymbol{\theta}/\tilde{\boldsymbol{\theta}}_0, \mathbf{Y}_m$ )*".

Indeed, Gelman [23] gives the first terms of the Taylor's expansion of

$$\begin{aligned}\log(p(\theta|y)) &= \log(p(\hat{\theta}|y)) + \frac{1}{2}(\theta - \hat{\theta})^2 \left[ \frac{d^2 \log(p(\theta|y))}{d\theta^2} \right] \Big|_{\theta=\hat{\theta}} + \dots \\ &= \log(p(\hat{\theta}|y)) - \frac{1}{2}(\theta - \hat{\theta})^2 I(\hat{\theta}) + \dots\end{aligned}$$

where  $\hat{\theta}$  is the posterior mode, which yield to the normal approximation  $p(\theta|y) \sim \mathcal{N}(\hat{\theta}, [I(\hat{\theta})]^{-1})$ , where  $I(\hat{\theta})$  is the **observed information**. With this consideration, the use of the second derivative in equation 2.4.1 appears naturally, as an asymptotic equivalent of the inverse of the variance. In this way, it is logical to think about the

use of variance instead of the second derivative, not always easy to use in Morita's way because of the integration over  $f(Y_m)$  in  $D_{q,+}$ .

### C.1.2 Non-asymptotic view for the variance: the natural exponential family

A non asymptotic consideration is possible to explain the use of the second derivative, but not the  $x$  derivative, the parameter derivative.

The exponential family is the set of the probability law for which the density could be written as

$$f(x|\theta) = h(x) \exp(\langle \eta(\theta), T(x) \rangle - A(\eta))$$

$\langle \cdot, \cdot \rangle$  being the inner product. There are several possible definitions of the natural exponential family. We adopted the approach of Morris et al.[56]. The natural exponential family is a subset of the exponential family with a special point: the  $\eta$  and  $T$  functions are both the identity. Furthermore it is established that in this case, the variance of a random variable  $X \sim f$  is given by the second derivative

$$\mathbb{V}[X] = A''(\theta)$$

Using this approach for the variance could be very interesting due to the fact that natural exponential families with quadratic variance function cover several distributions: Normal, Poisson, Gamma, Binomial, Negative Binomial and the hyperbolic secant distribution.

### C.1.3 Alternative variance approach

#### Algorithm

For  $m$  from 1 to  $N_{max}$ : [We assume that the ESS is in  $1:N_{max}$ ]

1. Set B and for k in 1:B:
  2. Simulate  $Y_k = (Y_1^k, \dots, Y_m^k)$ ,  $Y_i^k \sim Y|\theta$  [simulate  $\theta$  under  $p(\theta|\tilde{\theta})$  for each  $Y_i^k$ , or take the mean. Both give similar results]
  3. Estimate the posterior  $q_m(\theta)$  ( $m$ = size of the sample)
  4. Extract sample  $\Theta_k = (\theta_1^k, \dots, \theta_n^k)$  of  $\theta$  from  $q_m(\theta/D_0)$
  5. Set  $\mathbf{V} = (\mathbb{V}(\Theta_1), \dots, \mathbb{V}(\Theta_B))$  the vector of variances of each sample
6.  $V_m = \bar{V}$ , equivalent of integrating it over  $Y$

And the ESS  $m^*$  is  $m^* = \underset{m}{\operatorname{argmin}} |V_m - v|$  where  $v = \mathbb{V}(p(\theta/\tilde{\theta}))$  could be replaced if necessary by the empirical variance of a sample of observations obtained with the law  $p(\theta|\tilde{\theta})$ .

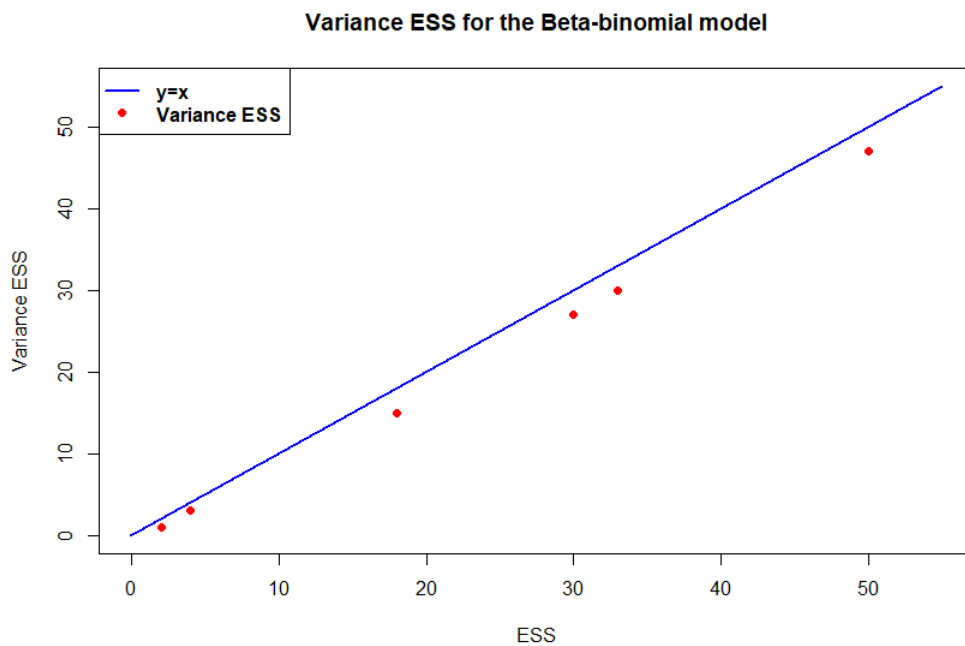
The advantages of this method are first the no need of the calculation of the second derivative, and the easy computation since only the simulation under  $p(\theta|\tilde{\theta})$  and  $Y_i|\theta$  are required.

### C.1.3.1 Simulations

To evaluate the performance of this method to compute the ESS, we performed a simulation studie with two cases.

#### Beta-binomial model

$\tilde{\alpha}$	$\tilde{\beta}$	real ESS	"Variance" ESS
1	1	2	1
1	3	4	3
3	1	4	3
7	11	18	15
15	15	30	27
13	20	33	30
30	20	50	47

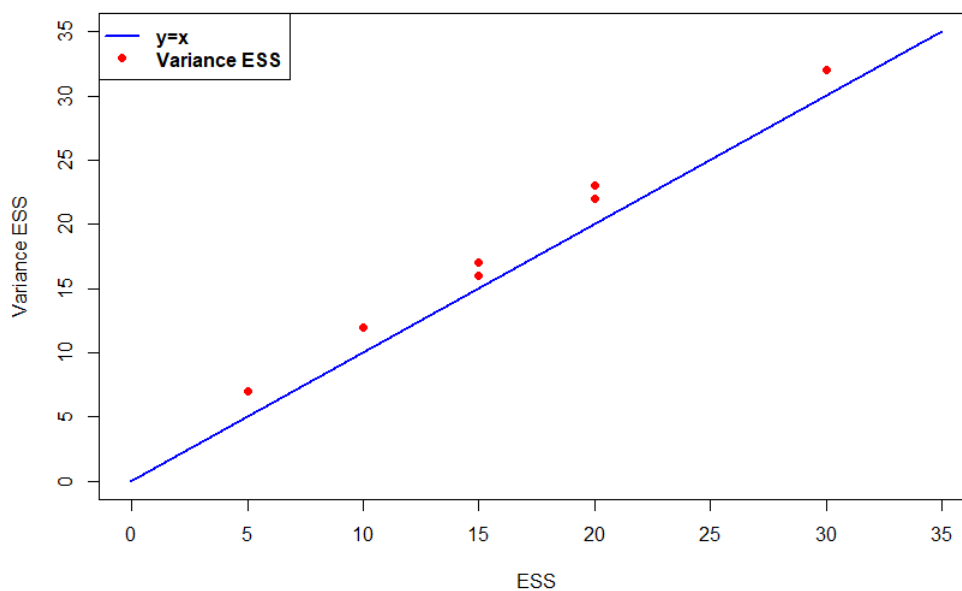




**Gamma-exponential model** In this model, the data are normally distributed  $Y|\theta \sim \mathcal{E}(\theta)$  and  $\theta \sim \mathcal{Gamma}(\tilde{\alpha}, \tilde{\beta})$ .

shape ( $\tilde{\alpha}$ )	rate( $\tilde{\beta}$ )	real ESS	variance ESS
5	2	5	7
5	5	5	7
10	5	10	12
10	10	10	12
15	3	15	17
15	5	15	17
15	10	15	16
20	5	20	22
20	7	20	23
20	10	20	22
30	5	30	32
30	10	30	32

Variance ESS for the gamma-exponential model



**Univariate normal with known variance** In this model  $Y|\theta \sim \mathcal{N}(\theta, \sigma^2)$  with  $\sigma$  known and  $\theta|\tilde{\theta} \sim \mathcal{N}(\tilde{\mu}, \tilde{\sigma}^2)$ . The ESS is  $\frac{\sigma^2}{\tilde{\sigma}^2}$  in that case. For all tried values, we obtained the exact ESS with our method.

## C.2 ESS for the power prior

In Ollier et al. [65] we used the approximation  $ESS[\pi^{PP}(\cdot|D_0, \alpha_0)] = \alpha_0 n_0$  with  $\pi^{PP}(\cdot|D_0, \alpha_0)$  the power prior for historical data  $D_0$  with  $n_0$  observations raised to  $\alpha_0$ . We propose here to justify this approximation, proposed first by Morita et al. [53]. We consider the power prior  $\pi^{PP}(\boldsymbol{\theta}) = \frac{L(\boldsymbol{\theta}/D_0)\pi_0(\boldsymbol{\theta})}{\int L(\boldsymbol{\theta}/D_0)\pi_0(\boldsymbol{\theta})d\boldsymbol{\theta}}$ , with  $L(\boldsymbol{\theta}/D_0) = \prod_{i=1}^{n_0} f(X_{i,0}/\boldsymbol{\theta})$  and focus on its ESS.

Here  $p(\boldsymbol{\theta}) = \pi^{PP}(\boldsymbol{\theta})$  is the prior of interest, and we assume  $\mathbb{E} \left[ \left| \frac{\partial^2 \log(f(X_i/\boldsymbol{\theta}))}{\partial \theta_j^2} \right| \right] < +\infty$  for all  $j$ .

After some calculations we obtain

$$\begin{aligned} D_{p,+} &= -\alpha_0 \sum_{j=1}^q \sum_{i=1}^{n_0} \frac{\partial^2 \log(f(X_{i,0}/\boldsymbol{\theta}))}{\partial \theta_j^2} - \sum_{j=1}^q \frac{\partial^2 \log(\pi_0(\boldsymbol{\theta}))}{\partial \theta_j^2} \\ D_{q,+} &= -\sum_{j=1}^q \int \left\{ \sum_{i=1}^n \frac{\partial^2 \log(f(X_i/\boldsymbol{\theta}))}{\partial \theta_j^2} + \frac{\partial^2 \log(q_0(\boldsymbol{\theta}))}{\partial \theta_j^2} \right\} f_n(\mathbf{X}) d\mathbf{X} \\ &= -n \sum_{j=1}^q \mathbb{E} \left[ \frac{\partial^2 \log(f(X_i/\boldsymbol{\theta}))}{\partial \theta_j^2} \right] - \sum_{j=1}^q \frac{\partial^2 \log(q_0(\boldsymbol{\theta}))}{\partial \theta_j^2} \end{aligned}$$

Then

$$\delta(\pi^{PP}, n) = \left| -\alpha_0 \sum_{j=1}^q \sum_{i=1}^{n_0} \frac{\partial^2 \log(f(X_{i,0}/\boldsymbol{\theta}))}{\partial \theta_j^2} \Big|_{\boldsymbol{\theta}=\bar{\boldsymbol{\theta}}} - \sum_{j=1}^q \frac{\partial^2 \log(\pi_0(\boldsymbol{\theta}))}{\partial \theta_j^2} \Big|_{\boldsymbol{\theta}=\bar{\boldsymbol{\theta}}} + n \sum_{j=1}^q \mathbb{E} \left[ \frac{\partial^2 \log(f(X_i/\boldsymbol{\theta}))}{\partial \theta_j^2} \right] \Big|_{\boldsymbol{\theta}=\bar{\boldsymbol{\theta}}} + \sum_{j=1}^q \frac{\partial^2 \log(q_0(\boldsymbol{\theta}))}{\partial \theta_j^2} \Big|_{\boldsymbol{\theta}=\bar{\boldsymbol{\theta}}} \right|$$

Since it is not analytically possible to minimise with  $n$   $\delta(\pi^{PP}, n)$ , we proceed by lower and upper bounds approximations to give an asymptotic value of the ESS of the power prior.

We directly obtain

$$\begin{aligned} \delta(\pi^{PP}, n) &\leq \left| \alpha_0 \sum_{j=1}^q \sum_{i=1}^{n_0} \frac{\partial^2 \log(f(X_{i,0}/\boldsymbol{\theta}))}{\partial \theta_j^2} \Big|_{\boldsymbol{\theta}=\bar{\boldsymbol{\theta}}} - n \sum_{j=1}^q \mathbb{E} \left[ \frac{\partial^2 \log(f(X_i/\boldsymbol{\theta}))}{\partial \theta_j^2} \right] \Big|_{\boldsymbol{\theta}=\bar{\boldsymbol{\theta}}} \right| \\ &\quad + \left| \sum_{j=1}^q \frac{\partial^2 \log(q_0(\boldsymbol{\theta}))}{\partial \theta_j^2} \Big|_{\boldsymbol{\theta}=\bar{\boldsymbol{\theta}}} - \sum_{j=1}^q \frac{\partial^2 \log(\pi_0(\boldsymbol{\theta}))}{\partial \theta_j^2} \Big|_{\boldsymbol{\theta}=\bar{\boldsymbol{\theta}}} \right| \end{aligned}$$

Then, we can assume that the second term is about zero or negligible since these are non-informative distributions, and with Law of Large Number,

$$\delta(\pi^{PP}, n) \leq \left| (\alpha_0 n_0 - n) \sum_{j=1}^q \mathbb{E} \left[ \frac{\partial^2 \log(f(X_i/\boldsymbol{\theta}))}{\partial \theta_j^2} \right] \Big|_{\boldsymbol{\theta}=\bar{\boldsymbol{\theta}}} \right| + \epsilon_0 \text{ if } n_0 \rightarrow +\infty, \text{ with } \epsilon_0 \rightarrow 0$$

which argmin is  $n = \alpha_0 n_0$ .

For the lower bound, we use  $\|x\| - \|y\| \leq \|x + y\|$ :

$$\begin{aligned} \delta(\pi^{PP}, n) \geq & \left| \left| \alpha_0 \sum_{j=1}^q \sum_{i=1}^{n_0} \frac{\partial^2 \log(f(X_{i,0}/\boldsymbol{\theta}))}{\partial \theta_j^2} \Big|_{\boldsymbol{\theta}=\bar{\boldsymbol{\theta}}} - n \sum_{j=1}^q \mathbb{E} \left[ \frac{\partial^2 \log(f(X_i \boldsymbol{\theta}))}{\partial \theta_j^2} \right] \Big|_{\boldsymbol{\theta}=\bar{\boldsymbol{\theta}}} \right| \\ & - \left| \sum_{j=1}^q \frac{\partial^2 \log(\pi_0(\boldsymbol{\theta}))}{\partial \theta_j^2} \Big|_{\boldsymbol{\theta}=\bar{\boldsymbol{\theta}}} - \sum_{j=1}^q \frac{\partial^2 \log(q_0(\boldsymbol{\theta}))}{\partial \theta_j^2} \Big|_{\boldsymbol{\theta}=\bar{\boldsymbol{\theta}}} \right| \end{aligned}$$

We same kind of hypothesis for the second derivative for non-informative distributions, the second term is also negligible, and asymptotically the first tend to be

$$\left| (\alpha_0 n_0 - n) \mathbb{E} \left[ \frac{\partial^2 \log(f(X_{i,0} \boldsymbol{\theta}))}{\partial \theta_j^2} \right] \Big|_{\boldsymbol{\theta}=\bar{\boldsymbol{\theta}}} \right|$$

So we finally get

$$\left| (\alpha_0 n_0 - n) \mathbb{E} \left[ \frac{\partial^2 \log(f(X_{i,0} \boldsymbol{\theta}))}{\partial \theta_j^2} \right] \Big|_{\boldsymbol{\theta}=\bar{\boldsymbol{\theta}}} \right| - \epsilon_1 \leq \delta(\pi^{PP}, n) \leq \left| (\alpha_0 n_0 - n) \mathbb{E} \left[ \frac{\partial^2 \log(f(X_i \boldsymbol{\theta}))}{\partial \theta_j^2} \right] \Big|_{\boldsymbol{\theta}=\bar{\boldsymbol{\theta}}} \right| + \epsilon_0$$

Since here  $X_{i,0} \sim X_i$ ,  $ESS[\pi^{PP}(\boldsymbol{\theta})] \xrightarrow{n_0 \rightarrow \infty} \alpha_0 n_0$ . ■

# Résumé détaillé

Les essais de phase I ou de recherche de dose en recherche clinique sont les premiers essais réalisés chez l'humain et ont pour objectif l'évaluation de la toxicité de la molécule sur l'homme. Dans le cas particulier de la recherche de dose en oncologie, le but de ces essais est l'identification de la dose maximale tolérée (DMT), c'est-à-dire la plus forte dose qu'il soit possible de donner à un patient en ayant défini un seuil de toxicité au préalable. En effet, l'hypothèse pour certaines familles de molécules est qu'une plus forte dose est plus efficace, mais aussi plus toxique pour les patients. On considèrera donc qu'une dose au-dessus de la dose maximale tolérée aura une probabilité de toxicité inacceptable. Par ailleurs, dans le cadre des essais cliniques de recherche de dose en oncologie, il est fréquent qu'une seule molécule soit testée plusieurs fois, dans différentes régions du monde entre différentes populations. Dans ce contexte, l'*international council for harmonisation of technical requirements for pharmaceuticals for human use* (ICH) définit les bridging-studies comme des essais cliniques complémentaires permettant de prendre en compte les différences entre plusieurs populations, en termes d'efficacité, de toxicité, de comorbidités, de pharmacocinétique notamment. En effet, des différences génétiques entre différentes populations entraînant des différences métaboliques, la relation dose-réponse pour cette même molécule peut différer entre différentes populations, en termes d'efficacité ou de toxicité par exemple. La conduite d'un essai clinique additionnel est donc une étape souvent nécessaire pour évaluer la relation dose-réponse dans la nouvelle population. La problématique qui se pose alors est l'utilisation d'une méthode statistique permettant la prise en compte pour un nouvel essai clinique des résultats d'un essai ayant été réalisé dans une autre population, pour la même molécule. En effet, la prise en compte de résultats antérieurs permettrait d'améliorer l'estimation de la dose maximale tolérée, dans des essais de recherche de dose où le nombre de patients est généralement très réduit. Dans un tel contexte, toute information extérieure permettant l'amélioration de l'estimation de la relation dose-toxicité est cruciale. Par ailleurs, il est important pour des autorités de santé régionales, lorsque deux essais de recherche de dose ont été effectués pour une même molécule dans deux populations différentes, d'avoir des critères

objectifs de comparaison afin de décider si l' «extrapolation» des résultats est possible ou non entre ces deux populations. Nos objectifs sont donc, dans un premier temps, de fournir une méthode permettant la prise en compte d'information historique pour un nouvel essai devant avoir lieu lorsque la molécule a déjà été testée dans une autre population, et dans un deuxième temps, de fournir des outils objectifs de comparaisons de résultats entre différents essais de recherche de dose réalisés pour une même molécule mais dans des populations différentes.

Nous avons dans un premier temps considéré le cas où un essai a été réalisé dans une population, et où un autre essai sur la même molécule doit avoir lieu mais pour une autre population. Pour l'essai en cours, nous considérons le cas d'un essai adaptatif, dans lequel les cohortes de patients sont enrôlées les unes après les autres à un niveau de dose estimé comme le plus proche de la dose maximale tolérée en fonction des résultats des cohortes de patients précédentes. Afin de permettre la prise en compte de résultats d'un essai clinique antérieur dans un essai en cours, nous avons proposé une adaptation de la méthode bayésienne de réévaluation séquentielle de recherche de dose (*continual reassessment method*) à l'aide du *power prior* comme distribution a priori sur le paramètre d'intérêt à estimer. Ce paramètre détermine ensuite la relation dose-toxicité, cette relation étant une régression logistique. Le *power prior* comme distribution a priori sur le paramètre d'intérêt permet en effet d'incorporer la vraisemblance des données historiques élevée à une puissance alpha, comprise entre zéro et un. Le cas où le coefficient est égal à zéro correspond au cas où on n'utilise aucune information historique et le cas où il est égal à un au cas où toute l'information historique est utilisée. Le choix de la valeur du coefficient alpha détermine donc le degré d'emprunt de l'information historique dans la distribution a priori du paramètre d'intérêt. Pour mesurer la quantité d'information dans la distribution a priori, paramétrique, la notion d'*effective sample size* a été utilisée. Différentes méthodes d'estimation de ce paramètre alpha du *power prior* ont été étudiées dans la littérature. Nous avons pour notre part proposé une méthode prenant en compte non seulement la possibilité d'un *data-prior conflict*, mais également de fixer un seuil maximal de quantité d'information désirée pour la distribution a priori. Pour prendre en compte ces deux éléments, la méthode proposée consiste à séparer le coefficient alpha du *power prior* en un produit de deux autres coefficients : l'un étant le critère de similarité entre les deux jeux de données et l'autre déterminant un seuil maximal d'information dans la distribution a priori.

Afin de prendre en compte la possibilité d'un *data-prior conflict*, la méthode proposée permet de pondérer l'information empruntée dans la distribution a priori en fonction d'un critère de similarité entre les deux relations dose-toxicité des deux populations.

Nous utilisons pour cela les deux vraisemblances : celle de l'essai historique et celle de l'essai en court. Ce critère de similarité proposé est basé sur la distance de Hellinger entre les deux vraisemblances normalisées, il est égal à zéro si les deux vraisemblances sont identiques, et tend vers un lorsque les deux vraisemblance diffèrent de plus en plus. Par ailleurs, la différence d'effectif entre les deux essais doit également être considérée dans la pondération de l'information de la distribution a priori sur le paramètre d'intérêt. En effet, l'essai historique comportera plus de sujets que celui en cours, adaptatif, ce qui aura pour conséquence une augmentation de la quantité d'information incorporée dans le *power prior*. Or, une distribution a priori trop informative peut gêner l'estimation, particulièrement dans notre cas où nous avons peu de patients, et notamment dans le cas où les données et cette distribution sont en désaccord. Nous définissons donc, à l'aide de l'*effective sample size*, un seuil d'information maximale que la distribution a priori peut contenir, ce seuil étant fonction du nombre de patients dans l'essai en cours. Cette quantité d'information maximale autorisée est ensuite repondérée par le critère de similarité introduit précédemment. Nous avons évalué cette méthode, l'*adaptive power prior*, à l'aide d'une étude de simulations. Nous avons pour cela fixé un essai clinique fictif dans une population avec des probabilité de toxicité par dose prédéterminées. Les différents scénarios proposés ensuite correspondent à des données d'essai en cours plus ou moins proches des données de l'essai historique en terme de probabilité de toxicité pour chaque dose. Le cas où les l'essai en cours était simulé selon les mêmes probabilités de toxicité par dose a été étudié, et nous avons ensuite fait varier ces probabilités considérant des cas de populations plus ou moins sensibles à la molécule. Nous avons également proposé une analyse de sensibilité sur la façon de pondérer cette quantité d'information dans le *power prior* en fonction du nombre de patients dans l'essai en cours, en faisant varier cette quantité d'information prédéfinie, et du critère de similarité. Nous avons ainsi montré la possibilité pour notre méthode d'améliorer le pourcentage de sélection de la dose maximale tolérée quand les deux relations dose-toxicité étaient similaires, tout en ayant une performance comparable à la méthode standard de réévaluation séquentielle dans le cas contraire. De plus, l'étude du coefficient permettant l'emprunt d'information obtenu dans le *power prior* par notre méthode a montré effectivement un emprunt d'information plus élevé quand les deux jeux de données étaient similaires que dans le cas contraire, illustrant ainsi le bon fonctionnement du critère de similarité proposé. Il est également à noter que bien que cette méthode ai été utilisée dans un contexte clinique de recherche de dose avec la méthode bayésienne de réévaluation séquentielle, elle peut être adapté dans tout contexte bayésien prenant en compte des données historiques et dans lequel la pondération de l'information et la possibilité de data-prior conflict doit

être prise en compte. L'extension à d'autres modèles de recherche de dose est également envisageable.

Dans un second temps, nous avons considéré le cas où les résultats de deux essais cliniques terminés sont à disposition des autorités de santé dans une région du monde. Dans ce cas, il leur revient de décider si l'«extrapolation» des résultats d'une population à une autre est possible. Il apparaît donc nécessaire d'avoir à disposition des outils objectifs de comparaison entre les relations dose-toxicité, et d'une façon générale des critères de similarité des données entre les différents résultats d'essais cliniques réalisés dans différentes populations. Nous avons proposé plusieurs critères, basés sur les fonctions de vraisemblances et les relations dose-toxicité notamment. Nous avons par exemple proposé de modifier le précédent critère de similarité en comparant directement les distributions a posteriori des paramètres d'intérêt, et aussi des distributions a posteriori des doses maximales tolérées. Par ailleurs, nous avons introduit plusieurs critères de comparaisons des distributions des doses maximales tolérées des deux essais cliniques, permettant des critères ne fonctionnant pas uniquement au niveau paramétrique mais se focalisant directement sur la quantité d'intérêt qu'est la dose maximale tolérée. Nous avons également montré la nécessité de disposer de plusieurs critères pour évaluer les différences entre deux relations dose-toxicité. En effet, deux essais de recherche de dose sur deux groupes de patients peuvent trouver la même dose maximale tolérée mais obtenir des relations dose-toxicité différentes, ou bien des relations dose-toxicité différentes et différentes doses maximales tolérées. La mise au point de critères différents travaillant soit au niveau du modèle soit au niveau de la distribution de la dose-maximale tolérée permettait donc de détecter ces différents types de similarités. Ces différents critères ont été évalués dans un premier temps sur trois jeux de données simulées, correspondant aux trois possibilités de similarité ou disimilarité. Chacun consistait en deux essais de recherche de dose d'une même molécule, mais dans deux populations. Le premier jeu de données considérait le cas où les relations dose-toxicité étaient similaires et où les doses maximales tolérées étaient identiques, le second jeu de données le cas où les relations dose-toxicité étaient différentes mais avec la même dose maximale tolérée et enfin le troisième jeu de données le cas où les relations dose-toxicité et la dose maximale tolérée étaient différents tous les deux. Les différents critères introduits ont montré leur efficacité à détecter différents types de similarités ou disimilarités. Nous avons ensuite évalué nos différents critères sur plusieurs molécules pour lesquelles les données d'essais de recherche de dose dans deux populations (caucasiennes et japonaises) étaient publiées, et fourni des outils graphiques d'aide à l'interprétation des résultats. Dans chacun des cas, les critères proposés ont permis de fournir une interprétation des différences de relations

---

dose-toxicité entre différentes populations. Par ailleurs, les outils graphiques proposés permettent également d'avoir une idée claire des différences dans les distribution des doses maximales tolérés.