UNIVERSITA' DEGLI STUDI DI PAVIA

FACOLTA' DI INGEGNERIA
DIPARTIMENTO DI INGEGNERIA INDUSTRIALE E DELL'INFORMAZIONE

DOTTORATO DI RICERCA IN BIOINGEGNERIA E BIOINFORMATICA
XXIX CICLO - 2016

MODELING AND SIMULATION, EXTRAPOLATION AND ALTERNATIVE STUDY DESIGNS AS TOOLS TO FACILITATE PEDIATRIC DRUG DEVELOPMENT

PhD Thesis by

GIOVANNI SMANIA

Advisor:
Prof. Paolo Magni

PhD Program Chair:
Prof. Riccardo Bellazzi

Ad Enrica,
che sa rendermi una persona migliore

Acknowledgments

Several people have been involved and contributed to this thesis.

My very first acknowledgment goes to Oscar Della Pasqua and Alessandra Bertoldo for introducing me to the PKPD world. I would also like to thank the entire CVBF staff, particularly Mariagrazia Felisi, who gave me the opportunity to pursue this PhD. A big thank goes to Massimo Cella, who greatly contributed to this thesis and supported me during the first two years of my PhD. I wish also to thank Gérard Pons, whose wise suggestions and advices were always welcome. Many thanks go to the whole Clinical Research department of Menarini Ricerche, particularly to Angela Capriati, Simone Baldini, Enrica Mezzalana and Paolo Mazzei. Paolo deserves a further special thank for his relentless support on the writing of part of this thesis since I joined Menarini. I owe much gratitude to my PhD tutor prof Paolo Magni, tireless worker, always available, with plenty of ideas. I learned a lot from him, from both a scientific and personal point of view.

Special thanks to all my friends, to Grazia and Pietro for their support, to Emanuela and Tiziano, unmatchable parents always on my side, Aurora for her brotherhood and Valter for his unique meat. Finally, there is no word to describe the gratitude I owe to my girlfriend Enrica Battagello, thank you.

Abstract (Italiano)

Durante la pianificazione di trial clinici pediatrici è necessario far fronte a numerosi problemi, in particolare a difficoltà di tipo etico, di fattibilità e di efficienza. Allo stesso tempo c'è l'esigenza di garantire ai pazienti pediatrici dei trattamenti farmacologici il più adeguati possibile, spesso non disponibili sul mercato.

Al fine di assicurare una protezione idonea a questa popolazione vulnerabile, è di grande importanza l'impiego di metodologie innovative come il modeling matematico, l'estrapolazione e i disegni di studio alternativi. Questa tesi è incentrata sull'utilizzo di queste tecniche innovative con lo scopo di facilitare lo sviluppo pediatrico del farmaco.

L'estrapolazione è un approccio strategico che consiste nell'estendere informazioni e conclusioni sul farmaco oggetto di studio da una particolare popolazione di pazienti (detta popolazione sorgente) ad un'altra popolazione di pazienti (detta popolazione target), allo scopo di giustificare in quest'ultima l'utilizzo del farmaco e permettendo quindi di ridurre la quantità di dati da generare nella popolazione target. L'elaborazione di un piano di estrapolazione dovrebbe basarsi su di una sintesi sistematica dell'informazione disponibile con l'intenzione di formulare delle ipotesi esplicite e quantitative riguardo le similarità tra la popolazione sorgente e quella target in termini di patofisiologia della malattia, relazione farmacocinetica-farmacodinamica e risposta clinica. Diversi approcci sono stati applicati per la definizione di un piano di estrapolazione nello sviluppo pediatrico di un farmaco. Tra questi, il modeling matematico è sicuramente uno dei più attraenti ed è riconosciuto avere un ruolo chiave in questo contesto, motivo per cui dovrebbe essere alla base – ogniqualvolta è possibile farlo – di un solido piano di estrapolazione.

Come caso di studio, in questa tesi viene presentato l'utilizzo del modeling matematico nella selezione della dose di un trial clinico pediatrico per la prevenzione della sindrome da lisi tumorale basato sull'estrapolazione di dati di uno studio di fase III nel paziente adulto. Prendendo in considerazione le similarità della sindrome da lisi tumorale tra la popolazione pediatrica e quella adulta, l'efficacia e la sicurezza del farmaco nella popolazione target possono essere in parte estrapolate a partire da quelle osservata nell'adulto, consentendo così di evitare un classico sviluppo clinico a favore di un più flessibile studio farmacocinetico-farmacodinamico di fase I/II, con una conseguente analisi di simulazione per l'ottimizzazione della dose se richiesto. Per la selezione delle dosi da testare nello studio viene proposto un approccio basato sul modeling matematico che consente di mirare, nella popolazione target, alle

esposizioni al farmaco dimostrate essere efficaci nell'adulto, assicurando così che il paziente pediatrico non sia sotto o sovra esposto al farmaco stesso.

Infine, in questa tesi vengono presentati disegni di studio alternativi che, rispetto al classico disegno parallelo, possono potenzialmente incrementare la fattibilità dei trial pediatrici. Viene proposto un framework basato sulla simulazione di trial clinici per la valutazione ed il confronto del disegno parallelo con disegni di studio alternativi quali i disegni sequenziali, il disegno crossover ed il disegno "randomized withdrawal". Oltre a questi disegni di stampo frequentista, vengono indagate anche le performance di disegni di studio Bayesiani come il Bayesiano parallelo e due implementazioni alternative del Bayesiano sequenziale, una di tipo non-gerarchico e una semi-gerarchica, nei quali l'informazione a priori è ricavata da studi nell'adulto e pesata sulla base delle similarità attese nella risposta al trattamento tra la popolazione pediatrica e quella adulta. I disegni di studio sono valutati in termini di: errore di tipo I e di tipo II, sample size per gruppo, durata totale del trial, esposizione ai vari tipi di trattamento (ovvero trattamento attivo, controllo o nessuno dei due) e precisione della stima del parametro di effetto del trattamento. I risultati ottenuti mostrano che il disegno crossover richiede il minor sample size e la minor durata, sebbene implichi una maggiore esposizione sia al placebo che a all'assenza di alcun trattamento. Il randomized withdrawal massimizza l'esposizione al trattamento attivo, minimizzando contemporaneamente quella al placebo, anche se richiede il maggior numero di pazienti. Il sample size dei disegni sequenziali può in qualche circostanza essere minore di quello del crossover, anche se in tali casi non è garantita una robusta stima dell'effetto del trattamento. Riguardo ai disegni Bayesiani, non si osservano differenze sostanziali tra quello sequenziale non- e semi-gerarchico, ed entrambi richiedono un sample size ed una durata minore rispetto al Bayesiano parallelo, che d'altra parte garantisce una stima più precisa dell'effetto del trattamento. In generale, gli approcci Bayesiani sembrano avere perfomance migliori delle loro controparti frequentiste anche quando viene dato poco peso all'informazione a priori dall'adulto.

Complessivamente, il framework farmacometrico proposto permette un confronto multilivello di disegni di studio alternativi che può essere utilizzato per la selezione di trial clinici futuri nella popolazione pediatrica.

Abstract (English)

When planning pediatric clinical trials, issues such as feasibility, ethical challenges and efficiency implications must be addressed, whilst ensuring that the unmet medical needs will still be satisfied. In order to guarantee adequate protection, innovative methodological approaches such as modeling and simulation, extrapolation and alternative study designs are of great importance. This thesis focuses on the implementation of these innovative techniques as a tool to facilitate pediatric drug development.

Extrapolation is a strategic approach which consists of extending drug-related information and conclusions available from one particular patient population (source population) to another patient population (target population), in order to justify the use of the drug in the latter, thus reducing the need to generate additional information in the target population. The development of an extrapolation plan should build upon a systematic synthesis of available information with the aim of providing explicit (quantitative) hypotheses regarding the similarity of the disease pathophysiology and the similarity of pharmacodynamics and clinical response to the intervention between the source and target populations. Various approaches have been applied to define an extrapolation plan within paediatric drug development programs. Among these, modeling and simulation is recognised as a mean of outstanding value upon which the extrapolation process should be underpinned whenever possible.

As case study, the use of modeling and simulation for dose selection in a pediatric trial for tumor lysis syndrome prevention based on the extrapolation of adult phase III data is presented in this thesis. Considering tumor lysis syndrome similarities between the pediatric and adult population, drug efficacy and safety in the target population can be partly extrapolated from those observed in adults and a standard development plan can be skipped in favor of a more flexible phase I/II pharmacokinetic-pharmacodynamic study, with a following modeling and simulation analysis for dose optimization if needed. A model-based approach is proposed for dose selection for the pediatric study, which allows targeting the efficacious drug exposure observed in adults whilst ensuring that children will be not under/over-exposed to the drug.

Finally, alternative study designs which can increase the feasibility of pediatric trials when compared to classical parallel designs are presented in this thesis. A model-based clinical trial simulation framework is proposed as a tool for the comparison of the parallel design with the alternative sequential, crossover and randomized withdrawal designs. Besides these frequentist designs, the performance of a fixed-sample Bayesian design and

two alternative Bayesian sequential designs, i.e. a non-hierarchical and a semi-hierarchical one – where prior information is elicited from adult trials and weighted based on the expected similarity of response to treatment between the pediatric and adult population – are also investigated. Study designs are evaluated in terms of: type I and II errors, sample size per arm, trial duration, treatments exposures and parameter estimate precision. The results obtained show that the crossover requires the lowest sample size and trial duration, although it implies higher placebo and no treatment exposures. The randomized withdrawal design maximizes exposure to active treatment while minimizing that to placebo, but requires the largest sample size. Sample size of sequential designs can sometimes be smaller than the crossover one, although with poorer estimate precision. With respect to Bayesian designs, no substantial differences were observed between non-hierarchical and semi-hierarchical Bayesian sequential designs. The sequential implementation of Bayesian designs requires on average smaller sample size and trial duration compared to the standard one, which on the other hand guarantees higher estimate precision. When large differences between children and adults are expected, Bayesian sequential designs can return very large sample size. Overall, Bayesian approaches appear to outperform their frequentist counterparts in the design of pediatric trials even when little weight is given to prior information from adults. In general, the proposed pharmacometric framework allows a multiscale comparison of alternative study designs which can be used for design selection in future pediatric trials.

Contents

Chapter 1

Introduction

The implementation of pediatric trials is challenging and often difficult to accomplish due to ethical, practical and even financial considerations. The design, analysis and interpretation of clinical studies in the pediatric population require specific techniques to ensure accurate decision-making regarding the pharmacokinetics, pharmacodynamics, safety and efficacy of drugs, as also supported by the guideline on clinical trials in small populations set by the European Medicine Agency, which states that "crude (simple) methods may often be adequate when we have huge amounts of data, but when there are very few data, it is imperative that the most efficient and informative analytical methods should be used".

Extrapolation is a strategic approach that may allow one to circumvent some of the aforementioned difficulties: it consists in extending information and conclusions available from studies in one or more subgroups of the patient population (source population), or in related conditions or with related medicinal products, to make inferences for another subgroup of the population (target population), or condition or product, thus reducing the need to generate additional information.

A very useful methodological tool that naturally fits into the context of extending information from a source population to make inferences for another population is Modeling and Simulation (M&S). The added value of M&S in pediatric clinical research has been extensively documented, and its weight at a regulatory level in supporting extrapolation has constantly been increasing in the last years.

Finally, the current way of designing efficacy trials in pediatrics is still largely traditional and do not go beyond standard approaches such as the parallel design. Due to ethical and methodological hurdles (e.g. the invasiveness related to pain/anxiety and blood loss, and the limited volume and number of blood samples that can be withdrawn in pediatric patients), identification, recruitment and enrolment of a number of children that could guarantee a sufficient statistical power are often difficult to accomplish.

Conversely, alternative study designs possess favourable features which can address part of the issues related to paediatric clinical research, especially when compared to the current standard practice.

This thesis deals with the use of M&S, extrapolation and alternative study designs to facilitate pediatric drug development, and it is structured as follows. Chapter 2 aims to review, explain and motivate the use of the extrapolation approach in pediatric drug development with a focus on the regulatory point of view. Chapter 3 presents a case study dealing with a model-based dose selection in a pediatric clinical trial for tumor lysis syndrome prevention based on the extrapolation of adult phase III data, with an emphasis on the use of pharmacokinetic-pharmacodynamic (PK-PD) M&S as a protocol optimization and data analsysis tool. Chapter 4 presents a PK-PD-based clinical trial simulation framework whose aim is to assess the performance of alternative study designs in pediatric trials across different metrics of comparison. While Chapter 4 deals with alternative designs of frequentist nature, Chapter 5 focuses on Bayesian approaches where prior information is elicited from historical adult data. The thesis ends with an overall conclusion in Chapter 6.

Chapter 2

The extrapolation approach

2.1. Definition and concept

Optimizing the development of medicinal products for children can lead to many dilemmas. On the one hand, there is a clear medical, ethical and regulatory need for rigorous evaluation of medicinal products for children. This is strongly supported in Europe by the "Regulation on medicinal products for pediatric use" [1]. On the other hand, children are vulnerable and must be protected without being exposed to unnecessary trials – this includes protection from the potential harm (where possible) of the clinical investigations required for appropriate evaluation. Hence, when planning pediatric clinical trials, issues such as feasibility, ethical challenges and efficiency implications must be addressed, whilst ensuring that the unmet medical needs will still be addressed.

To ensure adequate protection, different approaches may be adopted: for example, appropriate preventative or adapted technical procedures may be undertaken, or innovative methodological approaches can be used, such as extrapolation.

Extrapolation is a strategic approach, which consists of extending drug-related information and conclusions available from one particular patient population (source population) to another patient population (target population), in order to justify the use of this drug in the latter, thus reducing the need to generate additional information in the target population.

The ICH E 11 guideline on clinical investigation of medicinal products in the pediatric population [2] provides general recommendations on what type of studies should be performed in children based on the differences/similarities with older populations, with the aim of reducing the amount of data that needs to be generated in the pediatric population to provide adequate information for pediatric use. The Food and Drug Administration (FDA) initially translated such recommendations into a

decision tree (Figure 2.1) published in the FDA guidance on exposure-response relationship [3]. Subsequently, FDA released two pediatric-specific draft guidance, one on general clinical pharmacology considerations [4] and one on extrapolation for pediatric uses of medical devices [5]. However, despite providing an overview of the possible scenarios and the general clinical development to be followed, the decision tree does not address several important issues.

A slightly different approach from the FDA decision tree is adopted by the European Medicines Agency (EMA), which is trying to integrate more pharmacology and maturation, and delineate a broader strategy for the definition of an extrapolation plan [6]. For example, the FDA decision tree applies only to extrapolation between age-classes, whereas in the EMA concept paper extrapolation across conditions and drug classes is also considered.

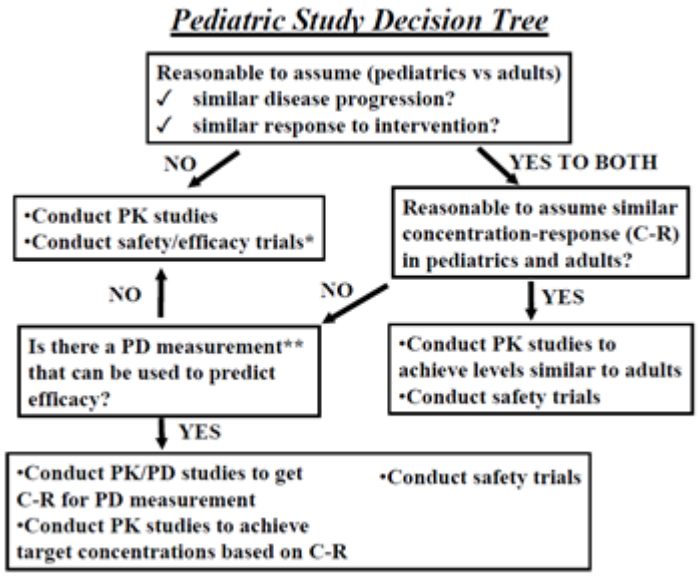


Figure 2.1: FDA pediatric decision tree [3].

The development of an extrapolation plan should build upon a systematic synthesis of all available information, which collects data from literature information, non-clinical studies, adult clinical studies and existing pediatric data - if already available in different age groups, conditions or the same class of drugs.

The extrapolation process must be developed with the aim of providing explicit (quantitative) hypotheses regarding the similarity of the disease pathophysiology and the similarity of pharmacodynamic response to the intervention between the source and target population. The uncertainty of each assumption should be qualified and quantified, for example in terms of its limitations, together with its potential clinical impact. The uncertainties should guide the extrapolation plan and will influence the set

of rules and methodological tools for the reduction of data requirements in the target population. In order to successfully complete the extrapolation process, the hypotheses must be pre-specified and based on the outcome of the assessment in the source and target populations, with consideration of the similarity/dissimilarity of the two populations, and the assumptions that should underpin the expected biological, pathophysiological and pharmacological differences.

Firstly, a clear understanding of the target disease, along with its subtypes, has to be documented in terms of etiology, pathophysiology and symptoms; moreover, similarities and differences in disease progression should be addressed where a spontaneous evolution pattern has already been identified in the natural history of the disease (e.g. in the evolution of cognitive or motor functions).

Secondly, the pharmacology of the drug has to be comprehensively described in relation to the mechanism of action, pharmacokinetics (PK), pharmacodymanics (PD) and dose-exposure-response relationship. Particular emphasis should be put on the developmental changes due to organ maturation and the ontogeny of enzymes, transporters, and receptors, especially in children younger than 2-3 years of age [7].

Finally, the degree of similarity/heterogeneity in PD and clinical response between the two populations has to be quantified. Should the nature of the target population be such that definite clinical endpoints cannot be measured, PD endpoints or biomarkers suitable to predict clinical response in children must be considered early during adult drug development whenever possible. In this way, extrapolation can become an integrative part of drug development based on prospectively collected data, contrasting with the tendency of retrospectively planning extrapolation ad-hoc based on available information.

The extrapolation process is primarily clinically-related and should be supported by the use of innovative tools such as modeling and simulation (M&S) and statistical methods. M&S is a tool of particular interest for extending information from a source population to make inferences about another population whenever applicable. The added value of M&S in pediatric clinical research has been extensively documented [8-12], and its weight in supporting extrapolation at regulatory level has steadily increased in recent years, both in Europe and the US [13-15].

However, M&S should not be identified as synonymous with extrapolation. Extrapolation is a strategic clinically-related approach to pediatric drug development, broader than M&S, which is a corroborating tool that can be utilized for optimal data analysis, predictions of results and dosing regimens recommendations [16].

Full, partial or no extrapolation

On the basis of the degree of similarity - or dissimilarity - between the source and target population and the uncertainty of hypotheses, three

general categories should be considered: full, partial, and no extrapolation [6].

Complete or full extrapolation of efficacy between the source and the target population (e.g. from adults to children) can be acceptable with quantified assumptions that are supported by robust data indicating small or negligible differences between the two populations, yet some supportive data to validate the extrapolation concept may still be necessary. This situation is likely when extrapolating between very similar populations [17].

If, on the contrary, it is not possible to assume that children have a similar natural history of the disease or response to intervention when compared to adults (e.g. for diseases specific to children) and/or for conditions in which a validated clinical endpoint cannot be assessed in the target population (e.g. 6-Minute Walking Test), a complete set of data in the target population is typically necessary; this could be described as a situation of "no extrapolation".

More frequently, the extent to which extrapolation can be applied lies in between these two extreme situations; efficacy is then extrapolated based on a reduced set of data in the target population, depending on the magnitude of expected differences and the uncertainty of the assumptions. This situation is referred to as "partial extrapolation", and two principal scenarios may be considered: (i) bridging using PK or PK/PD in the target population to extrapolate efficacy. This approach would be based on the concept that matching drug exposure-response to the source population will be associated with similar efficacy in the target population. (ii) Some efficacy data are considered necessary in the target population, the nature of which depends on the degree of extrapolation from the source population. Such a scenario could be supported by statistical approaches using prior information from the source population(s) [18].

For example, gabapentin was approved for the treatment of partial onset seizures (POS) in children aged 6 years and above although trial results did not show a statistically significant difference in the 50% responder rate [19-20]. Indeed, the clinical trial of adjunctive treatment of POS in pediatric subjects (aged 3 to 12 years) showed a numerical but not statistically significant superiority in the 50% responder rate of gabapentin over placebo. In spite of these results, taking into consideration the adequate safety profile and the medical need in this pediatric population, the following indication was granted: "Gabapentin is indicated as adjunctive therapy in the treatment of partial seizures with and without secondary generalization in adults and children aged 6 years and above."

On the basis of ethical and feasibility issues, as well as resource allocation, clear justification is always needed as to why extrapolation is being undertaken rather than a complete set of prospective studies. Extrapolation will increase the results uncertainty, which needs to be quantitatively and statistically estimated, and requires potential risks to be addressed in a risk management plan. For populations in which trial feasibility is not an obstacle, extrapolation may still be appropriate in order

to optimize the clinical development of the compound and/or to avoid unnecessarily exposing children and their families to the burden of a clinical trial.
Once a rationale is provided to support the selected extrapolation plan, there is the need to define a set of rules and methodological tools for the reduction of data requirements (types of studies, design modifications, number of patients) in accordance with the degree of expected similarities between the two populations (source and target). The data generated in the target population should validate the extrapolation concept and complement those data that may be extrapolated from the source population. Studies should focus on specific areas, e.g. age subsets, where the largest differences to the source population are expected.

2.2. Types of source data

Extrapolation for drug disposition data

Several methods are available for dose/clearance scaling of a drug from a source population to children [21], and many examples and reviews can be found in the literature, either of extrapolation from older subjects [22-26] or, less frequently, from animal studies [27, 28]. In principle, regardless of the method selected, it is important to have a clear understanding of the differences in drug absorption, distribution, metabolism and excretion between the two populations, in relation to the various aspects of developmental pharmacology [29].

Simple linear dose/clearance scaling on a mg/kg basis has been deemed suboptimal and may lead to sub-therapeutic drug exposures, especially in the very young patients [30, 31]. Allometric approaches offer a sound basis for scaling doses from older to younger patients, since they are supported by a well-established theory [30]. However, allometry can only capture differences in drug disposition and effect due to size variation, without addressing the influence of developmental changes related to target organs maturation. As a result, the allometric principle is usually coupled with a so-called maturation function aimed at describing the development of the relevant biological system (e.g. CYP450 enzyme activities, kidney, etc.) in relation to age [26]. Physiologically-based pharmacokinetic (PBPK) models are system-specific models characterizing human physiology that enable to incorporate distinct types of data (e.g. in vitro, preclinical, clinical), thereby naturally fitting in a context of data synthesis, especially for a “first in children” trial [32, 33]. Although it may be argued that, in contrast to allometric scaling, PBPK models do not have a theoretical foundation as they are based solely on empiricism [26], their inherent mechanistic nature allows the movement from one drug to another with more flexibility than the combination of allometric scaling with a maturation function [34]. Maharajand and Edginton [35] proposed a workflow for scaling adult PBPK models to children, while Johnson et al.

[36] predicted the clearance of 11 drugs in neonates, infants and children adapting information from the literature into different PBPK models.

In pediatric drug development these techniques are usually planned at the design stage. Exploiting available information at the time of trial design, the three assumptions of the extrapolation concept are assumed to be true and bridging takes place, i.e. the dose that will be used in the confirmatory pediatric trial is the one that leads to drug exposure levels considered to be efficacious in adults, assuming similar plasma exposure would translate into a similar clinical response [37-39]. Dose definition is usually supported by PK simulations with a PK or PBPK model (see Chapter 3). This can be viewed more as an optimization process rather than an actual extrapolation, which would instead be applied for drug registration.

Extrapolation for efficacy disposition

The source population from which efficacy is extrapolated may be the adult one [40]. In epilepsy, for example, POS are similar in children and in adults, and extrapolation can be achieved provided that the dose is adjusted according to age-specific drug disposition [41]. For the Pediatric Investigation Plan (PIP) of brivaracetam, the EMA has considered whether it would be acceptable not to request specific new data in children on the basis of extrapolation results. This type of epilepsy usually responds relatively well to drug therapy [42]. However, other epileptic syndromes in children do not exist in adults and extrapolation from adults is not possible, therefore pediatric data are required. These types of epilepsy are usually resistant to available drug therapy and are associated with poor psychomotor development in affected children.

There were some circumstances in which extrapolation was applied without the collection of new pediatric data in the target indication. Topiramate [43] and oxcarbazepine [44] monotherapy dosage regimens were bridged from the adjunctive setting in the treatment of POS. The evidence gap relating to monotherapy treatment in children was filled in using pediatric/adult data from adjunctive therapy and adult monotherapy data (left panel of Figure 2.2).

Similarly, approval of darunavir co-administered with ritonavir in 3 to <12 year-old HIV-1 infected patients naïve to antiretroviral (ARV) therapy leveraged data from >3 years old ARV-experienced patients and from 12 to <18 years old ARV-naïve patients [45] (right panel of Figure 2.2).

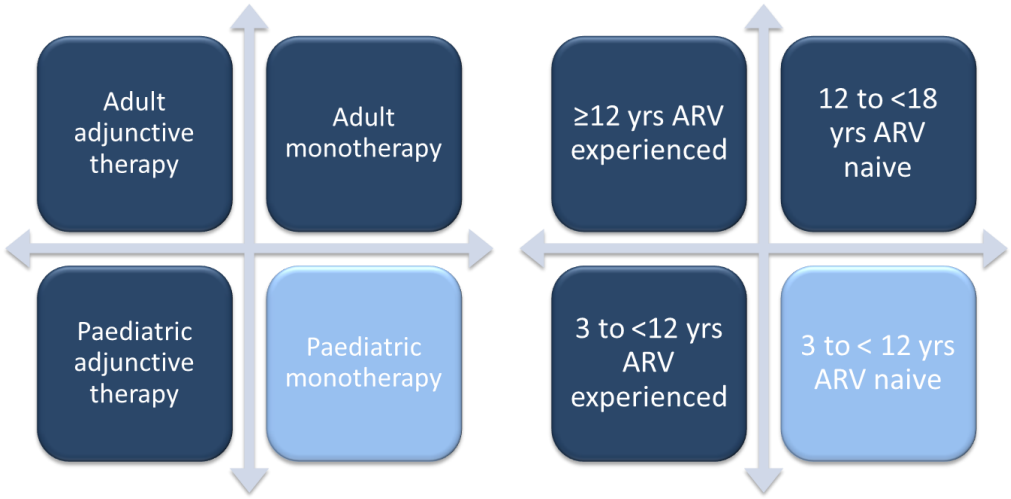


Figure 2.2: Examples for which extrapolation was applied without the collection of new pediatric data in the target indication (light blue boxes). Dark blue boxes represent data from the source population(s).

Extrapolation for demonstration of safety

The extrapolation of safety should always be viewed with caution. Although children's therapeutic (beneficial) responses to treatment may be similar to those of adults, adverse event (AE) profiles can vary substantially between these two populations [46]. An illustrative example is the increased sensitivity to sotalol-induced QTc interval prolongation in neonates compared with older children and adolescents [47]. Age-specific adverse events are present across the whole pediatric age-range, from conception to post-puberty (Table 2.1). Despite many safety concerns relate to neonates and infants (especially due to off-label use in these age subsets), excluding older children from appropriate safety evaluations may lead to serious consequences. This was exemplified by the 2007 FDA warning after ascertaining that children, adolescents and young adults taking antidepressant therapies were at increased risk of suicidal thinking and behavior compared to those taking placebo [48].

Examples presented in Table 2.1 show that extrapolation of safety from adults to children cannot be taken for granted, due to the potential for serious, adverse drug reactions which can affect both growth and maturation. However, in some instances, extrapolation may depend on the level of maturity of the physiological functions: extrapolation can be possible when the relevant physiological functions have reached full maturity. For example kidneys are fully matured at 3 years of age [85], thus safety data on the renal system may not specifically be collected if the target population is older than 3 years. Nevertheless, exemptions should always be justified based upon a clear understanding of the maturational time profile of the relevant physiological functions.

Table 2.1: Examples of age-specific adverse events across the whole pediatric age-range.

Drug/class of drugs	Adverse event	Time of exposure	Age-range at diagnosis	Ref.
Thalidomide	Phocomelia	In utero	Neonatal	[49]
Nonsteroidal anti-inflammatory drugs (NSAIDs)	Closure of ductus arteriosus	In utero	In utero (foetal death)	[50]
	Renal failure		Neonatal	[51, 52]
ACE inhibitors (ACEIs)	Renal failure	In utero	Birth	[53]
		Postnatal	Neonatal	[54]
Selective serotonin re-uptake inhibitors (SSRIs)	Withdrawal syndrome	Prenatal	Neonatal	[55, 56]
Antidepressants	Suicidality	Childhood	Childhood - adolescence	[48]
Methadone	Neonatal abstinence syndrome	Prenatal	Neonatal	[57, 58]
β-Blockers	Being born small for gestational age (SGA), preterm birth and perinatal mortality	Prenatal	Neonatal	[59]
	Neonatal hypoglycaemia	In utero	Neonatal	[60]
Anti-epileptic drugs (AEDs)	Haemorrhage	Prenatal	Neonatal	[61-63]
	Anatomic and cognitive impairments	Prenatal	Childhood - adolescence	[64, 65]
	Worsening of juvenile epilepsy	Postnatal	Childhood - adolescence	[66-68]
Glucocorticoids	Growth suppression	Postnatal	Childhood	[69, 70]
Propofol (prolonged infusion)	Rhabdomyolysis, hypoxia and myocardial failure	Postnatal	Childhood	[71, 72]
Oxygen	Retrolental fibroplasia	Postnatal	Infancy	[73]
Erythromycin	Pyloric stenosis	Postnatal	Infancy	[74, 75]
Diethylstilbestrol	Adenocarcinoma of the vagina and cervix	In utero	Childhood	[76]

Tetracyclines	Enamel hypoplasia	Postnatal-childhood	Children <8 years	[77]
Anthracyclines	Delayed symptoms of cardiotoxicity	Postnatal-childhood	Young adults	[78, 79]
Busulfan	Ovarian dysfunction	Postnatal-childhood	Young adults	[80, 81]
Chemotherapy	Impaired fertility	Postnatal-childhood	Young adults	[82]
Topiramate	Cognitive dysfunction	Postnatal-childhood	Children-adolescents	[83, 84]

Extrapolation from non-clinical studies

Non-clinical studies in the development of pediatric medicinal products offer an opportunity to start addressing potential differences in safety profiles between children and adults. However, in circumstances where developing organs may be affected, juvenile toxicity studies may be needed [86-88] in order to avoid both short- and long-term delayed adverse events such as the anatomic and cognitive impairments in children exposed to anti-epileptic drugs (AEDs) in utero [64, 65] and glucocorticoid-induced growth suppression in children [69, 70]. In addition, the use of PBPK models to address safety issues may be considered, as they facilitate the estimation of drug concentrations in target organs [89]. However, the assumptions regarding the dose-exposure-response relationship still need to be verified and this must not be forgotten at a time when the ongoing physiological development may modify the response of developing organs with regards to both efficacy and safety.

Moreover, animal data can be used to optimise dose selection during early development in pediatrics. Hope and Drusano [90] proposed a framework for animal in vivo-to-human bridging of antifungals PK-PD relationship. The authors used a case study to show the application of such an approach, based on echinocandin therapy in neonates with haematogenous Candida meningoencephalitis (HCME) [91]; this represents one of the very few examples of extrapolation of PK-PD from animal data. In this analysis, a PK-PD model of micafungin in rabbits and a neonatal PK model were integrated to estimate the doses leading to near-maximal antifungal effect in neonates. The fact that no a priori information was available on the appropriate dosage of micafungin for neonatal HCME permitted tolerance of strong assumptions regarding the similarity of exposure-response relationship and tissue penetration of the drug between rabbits and neonates. A similar approach for dose selection has been adopted for the PIP of vancomycin for the treatment of late-onset bacterial sepsis in neonates and infants aged under 3 months [92].

To further refine this approach, the influence of immaturity and the development of the immune system in early life should also be considered when defining a dosing rationale based on the PK-PD relationship (as was

indeed acknowledged by the authors) [93-95]. Unfortunately, however, while age-related differences in PK have been extensively studied and are increasingly well-characterized, a remarkable knowledge gap persists in the field of developmental PD. The scarce data available to date come predominantly from animal studies, and although they provide valuable insights, extrapolation to children is marred by uncertainty, especially in the absence of clinical data [96]. Therefore, there is an urgent need for new data on developmental PD in order to facilitate the expansion of optimal pharmacotherapy for children.

Extrapolation from another medicinal product

Another useful source of information is represented by data coming from another authorized medicinal product in the same condition. Jadhav et al. [97] presented a case study for leveraging prior quantitative knowledge for trial design of a new anti-hypertensive agent X for immediate blood pressure control, wherein the authors borrowed information on the placebo response and exposure-response relationship in children from pediatric labelling information for fenoldopam, another drug approved for a similar indication in both adults and children. The Sponsor and FDA used this information to perform clinical trial simulations (CTS) with the aim of enhancing the design of a trial for drug X with the correct sample size and dose ranges. In particular, placebo data from 16 pediatric patients in the fenoldopam trial were used to develop an empirical placebo model for CTS. As to drug effect simulations, Jadhav and colleagues took advantage of an Emax model for drug X based on adult data. Because the fenoldopam trial experience suggested that the pediatric population is less sensitive and less responsive than the adult one, the impact of different scenarios (in terms of different Emax and EC50 values) on the final sample size of the pediatric study was investigated. This led to an acceptable trial design, rational dosing recommendations and useful labelling information in pediatrics.

Extrapolation from another disease

If the treatment to be investigated in the target population is already approved for another disease, clinical data from studies in that disease may be used to support dosing extrapolation to the new indication.

A dose-response study of losartan in hypertensive children from 6 to 16 years of age concluded that a starting dose of 0.75 mg/kg given once a day effectively reduced diastolic blood pressure, and a once-daily dosage up to 1.44 mg/kg was generally well tolerated [98]. The design of a subsequent study for the use of losartan in children with proteinuria took advantage of this information [99]. In particular, dose selection was based on these previous findings and in the new trial children randomized to the losartan arm received a starting dose of approximately 0.7 mg/kg that was up-titrated to a maximum of 1.4 mg/kg.

2.3. Methods of use

2.3.1. Class-specific examples

This section now addresses some specific examples of therapeutic classes and the principles of extrapolation relevant to these medicines are considered in turn.

Anti-epileptic agents

As already mentioned, the treatment of POS has been deemed to be similar between children and adults and it may be reasonable to extrapolate efficacy from adults to children. In response to a regulatory request from the EMA for the approval of brivaracetam, Pellock et al. [42] performed a systematic review of published efficacy trials in the treatment of POS and primary generalised tonic-clonic seizures (PGTCS) in adults and children. They compared 52 efficacy trials of 5 different anti-epileptic drugs in children and adults by means of forest plots of the treatment effect size in terms of both the median percent seizure reduction and ≥ 50% responder rate. By showing that confidence intervals overlap each other in the two populations, the authors advocated that, overall, the effect measures were comparable between adult and pediatric studies. Although the evidence collected thus far suggests that extrapolation of efficacy is appropriate, safety data are still to be generated. In addition, such analysis is valid only for the adjunctive treatment of POS in children older than 2 years of age. Indeed, the paucity of historical data on adjunctive therapy for POS in children younger than 2 years, and on monotherapy for POS and monotherapy/adjunctive therapy for PGTCS in the whole pediatric age range did not allow the postulation of any similarities between the two populations in terms of response to treatment. Thus, collection of new clinical data is needed in these subsets of the pediatric population.

Besides showing no substantial differences between children and adults in terms of response to treatment, adjunctive therapy for POS was also deemed to satisfy the assumption on the similarity of the PK-PD relationship [43, 44]. The latter conclusion was reached by identifying two different PK-PD models in pediatric and adult patients and statistically testing whether the parameter describing the concentration-effect relationship was significantly different in the two models. As this was not the case, the PK-PD relationship was assumed to be independent of age, making the application of extrapolation in adjunctive treatment of POS straightforward.

Anti-infectious agents (antibiotics, antibacterials, antifungals and antivirals)

Traditionally, antimicrobial activity extrapolation was viewed as a relatively straightforward exercise. This perception arose because the PK-PD targets of therapy (summarized in Table 2.2 for antibacterials) were considered to be directly comparable between adults, children, and neonates. For many older antibiotics such as penicillins, the standard adult doses were scaled down for children and infants according to weight (in a simple linear fashion) or age, and frequently detailed pediatric or neonatal studies were omitted [100]. Since the clearance-body weight relationship is non-linear whereas the volume-body weight relationship is linear, the shape of the PK curve differs in children of different weights. In children aged under 2 years, maturation further complicates matters. This brings two potential problems in extrapolation: firstly, the dose to achieve the indices in Table 2 will not be linear mg/kg, and secondly, recent evidence suggests that PK curve shape - independent of AUC, for example - may be important in antimicrobial resistance development [101].

In addition to PK differences, extrapolation should ideally account for PD. Neonates, and particularly pre-term neonates, are functionally immunocompromised so the assumed dynamics of the "host – bug – drug" interactions may not hold. Likewise, the impact of ontogeny and developmental pharmacology also play an important role in drug safety. Some well-known examples, where the impact of antimicrobial pharmacology on safety has been apparent, include (i) the former use of chloramphenicol in neonates – which precipitated the potentially fatal gray baby syndrome – due to the immaturity of the neonatal UDP-glucuronyl transferase enzyme system; and (ii) the risks of sulphonamide use during the neonatal period, which can lead to kernicterus, because the drug displaces bilirubin from albumin binding sites; it is necessary to consider these historical examples when making recommendations for extrapolation of antimicrobial PK-PD data, in order to mitigate the risks of such adverse events.

A more recent example of antifungal pharmacology shows the application – and occasional limitation – of linear extrapolation of dosing regimens on a mg/kg basis, when voriconazole (a potent, broad-spectrum, tri-azole) was investigated in children. A 2004 study [103] aimed to generate a drug exposure in children similar to that achieved in adults, with the understanding that this should prove efficacious – assuming the PD aspects were equivalent. It became apparent during the study that pediatric patients actually had a greater voriconazole elimination capacity (per kg body weight) than healthy adult volunteers; so, to generate adequate exposure, children required 4 mg/kg (rather than the 3 mg/kg recommended in adults) [103]. However, a later study then suggested that 7 mg/kg twice daily were actually needed to reach the median AUC (area under the curve) values of adults (that was achieved with 3 to 4 mg/kg twice daily) [104]. Voriconazole's non-linear PK further complicated the development of

rational pediatric dosing guidelines, and provided a good example of where dedicated pediatric PK studies led to the evolution of dose guidelines, which were not predicted with extrapolation from adults – this example is discussed further below in the section on PK bridging and the evaluation of extrapolation.

Table 2.2: PKPD indices for major classes of antibacterial agents, reproduced from Barker et al. ADDR2014, (Open access under CC BY-NC-ND license) [102].

PD	Class of Antibacterial	Specific drug (where applicable)	PD parameter	PD Target of therapy
Time-dependent	Beta-lactams	-	%T>MIC	Max %T in the dosing interval >MIC
	Macrolides	Conventional macrolides	%T>MIC	Max %T in the dosing interval >MIC
		Azithromycin	AUC/MIC	Optimal daily amount
	Glycopeptides	-	AUC/MIC	Optimal daily amount
	Tetracyclines	-	AUC/MIC	Optimal daily amount
	Oxazolidinones	Conventional oxazolidinones	%T>MIC	Max %T in the dosing interval >MIC
		Linezolid	AUC/MIC	Optimal daily amount
Concentration-dependent	Aminoglycosides	-	C_{max}/MIC or AUC/MIC	Max peak concentration or optimal daily amount
	(Fluoro)quinolones	-	C_{max}/MIC or AUC/MIC	Max peak concentration or optimal daily amount
	Metronidazole	-	C_{max}/MIC or AUC/MIC	Max peak concentration or optimal daily amount

Another pivotal consideration for antimicrobial therapy extrapolation is whether the Minimum Inhibitory Concentration (MIC) distribution of the causative pathogen is the same in children and adults; this assumption has generally underpinned most previous studies. Two UK groups have now obtained data (H Hill et al (unpublished at the time of writing), and A Kent at al.) suggesting that this may always not be the case. In the study by Kent et al. [105], it was also shown that higher gentamicin MICs were associated with increased neonatal mortality – even when the MIC was within the

susceptible range according to current EUCAST susceptibility breakpoints. This reinforces the need for further research into PD in neonates. In the short term it is likely that researchers will continue to use the standard antimicrobial PK-PD target indices in children and adults, and aim to achieve similar drug exposures, after appropriate pediatric PK investigations. However, with improved understanding of antimicrobial PD including clinical factors to determine the resolution of infection (and thus appropriate duration of treatment) and advances in diagnostic technologies (e.g. to quantify bacterial DNA loads in real-time), the methods for extrapolation and dosing regimen design will have to become more sophisticated in the coming years.

Anticoagulants

Lala et al. [38] defined a genetic-based warfarin pediatric dosage regimen. They started from the evidence that drugs with a similar mechanism of action do not show a significantly different PK-PD relationship between pediatric and adult populations, thus justifying the use of an adult PK-PD model for bridging purposes. Nevertheless the authors highlighted that there are some data showing intrinsic developmental differences in the coagulation system in the very young (<2 years) [106], thereby making the previous hypothesis unreliable in this subset of the pediatric population. The PK was allometrically adjusted for body size in children, while the effect of enzymatic ontogeny on clearance was accounted for by linking clearance with age following a validated method based on warfarin data [36]. The final PK-PD model was then visually validated on data from 26 pediatric subjects, comparing the predicted individual International Normalised Ratio (INR) versus the observed one.

Based on these results the authors claimed that the model can be used for CTS, which ultimately led to the definition of a pediatric genetic-based warfarin dosing regimen (based on reaching the target INR).

The main limitation of the study, as acknowledged also by the authors, is depicted by the paucity of available clinical data. Accordingly, a multicentre pediatric warfarin pharmacogenetic trial sponsored by the FDA is now underway. Thus, an extrapolation concept has been formulated and partially validated on some clinical data. However, the quantity of clinical data did not allow complete confirmation of the assumptions made, so the extrapolation concept will be further evaluated and possibly refined with the collection of more clinical data. Nonetheless, the proposed methodology was suitable in helping to identify the right dose for the subsequent larger trial.

A similar analysis was done by Hamberg et al [107], who bridged an adult PK-PD warfarin model to children to account for maturational changes in PK parameters (both allometrically and through the same maturation function used by Lala and colleagues). Interestingly, both the adapted model by Lala et al and Hamberg et al's model appear to overestimate pediatric INR. More specifically, Hamberg's team highlighted

that the overestimation seen in children <2 years old is not negligible, confirming that the assumptions made do not hold in this age subset.

2.3.2. PD measurements that can be used to predict efficacy

There exist clinical endpoints that have been validated in adults which are not measurable in the pediatric population or a subset thereof, making extrapolation difficult. In order to address this issue, new clinical measures known to predict efficacy that can be assessed in children should be explored during adult development and subsequently validated in the pediatric population. Such an approach was successfully exploited in the definition of a pediatric sildenafil dosage for pulmonary arterial hypertension (PAH) [108]. In adults, treatment effect in terms of exercise capacity is assessed through the 6-minute walk distance test (6MWDT). The test cannot be consistently measured across age groups of children due to their variability in cooperation and/or capability to perform the test [109]. An FDA analysis using data from drugs approved for adult PAH found a linear relationship between the 6MWDT and the pulmonary vascular resistance index (PVRI), a measure that can be easily assessed in children. Based on these findings, the sponsor verified the consistency of exercise improvement and PVRI improvement between children and adults and defined the pediatric dosage regimen by determining the dose leading to a PVRI improvement corresponding to a target 6MWDT improvement.

2.3.3. Extrapolation process

Irrespective of the methodology adopted, the use of extrapolation for the approval of new medicines for pediatric use should be planned early on during adult development. This enables ad-hoc data collection which ultimately facilitates the formulation of the extrapolation concept and allows reliable conclusions to be reached on the safe and effective use of the medicine in children.

Even though the methodologies presented so far highlighted the benefits of the extrapolation approach for a specific aim within the entire drug development, the extrapolation process should be perceived as an iterative one where hypotheses are challenged and subsequently confirmed or refined. The approval of olmesartan medoxomil in hypertensive pediatric patients in the US appropriately reflects this concept and shows how extrapolation underpins the whole pediatric drug development roadmap [37]. First, a virtual population of pediatric patients with hypertension was simulated (in terms of age, height, weight and baseline blood pressure (BP)) based on published demographic and BP data. Second, two assumptions were made on the PK and PK-PD relationship of the drug in children: (i) an adult PK model was used on the grounds that body weight

was identified as an important covariate for clearance, which was allometrically scaled and (ii) the exposure-response relationship between children and adults was considered similar, thus an adult PK-PD model was used to describe the response to treatment in pediatric patients. Third, the development of olmesatran medoxomil in children foresaw a three-arm trial to estimate the dose-response relationship, as requested by the FDA. The sponsor therefore leveraged the aforementioned PK and PK-PD models to design the forthcoming trial: with the PK model the expected pediatric drug exposures were simulated for the dosages to be used in the trial, while the PK-PD model was used to perform CTS. The results of this first extrapolation were twofold: the PK simulation showed that the designated doses would have generated appropriate olmesartan exposures in children, while CTS indicated that two dose groups would have been sufficient to estimate a dose-response relationship, thus allowing the reduction of the total sample size by one-third.

The FDA accepted the sponsor's analyses and conclusions, and the actual trial was run with two arms, together with an accompanying phase I single-dose study. The clinical data obtained from these two studies validated the extrapolation concept. On the one hand, a new pediatric PK model was built based on the PK study, which confirmed the appropriateness of allometrically scaling adult clearance by body weight. Most importantly, the two-arm trial was able to detect a statistically significant dose-response relationship of olmesartan medoxomil in children, thereby leading to pediatric approval of the drug.

2.4. Evaluation of extrapolation applications

The extrapolation concept should ideally be evaluated on clinical data from the target population. If such data do not confirm the assumptions made, these need to be reformulated and the extrapolation concept adjusted accordingly. Thus, an iterative process consisting of a testing-learning loop takes place – similar to the standard, well-established 'learn-confirm' paradigm of clinical drug development [110].

An illustrative example of a testing-learning loop is depicted by voriconazole pediatric development for fungal infections, where PK bridging formed the basis of the extrapolation plan. Based on data from three open-label PK and safety pediatric studies, with a total of 82 patients, a PK model was built in order to define the optimal dosage regimen in children [104]. However, no consensus was reached on the model-based proposed pediatric dosages, and more data were requested by regulators to confirm that the pediatric dosages proposed would lead to the efficacious voriconazole exposures observed in adults. Thus, a further PK study was performed to better characterize the PK and safety of voriconazole in children and to assess the suitability of the proposed dosage regimens [111]. The results revealed that the previously proposed pediatric intravenous dosing regimens - based on population PK modeling - actually

led to lower exposures than those observed in adults, suggesting that such regimens needed to be further modified. Consequently, an integrated population PK M&S exercise was performed based on pooled data from previous studies in children, adolescents and healthy adults [112]. Findings from this M&S analysis facilitated the appropriate adjustment of the previous regimens and thus enabled provision of suitable voriconazole dosage recommendations for children aged 2 to <12 years.

The higher the degree of similarity assumed between source and target population the smaller the set of data that needs to be collected in the latter; consequently, the more the source data are extrapolated, the fewer the data available to evaluate the extrapolation concept. As a result, the risk of false conclusions concerning the efficacy and safety of the new treatment in the target population is potentially increased. In order to mitigate such risks, supporting measures should be implemented. Such measures could embrace both premarketing and post-marketing studies depending, among other factors, on the type of measure(s) to be assessed (e.g. long-term safety and efficacy based on real-world data in a post-marketing setting). Moreover, observational approaches such as case-control studies nested in parallel cohort studies can provide a unique insight for the evaluation of rare AEs in specific pediatric subpopulations [113, 114].

2.5. Discussion

2.5.1. Limitations, future developments and perspectives

An attempt to build a framework for extrapolation from adults to children is represented by the FDA pediatric decision tree [3]. It has provided an algorithm for the selection of the type of study to be conducted in pediatric patients on the basis of ICH E-11 guideline recommendations. However, the decision tree does not take into account PK and PD maturation per se. The assumption of a proportional relationship between the parameters of interest and age does not allow considering nonlinearities that are inherently present in developing children, primarily in neonates and infants. Consequently the lower age cut-offs for age down-extrapolation from adults and children have not been considered.

Second, the PD of the compounds is overlooked and the extrapolation is solely based on the exposure-response relationship assuming this relationship is constant between adults and children and within children regardless of age or other factors of PD variability.

Third, extrapolation is described when the source and target populations differ with respect to age only, whereas extrapolation across diseases (same mechanism of action of the drug) and/or across drug classes (same ‘disease’) is not considered.

Finally, no recommendations can be given when the disease is specific to children, when the medicine is a first-in-class (no known response), or

when clinical efficacy endpoints in adults cannot be assessed in children. For example, juvenile epileptic encephalopathies are specific to children and no data can be borrowed from adults because of more severe symptoms, drug resistance and severe cognitive prognosis. As a consequence, full efficacy studies should be performed in these populations due to unmet medical needs. Given the small numbers of children affected by each type of epilepsy and the high number of different types of epilepsy, resulting in recruitment problems, standard parallel designs are not feasible. Recommendations on different study designs suitable to overcome this obstacle are necessary [41]. Furthermore, some endpoints used in adults cannot be used in children through the whole pediatric age-span because the active participation of young children cannot be obtained. Below a certain age cut-off (which varies according to the endpoint) there is a need to replace the endpoint used in adults by another one specific to children, which in turn jeopardizes a monotonous quantitative description of the endpoint as a function of age, and consequently hamper the extrapolation process. For example pain, assessed in adults and in children above 6 years by auto-evaluation using the Visual Analogue Scale (VAS), can only be quantified in young infants and neonates using hetero-evaluation and age-appropriate and condition-appropriate validated pain scales in the form of lists of items filled in by health professionals. Moreover, although there are a number of measures available for the assessment of muscular strength in Duchene Muscular Dystrophy, regulators tempt to accept only the 6MWDT (primarily because of its reproducibility and clinical relevance), which cannot be reliably quantified in children.

Importantly, extrapolation efforts are limited by current insufficient knowledge on developmental PK-PD relevant to any specific medicine (physiology-receptor pharmacology). For example, the pathophysiology of Gastro-Esophageal Reflux Disease (GERD) in infants (related to lower oesophageal sphincter relaxation) is different from that of older children and adults (acid mediated), likely because of the maturation of the lower esophageal sphincter, which is fully reached at 13 months of age [115]. Misspecification of such a difference may have contributed to the failure of four clinical trials in infants with GERD [116].

Building up a final and comprehensive framework for extrapolation is not trivial and is fraught with issues that have yet to be solved such as: the definition of the impact of extrapolation in drug development and regulatory reviews; how to account for feasibility and ethical restrictions for studies in specific populations; how to formally define and quantify similarities and differences of disease, PK/PD and clinical response to treatment and of safety; how to decide on the quality and quantity of existing data, to weigh the strength of prior information; how to integrate expert judgement in the extrapolation concept; how to validate assumptions made in the extrapolation concept; how to deal with the uncertainty and risks of extrapolation assumptions, especially in the regulatory review and

approval process, and how to collect, analyse and report post-authorisation data to support the extrapolation concept.

The extrapolation process should not be confused with M&S. Although the two are undoubtedly intertwined, extrapolation is primarily a reflection process based upon available knowledge across the preclinical, clinical, biological, pathophysiological, pharmacological and regulatory domains.

In many cases, the extrapolation process should be underpinned by M&S. M&S can be used to quantify the magnitude of differences between the source and target population, as well as contributing to study protocol design and data analysis (Figure 2.3). Descriptive and quantitative information should be collected in order to enable rational model development.

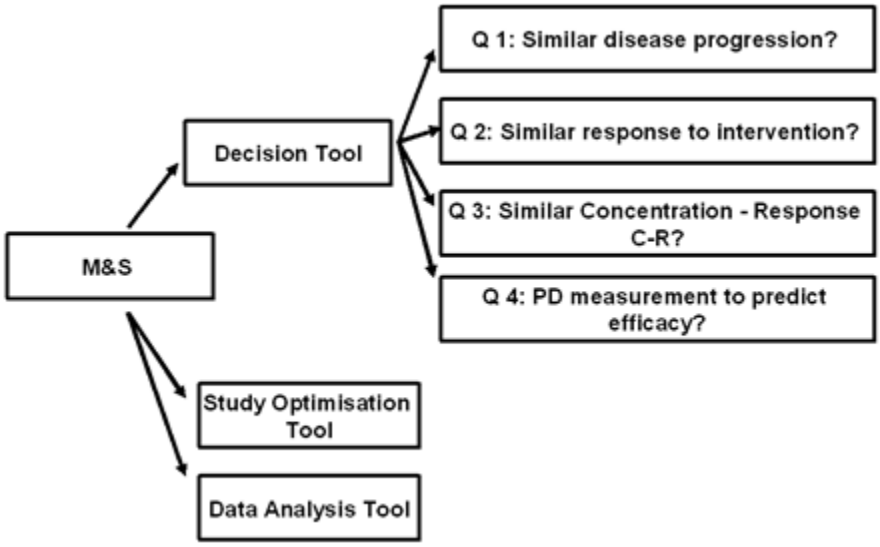


Figure 2.3: Role of modeling and simulation in pediatric drug development (adapted from Manolis and Pons [13]).

Collecting quantitative data to describe the maturational profiles of physiological processes involving drug disposition (PK) and drug effect (PD) as a function of age should allow knowledge gaps to be filled. Although drug disposition is routinely assessed, an alarming lack of PD information in the entire age range of the pediatric population continues to exist. As stated by Holford "testing for PK without testing for PD in children is unethical" [7]. Increasing the current knowledge on developmental pharmacodynamics would boost the building of exposure-response models specific for the pediatric population, a practice that has unfortunately been very limited so far (a 2008 review found that less than 10% of all published models were on PK-PD [9]). Most importantly, such models (including mechanistic PD models) would enable testing of the hypothesis of similar concentration-response relationship, which seems to be over-assumed and appears to be one of the major reasons for trial failures in pediatrics [116]. Surrogate markers to be used both in children and in adults regardless of age/maturation also need to be developed to be

able to bridge age-subsets in which only different clinical endpoints can be used (e.g. bridging above and below 6 years-old children for a similar quantification of pain).

Moreover, collecting quantitative longitudinal data on the natural history of some diseases in children from public database and disease/patients registries may identify evolution patterns in natural history of these diseases that can be helpful in modeling the evolution profile. Not only such models could be used for assessing similarities between source and target population, but also they could characterise diseases specific to children and therefore be used to inform trial design in these peculiar populations.

The extrapolation working group at EMA, including representatives from the Pediatric Committee (PDCO), the Committee for Medicinal Products for Human Use (CHMP), and the Scientific Advice Working Party (SAWP), is drafting a reflection paper on extrapolation of efficacy and safety in medicine development. A pivotal milestone would be the creation of a database of case examples from various therapeutic areas and, eventually, the creation of an algorithm (or set of approaches) for extrapolation and an inventory of methodological rules. Consultation of, and contribution from stakeholders (e.g. pharmacologists, methodologists, industry, academia) and discussion and harmonisation with the FDA shall also be implemented. The ultimate final goal is to develop common regulatory guidance on extrapolation for medicines development.

Extrapolation is a pivotal strategy to avoid unnecessary studies in children, to allocate resources to areas where studies are most needed and to overcome ethical and feasibility problems encountered in executing trials in very small populations such as the pediatric one. Although major regulators have not yet agreed on a shared view of the problem, FDA and EMA are working towards harmonization regarding the use of the extrapolation approach for the evaluation of medicinal products in children (as part of global harmonization of the regulatory pediatric drug development). Such collaboration has already led to positive results in a number of diseases such as diabetes [117], ulcerative colitis [118, 119] and Gaucher disease [120].

Although in this Chapter a number of issues to consider on the use of extrapolation in the evaluation of medicinal products for children were raised, it is likely that the acceptability of extrapolation will remain a case-by-case decision. Medicines developers are encouraged to engage in interaction with regulators as early as possible, in order to optimise pediatric development with respect to adult development.

References

[1] EU. Regulation on medicinal products for paediatric use Official Journal of the European Union. 2006.

[2] E11 I. Guideline on clinical investigation of medicinal products in the paediatric population. 2000.

[3] FDA. Guidance for Industry: Exposure-Response Relationships: Study Design, Data Analysis, and Regulatory Applications. 2003.

[4] FDA. General Clinical Pharmacology Considerations for Paediatric Studies for Drugs and Biological Products. 2014.

[5] FDA. Leveraging Existing Clinical Data for Extrapolation to Paediatric Uses of Medical Devices. 2015.

[6] EMA. Concept paper on extrapolation of efficacy and safety in medicine development 2012.

[7] Holford N. Dosing in children. Clinical pharmacology and therapeutics. 2010;87(3):367-70.

[8] Bellanti F, Della Pasqua O. Modelling and simulation as research tools in paediatric drug development. European journal of clinical pharmacology. 2011;67 Suppl 1:75-86.

[9] Tod M, Jullien V, Pons G. Facilitation of drug evaluation in children by population methods and modelling. Clinical pharmacokinetics. 2008;47(4):231-43.

[10] De Cock RF, Piana C, Krekels EH, Danhof M, Allegaert K, Knibbe CA. The role of population PK-PD modelling in paediatric clinical research. European journal of clinical pharmacology. 2011;67 Suppl 1:5-16.

[11] van Hasselt JG, van Eijkelenburg NK, Beijnen JH, Schellens JH, Huitema AD. Optimizing drug development of anti-cancer drugs in children using modelling and simulation. British journal of clinical pharmacology. 2013;76(1):30-47.

[12] Laer S, Barrett JS, Meibohm B. The in silico child: using simulation to guide paediatric drug development and manage paediatric pharmacotherapy. Journal of clinical pharmacology. 2009;49(8):889-904.

[13] Manolis E, Pons G. Proposals for model-based paediatric medicinal development within the current European Union regulatory framework. British journal of clinical pharmacology. 2009;68(4):493-501.

[14] Marier J, Mangum B, Reid B, Barrett JS. A Modeling and Simulation Framework to Support Global Regulatory Strategies for Pediatrc Drug Development Programs. Therapeutic Innovation & Regulatory Science. 2013;47(5):550-6.

[15] Harnisch L, Shepard T, Pons G, Della Pasqua O. Modeling and simulation as a tool to bridge efficacy and safety data in special populations. CPT: pharmacometrics & systems pharmacology. 2013;2:e28.

[16] Manolis E, Osman TE, Herold R, Koenig F, Tomasi P, Vamvakas S, et al. Role of modeling and simulation in paediatric investigation plans. Paediatric anaesthesia. 2011;21(3):214-21.

[17] Dunne J, Rodriguez WJ, Murphy MD, Beasley BN, Burckart GJ, Filie JD, et al. Extrapolation of adult data and other data in paediatric drug-development programs. Paediatrics. 2011;128(5):e1242-9.

[18] Schoenfeld DA, Hui Z, Finkelstein DM. Bayesian design using adult data to augment paediatric trials. Clinical trials. 2009;6(4):297-304.

[19] EMA. Neurontin Summary of Product Characteristics. 2006.

[20] Appleton R, Fichtner K, LaMoreaux L, Alexander J, Halsall G, Murray G, et al. Gabapentin as add-on therapy in children with refractory partial seizures: a 12-week, multicentre, double-blind, placebo-controlled study. Gabapentin Paediatric Study Group. Epilepsia. 1999;40(8):1147-54.

[21] Mahmood I. Dosing in children: a critical review of the pharmacokinetic allometric scaling and modelling approaches in paediatric drug development and clinical settings. Clinical pharmacokinetics. 2014;53(4):327-46.

[22] Cella M, Knibbe C, Danhof M, Della Pasqua O. What is the right dose for children? British journal of clinical pharmacology. 2010;70(4):597-603.

[23] Mahmood I. Allometric extrapolation of factors VII, VIII and IX clearance in children from adults. Journal of thrombosis and haemostasis : JTH. 2012;10(8):1609-13.

[24] De Cock RFW, Allegaert K, Sherwin CMT, Nielsen EI, De Hoog M, Van Den Anker JN, et al. A Neonatal amikacin covariate model can be used to predict ontogeny of other drugs eliminated through glomerular filtration in neonates. Pharmaceutical Research. 2014;31(3):754-67.

[25] Edginton AN, Schmitt W, Voith B, Willmann S. A mechanistic approach for the scaling of clearance in children. Clinical pharmacokinetics. 2006;45(7):683-704.

[26] Holford N, Heo YA, Anderson B. A pharmacokinetic standard for babies and adults. Journal of pharmaceutical sciences. 2013;102(9):2941-52.

[27] Knibbe CA, Zuideveld KP, Aarts LP, Kuks PF, Danhof M. Allometric relationships between the pharmacokinetics of propofol in rats, children and adults. British journal of clinical pharmacology. 2005;59(6):705-11.

[28] Zhao W, Le Guellec C, Benjamin DK, Jr., Hope WW, Bourgeois T, Watt KM, et al. First Dose in Neonates: Are Juvenile Mice, Adults and In Vitro-In Silico Data Predictive of Neonatal Pharmacokinetics of Fluconazole. Clinical pharmacokinetics. 2014.

[29] Kearns GL, Abdel-Rahman SM, Alander SW, Blowey DL, Leeder JS, Kauffman RE. Developmental pharmacology--drug disposition, action, and therapy in infants and children. The New England journal of medicine. 2003;349(12):1157-67.

[30] Holford NH. A size standard for pharmacokinetics. Clinical pharmacokinetics. 1996;30(5):329-32.

[31] Cella M, Gorter de Vries F, Burger D, Danhof M, Della Pasqua O. A model-based approach to dose selection in early paediatric development. Clinical pharmacology and therapeutics. 2010;87(3):294-302.

[32] Barrett JS, Della Casa Alberighi O, Laer S, Meibohm B. Physiologically based pharmacokinetic (PBPK) modeling in children. Clinical pharmacology and therapeutics. 2012;92(1):40-9.

[33] Leeder JS, Brown JT, Soden SE. Individualizing the use of medications in children: making goldilocks happy. Clinical pharmacology and therapeutics. 2014;96(3):304-6.

[34] Strougo A, Eissing T, Yassen A, Willmann S, Danhof M, Freijer J. First dose in children: physiological insights into pharmacokinetic scaling approaches and their implications in paediatric drug development. Journal of pharmacokinetics and pharmacodynamics. 2012;39(2):195-203.

[35] Maharaj AR, Edginton AN. Physiologically based pharmacokinetic modeling and simulation in paediatric drug development. CPT: pharmacometrics & systems pharmacology. 2014;3:e150.

[36] Johnson TN, Rostami-Hodjegan A, Tucker GT. Prediction of the clearance of eleven drugs and associated variability in neonates, infants and children. Clinical pharmacokinetics. 2006;45(9):931-56.

[37] Salazar DE, Song SH, Shi J, Rohatagi S, Heyrman R, Wada DR, et al. The use of modeling and simulation to guide clinical development of olmesartan medoxomil in paediatric subjects. Clinical pharmacology and therapeutics. 2012;91(2):250-6.

[38] Lala M, Burckart GJ, Takao CM, Pravica V, Momper JD, Gobburu JV. Genetics-based paediatric warfarin dosage regimen derived using pharmacometric bridging. The journal of paediatric pharmacology and therapeutics : JPPT : the official journal of PPAG. 2013;18(3):209-19.

[39] Johnson TN. Modelling approaches to dose estimation in children. British journal of clinical pharmacology. 2005;59(6):663-9.

[40] Hampson LV, Herold R, Posch M, Saperia J, Whitehead A. Bridging the gap: a review of dose investigations in paediatric investigation plans. British journal of clinical pharmacology. 2014;78(4):898-907.

[41] Chiron C, Dulac O, Pons G. Antiepileptic drug development in children: considerations for a revisited strategy. Drugs. 2008;68(1):17-25.

[42] Pellock JM, Carman WJ, Thyagarajan V, Daniels T, Morris DL, D'Cruz O. Efficacy of antiepileptic drugs in adults predicts efficacy in children: a systematic review. Neurology. 2012;79(14):1482-9.

[43] Girgis IG, Nandy P, Nye JS, Ford L, Mohanty S, Wang S, et al. Pharmacokinetic-pharmacodynamic assessment of topiramate dosing regimens for children with epilepsy 2 to <10 years of age. Epilepsia. 2010;51(10):1954-62.

[44] Nedelman JR, Rubin DB, Sheiner LB. Diagnostics for confounding in PK/PD models for oxcarbazepine. Statistics in medicine. 2007;26(2):290-308.

[45] Brochot A, Kakuda TN, Van De Casteele T, Opsomer M, Tomaka FL, Vermeulen A, et al. Model-Based Once-Daily Darunavir/Ritonavir Dosing Recommendations in Paediatric HIV-1-Infected Patients Aged >/=3 to <12 Years. Clinical Pharmacology & Therapeutics: pharmacometrics & systems pharmacology. 2015;4(7):406-14.

[46] Stephenson T. How children's responses to drugs differ from adults. British Journal of Clinical Pharmacology. 2005;59(6):670-3.

[47] Laer S, Elshoff JP, Meibohm B, Weil J, Mir TS, Zhang W, et al. Development of a safe and effective paediatric dosing regimen for sotalol based on population pharmacokinetics and pharmacodynamics in children with supraventricular tachycardia. Journal of the American College of Cardiology. 2005;46(7):1322-30.

[48] FDA. Antidepressant Use in Children, Adolescents, and Adults. http://wwwfdagov/Drugs/DrugSafety/InformationbyDrugClass/UCM096273. 2007.

[49] Kim JH, Scialli AR. Thalidomide: the tragedy of birth defects and the effective treatment of disease. Toxicological sciences : an official journal of the Society of Toxicology. 2011;122(1):1-6.

[50] Koren G, Florescu A, Costei AM, Boskovic R, Moretti ME. Nonsteroidal antiinflammatory drugs during third trimester and the risk of premature closure of the ductus arteriosus: a meta-analysis. The Annals of pharmacotherapy. 2006;40(5):824-9.

[51] Benini D, Fanos V, Cuzzolin L, Tato L. In utero exposure to nonsteroidal anti-inflammatory drugs: neonatal renal failure. Paediatric nephrology. 2004;19(2):232-4.

[52] Cuzzolin L, Dal Cere M, Fanos V. NSAID-induced nephrotoxicity from the fetus to the child. Drug safety. 2001;24(1):9-18.

[53] Tabacova SA, Kimmel CA. Enalapril: pharmacokinetic/dynamic inferences for comparative developmental toxicity. A review. Reproductive toxicology. 2001;15(5):467-78.

[54] Chevalier RL. Mechanisms of fetal and neonatal renal impairment by pharmacologic inhibition of angiotensin. Current medicinal chemistry. 2012;19(27):4572-80.

[55] Sanz EJ, De-las-Cuevas C, Kiuru A, Bate A, Edwards R. Selective serotonin reuptake inhibitors in pregnant women and neonatal withdrawal syndrome: a database analysis. Lancet. 2005;365(9458):482-7.

[56] Stiskal JA, Kulin N, Koren G, Ho T, Ito S. Neonatal paroxetine withdrawal syndrome. Archives of disease in childhood Fetal and neonatal edition. 2001;84(2):F134-5.

[57] Cleary BJ, Donnelly J, Strawbridge J, Gallagher PJ, Fahey T, Clarke M, et al. Methadone dose and neonatal abstinence syndrome-systematic review and meta-analysis. Addiction. 2010;105(12):2071-84.

[58] Cleary BJ, Donnelly JM, Strawbridge JD, Gallagher PJ, Fahey T, White MJ, et al. Methadone and perinatal outcomes: a retrospective cohort study. American journal of obstetrics and gynecology. 2011;204(2):139 e1-9.

[59] Meidahl Petersen K, Jimenez-Solem E, Andersen JT, Petersen M, Brodbaek K, Kober L, et al. beta-Blocker treatment during pregnancy and adverse pregnancy outcomes: a nationwide population-based cohort study. BMJ open. 2012;2(4).

[60] Davis R, Andrade S, Rubanowice D, McPhillips H, Boudreau D, Raebel M, et al. C-A4-03: Risks to the Newborn Associated With In-Utero Exposure to Beta-Blockers and Calcium-Channel Blockers. Clinical Medicine & Research. 2010;8(1):57-.

[61] Bleyer WA, Skinner AL. Fatal Neonatal Hemorrhage After Maternal Anticonvulsant Therapy. Jama. 1976;235(6):626.

[62] Kazmin A, Wong RC, Sermer M, Koren G. Antiepileptic drugs in pregnancy and hemorrhagic disease of the newborn: an update. Canadian family physician Medecin de famille canadien. 2010;56(12):1291-2.

[63] Mountain KR, Hirsh J, Gallus AS. Neonatal coagulation defect due to anticonvulsant drug treatment in pregnancy. Lancet. 1970;1(7641):265-8.

[64] Meador KJ. Neurodevelopmental effects of antiepileptic drugs. Current neurology and neuroscience reports. 2002;2(4):373-8.

[65] Dean JC, Hailey H, Moore SJ, Lloyd DJ, Turnpenny PD, Little J. Long term health and neurodevelopment in children exposed to antiepileptic drugs before birth. Journal of medical genetics. 2002;39(4):251-9.

[66] Gelisse P, Genton P, Kuate C, Pesenti A, Baldy-Moulinier M, Crespel A. Worsening of seizures by oxcarbazepine in juvenile idiopathic generalized epilepsies. Epilepsia. 2004;45(10):1282-6.

[67] Genton P, Gelisse P, Thomas P, Dravet C. Do carbamazepine and phenytoin aggravate juvenile myoclonic epilepsy? Neurology. 2000;55(8):1106-9.

[68] Horn CS, Ater SB, Hurst DL. Carbamazepine-exacerbated epilepsy in children and adolescents. Paediatric neurology. 1986;2(6):340-5.

[69] Allen DB. Growth suppression by glucocorticoid therapy. Endocrinology and metabolism clinics of North America. 1996;25(3):699-717.

[70] Mazziotti G, Giustina A. Glucocorticoids and the regulation of growth hormone secretion. Nature reviews Endocrinology. 2013;9(5):265-76.

[71] Bray RJ. Propofol infusion syndrome in children. Paediatric anaesthesia. 1998;8(6):491-9.

[72] Hanna JP, Ramundo ML. Rhabdomyolysis and hypoxia associated with prolonged propofol infusion in children. Neurology. 1998;50(1):301-3.

[73] Patz A. The role of oxygen in retrolental fibroplasia. Paediatrics. 1957;19(3):504-24.

[74] Cooper WO, Griffin MR, Arbogast P, Hickson GB, Gautam S, Ray WA. Very early exposure to erythromycin and infantile hypertrophic pyloric stenosis. Archives of paediatrics & adolescent medicine. 2002;156(7):647-50.

[75] Hauben M, Amsden GW. The association of erythromycin and infantile hypertrophic pyloric stenosis: causal or coincidental? Drug safety. 2002;25(13):929-42.

[76] Melnick S, Cole P, Anderson D, Herbst A. Rates and risks of diethylstilbestrol-related clear-cell adenocarcinoma of the vagina and cervix. An update. The New England journal of medicine. 1987;316(9):514-6.

[77] Witkop CJ, Jr., Wolf RO. Hypoplasia and Intrinsic Staining of Enamel Following Tetracycline Therapy. Jama. 1963;185:1008-11.

[78] Grenier MA, Lipshultz SE. Epidemiology of anthracycline cardiotoxicity in children and adults. Seminars in oncology. 1998;25(4 Suppl 10):72-85.

[79] Iarussi D, Indolfi P, Casale F, Martino V, Di Tullio MT, Calabro R. Anthracycline-induced cardiotoxicity in children with cancer: strategies for prevention and management. Paediatric drugs. 2005;7(2):67-76.

[80] Cho WK, Lee JW, Chung NG, Jung MH, Cho B, Suh BK, et al. Primary ovarian dysfunction after hematopoietic stem cell transplantation during childhood: busulfan-based conditioning is a major concern. Journal of paediatric endocrinology & metabolism : JPEM. 2011;24(11-12):1031-5.

[81] Teinturier C, Hartmann O, Valteau-Couanet D, Benhamou E, Bougneres PF. Ovarian function after autologous bone marrow transplantation in childhood: high-dose busulfan is a major cause of ovarian failure. Bone marrow transplantation. 1998;22(10):989-94.

[82] Brougham MF, Kelnar CJ, Sharpe RM, Wallace WH. Male fertility following childhood cancer: current concepts and future therapies. Asian journal of andrology. 2003;5(4):325-37.

[83] de Araujo Filho GM, Pascalicchio TF, Lin K, Sousa PS, Yacubian EM. Neuropsychiatric profiles of patients with juvenile myoclonic epilepsy treated with valproate or topiramate. Epilepsy & behavior : E&B. 2006;8(3):606-9.

[84] Kang HC, Eun BL, Wu Lee C, Ku Moon H, Kim JS, Wook Kim D, et al. The effects on cognitive function and behavioral problems of topiramate compared to carbamazepine as monotherapy for children with benign rolandic epilepsy. Epilepsia. 2007;48(9):1716-23.

[85] Mulberg AE, Silber SA, J.N. VdA. Paediatric Drug Development: John Wiley & Sons; 2009.

[86] EMA. Guideline on the need for non-clinical testing in juvenile animals on human pharmaceuticals for paediatric indications 2008.

[87] FDA. Nonclinical Safety Evaluation of Paediatric Drug Products. 2006.

[88] Soellner L, Olejniczak K. The need for juvenile animal studies--a critical review. Regulatory toxicology and pharmacology : RTP. 2013;65(1):87-99.

[89] Liao CM, Lin TL, Chen SC. A Weibull-PBPK model for assessing risk of arsenic-induced skin lesions in children. The Science of the total environment. 2008;392(2-3):203-17.

[90] Hope WW, Drusano GL. Antifungal pharmacokinetics and pharmacodynamics: bridging from the bench to bedside. Clinical microbiology and infection : the official publication of the European Society of Clinical Microbiology and Infectious Diseases. 2009;15(7):602-12.

[91] Hope WW, Mickiene D, Petraitis V, Petraitiene R, Kelaher AM, Hughes JE, et al. The pharmacokinetics and pharmacodynamics of micafungin in experimental hematogenous Candida meningoencephalitis: implications for echinocandin therapy in neonates. The Journal of infectious diseases. 2008;197(1):163-71.

[92] EMA. Vancomycin PIP decision. 2013.

[93] Erkeller-Yuksel FM, Deneys V, Yuksel B, Hannet I, Hulstaert F, Hamilton C, et al. Age-related changes in human blood lymphocyte subpopulations. The Journal of paediatrics. 1992;120(2 Pt 1):216-22.

[94] Hannet I, Erkeller-Yuksel F, Lydyard P, Deneys V, DeBruyere M. Developmental and maturational changes in human blood lymphocyte subpopulations. Immunology today. 1992;13(6):215, 8.

[95] Clapp DW. Developmental regulation of the immune system. Seminars in perinatology. 2006;30(2):69-72.

[96] Mulla H. Understanding developmental pharmacodynamics: importance for drug development and clinical practice. Paediatric drugs. 2010;12(4):223-33.

[97] Jadhav PR, Zhang J, Gobburu JV. Leveraging prior quantitative knowledge in guiding paediatric drug development: a case study. Pharmaceutical statistics. 2009;8(3):216-24.

[98] Shahinfar S, Cano F, Soffer BA, Ahmed T, Santoro EP, Zhang Z, et al. A double-blind, dose-response study of losartan in hypertensive children. American journal of hypertension. 2005;18(2 Pt 1):183-90.

[99] Webb NJ, Lam C, Loeys T, Shahinfar S, Strehlau J, Wells TG, et al. Randomized, double-blind, controlled study of losartan in children with proteinuria. Clinical journal of the American Society of Nephrology : CJASN. 2010;5(3):417-24.

[100] Ahmed U, Spyridis N, Wong IC, Sharland M, Long PF, improving Children's Antibiotic Prescribing UKRN. Dosing of oral penicillins in children: is big child=half an adult, small child=half a big child, baby=half a small child still the best we can do? Bmj. 2011;343:d7803.

[101] Rees VE, Bulitta JB, Nation RL, Tsuji BT, Sorgel F, Landersdorfer CB. Shape does matter: short high-concentration exposure minimizes resistance emergence for fluoroquinolones in Pseudomonas aeruginosa. The Journal of antimicrobial chemotherapy. 2015;70(3):818-26.

[102] Barker CI, Germovsek E, Hoare RL, Lestner JM, Lewis J, Standing JF. Pharmacokinetic/pharmacodynamic modelling approaches in paediatric infectious diseases and immunology. Advanced drug delivery reviews. 2014;73:127-39.

[103] Walsh TJ, Karlsson MO, Driscoll T, Arguedas AG, Adamson P, Saez-Llorens X, et al. Pharmacokinetics and safety of intravenous voriconazole in children after single- or multiple-dose administration. Antimicrobial agents and chemotherapy. 2004;48(6):2166-72.

[104] Karlsson MO, Lutsar I, Milligan PA. Population pharmacokinetic analysis of voriconazole plasma concentration data from paediatric studies. Antimicrobial agents and chemotherapy. 2009;53(3):935-44.

[105] Kent A, Kortsalioudaki C, Monahan IM, Bielicki J, Planche TD, Heath PT, et al., editors. Increasing MIC is associated with higher mortality in gentamicin treated neonatal gram negative infections. Paediatric Infectious Diseases Meeting; 2015.

[106] Yee DL, O'Brien SH, Young G. Pharmacokinetics and pharmacodynamics of anticoagulants in paediatric patients. Clinical pharmacokinetics. 2013;52(11):967-80.

[107] Hamberg AK, Friberg LE, Hanseus K, Ekman-Joelsson BM, Sunnegardh J, Jonzon A, et al. Warfarin dose prediction in children using pharmacometric bridging--comparison with published pharmacogenetic dosing algorithms. European journal of clinical pharmacology. 2013;69(6):1275-83.

[108] Chanu P, Gao X, Smith M, Bruno R, Harnisch L, editors. A dose selection rationale based on hemodynamics for sildenafil in paediatric patients with pulmonary arterial hypertension (PAH). PAGE 2011; Athens.

[109] Geiger R, Strasak A, Treml B, Gasser K, Kleinsasser A, Fischer V, et al. Six-minute walk test in children and adolescents. The Journal of paediatrics. 2007;150(4):395-9, 9 e1-2.

[110] Sheiner LB. Learning versus confirming in clinical drug development. Clinical pharmacology and therapeutics. 1997;61(3):275-91.

[111] Driscoll TA, Yu LC, Frangoul H, Krance RA, Nemecek E, Blumer J, et al. Comparison of pharmacokinetics and safety of voriconazole intravenous-to-oral switch in immunocompromised children and healthy adults. Antimicrobial agents and chemotherapy. 2011;55(12):5770-9.

[112] Friberg LE, Ravva P, Karlsson MO, Liu P. Integrated population pharmacokinetic analysis of voriconazole in children, adolescents, and adults. Antimicrobial agents and chemotherapy. 2012;56(6):3032-42.

[113] Mikaeloff Y, Kezouh A, Suissa S. Nonsteroidal anti-inflammatory drug use and the risk of severe skin and soft tissue complications in patients with varicella or zoster disease. British journal of clinical pharmacology. 2008;65(2):203-9.

[114] Schlienger RG, Jick SS, Meier CR. Inhaled corticosteroids and the risk of fractures in children and adolescents. Paediatrics. 2004;114(2):469-73.

[115] Wang J, Avant D, Green D, Seo S, Fisher J, Mulberg AE, et al. A survey of neonatal pharmacokinetic and pharmacodynamic studies in paediatric drug development. Clinical pharmacology and therapeutics. 2015.

[116] Momper JD, Mulugeta Y, Burckart GJ. Failed paediatric drug development trials. Clinical pharmacology and therapeutics. 2015.

[117] Karres J, Pratt V, Guettier JM, Temeck J, Tamborlane WV, Dunger D, et al. Joining forces: a call for greater collaboration to study new medicines in children and adolescents with type 2 diabetes. Diabetes care. 2014;37(10):2665-7.

[118] Sun H, Vesely R, Nelson RM, Taminiau J, Szitanyi P, Isaac M, et al. Steps toward harmonization for clinical development of medicines in paediatric ulcerative colitis-a global scientific discussion, part 2: data extrapolation, trial design, and pharmacokinetics. Journal of paediatric gastroenterology and nutrition. 2014;58(6):684-8.

[119] Sun H, Vesely R, Taminiau J, Szitanyi P, Papadopoulos EJ, Isaac M, et al. Steps toward harmonization for clinical development of medicines in paediatric ulcerative colitis-a global scientific discussion, part 1: efficacy endpoints and disease outcome assessments. Journal of paediatric gastroenterology and nutrition. 2014;58(6):679-83.

[120] EMA. Gaucher disease: A strategic collaborative approach from EMA and FDA. (EMA/44410/2014). 2014.

Chapter 3

Model-based dose selection in a pediatric study of Febuxostat for the prevention of tumor lysis syndrome

3.1. Introduction

The role of modeling and simulation (M&S) as a tool to bridge efficacy and safety data from adults to children and adolescents in support of the extrapolation approach has been introduced in Chapter 2. Indeed, M&S can be used to quantify the magnitude of differences between the pediatric and adult population, as well as to optimize study protocol design and data analysis. The current Chapter presents a case study on the use of a model-based approach for dose selection in a pediatric trial of Febuxostat (FBX) for tumor lysis syndrome (TLS) prevention. A comparison of this approach with the traditional linear scaling of the dose with body weight is also presented.

Tumor lysis syndrome represents a critical and possibly fatal complication resulting from the rapid lysis of large amounts of tumor cells, observed most often after initial treatment with chemotherapy [1]. Cancer cells lysis is followed by the release of potassium, phosphorus and nucleic acids, which are metabolized into hypoxanthine, then in xanthine, and finally in uric acid, which can induce acute kidney injury. In both adults and children, TLS can ultimately results in nausea, vomiting, cardiac dysrhythmias, seizures and possibly sudden death. TLS diagnostic criteria are similar between the pediatric and adult population, with the exception of some threshold laboratory values which take into account the physiological age-related differences [2].

FBX is a 2-arylthiazole derivative that achieves its therapeutic effect of decreasing serum uric acid activity by selective inhibition of xanthine oxidase. In adults, the pharmacokinetic (PK) profile of FBX is well

characterized. After oral administration, 84% of the dose is absorbed, with maximum plasma concentrations (Cmax) observed between 1 and 1.5 h. Mean values for plasma Cmax and area under the concentration-time curve (AUC) increase linearly with dose; however, a greater-than proportional increase in both parameters has been observed for doses higher than 120 mg. FBX is primarily biotransformed by UGT1A1, 1A8, and 1A9 and by CYP1A1, CYP1A2, CYP2C8 or CYP2C9. Approximately 49% of the FBX dose is excreted renally [3].

FBX has been recently approved in Europe at the dose of 120 mg QD for the prevention and treatment of hyperuricemia in adult patients undergoing chemotherapy for hematologic malignancies at intermediate to high TLS risk. No clinical data are available on the use of FBX in pediatric population. On the grounds of the unmet clinical need, the subset of the pediatric population suffering from hematologic malignancies at intermediate to high risk of TLS which would most benefit from the availability of FBX is the one aged 6 to 17 years (hereafter called target population). Considering TLS similarities between the target and adult population in terms of pathophysiology, clinical manifestations, diagnostic criteria, grading system for disease severity, as well as the equality in FBX mechanism of action, it is reasonable to assume that children aged 6 to 17 years old will respond to FBX similarly to adults, showing a similar pharmacokinetic-pharmacodynamic (PK-PD) relationship. Moreover, since FBX biotransformation and elimination pathways are fully mature at 6 years of age [4-7], it is expected that the PK between the two populations can be scaled based only upon body weight. Therefore, FBX efficacy and safety in the target population can be extrapolated from that observed in adults and a full clinical development plan can be skipped in favor of a more flexible phase I/II PK-PD study, followed by a M&S analysis for dose optimization if needed [8]. This will allow to confirm the hypotheses on the similarities between the target and adult population, in particular those related to the PK and PK-PD of FBX in children $\geq$ 6 years of age, while minimizing the number of pediatric patients exposed to clinical investigations.

3.2. Methods

Dose selection for the pediatric PK-PD study: Linear VS Allometric scaling

The common assumption underlying linear and allometric scaling of adult doses to pediatrics is that differences in elimination capacity (i.e. clearance) between the two populations are completely captured by differences in body size, with body weight typically used as a surrogate.

Clearance scaling by body weight can be formulated as in equation (1), where a=1 for linear scaling and a=0.75 for allometric scaling.

$$CL_{pediatrics} = CL_{adults}\left(\frac{BW_{pediatrics}}{BW_{adults}}\right)^{a} (1)$$

The 0.75 exponent comes from Kleiber's law [9], which states that animal's basal metabolic rate scales to the ¾ power of the animal's mass. A naive interpretation of such law is that small organisms respire more per unit of weight than bigger organisms do because a larger fraction of their body mass consists of structural elements (e.g. bone) rather than reserves (e.g. blood vessels). Therefore, as mass increases, the overall metabolic rate does not increase in direct proportion.

With respect to the FBX pediatric trial, the dose selection has been primarily driven by safety considerations. For linear scaling the upper bound for the doses to be tested in the target population was the highest safe dose in mg/kg administered to adults in phase III trials (FBX 40, 80, 120 and 240 mg was studied in 2690 patients >18 years old in phase III studies C02-009, C02-010 and F-GT06-153, see Table 3.1), that is, 3.81 mg/kg. On the other hand, the goal of allometric scaling is to ensure that FBX exposures (Cmax and AUC) in pediatric patients do not exceed the highest exposure observed in adults at the safe dose of 240 mg. For both approaches, an additional efficacy criteria is also applied for final dose selection, in order to guarantee that FBX levels in the target population are not lower than those observed at the efficacious dose in adults (i.e. 120 mg).

Table 3.1. Summary of demographic characteristics of Febuxostat adult clinical trials used as historical data.

Study C02-009						
Variable	**Placebo (N=134)**	**Febuxostat 80 mg QD (N=267)**	**Febuxostat 120 mg QD (N=267)**	**Febuxostat 240 mg QD (N=267)**	**Allopurinol 300/100 mg QD (N=268)**	**All subjects (N=1072)**
Gender n (%)						
Female	11 (8%)	16 (6%)	13 (5%)	8 (6%)	19 (7%)	67 (6%)
Male	123 (92%)	251 (94%)	256 (95%)	126 (94%)	249 (93%)	1005 (94%)
Age (years)						
Mean (SD)	51.5 (12.18)	50.6 (12.24)	51.2 (11.57)	54.3 (12.83)	51.8 (12.25)	51.6 (12.17)
Weight (pounds)						
Mean (SD)	215.2 (43.05)	227.6 (43.77)	230.3 (48.70)	227.2 (49.04)	224.1 (43.01)	225.8 (45.61)

Study C02-010				
Variable	**Febuxostat 80 mg QD (N=256)**	**Febuxostat 120 mg QD (N=251)**	**Allopurinol 300 mg QD (N=253)**	**All subjects (N=760)**
Gender n (%)				
Female	13 (5%)	8 (3%)	10 (4%)	31 (4%)
Male	243 (95%)	243 (97%)	243 (96%)	729 (96%)
Age (years)				
Mean (SD)	51.8 (11.69)	52.0 (12.12)	51.6 (12.63)	51.8 (12.13)
Weight (pounds)				
Mean (SD)	224.7 (44.02)	223.9 (44.63)	224.8 (45.14)	224.5 (44.54)

Study F-GT06-153				
Variable	**Febuxostat 40 mg QD (N=757)**	**Febuxostat 80 mg QD (N=756)**	**Allopurinol 300/200 mg QD (N=756)**	**All subjects (N=2269)**
Gender n (%)				
Female	35 (5%)	46 (6%)	47 (6%)	128 (5%)
Male	722 (95%)	710 (94%)	709 (94%)	2141 (95%)
Age (years)				
Mean (SD)	52.5 (11.68)	53.0 (11.79)	52.9 (11.73)	52.8 (11.73)
Weight (pounds)				
Mean (SD)	229.9 (48.58)	227.3 (47.7)	225.5 (46.09)	227.6 (47.48)

Study TMX-99-001			
Variable	**Febuxostat 120 mg QD (N=10)**	**Febuxostat 240 mg QD (N=10)**	**All subjects (N=154)**
Gender n (%)			
Female	5 (50%)	5 (50%)	67 (44%)
Male	5 (50%)	5 (50%)	87 (56%)
Age (years)			
Mean (SD)	33.8 (12.09)	29.8 (10.54)	32.5 (9.96)
Weight (pounds)			
Mean (SD)	174.2 (23.32)	173.0 (36.97)	173.2 (32.92)

Based on Phase I study TMX-99-001 (Table 3.1), target adult AUC and Cmax values at 120 mg were 11960 ng*hr/mL and 5308 ng/mL, respectively, while AUC and Cmax at 240 mg in adults were 34976 ng*hr/mL and 11263 ng/mL, respectively. Pediatric exposures were simulated using the in-house adult PK model of Figure 3.1, whose parameters values are reported in Table 3.2.

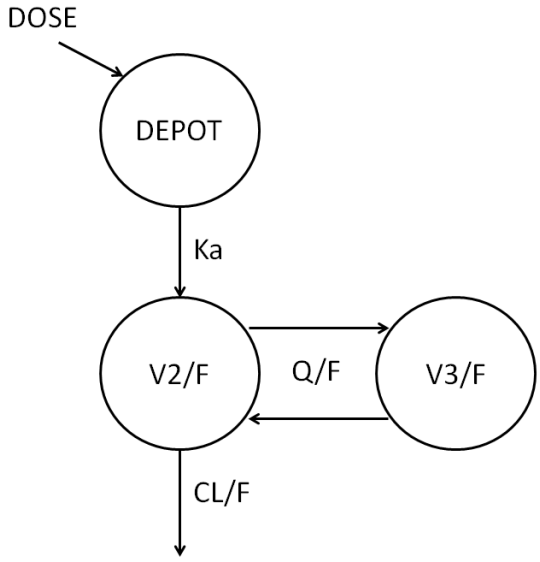


Figure 3.1. Compartmental pharmacokinetic (PK) model of Febuxostat in adults used for pediatric PK simulations. Ka: absorption rate constant; V2/F: apparent volume of central compartment; V3/F: apparent volume of peripheral compartment; Q/F: apparent inter-compartmental clearance; CL/F: apparent systemic clearance. Parameter values are reported in Table 3.2.

Allometric equations based on weight were used to scale clearances (CL/F and Q/F) and volumes (V2/F and V3/F) with an exponent of 0.75 and 1, respectively. The target population was simulated in terms of body weight and age: a total of 1000 virtual pediatric patients per dose-group were randomly generated using weight-for-age tables from World Health Organization statistics [10]. In agreement with the patient population to be enrolled in the study, age was assumed to be uniformly distributed in the interval 6-17 years, while weight, within each year of age, was assumed to follow a log-normal distribution. NONMEM version 7.2.0 [11] and R software version 3.0.1 [12] were used for PK and covariate simulations, respectively.

Two age groups were considered: children (6-11 years) and adolescents (12-17 years); for each age group the candidate doses were 40 mg, 60 mg, 80 mg and 120 mg. The 5^{th} percentile of weights was taken into account in order to obtained a conservative dose-selection in terms of safety.

Table 3.2. Population pharmacokinetic parameter values for Febuxostat in adult patients (in-house data).

Parameter	Value	Between-subject variability (CV%)
Ka [1/hr]	8.66	230
Absorption lag-time [hr]	0.465	-
V2/F [L]	49.4	54
V3/F [L]	22.3	-
CL/F [L/hr]	7.76	30
Q/F [L/hr]	2.74	-

3.3. Results

3.3.1. Linear scaling

Figure 3.2 shows the relationship between the body weight-normalized dose and age for the four candidate doses. At a given age, the doses in mg which lie within the grey shaded area exceed the safe threshold of 3.81 mg/kg and are therefore not allowed by linear scaling.

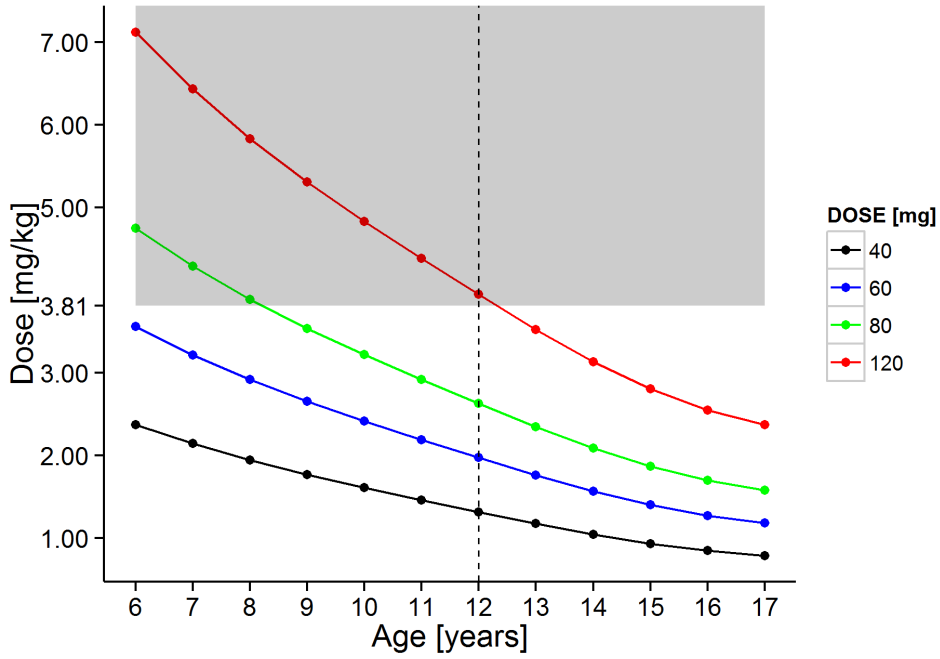


Figure 3.2. Dose (mg/kg)-age relationship obtained with linear scaling for a 40 mg (black line), 60 mg (blue line), 80 mg (green line) and 120 mg (red line) dose. The gray shaded area depicts the subset of doses not allowed by linear scaling because greater than the safety threshold of 3.81 mg/kg. The dotted vertical line splits the x-axis in children (6-11 years) and adolescents (12-17 years).

The suitability of the candidate doses in each age based on linear scaling is summarized in Table 3.3. It can be seen that linear scaling allows the doses of 40 mg and 60 mg to be administered to the entire target population, whereas the doses of 80 mg and 120 mg are admitted only for children older than 8 and 12 years of age, respectively.

Table 3.3. Suitability of the candidate doses in each age based on linear scaling: V = dose does not exceed the safety threshold of 3.81 mg/kg; X = dose exceeds the safety threshold.

Age (years)	Age group	5th percentile of weights (kg)	40 mg	60 mg	80 mg	120 mg
6	Children	16.86	V	V	X	X
7	Children	18.66	V	V	X	X
8	Children	20.58	V	V	X	X
9	Children	22.62	V	V	V	X
10	Children	24.85	V	V	V	X
11	Children	27.39	V	V	V	X
12	Adolescents	30.41	V	V	V	X
13	Adolescents	34.06	V	V	V	V
14	Adolescents	38.29	V	V	V	V
15	Adolescents	42.83	V	V	V	V
16	Adolescents	47.15	V	V	V	V
17	Adolescents	50.68	V	V	V	V

3.3.2. Allometric scaling

The results of the simulations with the weight-adjusted adult PK model are depicted in Figure 3.3 and Figure 3.4 for AUC and Cmax, respectively. Two different boxplots are reported for each dose- and age-group: one represents the 5th percentile of weights of the target population and the other one represents the remainder.

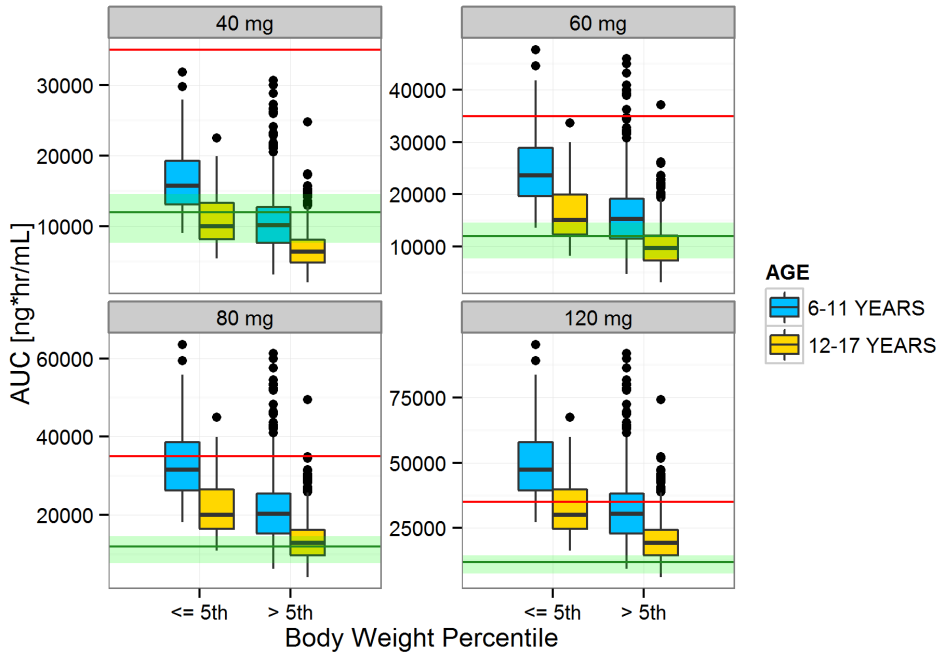


Figure 3.3. Boxplots of Febuxostat Area Under the Curve (AUC) obtained from 1000 virtual pediatric patients aged 6-11 years (blue) and 12-17 years (yellow) for doses of 40 mg (top-left panel), 60 mg (top-right panel), 80 mg (bottom-left panel) and 120 mg (bottom-right panel). The boxplots of the 5th percentile and of the rest of the target population are visualized for each dose-group. The green and red solid lines represent the mean AUC in adults at 120 mg and 240 mg, respectively, while the green shaded area covers the range of AUCs observed in adults at 120 mg.

Blue boxplots depicted in Figures 3.3 and 3.4 indicate that children dosed up to 60 mg would achieve the target AUC and Cmax in adults (green shaded area) and, at the same time, would not exceed the highest mean exposure at 240 mg observed to be safe in adults (red line). Moreover, less than 50% of children in the 5th percentile would be overexposed to FBX at the dose of 80 mg (bottom-left panel of Figure 3.3), whereas more than 75% would exceed the safety threshold for the 120 mg dose (bottom-right panel of Figure 3.3).

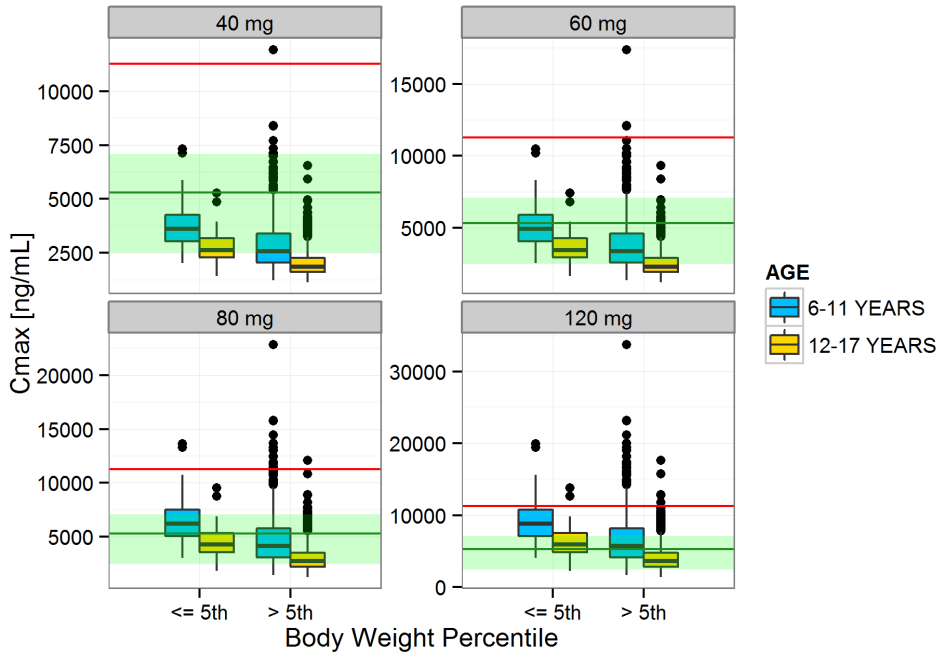


Figure 3.4. Boxplots of Febuxostat peak concentrations (Cmax) obtained from 1000 virtual pediatric patients aged 6-11 years (blue) and 12-17 years (yellow) for doses of 40 (top-left panel), 60 mg (top-right panel), 80 mg (bottom-left panel) and 120 mg (bottom-right panel). The boxplots of the 5th percentile and of the rest of the target population are visualized for each dose-group. The green and red solid lines represent the mean Cmax in adults at 120 mg and 240 mg, respectively, while the green shaded area covers the range of Cmaxs observed in adults at 120 mg.

With respect to the adolescent subset of the target population, yellow boxplots of upper panels in Figures 3.3 and 3.4 show that the doses of 40 mg and 60 mg are not sufficiently high to guarantee the attainment of the efficacious exposure obtained at 120 mg in adults both in terms of AUC and Cmax. On the contrary, doses of 80 mg and 120 mg reach the target exposure and, apart from a small percentage of the 5th percentile of the adolescent population, do not exceed the safety threshold.

Overall, the suitability of the tested doses in each age-group based on allometric scaling is summarized in Table 3.4.

Table 3.4. Suitability of the candidate doses in each age-group based on allometric scaling: V = dose allowed by allometric scaling; X = dose not allowed by allometric scaling because of safety reasons. XX = dose not allowed by allometric scaling because of efficacy reasons.

Age group	40 mg	60 mg	80 mg	120 mg
Children	V	V	V	X
Adolescents	XX	XX	V	V

3.4. Discussion

The present Chapter dealt with two alternative approaches - linear versus allometric - for dose selection in a pediatric study of FBX, a xanthine-oxidase inhibitor approved in the adult population for the prevention and treatment of hyperuricemia in adult patients undergoing chemotherapy for hematologic malignancies at intermediate to high TLS risk.

Several methods are available for dose/clearance scaling of a drug from a source population to pediatrics [13]. In principle, regardless of the method selected, it is important to have a clear understanding of the differences in drug absorption, distribution, metabolism and excretion between the two populations, in relation to the various aspects of developmental pharmacology [14]. Because pathways of FBX biotransformation and elimination are fully mature at 6 years of age, PK scaling based solely on body weight can be considered reasonably accurate in this context.

Linear dose/clearance scaling on a mg/kg basis may lead to subtherapeutic drug exposures [15, 16]. Conversely, allometric approaches, besides being endorsed by US and EU regulators for dose selection in pediatric trials [18], offer a sound basis for scaling doses from older to younger patients since they are supported by a well-established theory [17]. Nonetheless, the results obtained in the current analysis show that, for the age-range and dose-range under consideration, the two methods do not remarkably disagree when it comes to dose selection for the pediatric trial. Considering the PK-PD nature of the trial, two different doses per age-group will be tested in order to adequately characterize the PK and PK-PD of FBX in the target population; for safety reasons, doses will be administered sequentially within each age-group. Linear and allometric scaling agree in that adolescents will receive 80 mg and 120 mg. With respect to children, allometric scaling would allow the dose of 80 mg for the entire subset of the pediatric population, whilst linear scaling would not for children younger than 8 years old. However, considering that children from 6 to 11 years old belong to the same age-group and in order to guarantee an additional margin of safety, the two selected doses in children are 40 mg and 60 mg to be tested sequentially (from the lower to the high dose).

Higher doses allowed by allometric scaling compared to linear scaling were expected: Figure 3.5 shows that clearance predictions based on allometric scaling (orange line) are higher than those based on linear scaling (blue line). In particular, the dotted line of Figure 3.5, which represents the ratio between linear and allometric scaling prediction, shows that the discrepancy between the two methods decreases with increasing body weights.

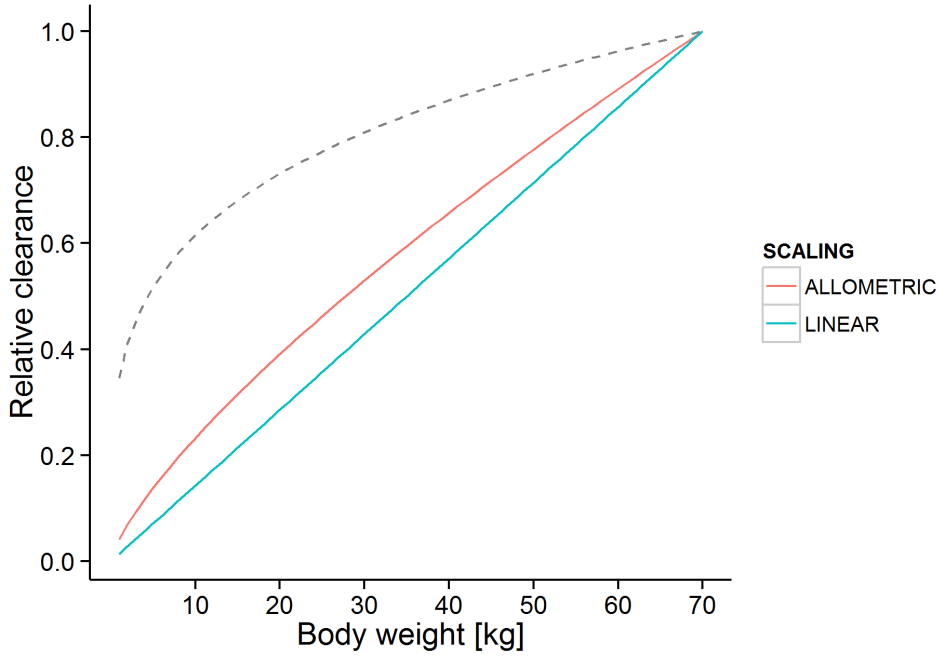


Figure 3.5. Clearance prediction (relative to the clearance of a 70 kg adult) of allometric (orange) versus linear (blue) scaling. The dotted line is the ratio of the linear to allometric predictions (adapted from Holford, 1996 [15]).

It should be pointed out that allometry can only capture differences in drug disposition due to size variation, without addressing the influence of developmental changes related to organs maturation. Consequently, clearance predictions depicted in Figure 3.5 are not reliable for weight-ranges related to ages during which drug-specific elimination pathways are still under development.

Dose selection based on matching adult efficacious exposure is built upon the implicit assumption that similar plasma exposures would translate into similar clinical responses, i.e. that the PK-PD relationship is similar in the target and adult population. Accordingly, the primary objective of the pediatric PK-PD study is to assess and compare the PK of FBX in the target population and adults at intermediate to high risk of TLS. In addition, the secondary objective of the trial is to evaluate and compare the safety and the PK-PD relationship of FBX between pediatric and adult patients in order to validate the extrapolation concept. For this aim, an indirect response PK-PD model linking FBX plasma concentrations to serum uric acid (sUA) levels will be employed (Figure 3.6). Such a model possesses a sound mechanistic basis: FBX lowers sUA levels by inhibiting xanthine-oxidase activity; consequently, FBX exerts its effect by decreasing the zero-order rate constant (Kin) which quantifies the production of sUA by the human body.

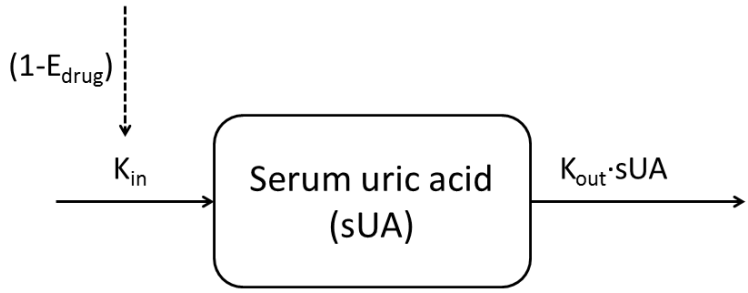


Figure 3.6. Schematic representation of the indirect pharmacokinetic-pharmacodynamic model that will be used to analyze serum uric acid (sUA) data in the Febuxostat pediatric study. K_{in} is the zero-order rate constant of sUA production, while K_{out} is the first-order rate constant of sUA elimination. According to Febuxostat mechanism of action, the drug exerts its effect by inhibiting sUA production. E_{drug} represents the drug effect and can be a linear or sigmoidal E_{max} function of Febuxostat plasma concentration.

Should the tested doses lead to pediatric exposures which either exceed or do not achieve adult ones, the PK-PD analysis together with the collected safety data will support the findings from the PK analysis, and confirm that potential deviations in PK will not be clinically relevant. If required, dose recommendations will be provided solely by means of PK simulation of alternative dosages in the target population, without the need to run further clinical trials and thus to expose children to unnecessary clinical investigations.

In conclusion, the present Chapter illustrated an application of PK-PD M&S in the development of pediatric medicines based on the extrapolation of adult data, emphasizing how a model-based approach can support the selection of a safe, efficacious and informative dose regimen to be tested in early clinical trials in children.

References

[1] Coiffier B, Altman A, Pui CH, Younes A, Cairo MS. Guidelines for the management of paediatric and adult tumour lysis syndrome: an evidence-based review. J Clin Oncol 2008;26(16):2767-2778.

[2] Cairo MS, Bishop M. Tumour lysis syndrome: new therapeutic strategies and classification. Br J Haematol 2004;127(1):3-11.

[3] Adenuric European Public Assessment Report (http://www.ema.europa.eu/docs/ en_GB/document_library/EPAR_-_Product_Information/human/000777/ WC500021812.pdf).

[4] Alcorn J, McNamara PJ. Pharmacokinetics in the newborn. Adv Drug Deliv Rev. 2003;55(5):667-86.

[5] Salem F, Johnson TN, Abduljalil K, Tucker GT, Rostami-Hodjegan A. A re-evaluation and validation of ontogeny functions for cytochrome P450 1A2 and 3A4 based on in vivo data. Clin Pharmacokinet. 2014;53(7):625-36.

[6] Johnson TN, Rostami-Hodjegan A, Tucker GT. Prediction of the clearance of eleven drugs and associated variability in neonates, infants and children. Clin Pharmacokinet. 2006;45(9):931-56.

[7] Koukouritaki SB, Manro JR, Marsh SA, Stevens JC, Rettie AE, McCarver DG, Hines RN. Developmental expression of human hepatic CYP2C9 and CYP2C19. J Pharmacol Exp Ther. 2004;308(3):965-74.

[8] EMA. Febuxostat PIP decision. 2014.

[9] Kleiber M. Body size and metabolism. Hilgardia (1932); 6: 315–351.

[10] WHO. Weight-for-age percentiles (2006) [cited www.who.int March 2015].

[11] Beal S., Sheiner L.B., Boeckmann A. & Bauer R.J. NONMEM 7.2.0 User's Guides (1989-2011). Icon Development Solutions, Ellicott City.

[12] R Development Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing (2013).

[13] Mahmood I. Dosing in children: a critical review of the pharmacokinetic allometric scaling and modelling approaches in paediatric drug development and clinical settings. Clinical pharmacokinetics. 2014;53(4):327-46

[14] Kearns GL, Abdel-Rahman SM, Alander SW, Blowey DL, Leeder JS, Kauffman RE. Developmental pharmacology--drug disposition, action, and therapy in infants and children. The New England journal of medicine. 2003;349(12):1157-67.

[15] Holford NH. A size standard for pharmacokinetics. Clinical pharmacokinetics. 1996;30(5):329-32.

[16] Cella M, Gorter de Vries F, Burger D, Danhof M, Della Pasqua O. A model-based approach to dose selection in early paediatric development. Clinical pharmacology and therapeutics. 2010;87(3):294-302.

[17] West GB, Brown JH, Enquist BJ. A general model for the origin of allometric scaling laws in biology. Science 1997;276:122–6.

[18] http://www.fda.gov/downloads/NewsEvents/MeetingsConferencesWorkshops/UCM415211.pdf

Chapter 4

Model-based assessment of alternative study designs in pediatric trials: frequentist approaches

Part of this Chapter was published in
Smania G, Baiardi P, Ceci A, Magni P, Cella M. Model-Based Assessment of Alternative Study Designs in Pediatric Trials. Part I: Frequentist Approaches. *CPT: Pharmacometrics & Systems Pharmacology*;5(6): 305-312, 2016.

4.1. Introduction

Pediatric drug development faces several difficulties due to ethical, practical and financial considerations. Despite the EU Pediatric Regulation (EC No. 1901/2006) [1], the US Best Pharmaceuticals for Children Act [2] and the US Pediatric Research Equity Act [3] partially saved children from their role of therapeutic orphans by facilitating the execution of pediatric clinical trials, a number of obstacles still remain in providing children with safe and effective drugs [4-6].

Consequently, the design and analysis of pediatric clinical trials necessitate the most efficient and informative analytical methods [7]. The gold standard method for assessing the efficacy and safety of a new drug in patients is the Randomized Controlled Trial (RCT), which minimizes bias and provides a clear and reliable understating of the risk/benefit ratio of a new experimental treatment. Conventional confirmatory RCTs are mostly executed using a parallel-arms design with hundreds or even thousands of patients enrolled. Since the number of patients that can be enrolled in pediatric studies is limited, trials of such sizes are often unfeasible. Scientifically, clinically and logistically plausible alternatives to the classical parallel design are therefore needed [8].

The aim of this Chapter is to compare the performance of the classical parallel design with that of the alternative crossover, randomized withdrawal and sequential designs by means of pharmacokinetic-pharmacodynamic (PK-PD) based Clinical Trial Simulation (CTS). Bayesian approaches are investigated in Chapter 5. Known advantages and disadvantages of the study designs which will be evaluated in the present Chapter when compared to the standard parallel design are outlined in Table 4.1.

A published paediatric PK-PD model of topiramate (TPM) [14] was used as a paradigm for CTS in epileptic children. Designs were evaluated in terms of: type I and type II errors; sample size per arm; total trial duration; relative extent of placebo, active treatment and no-treatment exposure (due to periods during which the patients do not take either TPM or placebo, e.g. baseline and/or washout periods) and precision of treatment difference estimate. For some of the investigated designs part of these measures have been computed analytically without the need for CTS.

CTS has been successfully used in pediatrics to help trial design, not only for dose selection [15-19] but also to set other trial features such as number of dose groups and number of patients per group [20-23]. However, no attempts were made on the simultaneous investigation of a battery of alternative designs different from parallel/crossover. Moreover, comparisons are normally built upon purely statistical criteria such as sample size and mean square error of estimates, whereas in this work the additional use of total trial duration and treatments exposure are proposed to evaluate the overall performance of a particular design.

Table 4.1: Pros and cons known from the literature of alternative designs considered in the present chpater when compared to the standard parallel design.

Study design	Pros	Cons	Ref.
Cross-over	• Smaller sample size • All participants receive treatment	• Carryover effect • Longer duration • Not suitable if the disease is not stable over time • Not suitable in case of treatment with permanent effect • All participants receive placebo • Need for washout period	[9]
Randomized Withdrawal	• Subjects continue receiving study drug only if they respond to it • Lower exposure to placebo • Enrichment of study population	• Carryover effect • Treatment effect estimate is biased towards responders • Suitable only for stable chronic diseases • Ethical concerns with depriving patients of the benefit they had already obtained from the active drug	[8], [10, 11]
Sequential designs	• Allows to terminate a trial when evidence has emerged that one treatment is clearly either superior or inferior to the other • Sample size is on average smaller	• Treatment outcomes should be available quickly in relation to patients recruitment rate • Maximum sample size can be larger • Increased logistic complexity	[12, 13]

4.2. Methods

4.2.1. Case study: pharmacokinetic-pharmacodynamic model of topiramate adjunctive therapy in children with epilepsy

A literature search was performed in order to identify a pediatric PK-PD model suitable for the analysis. Among the few models that resulted from the review, many were inadequate because of insufficient details to allow the use of the model in a CTS setting, lack of the PK component, unsatisfactory model evaluation and modeling of a safety PD measure rather than an efficacy one. The final choice was on a PK-PD model of

TPM in pediatric patients between 2 and 10 years of age with partial-onset or primary generalized tonic–clonic seizures [14]. PK data are described by a two-compartment model with first order absorption where weight and age were found to be significant descriptors of the population PK profile (Table 4.2). PD data used to build the model came from eight clinical trials in refractory epileptic patients (pediatric and adult).

The PK-PD model is shown in eq. (1) and it relates TPM steady-state trough plasma concentrations (C_{min}) to the log-transformed percent reduction in seizure frequency from baseline $Y = log\left(\frac{S-B}{B}100 + 110\right)$, being S the average seizure frequency per 28 days during the treatment phase and B the corresponding average during the baseline phase. β_0 and β_1 depict the placebo and drug effect, respectively, whereas ε represents the unexplained variability in Y and is assumed to be normally distributed with mean 0 and variance σ^2. Parameters values for the PK-PD model are summarized in Table 4.2.

$$Y = \beta_0 + \beta_1 \cdot C_{min} + \varepsilon \qquad (1)$$

Model (1) describes the anti-epileptic effect of TPM adjunctive therapy in treatment refractory children.

Table 4.2: Parameter values from the PK and PK-PD models (adapted from Grigis et al. [14]).

Parameter	Typical value (%SE)	Between-subject variability (%SE)
PK MODEL		
Clearance (L/h)		
CLSTM (baseline clearance monotherapy) (θ_1)	1.21 (1.2)	27.28 (10.2)
CLSTA (effect of adjuvant) (θ_2)	0.479 (25.3)	
FCWT (effect of weight) (θ_3)	0.453 (9.0)	
FCAGE (effect of age) (θ_4)	-0.00306 (30.9)	
Central volume of distribution (L)		
VST (θ_8)	4.61 (33.2)	116.2 (35.0)
FVWT (effect of weight) (θ_9)	1.14 (19.1)	
Ka (h-1) (θ_{10})	0.105 (27.0)	22.34 (88.2)
K23 (h-1) (θ_{11})	0.577 (16.7)	NE
K32 (h-1) (θ_{12})	0.0586 (23.6)	NE
CCV residual error (%CV)	25.46 (7.8)	
Additive residual error (mg/L)	0.1797 (39.9)	
PK-PD MODEL		
Placebo effect		
β_0	4.4830 (9.16)	NE
Concentration effect		
β_1	-0.0579 (3.05)	NE
σ_ε	0.751664	

%SE: percent standard error. NE: not evaluated.

4.2.2. Study designs description

All study designs presented in this Chapter are alternative implementations of a two-arm RCT. Patients in the control group received placebo (i.e., their current anti-epileptic treatment plus placebo) while patients in the treatment group received 3.5 mg/kg B.I.D. of TPM (i.e., their current anti-epileptic treatment plus TPM), that is the average FDA/EMA recommended TPM dosage regimen for the adjunctive treatment of epileptic children [24, 25]. The clinical endpoint of the trial was the log-transformed translated percent reduction in seizure frequency from baseline (i.e., Y). Coherently with the model and in agreement with previous findings, the length of the baseline and the treatment phase was set at 1 month each, for an overall duration of the trial per child (τ) of 2 months [26].

In order to design the studies, an initial estimate of both the improvement of TPM over placebo ($\delta=\mu_{PCB}-\mu_{TPM}$, where μ_{PCB} and μ_{TPM} are the expected placebo and TPM responses in terms of Y) and the variability of Y (σ) have to be formulated. σ was derived from the original publication and set to 0.7517, whereas Monte Carlo methods were used to compute δ from $10^{\wedge 6}$ samples, leading to a value of $\delta=0.2467$. An improvement of 0.2467 in the Y scale corresponds to approximately a 19% further decrease in seizure reduction for TPM 7 mg/kg/day against placebo, considering an average placebo seizure reduction of 21.5% (obtained from the PK-PD model). The superiority of TPM over placebo was assessed through standard one-sided statistical testing on the null hypothesis of no treatment difference H0: $\delta<0$ with 5% significance (i.e., $\alpha=0.05$) and 80% power (i.e., $\beta=0.20$).

Parallel design (PaD)

In a two arm PaD, patients are randomized into two parallel groups to receive either placebo or TPM, with the number of patients to be randomized in each group fixed a priori. In agreement with the PK-PD model, responses were assumed to be normally distributed with the same variance σ^2 in the TPM and placebo arm. Accordingly, normal-approximation for a one-sided Student's t-test was assumed to obtain the number of patients to be enrolled in each group

$$n = 2\left[\frac{(z_{1-\alpha}+z_{1-\beta})\cdot\sigma}{\delta}\right]^2,$$

with z_x being the x-th quantile of the standard normal distribution.

Crossover design (XD)

In a XD, patients are randomized to one of two treatment sequences, one where they receive first TPM and then placebo and one where they receive first placebo and then TPM. The length of the washout period between the two treatment sequences was set to 1 month, in agreement with previous crossover studies in pediatric epileptic patients and TPM elimination half-life [27].

Sample size calculation for the XD was adapted from Wellek and Blettner [9]. As for the PaD, responses were assumed to be normally distributed with the same variance σ^2 in patients receiving TPM and placebo. In particular, in order to obtain the number of patients to be randomized in each sequence, the following formula (assuming normal-approximation of t distribution) was used

$$n = 2\left[\frac{(z_{1-\alpha}+z_{1-\beta})\cdot\sigma}{2\delta}\right]^2 [2(1-\rho)],$$

where ρ is the correlation of Y between the two periods of the XD (ρ can be thought of as the proportion of PD BSV contained in ε: if ρ=0 then all the variability contained in ε is intra-individual variability (IIV), if ρ=1 then all the variability contained in ε is BSV). In order to evaluate the sensitivity of design performance with respect to such parameter, simulations were carried out for ρ = 0, 0.25, 0.5, 0.75.

Randomized withdrawal design (RWD)

In a RWD, after an initial open-label period in which all patients receive TPM, only patients who positively responded to TPM (defined as patients whose percentage seizure reduction from baseline is greater than the corresponding average placebo response) enter the double-blind phase and are randomized to receive either placebo or TPM, whereas the non-responders discontinue the trial. The same washout period and correlation ρ defined in the XD were assumed in the RWD between the open-label and double blind phase.

In order to maintain the desired statistical properties of the analysis, the sample size of the RWD in the double blind phase (whose collected measures will be subject to statistical testing) should be similar to the PaD one. Therefore, an initial estimate of the percentage of responders (θ) is needed in order to obtain the total sample size at the open-label phase, which is defined as

$$4\left[\frac{\left(z_{1-\alpha}+z_{1-\beta}\right)\cdot\sigma}{\delta}\right]^2\cdot\frac{1}{\theta}$$

Following the same procedure used to obtain δ, PK-PD simulations allowed deriving an estimate of the responder rate θ of 0.627, suggesting that about 62.7% of children have a response to TPM greater than the average placebo response.

Group sequential designs: Sequential Probability Ratio Test (SPRT) and Triangular Test (TT)

In group sequential designs, statistical analyses are sequentially performed after the enrolment of groups of patients of predetermined size *G*. This allows early stopping of the trial for either efficacy or futility. Several statistical approaches have been proposed for the design and analysis of group sequential trials (e.g. O'Brien-Fleming method [28] and Pocock method [29]). In this Chapter, two alternative implementations of group sequential designs were considered, namely the SPRT and TT. Despite such designs have been rarely applied, they appear to have favorable properties for pediatric trials [30]. The statistical framework for these two designs was adapted from Whitehead [31]. These methods are also known as boundary methods since, at each interim analysis, a sample

statistics Z (which can be thought of as the accumulated evidence of δ) is plotted against a second sample statistics V (which can be thought of as the amount of information about δ contained in Z), and when the value of Z exits a so-called continuation region delimited by two boundaries in the V-Z plane, H_0 is either accepted or refused (Figure 4.1). The two methods differ for the equations of the boundaries: in the SPRT these are parallel and the continuation region is open, while in the TT they converge defining a close continuation region. On the one hand the TT may thereby appear more relevant a-priori, because the sample size could theoretically be infinite by using the SPRT. On the other hand sample size reductions in case of clear evidence of efficacy/futility are larger with the SPRT when compared to the TT.

Formally, considering the case of normally distributed responses with common standard deviation σ, if at a certain point during the trial m responses in the placebo group ($y_{1p}, y_{2p}, \ldots, y_{mp}$) with sum Y_P and m responses in the TPM group ($y_{1t}, y_{2t}, \ldots, y_{mt}$) with sum Y_T have been observed, $Z = \frac{Y_P - Y_T}{2\sigma^2}$ and $V = \frac{m}{2\sigma^2}$. In order to define the boundaries, a further variable I has to be introduced: this is called "inspection interval", which defines the ideal constant increase of V between two consecutive inspections, i.e. $V = I, 2I, 3I, \ldots$. The inspection interval can be computed as $I = \frac{G}{4\sigma^2}$.

The continuation region is defined as $\Omega = \{(V, Z) | Z \in (-q + kV, q + rV)\}$, where for the SPRT $q = \frac{1}{\bar{\delta}} \log \frac{1-\alpha}{\alpha} - 0.583\sqrt{I}$ and $k = r = \frac{1}{2}\bar{\delta}$ (p. 84 of Whitehead [31]), whereas for the TT $q = \frac{2}{\bar{\delta}} \log \frac{1}{2\alpha} - 0.583\sqrt{I}$, $k = \frac{3}{4}\bar{\delta}$ and $r = \frac{1}{4}\bar{\delta}$ (p. 72 of Whitehead [31]), with $\bar{\delta} = \delta \frac{2z_{1-\alpha}}{z_{1-\alpha} + z_{1-\beta}}$. At each interim analysis three alternative decisions can be taken based on the current value of Z: (i) if Z quits Ω through the lower boundary, H_0 is accepted; (ii) if Z quits Ω through the upper boundary, H_0 is rejected; (iii) if $Z \in \Omega$, G more patients have to be enrolled in the trial.

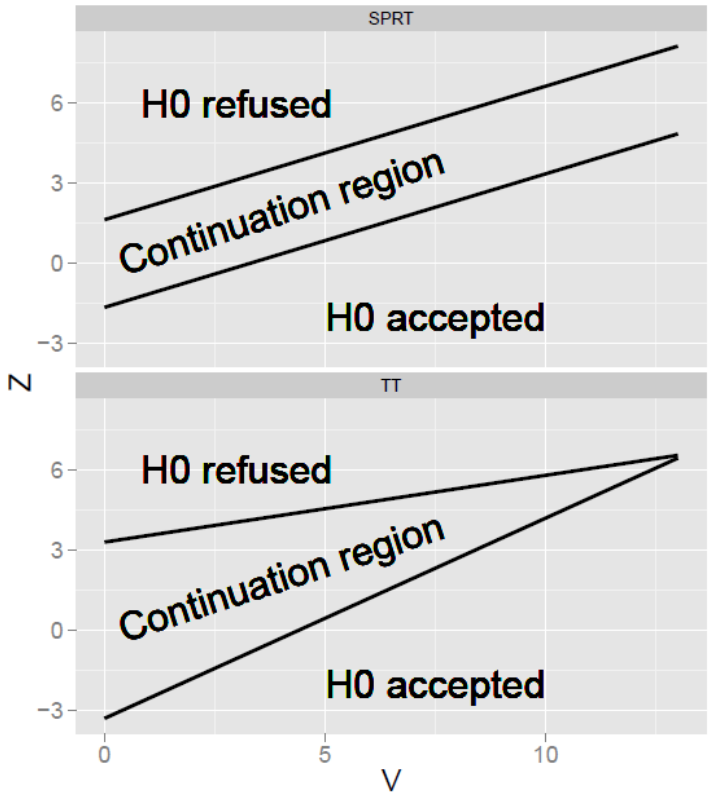


Figure 4.1: Example of acceptance/rejection boundaries of the Sequential Probability Ratio Test (SPRT, upper panel) and the Triangular Test (TT, lower panel) for $\delta=\sigma=1$ and $\alpha=\beta=0.05$. During the trial the value of the Z statistics computed at each analysis is plotted against the associated V-value, building up a path on the Z-V plane. If such path crosses the upper boundary, H_0 is refused; if it crosses the lower boundary, H_0 is accepted; if it stays within the continuation region, patients recruitment goes on.

4.2.3. Study designs simulation

The simulation of each design was based on the following step-wise procedure:

1. Patients population was simulated in terms of weight and age (as these were the only two significant covariates of the PK model) using weight-for-age tables from World Health Organization statistics [32]. After having uniformly generated 2000 age values between 2 and 10 years, body weights (BWs) were simulated within each year of age by assuming them to follow a log-normal distribution whose mean and standard deviation were derived by fitting a log-normal distribution to the BW percentiles. Despite gender was not identified as a significant covariate, the age-BW relationship varies between male and female subjects; consistently, male and female patients had the same chance to enter the trial.

2. For each virtual child, the individual C_{min} was extracted from 500 PK profiles generated by using NONMEM version 7.2.0 [33] (see Appendix 7.1 for the NONMEM code used).
3. Patients to be enrolled following the design-specific rules were randomly selected from the population simulated at step 1 and, for each child, a single C_{min} was further sampled from its 500 values simulated at step 2. These values were fed into the PK-PD model in order to simulate the clinical endpoint. For subjects randomized to the control arm C_{min} was set to zero.
4. The study design-specific statistical analysis was applied to the simulated endpoints.

Steps 3 and 4 were re-iterated 1000 times per study design, i.e. 1000 CTS were performed. A visual description of this step-wise procedure can be found in Figure 4.2. R software version 3.0.1 [34] was used for steps 1 (see Appendix 7.2 for R code used), 3 and 4 (see Appendix 7.3 for R code used).

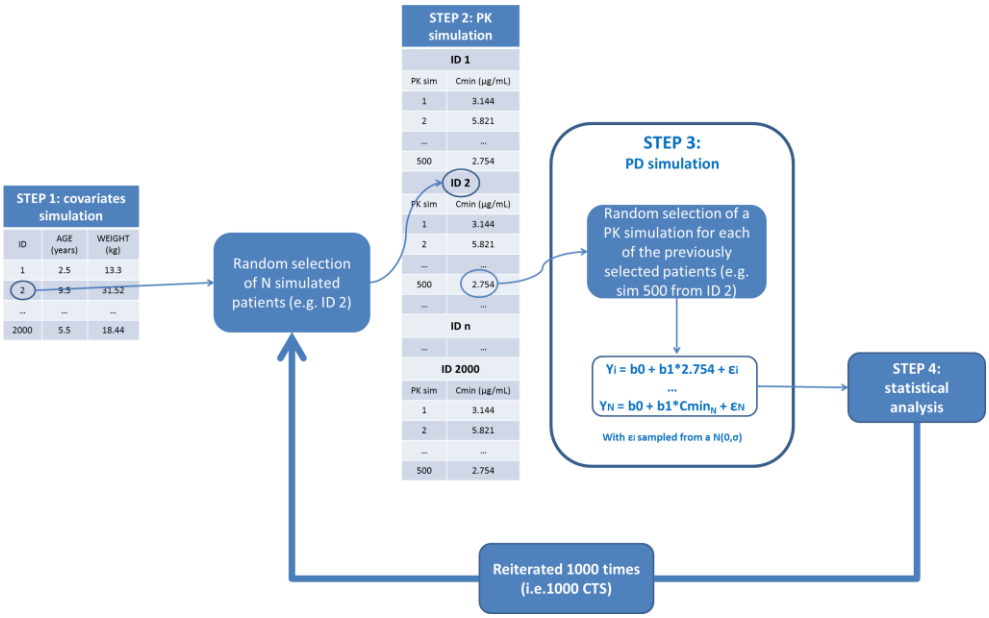


Figure 4.2: Step-wise procedure used for clinical trial simulations of the different study designs.

The following metrics were adopted for designs comparison:

a) *Type I error ($\hat{\alpha}$)*: the proportion of statistical analyses leading to the rejection of H_0 when simulating under H_0 (i.e., when in the simulations described at step 3 β_1=0).

b) *Type II error ($\hat{\beta}$)*: the proportion of statistical analyses leading to the acceptance of H_0 when simulating under H_1 (i.e., when in the simulations described at step 3 β_1 is set to its estimated value).

c) *Sample size per arm (SS)*: number of children enrolled in each arm of the trial.

d) *Total trial duration (TD)*: the duration in months of the entire trial as a function of enrollment rate (ER; ER up to 10 patients/month were considered). Although certain features of a trial design (e.g. treatment blinding, inclusion/exclusion criteria) can influence ER [35], these were not considered in this analysis and ER was therefore assumed to be design-independent.

e) *Average extent of placebo, TPM and no-treatment exposure*: percentage of exposure to placebo, TPM and no-treatment relative to total trial exposure.

f) *Treatment difference estimate precision (EP)*: the precision of the estimate of δ ($\hat{\delta}$) expressed in terms of the width of its 95% confidence interval.

A comprehensive description of the calculation of such metrics for each of the investigated designs is given hereafter.

Parallel Design

According to the calculated total sample size, at each iteration of step 3 half of the patients were assigned to the control group and half to the treatment group. Step 4 consisted in performing a Student's t-test between these two groups and computing the estimated difference with its confidence interval. The comparison metrics were thus obtained as follows:

a) $\hat{\alpha}$= percentage of simulated trials with p-values<0.05 out of 1000 CTS when simulating under H_0.

b) $\hat{\beta}$= percentage of simulated trials with p-values>0.05 out of 1000 CTS when simulating under H_1.

c) SS = $2\left[\frac{(z_{1-\alpha}+z_{1-\beta})\cdot\sigma}{\delta}\right]^2$ (fixed a priori) where $z_x$is the x-th quantile of the standard normal distribution, δ is the difference of Y, the primary endpoint of the trial, between TPM and placebo, and σ is the variability of Y.

d) Trial duration is given by the months that need to be waited before recruiting all 2SS patients, plus the time needed to measure the endpoint of the last patient enrolled. In other words: TD = $\frac{2SS}{ER}+\tau$, where ER is the enrollment rate and τ the duration of the trial per child (i.e., 2 months).

e) Extent of placebo and TPM exposure = $\frac{\tau}{2}SS/$TTE;
Extent of no-treatment exposure = $\tau \cdot SS/$TTE;
Where TTE stands for Total Trial Exposure = $\tau \cdot 2 \cdot SS$.

f) EP = distribution of the 95% confidence interval widths.

Crossover design

According to the calculated total sample size, at each iteration of step 3 half of the patients were assigned to the TPM-placebo sequence and half to

the placebo-TPM sequence. Step 4 consisted in performing a two-sample Student's t-test between the within-subject differences of responses computed in the two sequences (i.e. TPM response minus placebo response in the TPM-placebo sequence and placebo response minus TPM response in the placebo-TPM sequence) [9] and computing the estimated difference with its confidence interval. Carryover effects were assumed to be negligible.

Metrics a), b) and f) were obtained as in the PaD, whereas the remainders as follows:

c) SS $= 2\left[\frac{(z_{1-\alpha}+z_{1-\beta})\cdot\sigma}{2\delta}\right]^2 [2(1-\rho)]$ (fixed a priori), where ρ is the correlation of Y between the two periods of the XD and has been fixed to ρ=0, 0.25, 0.50, 0.75.

d) Trial duration is given by the time required to recruit all 2SS patients, plus the time needed to measure the endpoint in the first period of the last patient enrolled (τ), plus the washout period and plus an additional τ/2 (because the baseline had already been measured) needed to obtain the endpoint in the second period of the last patient enrolled. In other words: TD $= 2\frac{SS}{ER} + \frac{3}{2}\tau + washout$, where *washout* is the washout period between the two periods of the XD (set to 1 month).

e) Extent of placebo and TPM exposure = $(\tau \cdot SS)/$TTE;
extent of no-treatment exposure = $2 \cdot SS(\frac{\tau}{2} + washout)/$TTE;
where TTE $=2 \cdot SS(\frac{3\tau}{2} + washout)$.

Randomized withdrawal design

According to the calculated total sample size, at each iteration of step 3 all children received TPM in an open-label phase and their response (Y_{OL}) was simulated according to the model Y_{OL}=b0 + b1Cmin + ε_{OL} (where ε_{OL} is the within subject variability sampled in the open-label phase). Only those patients whose response was greater than the average placebo response (i.e., for whom Y_{OL}<b0) entered the double blind phase and were randomized to either TPM or placebo. In the post-randomization simulations of step 3, responses were not assumed to be completely independent to the ones in the open-label phase, but a correlation ρ (ρ=0, 0.25, 0.50, 0.75) between ε_{OL} and ε_{DB} (i.e., the within subject variability in the double blind phase) was imposed. In order to implement this in the simulations, the pair (ε_{OL}, ε_{DB}) was sampled from a multivariate normal distribution with mean zero and variance-covariance matrix equal to

$$\sigma_\varepsilon \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}$$

Ultimately, the response in the double blind phase was obtained as Y_{DB}=b0 + b1Cmin + ε_{DB}, where Cmin was left equal to the one in the open-label simulations. Step 4 consisted in performing a Student's t-test between the two post-randomization groups and consequently obtaining the p-value of such test and computing an estimate of the treatment effect with its 95% confidence interval.

Metrics a), b) and f) were obtained as in the PaD, whereas the reminders as follows:

c) Although in the open-label phase of a RWD there is only one arm and it is not realistic to talk about "sample size per arm", in order to be consistent with the SS measured for other designs the sample size was calculated as half of the sample size in the open-label phase, i.e. SS = $2\left[\frac{(z_{1-\alpha}+z_{1-\beta})\cdot\sigma}{\delta}\right]^2 \cdot \frac{1}{\theta}$ (fixed a priori), where θ is the estimated percentage of responders during the open-label phase.

d) Trial duration is calculated as the one of the XD. The worst case scenario for trial duration was reported, i.e. it was assumed that the last recruited patient was a responder and thus there is the need to wait for his/her response before completing the study.

TD = $2\frac{SS}{ER} + \frac{3}{2}\tau + washout$ (worst case scenario).

e) (Expected) extent of placebo exposure = $(\frac{\tau}{2} \cdot E[SS_{DB}])/$TTE;

(expected) extent of TPM exposure = $\frac{\tau}{2}(2 \cdot SS + E[SS_{DB}])/$TTE;

(expected) extent of no-treatment exposure = $2(\frac{\tau}{2}SS + washout \cdot E[SS_{DB}])/$TTE;

where TTE =$\tau \cdot 2 \cdot SS + 2 \cdot E[SS_{DB}](\frac{\tau}{2} + washout)$ and $E[SS_{DB}]$ is the average sample size per arm in the double-blind phase.

Group sequential designs: Sequential Probability Ratio Test and Triangular Test

SPRT and TT were simulated with G = 20. In agreement with the sequential nature of these designs, for each CTS steps 3 and 4 were sequentially performed until acceptance/rejection of H_0. In particular, in step 3 G children were randomized to TPM and placebo in a 1:1 ratio and their simulated responses used to compute values of Z and V statistics, whereas in step 4 the membership of Z to Ω was tested.

Metrics were obtained as follows:

a) $\hat{\alpha}$= percentage of simulated trials with Z values crossing the upper boundary out of 1000 CTS when simulating under H_0.

b) $\hat{\beta}$= percentage of simulated trials with Z values crossing the lower boundary out of 1000 CTS when simulating under H_1.

c) SS =distribution of the sample sizes obtained at each trial simulation.

d) Total trial duration of a sequential design is given by the sum of the following four addends:

- $\frac{G}{ER}$: time required to enroll the first G patients;
- $\left\lceil\frac{2SS}{G}\right\rceil\tau$, where ⌈ ⌉ denotes the ceiling function: total time required to wait for all the sequential analyses to be completed;
- $\max\left(0,\frac{G-ER\cdot\tau}{ER}\left(\left\lceil\frac{2SS}{G}\right\rceil-2\right)\right)$: additional time required for all interim analyses (except for the first and the last one) if the number of patients that can be enrolled during τ (i.e. ER·τ) is less than G;
- $\max\left(0,\frac{2SS-\left(\left\lceil\frac{2SS}{G}\right\rceil-1\right)\cdot G-ER\cdot\tau}{ER}\right)$: additional time required for the last interim analysis if the number of patients that can be enrolled during τ is less than the last set of patients to be enrolled $(2SS-\left(\left\lceil\frac{2SS}{G}\right\rceil-1\right)\cdot G)$

Therefore the total trial duration of a sequential design is defined as:

$$TD=\frac{G}{ER}+\left\lceil\frac{2SS}{G}\right\rceil\tau+\max\left(0,\frac{G-ER\cdot\tau}{ER}\left(\left\lceil\frac{2SS}{G}\right\rceil-2\right)\right)+\max\left(0,\frac{2SS-\left(\left\lceil\frac{2SS}{G}\right\rceil-1\right)\cdot G-ER\cdot\tau}{ER}\right).$$

e) (Expected) extent of placebo exposure = $(\frac{\tau}{2}E[SS])/\text{TTE}$;
(expected) extent of TPM exposure = $(\frac{\tau}{2}E[SS])/\text{TTE}$;
(expected) extent of no-treatment exposure = $(\tau\cdot E[SS])/\text{TTE}$;
where TTE= $\tau\cdot 2\cdot E[SS]$ and $E[SS]$ is the average sample size per arm.

f) EP: at the termination of a sequential trial, maximum likelihood estimation of δ is a biased estimator of treatment difference "because of the dependence of the stopping rule on the nature of the evidence collected" [31]. Whitehead [31] proposed an alternative estimator of δ which is median unbiased. An ad-hoc R functions was built to compute such estimate and its 95% confidence interval (see Appendix 7.4).

4.3. Results

4.3.1. Type I and type II errors

$\hat{\alpha}$ and $\hat{\beta}$ are close to their predetermined levels of 5% and 20% for all designs except for the RWD with ρ>0, where $\hat{\beta}$ appears to decrease when correlation increases (Table 4.3). This increasing power with increasing ρ is due to a decrease in the variability of responses in the double-blind phase of the study further given by the fact that drug effect is evaluated in a specific subset of the pediatric population (i.e., those that respond to TPM).

Sequential designs show a slightly higher $\hat{\alpha}$, although for the TT the target value of 5% is contained in the 95% confidence interval of the estimated $\hat{\alpha}$. Conversely, $\hat{\alpha}$ of the SPRT is significantly higher than 5%.

4.3.2. Sample size per arm

Table 4.3 shows that the XD with ρ=0.75 leads to a SS of 15 children, which is the minimum SS among all designs. Even for lower values of ρ the XD requires a lower SS compared with other designs, and when the PD BSV is negligible (ρ=0) SS of XD is half of the SS in the PaD.

SS for sequential designs are not deterministic. The histograms of the SS obtained at each simulation of the two sequential designs are depicted in Figure 4.3. It can be seen that on average both the SPRT and TT requires less patients than the PaD (around 76 per arm).

Finally, the estimated probability of terminating the sequential designs with a SS greater than the PaD is fairly low (19% for the SPRT and 13.9% for the TT).

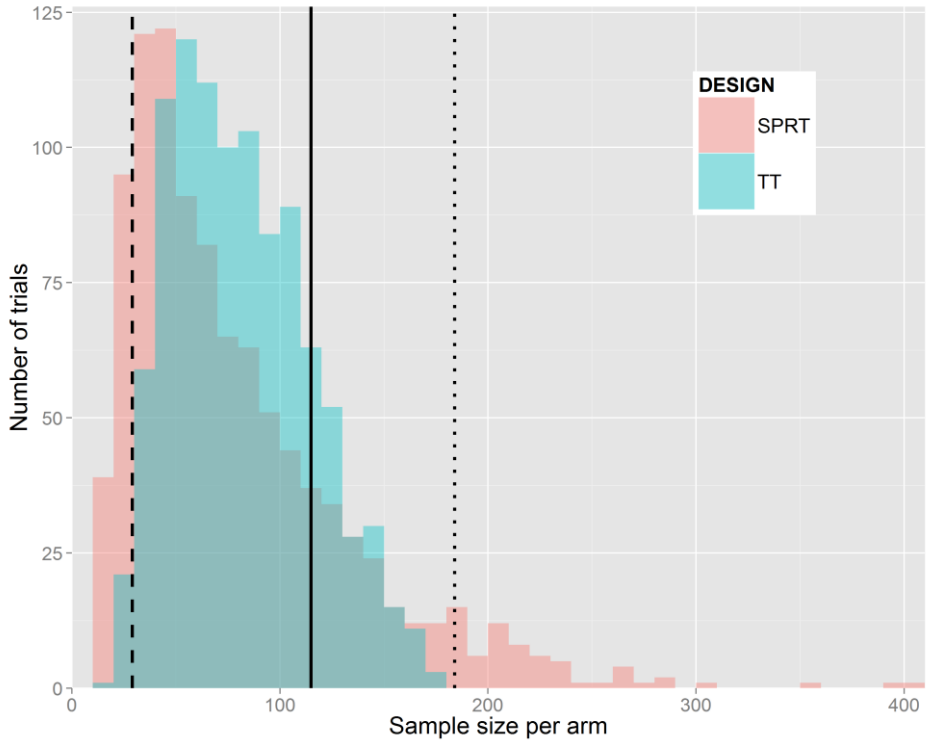


Figure 4.3: Histograms of the sample sizes per arm obtained at each of the 1000 clinical trial simulation of the Sequential Probability Ratio Test (SPRT, pink histogram) and the Triangular Test (TT, green histogram). Black vertical lines indicate the sample sizes per arm of the Parallel design (solid line), Crossover design with ρ=0.5 (dashed line) and Randomized Withdrawal design (dotted line).

4.3.3. Total trial duration

TD reflects the required SS: the higher the sample size, the higher the duration (for a given ER). Accordingly the XD has the lowest median TD among all investigated designs (Figure 4.4). The same may not be true for sequential designs though, wherein the time needed before obtaining the primary endpoint for the sequential analysis can remarkably increase TD. However, results show that for both pessimistic (4 patients/months) and optimistic (10 patients/months) enrollment rates, this was not the case (Table 4.3). In fact, the median TDs of SPRT and TT are lower than TD of the PaD, suggesting that for this magnitude of ER TD reduction due to a lower SS outweighs the TD increasing that could have been arisen because of sequential enrollment.

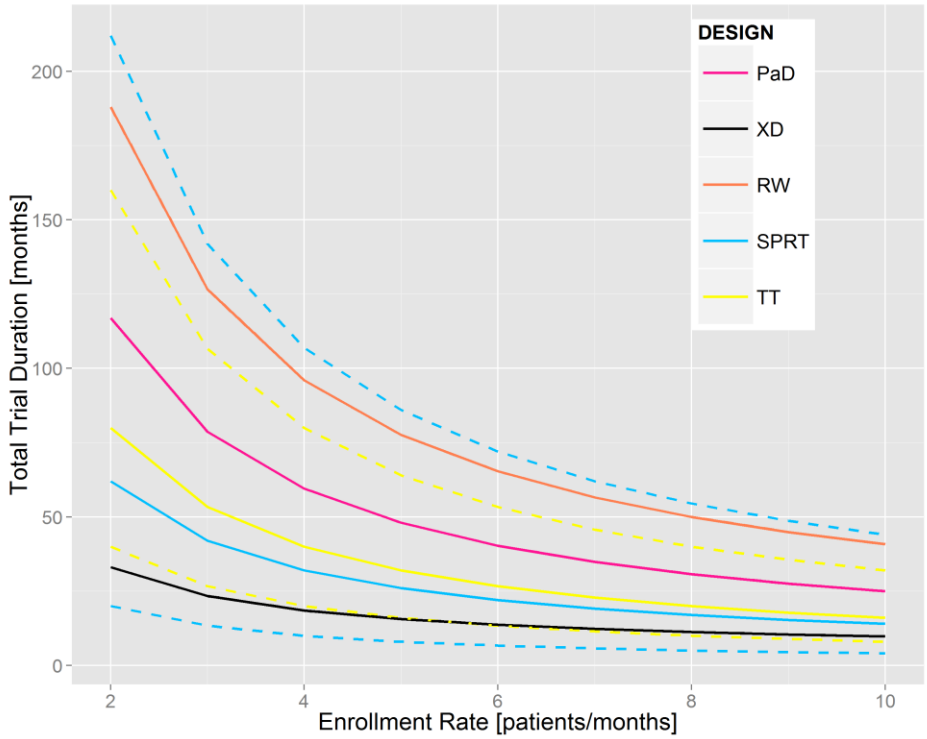


Figure 4.4. Total trial duration as a function of enrollment rate for each of the investigated designs (PaD: Parallel Design; XD: Crossover Design with ρ=0.5; RWD: Randomized Withdrawal Design with ρ=0.5; SPRT: Sequential Probability Ratio Test; TT: Triangular Test). Solid lines represent the median duration whereas dotted lines depict 95% prediction intervals.

4.3.4. Extent of placebo, topiramate and no-treatment exposure

Left panel of Figure 4.5, which quantifies the extent of exposure to placebo (black bar), TPM (red bar) and no-treatment (cyan bar) for each of the investigated designs in terms of the proportion of exposure relative to total trial exposure, clearly shows that the RWD allows to minimize exposure to placebo while maximizing exposure to TPM, no-treatment exposure being equal at 50% for all designs.

As to no-treatment exposures, these are comparable among all designs. However, if a washout of 2 months had been required, percentage of exposure to no-treatment in the XD and RWD would have risen to 60% and 58%, respectively (right panel of Figure 4.5).

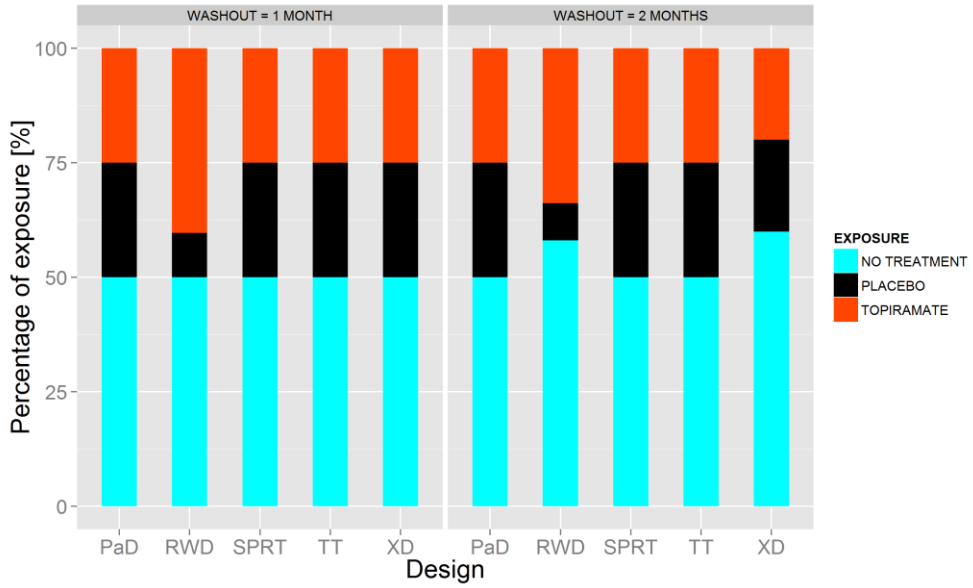


Figure 4.5. Percentage of exposure to no-treatment (cyan bar), placebo (black bar) and topiramate (red bar) relative to total trial exposure for each of the investigated designs (PaD: Parallel Design; XD: Crossover Design; RWD: Randomized Withdrawal Design; SPRT: Sequential Probability Ratio Test; TT: Triangular Test) for a washout of 1 (left panel) and 2 (right panel) months. For the XD and RWD this metric is equal across all values of ρ.

4.3.5. Treatment difference estimate precision

Although the value of $\hat{\delta}$ was similar for all designs (Figure 4.6a), its precision may substantially vary across them. In fact, Figure 4.6b shows that the sequential designs can lead to precisions much lower than those obtained with a PaD or RWD in terms of width of 95% confidence interval. This is partly explained by the number of samples used to compute the estimate: the SPRT and TT are those designs that allow on average to keep the sample size low thereby increasing the standard error of $\hat{\delta}$ and consequently the width of its 95% confidence interval.

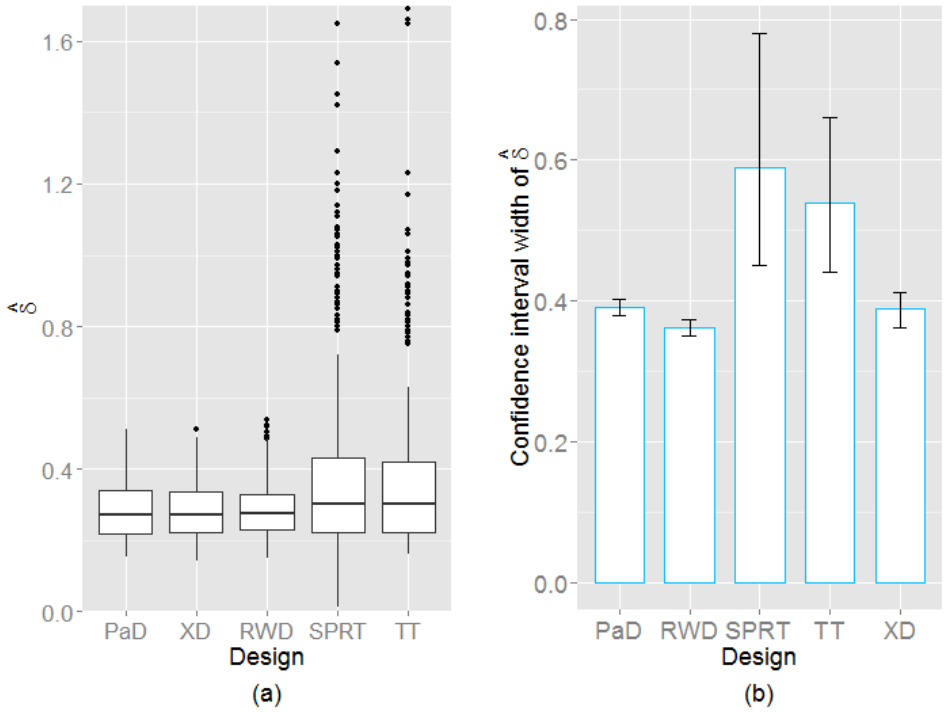


Figure 4.6. Boxplot of treatment difference estimates ($\hat{\delta}$) **(a)** and bar chart plot of the median confidence interval width of $\hat{\delta}$ **(b)** obtained at each of the 1000 clinical trial simulation of the Parallel Design (PaD), Crossover Design with ρ=0.5 (XD), Randomized Withdrawal Design with ρ=0.5 (RWD), Sequential Probability Ratio Test (SPRT) and Triangular Test (TT). 9 simulations in the SPRT and 1 in the TT were not able to compute $\hat{\delta}$ and its precision because of negative values of V due to very early stopping (i.e. final sample size per arm = 10 patients) and were therefore excluded from the figure.

Table 4.3: Metrics ($\hat{\alpha}$: type I error; $\hat{\beta}$: type II error; SS: sample size per arm; SS_{50}: median sample size per arm; SS_{75}, SS_{90}, SS_{95}: 75th, 90th and 95th percentiles of sample size distribution, respectively; TD: total trial duration) obtained from clinical trial simulation for each of the simulated designs and for different values of ρ. PaD: Parallel Design; XD: Crossover Design; RWD: Randomized Withdrawal Design; SPRT: Sequential Probability Ratio Test; TT: Triangular Test. ER: Enrollment Rate.

Metric \ Design		PaD	XD				RWD				SPRT	TT
			ρ=0	ρ=0.25	ρ=0.5	ρ=0.75	ρ=0	ρ=0.25	ρ=0.5	ρ=0.75		
$\hat{\alpha}$ (%) (95% CI)		5.1 (3.7-6.5)	4.7 (3.4-6.0)	5.4 (4.0-6.8)	5.2 (3.8-6.6)	4.4 (3.1-5.7)	5.4 (4.0-6.8)	4.9 (3.6-6.2)	4.9 (3.6-6.2)	4.2 (3.0-5.4)	7.0 (5.4-8.6)	5.9 (4.4-7.4)
$\hat{\beta}$(%) (95% CI)		21.2 (18.7-23.7)	19.1 (16.7-21.5)	22.8 (20.2-25.4)	22.5 (19.9-25.0)	25.0 (22.3-27.7)	13.4 (11.3-15.5)	13.3 (11.2-15.4)	10.7 (8.8-12.6)	7.5 (5.9-9.1)	18.3 (15.9-20.7)	20.3 (17.8-22.8)
E[SS] (patients)		115	58	44	29	15	116	116	116	116	75	77
SS_{50} (patients)							115	116	115	116	60	70
SS_{75} (patients)							118	118	118	119	100	100
SS_{90} (patients)							121	121	121	121	150	120
SS_{95} (patients)							122	123	123	123	181	140
MEDIAN TD (months)	ER=4 patients/ month	59.5	33.0	26.0	18.5	11.5	96.0				32.0	40.0
	ER=4 patients/ month	25	15.6	12.8	9.8	7.0	40.0				14.0	16.0

*SS reported in the table is the one of the double-blind phase, SS as defined in 4.2.3 (i.e. as half of the sample size in the open-label phase) is 184.

4.4. Discussion

The use of alternative study designs in pediatric trials can significantly improve their feasibility by reducing their sample size and duration and by increasing their acceptability.

Examples on the use of PK-PD based CTS for the assessment of trial design performance exist [20, 21], though no attempts were made on simultaneously exploring alternative study designs such as the RWD and the sequential designs. Investigations of sequential designs by CTS can also be found in the literature [36, 37]; however these are based on purely statistical models thereby neglecting the PK-PD component. As a result, the influence of patient demographics and dosage regimens on trial performance could have not been explicitly taken into account. The present Chapter provided a pharmacometric-based framework for a multi-dimensional comparison of alternative study designs.

Overall, the outcomes of this analysis are in line with the known pros and cons of the investigated designs introduced in Table 4.1. Results clearly show that for a pediatric trial the XD, irrespective of the value of ρ, allows to minimize the SS required while maintaining desired type I and type II errors. The minimization of SS translates also in a very low TD. However, the XD may not be easily accepted by parents/children because the washout period implies that children have to spend a higher period of time without taking any treatment when compared to other designs (Figure 4.5). In addition, despite all children enrolled in the trial will certainly receive the active treatment, they will certainly receive placebo as well, posing further ethical issues. Finally, although negligible carryover effects were assumed, these have to be considered when designing a pediatric trial with a crossover scheme because they can eventually compromise the analysis and interpretation of the results [9, 38].

The RWD ensures that all children enrolled in the trial will receive the new treatment and those not responding to the treatment will be quickly withdrawn from the trial. As shown in Figure 4.5, for a washout of 1 month, the percentage of exposure to TPM in the RWD is 40%, compared with 25% of other designs. At the same time the percentage of exposure to inactive treatment (placebo) is less than half that of other designs. Such properties are still valid if the washout period is doubled. These parent/patient friendly features of the RWD, along with an acceptable level of scientific rigor, contributed to its increased popularity in the design of juvenile arthritis trials [10, 11]. Nonetheless the results suggest that the RWD would require a higher SS compared to other designs (Figure 4.3) and, consequently, a higher TD (Figure 4.4). However, if it is reasonable to assume that patients response to treatment does not remarkably change between the open-label and double-blind period (i.e., ρ=0.75), the RWD leads to a greater power (92.5%, Table 4.3). From a SS perspective, by maintaining the type II error to approximately 20%, the SS of the RWD

would drop to values around that of PaD (115, results not shown). Because in RWD a slightly different population is studied compared to the other designs, a different $\hat{\delta}$ is typically expected. Figure 4.6a shows that this was not the case in this analysis. This is primarily due to the fact that a certain patient is classified as a responder in the open-label phase mainly because a large ε was sampled for that patient, and not because that patient had an increased exposure to TPM in terms of C_{min}. This implies that: *(i)* for large values of ρ (>~0.5), patients in the double-blind phase have a high response in both the placebo and TPM arm which prevents an increase of $\hat{\delta}$ (e.g. for ρ=0.75 mean $\hat{\delta}$ turned out to be around 0.270 in the RWD, compared to the value of 0.277 observed in the PaD); *(ii)* for low values of rho (<~0.5), the effect of a larger C_{min} on $\hat{\delta}$ is partially masked by high (or low) values of ε (mean $\hat{\delta}$ was estimated around 0.286 in the RWD, compared to the value of 0.277 observed in the PaD). Consequently, the results presented in this Chapter are likely to be observed in a RWD in patients with a large placebo response *(i)* or when the magnitude of the BSV is negligible comparted to IIV *(ii)*.

Sequential designs are of great interest for pediatric trials essentially because they allow stopping early for efficacy or futility. Previous pediatric sequential trials have shown a median SS decrease of 35% compared to standard PaDs [30]. The results presented in this Chapter confirm that on average both the SPRT and the TT determine a SS reduction compared to the PaD (between 33% and 50%), without compromising the desired statistical properties (although $\hat{\alpha}$ appears to be slightly higher, in agreement with earlier findings [36]). Moreover, the simulations show that the SPRT and TT have a 13.4% and 2.2% probability of demonstrating drug efficacy with a SS lower than 21, respectively. Accordingly TD for an enrollment rate of just 2 patients/month would fall to 20 months. Despite treatment effect estimate precision associated with these low SS cannot be considered acceptable, these designs may be of interest when very limited subjects can be recruited. On the other hand, since in sequential designs the SS is not fixed a-priori, final SS and TD of the SPRT and TT may turn out to be greater than those that would have been required by a fixed sample size approach. This is demonstrated by a 90^{th} percentile of SS distribution greater than the PaD SS (Table 4.3). Within sequential designs, the TT appears to outperform the SPRT in the unfortunate scenario of a late study termination, with a 90^{th} percentile in SS distribution 30 patients lower than the SPRT. These results are in line with those from Sebille and Bellissant [36].

The added value of this analysis compared to that of Sebille and Belissant is that the use of PK-PD based CTS enables to contextualize the analysis in the clinical condition under study and to investigate the impact of PK variability and patients characteristics on the possible results of the trial. CTS becomes then a tool for a sound evaluation of candidate designs by enabling for example to assess the impact of patient population

characteristics on the probability of terminating a sequential trial with a SS greater than that of a traditional design.

It is acknowledged that the present framework is underpinned by a robust pediatric PK-PD model which may not always be available at the design stage. If this is the case, extrapolating using an adult PK-PD model in lieu of the pediatric one can be considered provided that the following conditions can reasonably be assumed [39]: *(i)* the pathophysiology of the disease is the same between pediatrics and adults; *(ii)* either the PK in pediatric patients is known or extrapolation of the PK from adult data is suitable (i.e., differences in PK are explained solely by differences in body weight, a reasonable assumption in children older than two years [40]); *(iii)* the PK-PD relationship is similar between the two populations; *(iv)* the response to treatment is assessed in terms of the same PD measure in pediatrics and adults. In general, the degree of uncertainty of assumptions *(i-iv)* should guide the design of the pediatric study. Halvin et al. [41] proposed a statistical framework to quantitatively accommodate assumptions uncertainty by enlarging the significance level of the pediatric trial based on experts skepticism about the expected similarities and differences between the adult and pediatric population. The use of an adult PK-PD model in the CTS framework would implicitly convey a certain magnitude of skepticism (or rather belief) in the extrapolation process. The dependency of trial results and design performance on this magnitude of skepticism can potentially be investigated through the integration of the current framework with that proposed by Halvin et al. This would ultimately enable researchers to select the study design which best suits their current extrapolation concept.

Since this work is focused on pediatric efficacy trials, strengths and weaknesses of study designs with respect to ancillary trial objectives have not been explicitly investigated. One important aspect relates to the support of dose regimens in the pediatric population. From this point of view, the RWD is expected to be one of the more robust in justifying pediatric dosage because it allows emphasizing the effect of the tested dose compared to placebo by ruling out the confounding element that would be introduced by the randomization of non-responders. In addition, the XD owns the favorable property of estimating the true drug effect in each patient, leading to more precise estimates of BSV of drug effect parameter; nevertheless, its small sample size might jeopardize the reliability of the PK analysis. The same would also apply for sequential designs terminating with a very low sample size. PaD and sequential designs do not exhibit any particular advantage when it comes to supporting dose regimens in the pediatric population, the only difference between the two being attributable to differences in sample sizes.

Noteworthy, it has to be pointed out that the present analysis is based on the effect of TPM in children with partial onset seizures refractory to their current antiepileptic treatment, and the extrapolation of the results to different compounds/diseases/subpopulations should be further explored.

To conclude, there is no best study design for children with refractory epilepsy that performs better in all the metrics that have been monitored. Sequential designs are probably more appealing because they appear to considerably reduce the SS when large effect sizes are expected. This is particularly important if patients recruitment is the primary obstacle, as TD is not inflated by the sequential procedure of the design and low precisions in $\hat{\delta}$ may be tolerated. On the other hand if major concerns are on ethical acceptability of the trial, a RWD may be preferable because of shortened placebo exposure and simultaneously increased exposure to the active compound, especially if it is reasonable to assume that the individual response to treatment does not significantly change between the open-label and double-blind phase.

In general, pediatric design selection would largely benefit from a pharmacometric approach as the one described in this Chapter, which leverages prior information available and allows to test different "what if" scenarios by assessing the characteristics of the design across multiple levels.

References

[1] EU. Regulation on medicinal products for paediatric use. Official Journal of the European Union (2006).

[2] FDA. Best Pharmaceuticals for Children Act (2002).

[3] FDA. Pediatric Research Equity Act (2003).

[4] Tishler C.L. & Reiss N.S. Pediatric drug-trial recruitment: enticement without coercion. Pediatrics.127, 949-954 (2011).

[5] Bates K.E. et al. Pediatric cardiovascular safety: challenges in drug and device development and clinical application. Am. Heart. J.164, 481-492 (2012).

[6] Sammons H. Ethical issues of clinical trials in children: a European perspective. Arch. Dis. Child.94, 474-477 (2009).

[7] EMA. Guideline on clinical trials in small populations. 2006.

[8] Baiardi P., Giaquinto C., Girotto S., Manfredi C. & Ceci A. Innovative study design for paediatric clinical trials. Eur. J. Clin. Pharmacol. 67Suppl 1, 109-115 (2011).

[9] Wellek S. & Blettner M. On the proper use of the crossover design in clinical trials: part 18 of a series on evaluation of scientific publications. Dtsch. Arztebl. Int.109, 276-81(2012).

[10] Ruperto N. et al. Abatacept in children with juvenile idiopathic arthritis: a randomised, double-blind, placebo-controlled withdrawal trial. Lancet.;372), 383-591 (2008).

[11] Lovell D.J. et al. Adalimumab with or without methotrexate in juvenile rheumatoid arthritis. N. Engl.J.Med.359, 810-820 (2008).

[12] Todd S., Whitehead A., Stallard N. & Whitehead J. Interim analyses and sequential designs in phase III studies. Br. J. Clin. Pharmacol.51, 394-399 (2001).

[13] Sébille V. & Bellissant E. Sequential methods and group sequential designs for comparative clinical trials. Fundam. Clin. Pharmacol.17, 505-516 (2003).

[14] Girgis I.G. et al. Pharmacokinetic-pharmacodynamic assessment of topiramate dosing regimens for children with epilepsy 2 to <10 years of age. Epilepsia.51, 1954-1962 (2010).

[15] Nehus E.J., Mouksassi S., Vinks A.A. & Goldstein S. Meropenem in children receiving continuous renal replacement therapy: Clinical trial simulations using realistic covariates. J. Clin. Pharmacol.54, 1421-1428 (2014).

[16] Piana C.,et al.. A model-based approach for the evaluation of once daily dosing of lamivudine in HIV-infected children. Br. J. Clin. Pharmacol.77, 852-860 (2014).

[17] Willmann S. et al. Development of a paediatric population-based model of the pharmacokinetics of rivaroxaban. Clin. Pharmacokinet.53, 89-102 (2014).

[18] Cella M., Danhof M. & Della Pasqua O. Adaptive trials in paediatric development: dealing with heterogeneity and uncertainty in pharmacokinetic differences in children. Br. J. Clin. Pharmacol. 74,346-353 (2012).

[19] Brochot A. et al. Model-Based Once-Daily Darunavir/Ritonavir Dosing Recommendations in Pediatric HIV-1-Infected Patients Aged ≥3 to <12 Years. CPT Pharmacometrics Syst. Pharmacol.4, 406-414 (2015).

[20] Jadhav P.R., Zhang J.& Gobburu J.V. Leveraging prior quantitative knowledge in guiding pediatric drug development: a case study. Pharm.Stat.8, 216-224 (2009).

[21] Salazar D.E. et al. The use of modeling and simulation to guide clinical development of olmesartan medoxomil in pediatric subjects. Clin.Pharmacol.Ther.91, 250-256 (2012).

[22] Reif S., Schultze-Mosgau M. & Sutter G. From adults to children: simulation-based choice of an appropriate sparse-sampling schedule. Clin. Pharmacol. Ther.14, 189-200 (2012).

[23] Cella M., Danhof M. & Della Pasqua O. Bridging strategies for drug combinations in pediatric indications. Clin. Pharmacol. Ther.91, 726-733 (2012).

[24] FDA. Topomax Label (2012).

[25] EMA. Topomax package leaflet (2009).

[26] French J.A., Cabrera J., Emir B., Whalen E. & Lu F. Designing a new proof-of-principle trial for treatment of partial seizures to demonstrate efficacy with minimal sample size and duration-a case study. Epilepsy Res.106, 230-236 (2013).

[27] Eriksson A.S., Nergardh A. & Hoppu K. The efficacy of lamotrigine in children and adolescents with refractory generalized epilepsy: a randomized, double-blind, crossover study. Epilepsia.39, 495-501. (1998).

[28] O'Brien P.C. & Fleming T.R. A multiple testing procedure for clinical trials. Biometrics. 35, 549-556 (1979).

[29] Pocock S.J. Group sequential methods in the design and analysis of clinical trials. Biometrika. 64, 191–199 (1977).

[30] van der Lee J.H., Wesseling J., Tanck M.W. & Offringa M. Sequential design with boundaries approach in pediatric intervention research reduces sample size. J.Clin.Epidemio.63, 19-27 (2010).

[31] Whitehead J. The design and analysis of sequential clinical trials revised 2nd edn. (Jhon Wiley & Sons, Chichester UK, 1997).

[32] WHO. Weight-for-age percentiles (2006) [cited www.who.int March 2015].

[33] Beal S., Sheiner L.B., Boeckmann A. & Bauer R.J. NONMEM 7.2.0 User's Guides (1989-2011). Icon Development Solutions, Ellicott City.

[34] R Development Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing (2013).

[35] Bower P. et al. Interventions to improve recruitment andretention in clinical trials: a survey and workshopto assess current practice and future priorities. Trials. 15:399 (2014).

[36] Sebille V. & Bellissant E. Comparison of four sequential methods allowing for early stopping of comparative clinical trials. Clin. Sci.98, 569-578 (2000).

[37] Emerson S.S., Kittelson J.M. & Gillen D.L. Frequentist evaluation of group sequential clinical trial designs. Stat. Med.26, 5047-5080 (2007).

[38] Perucca E. What clinical trial designs have been used to test antiepileptic drugs and do we need to change them? Epileptic Disord.14, 124-131 (2012).

[39] EMA, Concept paper on extrapolation of efficacy and safety in medicine development (2012).

[40] Anderson B.J. & Holford N.H. Understanding dosing: children are small adults, neonates are immature children. Arch. Dis. Child.98, 737-744 (2013).

[41] Hlavin G., Koenig F., Male C., Posch M. & Bauer P. Evidence, eminence and extrapolation. Statist. Med. In press. (2016)

Chapter 5

Model-based assessment of alternative study designs in pediatric trials: Bayesian approaches

Part of this Chapter was published in
Smania G, Baiardi P, Ceci A, Cella M, Magni P. Model-Based Assessment of Alternative Study Designs in Pediatric Trials. Part I: Bayesian Approaches. *CPT: Pharmacometrics & Systems Pharmacology*;5(8): 402-410, 2016.

5.1. Introduction

Chapter 4 introduced the need for alternative study designs in the implementation of Randomized Controlled Trials (RCTs) for the evaluation of efficacy and safety of an experimental treatment in the pediatric population and frequentist approaches of alternative designs were presented.

The current Chapter focuses on study designs for pediatric trials based on Bayesian approaches. Bayes' theorem allows one to make inference on observed data by incorporating a priori beliefs (usually defined in terms of a prior probability distribution) on the phenomenon being observed. From an RCT perspective, historical information on treatment effect (e.g., from previous studies) can be leveraged to infer the efficacy of the treatment being studied in the new RCT [1]. Consequently, compared to classical frequentist approaches, the amount of data to be collected in the new study is reduced because these data are augmented by historical ones. This ultimately allows reducing the sample size of the study.

Such property of Bayesian designs is of tremendous importance for pediatric trials, where the number of patients that can be recruited is often very limited. In particular, if the disease being studied in the pediatric

population is similar to the corresponding disease in an older population, available RCTs in the latter can be leveraged to elicit a prior distribution for treatment effect to be used in the analysis of the pediatric trial (e.g., adult data as prior in pediatric trials, adolescents data as prior for trials in children, children data as prior for trials in infants, etc.). Bayesian techniques are also endorsed by the European Medicine Agency (EMA) for the use of the extrapolation approach in pediatric drug development programs [2]. Nonetheless, few examples can be found in the literature on the application of Bayesian approaches in pediatric RCTs borrowing prior information from adults data [3-5]. Among Bayesian designs, those of sequential nature are of further potential interest because of their inherent flexibility, especially if compared with their frequentist counterparts [6]. Although limited, applications of Bayesian sequential designs spanned from early phase II anticancer trials [7, 8], safety monitoring [9, 10] and dose-finding studies [11], whilst applications to pediatric efficacy trials remains scant.

The scope of the present Chapter is to compare the performance of a standard Bayesian design with that of two Bayesian sequential designs by means of pharmacokinetic-pharmacodynamic (PK-PD) based clinical trial simulation (CTS). No attempts to evaluate Bayesian study designs through PK-PD CTS could have been identified in the literature. Designs are evaluated in terms of: type I and type II errors; sample size per arm (SS); total trial duration (TD) and precision of treatment effect estimate. Finally, results of these Bayesian designs are cross-compared with those of the frequentist designs presented in Chapter 4.

5.2. Methods

For information on the PK-PD model used to simulate data, the general study design (treatment groups, doses, primary endpoints and trial duration) and the general framework for CTS the reader should refer to Chapter 4. Only methods differentiating from such Chapter are reported thereafter.

In agreement with the Bayesian approach, the null hypothesis of no treatment difference H_0 was tested through the posterior probability of the improvement provided by topiramate (TPM, the drug under study as additional therapy to the current patient specific anti-epileptic treatment) over placebo (in addition to the current patient specific anti-epileptic treatment) in epileptic children (δ_P) after having observed the clinical trial data, i.e., $p(\delta_P|Data)$.

5.2.1. Study designs description

Bayesian design (BD)

In a two arm BD, patients are randomized to two parallel groups to receive either placebo or TPM, with the number of patients to be randomized in each group fixed a priori.

The statistical framework for the BD was adapted from Schoenfeld et al. [4]. Formally, let δ_A and δ_P be the true improvements of TPM over placebo in the adult and pediatric population, respectively. Their prior distribution is $\delta_A, \delta_P \sim N(\mu, \nu^2)$, where μ has a non-informative prior ($\mu \sim N(0, \sigma_\mu^2)$, $\sigma_\mu^2 \to +\infty$), while ν is given a fixed number reflecting the supposed difference in improvement of TPM over placebo between children and adults. The parameter ν plays a pivotal role in the design and analysis of the trial. Schoenfeld and coworkers suggest eliciting ν from clinical judgment or, if available, from previous pediatric and adult studies as $\nu = |\bar{\delta}_A - \bar{\delta}_P|/\sqrt{2}$, i.e. as an approximate estimate of the sample standard deviation of treatment difference, where $\bar{\delta}_A$ and $\bar{\delta}_P$ represents estimates of δ_A and δ_P obtained from historical data. Because in the PK-PD model used for CTS [12] both pediatric and adult data were modelled (Table 5.1), plausible values of ν were deduced from parameter estimates of the final PK-PD models in the two populations. In particular, Monte Carlo methods were used to obtained $\bar{\delta}_P$ from 10^6 samples, which was set equal to 0.2467 as in Chapter 4, whereas $\bar{\delta}_A$ (set to 0.5016) was obtained in the same way of $\bar{\delta}_P$ but using adult PK-PD parameters and an average adult TPM dose regimen of 150 mg BID [13, 14]. This lead to set ν=0.18 (hereafter called Scenario 1). In order to explore different scenarios and to take into account the plausible situation of a larger difference in TPM improvement over placebo between children and adults, ν was set to 0.4 (hereafter called Scenario 2, see Table 5.2).

Table 5.1: Parameter values from the PK-PD model used for clinical trial simulations in adult and pediatric epileptic patients, adapted from Girgis et al. [12].

	Pediatric	**Adults**
Placebo effect β_0 (%SE)	4.4830 (9.16)	4.4469 (3.13)
Concentration effect β_1 (%SE)	-0.0579 (3.05)	-0.0627 (0.97)
σ_ε	0.7517	

Furthermore, let $\hat{\delta}_A$ be the maximum likelihood estimate of TPM effect over placebo based on an adult trial with $m_A/2$ patients per arm, and $\hat{\delta}_P$ the same estimate in the new pediatric trial with $m_P/2$ patients per arm. $\hat{\delta}_A$ and $\hat{\delta}_P$were assumed to follow a normal distribution:

$$\hat{\delta}_A \sim N(\delta_A, {s_A^2}/{m_A})$$

$$\hat{\delta}_P \sim N(\delta_P, {s_P^2}/{m_P}),$$

with $s_A=2\sigma_A$ and $s_P=2\sigma_P$ (σ_A and σ_P are the standard deviations of Y in the adult and pediatric population, respectively). It was assumed $\sigma_A=\sigma_P=\sigma$, with the value of σ given by the PK-PD model and reported in Table 5.1. Quantitatively, adult prior information was incorporated by setting $\hat{\delta}_A$ to the model-derived value of $\bar{\delta}_A$ and m_A to 663, which corresponds to the number of adult patients used to identify the adult PK-PD model.

According to the statistical framework from Schoenfeld et al., since δ_P, $\hat{\delta}_P$ and $\hat{\delta}_A$ follow a multivariate normal distribution with mean 0 and covariance matrix

$$\begin{bmatrix} \sigma_\mu^2+\nu^2 & \sigma_\mu^2+\nu^2 & \sigma_\mu^2 \\ & \sigma_\mu^2+\nu^2+{s_P^2}/{m_P} & \sigma_\mu^2 \\ & & \sigma_\mu^2+\nu^2+{s_A^2}/{m_A} \end{bmatrix},$$

the posterior distribution of δ_P is its conditional distribution given $\hat{\delta}_P$ and $\hat{\delta}_A$. By letting $\sigma_\mu^2 \to +\infty$ it turns out that $\delta_P|\hat{\delta}_P,\hat{\delta}_A \sim N\left(\mu_{\delta_P}, {\sigma_{\delta_P}}^2\right)$ where

$$\mu_{\delta_P} = \frac{\frac{m_P}{s_P^2}\hat{\delta}_P+\frac{\omega}{s_A^2}\hat{\delta}_A}{\frac{m_P}{s_P^2}+\frac{\omega}{s_A^2}} \qquad (1)$$

$${\sigma_{\delta_P}}^2 = \frac{s_P^2 s_A^2}{m_P s_A^2+\omega s_P^2} \qquad (2),$$

with $\omega = \frac{m_A s_A^2}{s_A^2+2\nu^2 m_A}$. μ_{δ_P} depicts the Bayesian estimator of the improvement of TPM over placebo in the pediatric population.

In their work, Schoenfeld and colleagues provide a method to define a Bayesian analogue of classical frequentist power given by the following formula:

$$Bayesian\_Power(\delta_P^*) = \Phi\left(\frac{\sqrt{m_P}}{s_P}\left[\delta_P^* - \frac{{s_P}^2}{m_P}\left(z_{1-\alpha}\sqrt{\frac{m_P}{{s_P}^2}+\frac{\omega}{{s_A}^2}} - \frac{\omega}{{s_A}^2}\hat{\delta}_A\right)\right]\right) \quad (3)$$

with z_x being the x-th quantile of the standard normal distribution, δ_P^* the minimum clinically important difference in TPM versus placebo in the

pediatric population and Φ the normal cumulative distribution function. If $\nu \rightarrow \infty$ (no data are borrowed from adults), $\omega \rightarrow 0$ and equation (3) collapses to the classical frequentist power. Similarly to the approach used in the frequentist setting, the sample size of the study was identified by exploiting equation (3), fixing Bayesian power to 0.8, α to 0.05 and δ_P^* to $\bar{\delta}_P$, that is, to 0.2467. According to the calculated total sample size, at each iteration of step 3 of the CTS framework (see section 4.2.3) half of the patients were assigned to the placebo group and half to the TPM group. Step 4 consisted in H_0 acceptance/rejection based on the posterior probability of treatment effect: if $p\left(\delta_P > 0 \middle| \hat{\delta}_P, \hat{\delta}_A\right) \leq 1 - \alpha$ H_0 is accepted, otherwise it is rejected.

Table 5.2: Investigated scenarios for the evaluation of the performance of the Bayesian design and of the two Bayesian sequential designs.

Scenario	Bayesian design	Bayesian sequential design	
		Non-hierarchical	Semi-hierarchical
1	ν=0.184	n_T=16 n_P=16 p_s=0.99 p_f=0.5	ν=0.184 ω=32 p_s=0.99 p_f=0.5
2	ν=0.4	n_T=3.5 n_P=3.5 p_s=0.99 p_f=0.75	ν=0.4 ω=7 p_s=0.99 p_f=0.75

Bayesian sequential designs (BSD): non-hierarchical (NON-H) and semi-hierarchical (SEMI-H) framework

In a sequential design, statistical analyses are sequentially performed after the enrollment of groups of patients of predetermined size *G*. This allows early stopping of the trial for either efficacy or futility. In Chapter 4, two alternative implementations of frequentist sequential designs were presented. In the present Chapter, two Bayesian implementations of sequential designs adapted from Gsponer et al. [15] are considered: one in a non-hierarchical (NON-H) and one in a semi-hierarchical (SEMI-H) framework.

In the NON-H, prior information on δ_P is indirectly defined through prior distributions on the placebo and TPM response as $\theta_P \sim N(\theta_{P0}, {\sigma_P}^2/n_p)$ and $\theta_T \sim N(\theta_{T0}, {\sigma_T}^2/n_T)$, respectively. These could be elicited from historical trials where a placebo (or TPM) response of mean θ_{P0} (θ_{T0}) and standard deviation $\sigma_P(\sigma_T)$ was estimated from $n_p(n_T)$ patients. Consequently, the posterior distribution of δ_P will be normal with mean ($\mu_{\delta_P,i}$) and variance (${\sigma_{\delta_{P,i}}}^2$) at the *i*-th step of the analysis given by:

$$\mu_{\delta_P,i} =$$

$$\left[\frac{V_{Pi}/n_i}{{\sigma_P}^2 + V_{Pi}/n_i}\theta_{P0} + \left(\frac{{\sigma_P}^2}{{\sigma_P}^2 + V_{Pi}/n_i}\right)\bar{Y}_{Pi}\right]$$
$$-\left[\frac{V_{Ti}/n_i}{{\sigma_T}^2 + V_{Ti}/n_i}\theta_{T0} + \left(\frac{{\sigma_T}^2}{{\sigma_T}^2 + V_{Ti}/n_i}\right)\bar{Y}_{Ti}\right]$$

$${\sigma_{\delta_{P,i}}}^2 = \frac{{\sigma_P}^2 V_{Pi}/n_i}{{\sigma_P}^2 + V_{Pi}/n_i} + \frac{{\sigma_T}^2 V_{Ti}/n_i}{{\sigma_T}^2 + V_{Ti}/n_i}$$

where $\bar{Y}_{Ti}$ and $\bar{Y}_{Pi}$ are the aggregated sample means in the TPM and placebo group at *i*-th step (i.e., after n_i patients have been recruited in each group) and V_{Ti} and V_{Pi} are the corresponding sample variances. In agreement with the BD, it was assumed $\sigma_P=\sigma_T=\sigma_\varepsilon$.

The SEMI-H shares the same framework of the BD, but, since it is a sequential design, inferences from the posterior distribution of δ_P are sequentially made at each interim analysis. Accordingly, $\hat{\delta}_P$ and s_P are computed at each interim analysis rather than being estimated once at the end of the trial. The decisional criteria for trial success/failure (i.e., rejection/acceptance of H_0) used for BSD were the following:

$$\begin{cases} Success\ (H0\ rejected): p\left(\delta_P > 0 \middle| \hat{\delta}_P, \hat{\delta}_A\right) > p_s \\ Failure\ (H0\ accepted): p\left(\delta_P < \delta_{min} \middle| \hat{\delta}_P, \hat{\delta}_A\right) > p_f \end{cases} \quad (4)$$

where δ_{min} was set to 0.12, which corresponds to a 10% further decrease in seizure reduction for TPM 7 mg/kg/day against placebo, considering an average placebo seizure reduction of 21.5% (obtained from the PK-PD model). The parameters of the posterior distribution used to evaluate criteria (4) in the SEMI-H correspond to that of BD (equations (1) and (2)). Consistently with the BD, NON-H and SEMI-H performance have been investigated under two alternative scenarios in terms of ν, p_s and p_f (Table 5.2). Despite for NON-H ν is not explicitly defined, in the framework presented by Schoenfeld et al. [4] the number ω can be thought of as the number of patients that are borrowed from the adult data and used in the analysis of pediatric data. This allowed obtaining the values of n_P and n_T in the two scenarios of the NON-H starting from the corresponding values of ω in the SEMI-H.

BSD were simulated with *G*=20. In agreement with the sequential nature of these designs, for each CTS steps 3 and 4 of the procedure described in 4.2.3 were sequentially performed until trial success/failure was detected according to criteria (4). In particular, in step 3 *G* children were randomized to TPM and placebo in a 1:1 ratio and their simulated responses used to compute the posterior probabilities in (4) based on the NON-H and the SEMI-H, while in step 4 such probabilities were compared with their corresponding thresholds.

The same metrics used in Chapter 4 were used in the current Chapter, except for the percentage of exposure to placebo, TPM and no-treatment relative to total trial exposure, as this measure is equal to that of PaD considering the equal randomization to the two treatment arms. A description of the calculation of such metrics for each of the investigated designs is given hereafter.

Bayesian Design

a) $\hat{\alpha}$ = percentage of simulated trials with $p(\delta_P > 0 | \hat{\delta}_P, \hat{\delta}_A) > 0.95$ out of 1000 CTS when simulating under H0.
b) $\hat{\beta}$ = percentage of simulated trials with $p(\delta_P > 0 | \hat{\delta}_P, \hat{\delta}_A) \leq 0.95$ out of 1000 CTS when simulating under H1.
c) SS = m_P for which Bayesian power defined in (3) equals 80% (fixed a priori).
d) TD = $\frac{SS}{ER} + \tau$. ER= enrollment rate. In this study, τ= 2 months (trial duration per child).
e) EP = distribution of the 95% credible interval widths obtained at each trial simulation.

Bayesian sequential designs

a. $\hat{\alpha}$ = percentage of simulated trials with $p(\delta_P > 0 | \hat{\delta}_P, \hat{\delta}_A) > p_s$ out of 1000 CTS when simulating under H0.
b. $\hat{\beta}$ = percentage of simulated trials with $p(\delta_P < \delta_{min} | \hat{\delta}_P, \hat{\delta}_A) > p_f$ out of 1000 CTS when simulating under H1.
c. SS = distribution of the sample sizes obtained at each trial simulation.
d. TD= $\frac{G}{ER} + \max\left(0, \frac{G-ER\cdot\tau}{ER}\left(\left\lceil\frac{2SS}{G}\right\rceil - 2\right)\right) + \max\left(0, \frac{2SS-\left(\left\lceil\frac{2SS}{G}\right\rceil-1\right)\cdot G-ER\cdot\tau}{ER}\right) + \left\lceil\frac{2SS}{G}\right\rceil\tau$, where ⌈ ⌉ denotes the ceiling function.
e. EP = distribution of the 95% credible interval widths obtained at each trial simulation.

R code for simulation of BD, NON-H and SEMI-H under Scenario 1 can be found in Appendix 7.3.6, 7.3.7 and 7.3.8, respectively.

5.3. Results

5.3.1. Type I and type II errors

Scenario 1

Due to the inherently different philosophy of the Bayesian approach compared to the frequentist one, there is no control of type I error in the design of Bayesian trials and its value depends upon the weight of prior information. Because in the current analysis prior information came from a positive adult trial, type I error resulted in 22.2% for the BD, 26.3% for the NON-H and 24.1% for the SEMI-H. As expected, type II error for the BD is around its predetermined value of 20% (Table 5.3). In BSD, probably due to the increased type I errors, it is slightly smaller: 15.1% for the NON-H and 16.0% for the SEMI-H, despite it was not planned at the design stage.

Scenario 2

In the second scenario, assuming a larger difference in treatment response between children and adults implies that prior information from adult data has less weight. Consequently, since adult data tend to favor H_1, decisions will shift more towards H_0 rather than H_1 when compared to Scenario 1, especially when data are simulated under H_0. As a result, type I errors drop to about 7% in both the BD and the SEMI-H and to 9.4% in the NON-H (Table 5.3).

On the contrary, type II errors are only slightly increased for all investigated designs. For the BD this was expected because of the design of the trial, whereas for the two Bayesian sequential designs this is the result of an increased p_f.

Table 5.3: Performance metrics ($\hat{\alpha}$: type I error; $\hat{\beta}$: type II error; SS: sample size per arm; SS_{50}: median sample size per arm; SS_{75}, SS_{90}, SS_{95}: 75th, 90th and 95th percentiles of sample size distribution; TD: total trial duration) obtained from 1000 clinical trial simulation of the Bayesian design and of the two Bayesian sequential designs (non-hierarchical and semi-hierarchical). CI: confidence interval. ER: enrollment rate.

Metric \ Design		Bayesian design		Bayesian sequential design			
				Non-hierarchical		Semi-hierarchical	
		Scenario 1	Scenario 2	Scenario 1	Scenario 2	Scenario 1	Scenario 2
$\hat{\alpha}$ (%) (95% CI)		22.2 (19.6-24.8)	7.3 (5.7-8.9)	26.3 (22.6-29.0)	9.4 (7.6-11.2)	24.1 (21.4-26.8)	7 (5.4-8.6)
$\hat{\beta}$ (%) (95% CI)		20.3 (17.8-22.8)	21.1 (18.6-23.6)	15.1 (12.9-17.3)	19.5 (17.0-22.0)	16.0 (13.7-18.3)	19.4 (16.9-21.9)
E[SS] (patients)		49	103	37	67	37	66
SS_{50} (patients)				20	50	20	50
SS_{75} (patients)				50	90	50	90
SS_{90} (patients)				81	150	80	150
SS_{95} (patients)				110	190	110	190
MEDIAN TD (months)	ER=4 patients/ month	26.5	53.5	12	27	12	27
	ER=10 patients/ month	11.8	22.6	6	12	6	12

5.3.2. Sample size per arm

Scenario 1

SS for the BD is determined a priori on the basis of equation (3) and its relationship with ν is showed in Figure 5.1a. Unlike BD, SS of BSD is not known a priori and the histograms of SS achieved at each simulation in the two scenarios are depicted in Figure 5.1b.

For all designs the average SS required in Scenario 1 is lower than that of Scenario 2 because in the latter less weight is given to prior information from adult data. In the BD the required SS resulted in 49 children per arm, which approximately corresponds to the 75th percentile of SS distributions of both BSD (Table 5.3), distributions which do not significantly differ between each other (Figure 5.1b).

Average SS for the NON-H and SEMI-H is 37 (Table 5.3). Despite mean and median SS of BSD are lower than the BD one, sequential recruitment of children may be remarkably prolonged if treatment effect signals are captured later on during the trial, as shown by the right tail of the histograms in upper panel of Figure 5.1b. The probability of the NON-H and SEMI-H requiring a higher SS than the BD is about 27%.

Scenario 2

Similarly to Scenario 1, the distributions of the SS in the NON-H and SEMI-H do not remarkably differ (lower panel of Figure 5.1b). In the BD the SS resulted in 103 children per arm. This approximately corresponds to the 80th percentile of SS distribution of the Bayesian sequential designs (i.e., 110 patients), which require on average 67 (NON-H) and 66 (SEMI-H) patients per arm (Table 5.3). As in Scenario 1, sequential designs may go on very long before a decision can be taken (right tail of lower panel of Figure 5.1b). The magnitude of such "prolongation" is similar to the one observed under Scenario 1, as indicated by a probability of the NON-H and SEMI-H requiring a higher number of children than the BD of 21.5% and 21.2%, respectively.

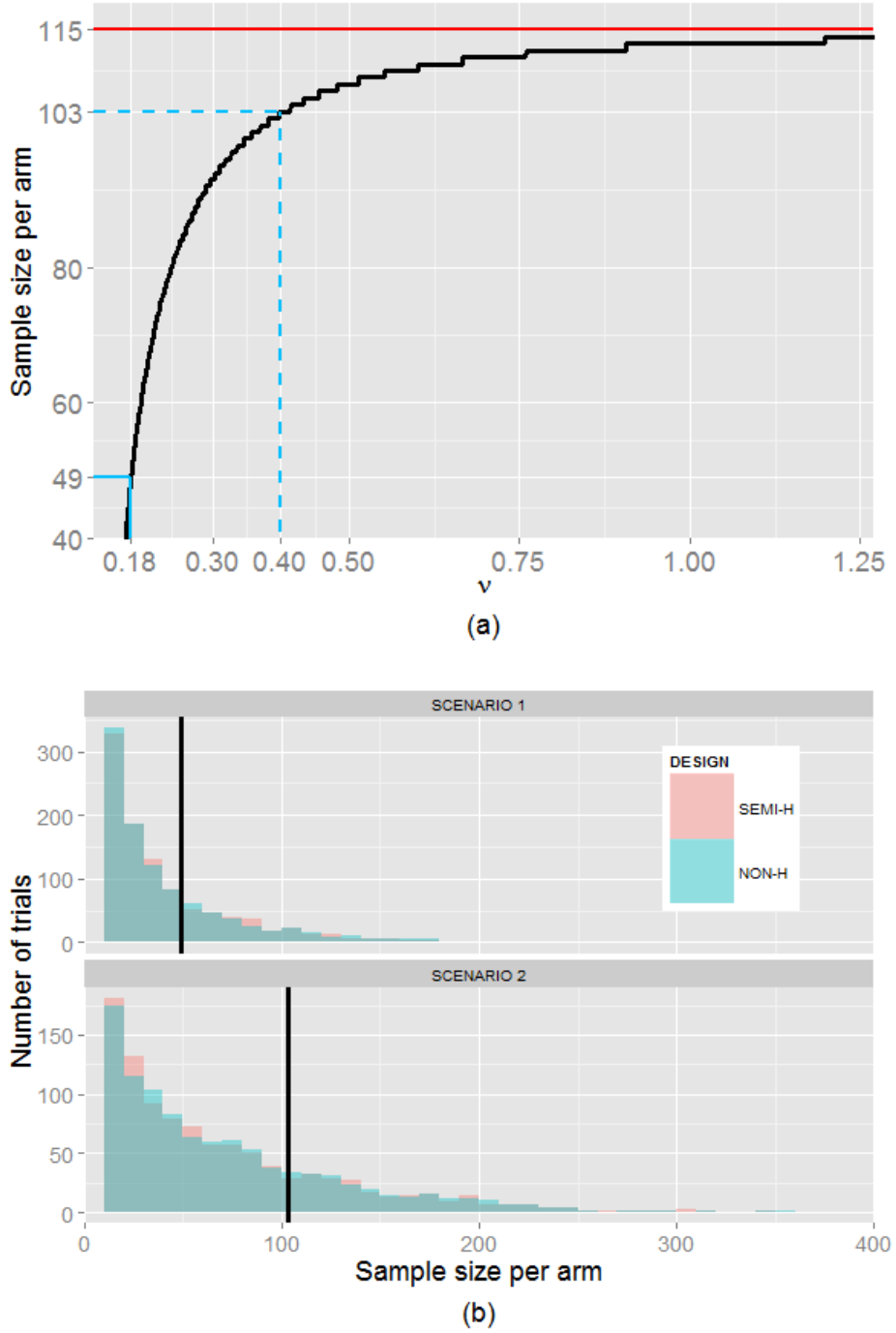


Figure 5.1. **(a)** Sample size per arm (SS) versus difference in improvement of TPM over placebo between children (δ_P) and adults (δ_A) expressed in terms of standard deviation of the prior distribution on δ_P and δ_A (ν). The red line represents the sample size per arm of a classical parallel frequentist design, azure lines indicate the value of ν and the corresponding SS of the Bayesian design in Scenario 1 (solid line) and 2 (dotted line). **(b)** Histograms of SSs obtained at each of the 1000 clinical trial simulation of the Bayesian sequential design in the non-hierarchical (NON-H, green histogram) and semi-hierarchical (SEMI-H, pink histogram) framework for Scenario 1 (upper panel) and 2 (lower panel). The black vertical lines indicate SS of the Bayesian design in the two scenarios.

5.3.3. Total trial duration

Scenario 1

TD as a function of enrollment rate is shown in Figure 5.2. The NON-H and SEMI-H have the lowest median duration among the three designs, which reflects the lowest median SS required (Table 5.3). Likely due to the lack of considerable differences between SS distributions in NON-H and SEMI-H, their medians and 95% prediction intervals in TD perfectly overlap.

Scenario 2

Similarly to what has been observed for Scenario 1, not only the 95% prediction intervals of TD are equal for the NON-H and the SEMI-H, but also the median TDs are. Median TDs of the Bayesian sequential designs are lower than the BD one (Table 5.3).

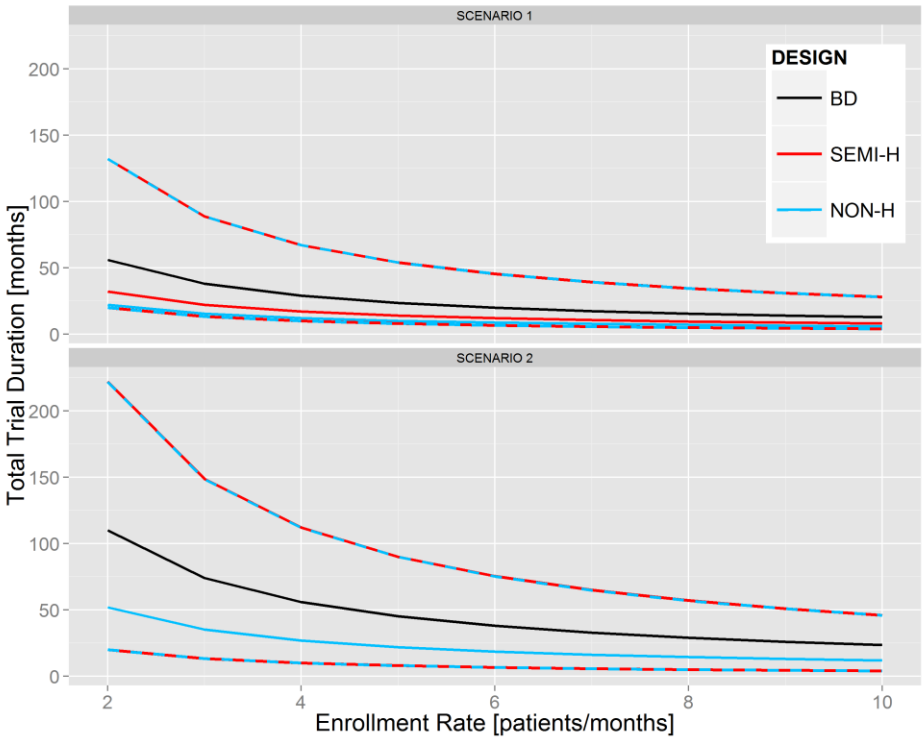
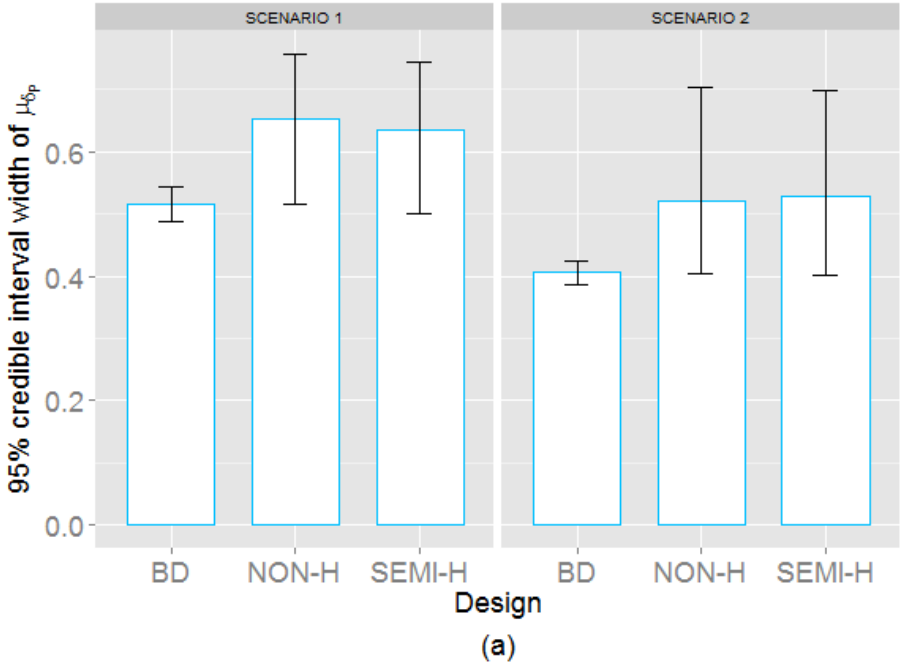


Figure 5.2. Total trial duration as a function of enrollment rate for the Bayesian design (black line) and the Bayesian sequential design in the non-hierarchical (NON-H, light blue lines) and semi-hierarchical (SEMI-H, orange lines) framework for Scenario 1 (upper panel) and 2 (lower panel). Solid lines represent the median duration whereas dotted lines depict 95% prediction intervals. Median and 95% prediction intervals of NON-H and SEMI-H are on top of each other in both scenarios.

5.3.4. Treatment difference estimate precision

Figure 5.3a shows that the BD leads on average to the shortest width of the 95% credible intervals thereby guaranteeing the highest precision in the Bayesian estimate of δ_P in both evaluated scenarios. When less weight is given to prior information on adult (right panel of Figure 5.3a), the precision appears to increase in all investigated designs.

No significant differences can be detected between the precision assured by NON-H and SEMI-H in both scenarios.



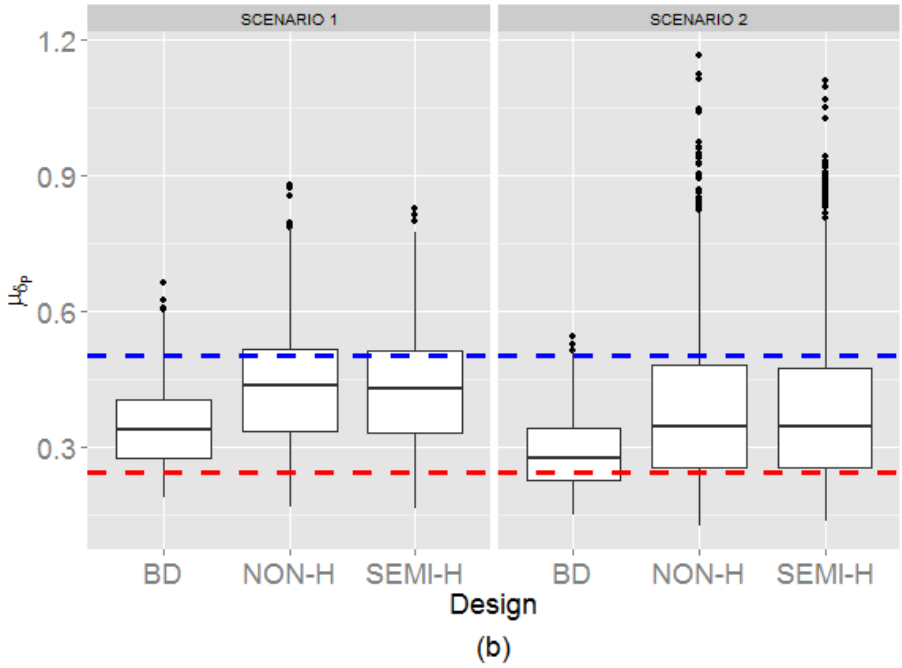


Figure 5.3. **(a)** Bar chart plot of the median 95% credible interval width of treatment difference estimates (μ_{δ_P}) obtained at each of the 1000 clinical trial simulation of the Bayesian design (BD), the Bayesian sequential design with a non-hierarchical (NON-H) and a semi-hierarchical (SEMI-H) framework for Scenario 1 (left panel) and 2 (right panel). The upper and lower "hinges" correspond to the first and third quartiles of 95% confidence intervals widths. **(b)** Boxplots of μ_{δ_P} obtained at each simulation of the BD, the NON-H and the SEMI-H for Scenario 1 (left panel) and 2 (right panel). The blue and red dotted lines represent the adult and pediatric treatment effects difference between topiramate and placebo (obtained from the PK-PD model), respectively.

5.4. Discussion

5.4.1. Comparison within Bayesian designs

General

PK-PD CTS provides a favorable tool to integrate prior information on drug disposition and effect to evaluate the performance of candidate study designs. In this Chapter, CTS was underpinned with a PK-PD model which was separately identified from both pediatric and adult data [12]. This enabled to formalize CTS of Bayesian designs by using prior information from adults data.

In particular, the difference in improvement of TPM over placebo between children and adults (ν) could be identified starting from parameter estimates of the PK-PD model. The decision on the value of ν is critical because it ultimately affects both the required SS and the final estimate of δ_P and the associated inference. However, the value of ν depends upon the specific problem and is not universal for all compounds and/or diseases. In the present analysis, for ν>1 the required SS of the BD tends towards that required by a standard frequentist design (Figure 5.1a). On the other hand, the value of ν obtained from PK-PD model parameters (Scenario 1) leads to a SS approximately 60% lower than that required in a frequentist setting, with clear advantages from a patients recruitment perspective.

Comparing both investigated scenarios, Figure 5.3b shows that, for all designs, the estimate of δ_P (i.e. μ_{δ_P}, the posterior mean of TPM improvement over placebo in seizures reduction) shifts towards the pediatric value given by the PK-PD model if a greater ν is considered. Also, Figure 5.3b suggests that BSD lead to an estimate of δ_P closer to the adult value when compared to their fixed-sample counterpart, partly because of the lower SS required by sequential designs, which makes μ_{δ_P} to rely more on prior (adult) information.

Should the pediatric PK-PD model not be available at the design stage, a model-based approach still provides considerable benefits in eliciting prior information. If for example children are expected to be twice as sensitive as adults (assumption that can be supported for instance by historical data from drugs with similar mechanism of action in children), ν can be derived by using the adult PK-PD model with a doubled drug-effect parameter. Consequently, the impact of such an assumption on designs performance can be quantitatively evaluated by means of the framework presented in this Chapter.

NON-H vs SEMI-H

Results show that there are no significant differences between the NON-H and the SEMI-H in both investigated scenarios and across all analyzed

metrics (Table 5.3). Such results were not totally unexpected as the weight of prior information in NON-H and SEMI-H is given in an equivalent manner. Nonetheless, because SEMI-H explicitly enables to weight prior information on adults on the grounds of clinical and scientific plausibility, it may be preferable to the NON-H where the weight of prior information is to be assigned based on an "equivalent sample size".

BSD vs BD

BSD, under comparable type I errors, require on average a lower SS than the BD (Table 5.3). With respect to mean SS, the reduction seen under Scenario 2 (~35%) is slightly higher than that observed under Scenario 1 (~24%). Though limited by only two scenarios, this suggests that the more the adult and pediatric populations are different the greater is the advantage in terms of mean SS brought by BSD compared to a fixed-sample BD. On the other hand, SS distributions of BSD become more skewed when passing from Scenario 1 to Scenario 2 (Figure 5.1b) and more caution is needed since there is a higher probability of ending the trial with non-practical SSs, which would ultimately jeopardize the conclusions that can be drawn from the study.

Differences in TD between BD and BSD can be explained by the aforementioned differences in SS. Figure 5.2 shows that, for the investigated range of enrolment rates, median TD of the NON-H and SEMI-H is lower than the duration of the BD. Like for SS, upper limit of 95% prediction interval of TD (upper dotted lines of Figure 5.2) clearly suggests that there exists a low probability of the trial lasting more than 100 months, which would thereby compromise its feasibility.

Regarding the precision of μ_{δ_P}, the BD performs better than the BSD with a median width of its 95% credible interval about 20% lower than the corresponding median of BSD (Figure 5.3a). The better precision of the BD is likely due to the higher number of samples used to compute μ_{δ_P}. Moving from Scenario 1 (left panel of Figure 5.3a) to Scenario 2 (right panel of Figure 5.3a), 95% median credible interval width consistently decreases for all designs, suggesting that on average the increase in precision due to a higher SS outweighs the decrease caused by a less informative prior. However, outliers of credible interval widths in BSD are higher in Scenario 2 than in Scenario 1 (results not shown). These less precise estimates are obtained when H_0 is rejected at the first interim analysis (i.e., when 10 patients per arm have been enrolled), revealing that for such low SS prior information is not strong enough to guarantee an acceptable precision of μ_{δ_P}.

Sensitivity analysis

In order to investigate the impact of non-negligible model misspecification on BD performance, a sensitivity analysis was performed by simulating the BD with a TPM effect parameter in children (β_1, Table

5.1) 5 times lower (i.e., β_1 =-0.01158; Modified Bayesian Design 1 (MBD1)) and 5 times higher (i.e., β_1 =-0.2895; Modified Bayesian Design 2 (MBD2)) than the estimated one.

Hypotheses made for the design and analysis of modified Bayesian designs were left unchanged, hence the SS of the MBD1 and MBD2 remained equal to the BD one. Table 5.4 indicates that type I errors are not affected by a misspecification of the drug effect parameter and can be considered equal to type I error of BD. This was not totally unexpected as type I errors are obtained by CTS under the null model, i.e., with β_1 set to zero. On the contrary, type II errors significantly shift towards opposite directions, depending on the side towards which β_1 was modified. In MBD1 type II error increases because simulations under H_1 assumed a lower effect with respect to the expected one, thus H_0 is accepted at higher rates compared to the BD. Vice versa, for the opposite reason, type II error of MBD2 approaches zero.

Table 5.4. Type I error ($\hat{\alpha}$) and type II error ($\hat{\beta}$) obtained from clinical trial simulation of the Bayesian design using a pediatric drug effect parameter (β_1) 5 times lower (MBD1) and 5 times higher (MBD2) than the estimated one (Table 5.1) under the two scenarios presented in Table 5.2.

	MBD1		**MBD2**	
	Scenario 1	**Scenario 2**	**Scenario 1**	**Scenario 2**
$\hat{\alpha}$ (%) (95% CI)	20.9 (18.4-23.4)	7.3 (5.7-8.9)	20.9 (18.4-23.4)	7.3 (5.7-8.9)
$\hat{\beta}$ (%) (95% CI)	71.4 (68.6-74.2)	84.3 (82.0-86.6)	0 (0-0)	0 (0-0)

As expected, the final estimate of treatment effect is slightly lower in MBD1 and remarkably higher in MBD2 (Figure 5.4b). This behavior is owed to the likelihood function being towards smaller treatment effect in MBD1 and towards much greater effects in MBD2. Importantly, treatment effect precision does not appear to be affected by model misspecification in MBD1, whereas a significant increase in credible interval widths is seen in MBD2 (Figure 5.4a). Such behavior of MBD2 can be explained again by means of the likelihood function: because the data are simulated under a larger treatment effect, the likelihood is far from the prior distribution (adult effect), which implies that the two only partially overlap between each other, leading to a spread posterior distribution.

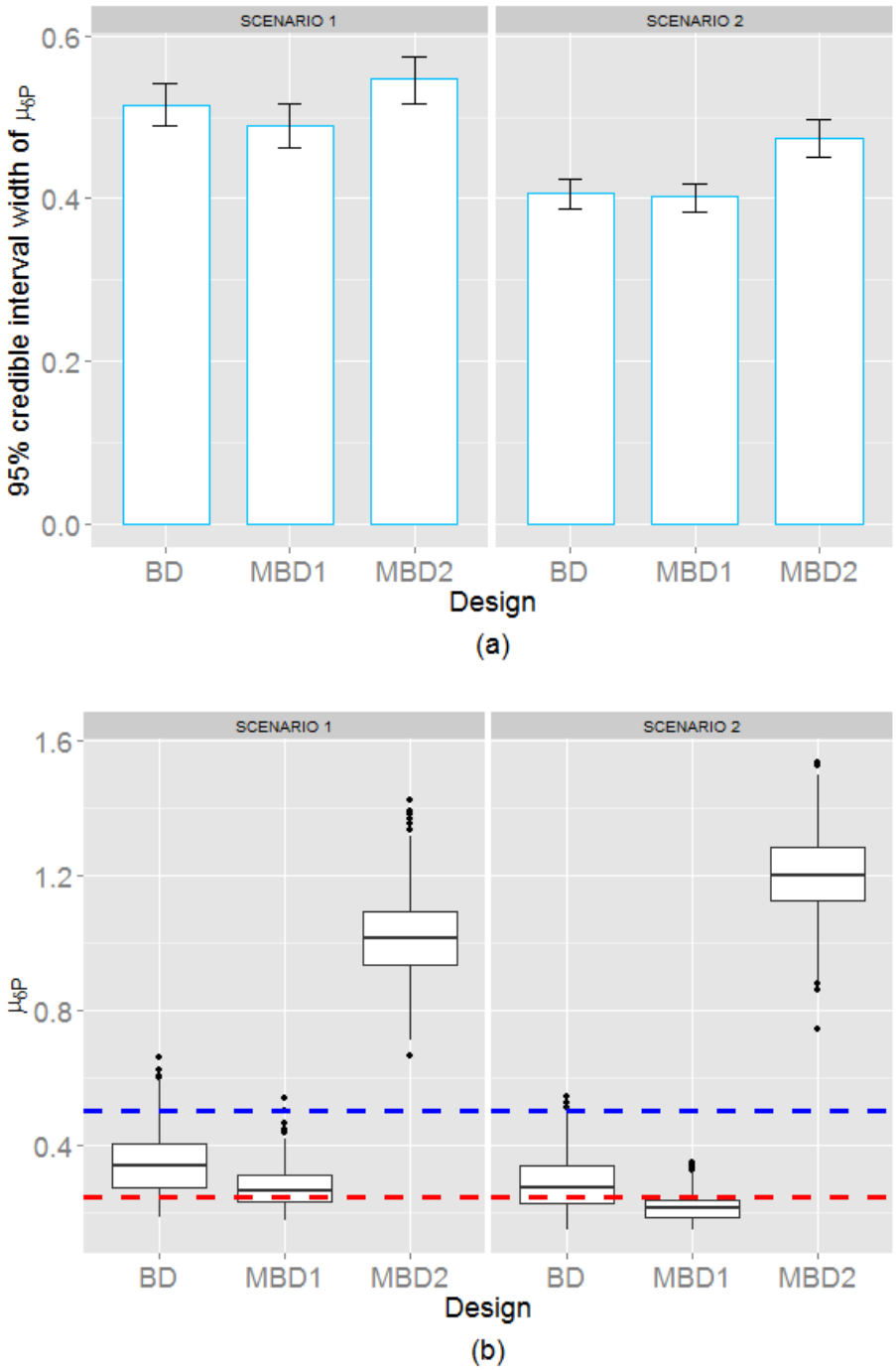


Figure 5.4. Bar chart plot of the 95% credible interval width of treatment difference estimates in pediatrics (μ_{δ_P}) **(a)** and boxplot of μ_{δ_P} **(b)** obtained at each simulation of the Bayesian design (BD), the Modified Bayesian Design 1 (MBD1) and the Modified Bayesian Design 2 (MBD2). The upper and lower "hinges" in Figure (a) correspond to the first and third quartiles of confidence intervals widths.

5.4.2. Comparison between Bayesian and frequentist approaches

Chapter 4 presented the performance of a set of alternative frequentist study designs (crossover, randomized withdrawal, sequential probability ratio test (SPRT) and triangular test (TT)) for pediatric trials and compared them with the standard parallel design (PaD). The present Chapter deals with the evaluation of Bayesian designs, whose comparison was based on the same metrics used for frequentist designs except for the percentage of exposure to placebo, TPM and no-treatment relative to total trial exposure.

One of the pivotal issues addressed by this work is the simultaneous comparison of a battery of alternative designs based on a pharmacometric model of the compound and the related placebo effect. Although interesting, comparing the goodness of Bayesian and frequentist approaches is not trivial because of the inherently different philosophy of these two methodologies and is still an open debate [16]. Scenario 1 highlights why Bayesian designs are appealing in pediatrics: the required SS is significantly reduced compared to that of their frequentist counterparts (BD vs PaD (fixed-sample designs) and SEMI-H vs SPRT/TT (sequential designs)). In particular, SS of BD (49 patients) is nearly 60% lower than that of the PaD (115 patients), while the SS distribution of the SEMI-H is squeezed towards lower SSs compared to that obtained with the SPRT/TT (left panel of Figure 5.5). Figure 5.6a shows that the reduced SS implies a remarkable lower precision of the BD estimate compared to that of the PaD (i.e. $\hat{\delta}$), whereas for BSD such difference is less pronounced because of the low precisions associated with the SPRT and TT; moreover, the estimated treatment effect is shifted towards the adult value for both fixed-sample (mean $\mu_{\delta_P}$of 0.3498) and sequential (mean μ_{δ_P}equals 0.4330 and 0.4265 for the NON-H and SEMI-H, respectively) Bayesian designs when compared to the corresponding fixed-sample (mean $\widehat{\boldsymbol{\delta}}$ of 0.2821) and sequential (mean $\hat{\delta}$ equals 0.3717 and 0.3588 for the SPRT and TT, respectively) frequentist designs (Figure 5.6b).

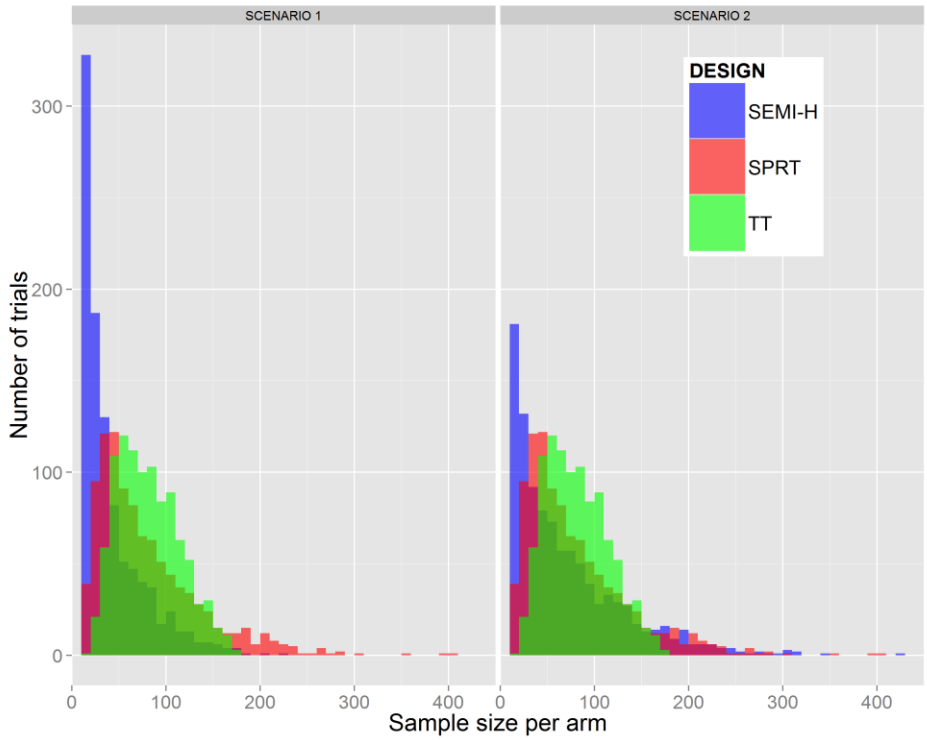
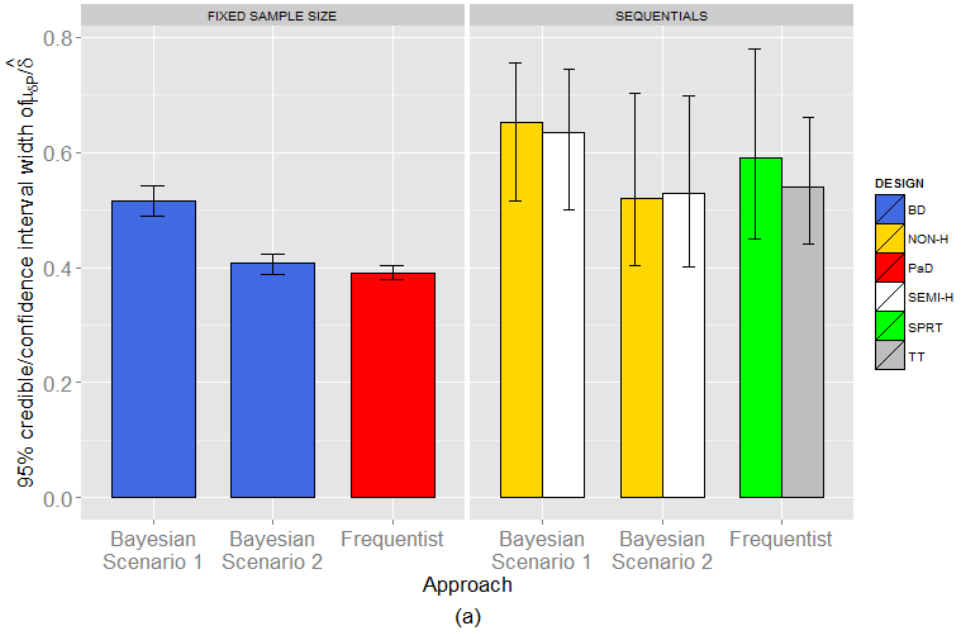


Figure 5.5. Histograms of the sample sizes per arm obtained at each of the 1000 clinical trial simulation of the semi-hierarchical Bayesian Sequential Design (SEMI-H, blue histogram), Sequential Probability Ratio Test (SPRT, red histogram) and the Triangular Test (TT, green histogram).

Different considerations can be made if frequentist designs are compared with Bayesian designs under Scenario 2.

In terms of SS, the BD allows to reduce the number of children to be enrolled by almost 10% when compared to the PaD while maintaining similar estimates of treatment effect and associated precisions (Figure 5.6), suggesting that the estimate obtained with a BD (mean μ_{δ_P} of 0.2905) is not significantly influenced by the adult prior distribution under Scenario 2.

Similarly to what has been observed when comparing fixed-sample designs, median SS of the SEMI-H (50 patients) is lower than the corresponding value of the SPRT (60 patients) and TT (70 patients). Better performance provided by Bayesian designs with respect to this metric can also be deduced from Figure 5.5, where it can be seen that SS histograms of the SPRT ant TT are shifted towards higher sample sizes compared to SEMI-H. However, SEMI-H seems to behave similarly to the SPRT in terms of very late stopping recruitment, i.e., low probabilities exist that the trial goes on very long, as indicated by a 95th percentile in SS distribution of 190.



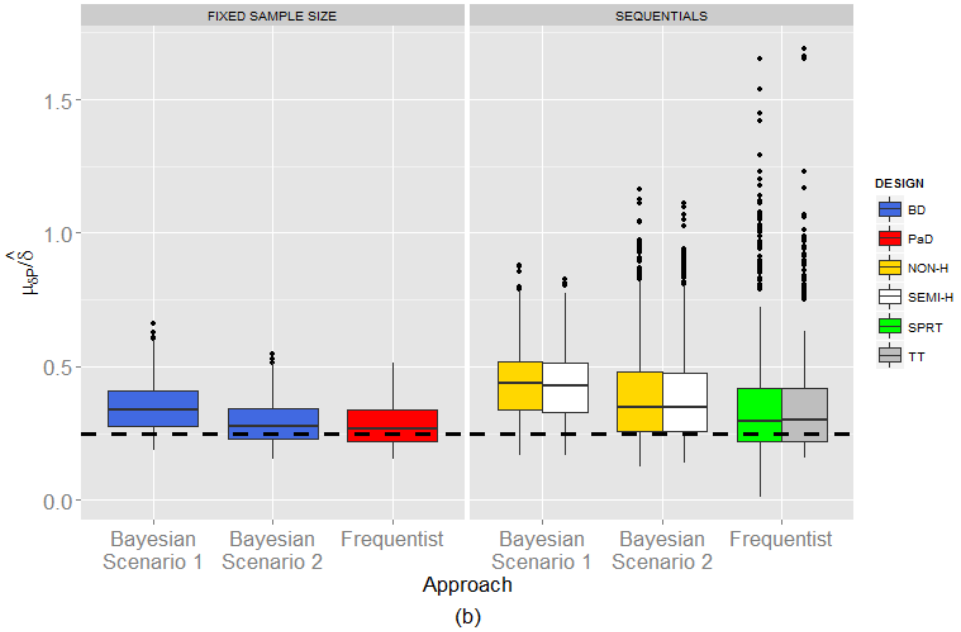


Figure 5.6. Bar chart plot of the 95% credible (Bayesian) and confidence (frequentist) interval width of treatment difference estimates in pediatrics (μ_{δ_P} for Bayesian designs and $\hat{\delta}$ for frequentist ones) **(a)** and boxplot of μ_{δ_P} and $\hat{\delta}$ **(b)** obtained at each of the 1000 clinical trial simulation of the Bayesian design (BD, blue bar), Parallel design (PaD, red bar), Bayesian sequential design with a non-hierarchical (NON-H, yellow bar) and a semi-hierarchical (SEMI-H, white bar) framework, Sequential Probability Ratio Test (SPRT, green bar) and Triangular Test (TT, gray bar). The upper and lower "hinges" in subfigure (a) correspond to the first and third quartiles of 95% credible/confidence intervals widths. The dashed horizontal black line represents the pediatric treatment effect difference between topiramate and placebo obtained from the PK-PD model (0.2467).

Right panel of Figure 5.6a reveals that no considerable differences in precision are seen between BSD and the TT (the frequentist sequential approach with the highest precision), even though Bayesian approaches seem to be slightly more robust to outliers (results not shown). In addition, equivalently to fixed-sample designs, the adult prior distribution does not remarkably influence the estimated effect in pediatrics in the NON-H (mean μ_{δ_P} of 0.3964) and SEMI-H (mean μ_{δ_P} of 0.3939) compared to the SPRT and TT (Figure 5.6b).

In a way, the comparison made under Scenario 2 could be considered fairer because type I errors obtained under such scenario (around 7-9%) are closer to those obtained in frequentist designs (around 5-7%). On the other hand the increased type I error rate of Bayesian designs observed under Scenario 1 is inherently due to the inclusion of a positive adult study and should be accepted as such.

With respect to the extrapolation of adult results to pediatric trials, Hlavin et al. [17] proposed a statistical framework to quantitatively accommodate the uncertainty about the assumptions on the similarity between the adult and pediatric population by enlarging the significance level of the pediatric trial based on experts skepticism. Although in the framework by Hlavin and colleagues Bayesian arguments are applied to calibrate the increase in the significance level, their approach is frequentist by nature and it is does not provide a clear way on how to quantitatively derive the skepticism factor on the basis of the expected similarities/differences between the two populations. Nevertheless, the condition of no skepticism can be translated into the condition $\nu=0$ of Schoenfeld et al. [4] (children and adults respond in the same way to the drug under study and no pediatric efficacy trial would be needed). Similarly, full skepticism can be converted into $\nu \rightarrow \infty$. Accordingly, for values of ν approaching the standard frequentist method ($\nu \sim 2$), the type I error obtained in this analysis corresponds to the adjusted α value proposed by Hlavin et al. in case of full skepticism, that is, 0.05 (i.e., no adjustment).

Although advantages and disadvantages of the investigated designs concerning the evaluation of dose regimens in the pediatric population were not explicitly considered, some general properties can still be outlined. Since BSD and BD are parallel in nature, they differ solely in terms of SS with respect to the estimation of the PK and/or PK-PD in children; as a result, on average, BD is expected to provide more precise PK/PK-PD estimates compared to BSD. Similarly, when significant weight is given to prior information on adult treatment effect, Bayesian designs would lead to estimates with poorer precision in contrast with frequentist ones; however, if also the PK/PK-PD in children is expected to be similar to that observed in adults, prior information on adult PK/PK-PD parameters can be leveraged to improve the precision of the estimates in the pediatric population and to ultimately provide an optimal dose selection [5].

Regulatory endorsements on the use of Bayesian designs for pediatric trials are present: EMA suggests using Bayesian approaches in Pediatric

Investigation Plans [2], whereas the Food and Drug Administration published a "Guidance for the Use of Bayesian Statistics in Medical Device Clinical Trials" [18]. Nevertheless few examples can be found in the literature. According to Gönen [19], the barriers to entry are many, but three stand out: prior, software, and motivation, where motivation seems to be the major one. Tradition also represents an additional hurdle, which is anyway related to motivation. Pediatric trials call for innovation and may therefore offer the opportunity to overcome these motivational issues and increase usage of Bayesian approaches.

It has to be pointed out that this analysis is based on the effect of TPM in children with partial onset seizures refractory to their current antiepileptic treatment, and the extrapolation of the results to different compounds/diseases/subpopulations should be further explored.

In conclusion, this Chapter provided a pharmacometric framework able to formalize PK-PD based CTS for Bayesian designs in pediatric trials using prior information from adult data, thereby allowing to investigate the influence of a specific study design on success/failure of a pediatric trial. With respect to the selection of a particular design, if prior information is available from adult studies but children are expected to respond substantially different from adults (Scenario 2), the performance of frequentist and Bayesian approaches can be assumed comparable, with slight advantages for the latter. However, when the pediatric population is expected to respond similarly to adults (Scenario 1), Bayesian designs would allow smaller, shorter, more reliable and more efficient trials in children. Among Bayesian designs, those of sequential nature, irrespective of their level of hierarchy, seem to require lower SS compared to the BD if larger treatment effects are expected, and could therefore represent an appealing options for trials in very small populations.

References

[1] Berry D.A. Bayesian clinical trials. Nat. Rev. Drug. Discov. 5, 27-36 (2006).

[2] EMA, Concept paper on extrapolation of efficacy and safety in medicine development.(2012).

[3] Goodman S.N. & Sladky J.T. A Bayesian approach to randomized controlled trials in children utilizing information from adults: the case of Guillain-Barre syndrome. Clin. Trials. 2, 305-310 (2005).

[4] Schoenfeld D.A., Hui Z. & Finkelstein D.M. Bayesian design using adult data to augment pediatric trials. Clin. Trials. 6, 297-304 (2009).

[5] Cella M., Gorter de Vries F., Burger D., Danhof M. & Della Pasqua O. A model-based approach to dose selection in early pediatric development. Clin. Pharmacol. Ther. 87, 294-302 (2010).

[6] Berry D.A. Interim analyses in clinical trials: classical vs. Bayesian approaches. Stat. Med.4, 521-526 (1985).

[7] Cronin K.A., Freedman L.S., Lieberman R., Weiss H.L., Beenken S.W. & Kelloff G.J. Bayesian monitoring of phase II trials in cancer chemoprevention. J. Clin. Epidemiol. 52, 705-711 (1999).

[8] Thall P.F. & Simon R. Practical Bayesian guidelines for phase IIB clinical trials. Biometrics. 50, 337-349 (1994).

[9] Chen C. & Chaloner K. A Bayesian stopping rule for a single arm study: With a case study of stem cell transplantation. Stat. Med. 25, 2956-2966 (2006).

[10] Yutaka M. & Yasuo O. A Bayesian Stopping Rule for Sequential Monitoring of Serious Adverse Events. Ther. Innov. Regul. Sci. 48, 444-452 (2014).

[11] Guede D. et al. Bayesian adaptive designs in single ascending dose trials in healthy volunteers. Br. J. Clin. Pharmacol. 78, 393-400 (2014).

[12] Girgis I.G. et al. Pharmacokinetic-pharmacodynamic assessment of topiramate dosing regimens for children with epilepsy 2 to <10 years of age. Epilepsia. 51, 1954-1962 (2010).

[13] FDA. Topomax Label (2012).

[14] EMA. Topomax package leaflet (2009).

[15] Gsponer T., Gerber F., Bornkamp B., Ohlssen D., Vandemeulebroecke M. & Schmidli H.A. Practical guide to Bayesian group sequential designs. Pharm. Stat. 13, 71-80 (2014).

[16] Lee J.J. & Chu C.T. Bayesian clinical trials in action. Stat. Med. 31, 2955-2972 (2012).

[17] Hlavin G., Koenig F., Male C., Posch M. & Bauer P. Evidence, eminence and extrapolation. Statist. Med. In press. (2016)

[18] FDA. Guidance for the Use of Bayesian Statistics in Medical Device Clinical Trials (2010).

[19] Gönen M. Bayesian clinical trials: no more excuses. Clin. Trials. 6, 203-204 (2009).

Chapter 6

Overall conclusion

Pediatric drug development encounters many difficulties due to ethical, practical and financial considerations. In order to overcome these obstacles and to ensure that unmet medical needs are addressed in a vulnerable population as the pediatric one, the most efficient and informative analytical methods should be used when designing, analyzing and interpreting pediatric clinical studies.

Development of a medicine in adults provides a rich source of data to inform pediatric programs and, given reasonable similarities between the two populations, extrapolation from adults allows reducing pediatric data requirements to make conclusions for drug usage in children. Contextually, Bayesian approaches offer a scientifically sound methodology for quantitatively incorporate adult data in the design and analysis of pediatric trials. Although in certain circumstances, such as for diseases specific to children, no information can be borrowed from adults or other source populations, feasibility of pediatric trials can still be increased by more classical yet alternative frequentist designs such as the randomized withdrawal and sequential designs.

As a common background, pharmacokinetic-pharmacodynamic (PK-PD) modelling and simulation (M&S) should be employed throughout pediatric development programs. Indeed, as illustrated in Chapter 3, M&S enables rationale synthesis of available evidence and allows optimizing the design of the study thereby minimizing the set of data to be generated in the pediatric population. Moreover, Chapter 4 and 5 demonstrated the usefulness of PK-PD-based clinical trial simulation to quantitatively inform and guide the selection of the best design for pediatric trials.

These particularly innovative methodological approaches are necessary tools to provide children with new treatment options that would allow reducing the large number of unmet medical needs that are currently affecting this population.

Chapter 7

Appendix

7.1. NONMEM code for pharmacokinetic simulation

NONMEM code used for PK simulation (step 2 of section 4.2.3).

```
$PROBLEM    TPM GIRGIS PK MODEL: 500 steady state
simulations of 1000 subjects used for step 2 of CTS
framework
$INPUT ID TIME DOSE AMT SS II CONC=DV MDV EVID SEX BW
AGE ADJ


$DATA PKsim_dataset1SS.csv IGNORE=@
$SUBROUTINE ADVAN4 TRANS1

$PK  ;from supplemental material Girgis et al

SIMUL=IREP

CLST=THETA(1)*(1+ADJ*THETA(2))
FCWT=(BW/69.9)**THETA(3)
FCAGE=EXP(THETA(4)*(AGE-31.4))
FCIN= THETA(5)**0
FCVP= THETA(6)**0
FCNE= THETA(7)**0
TVCL = CLST*FCWT*FCAGE*FCIN*FCVP*FCNE
CL=TVCL*EXP(ETA(1))
FVWT=(BW/69.9)**THETA(9)
V=THETA(8)*FVWT*EXP(ETA(2))
S2=V
KA=THETA(10)*EXP(ETA(3))
K23=THETA(11)
K32=THETA(12)
K=CL/V
```

```
$ERROR

IPRED = F
W = F
IF (W.LE.0) W = 0.00001
      IRES = DV-IPRED
      IWRES = IRES/W

      Y = F*(1+EPS(1)) + EPS(2)

$THETA
1.21         ;CLSTM
0.479        ;CLSTA
0.453        ;FCWT
-0.00306     ;FCAGE
1.94         ;FCIN
0.686        ;FCVP
0.635        ;FCNE
4.61         ;VST
1.14         ;FVWT
0.105        ;KA
0.577        ;K23
0.0586       ;K32

$OMEGA
0.07441984   ;CL --> (27.28/100)^2
1.350244     ;V  --> (116.2/100)^2
0.04990756   ;KA --> (22.34/100)^2

$SIGMA
0.06482116 ; sigma1^2 = (25.46/100)^2
0.03229209 ; sigma2^2 = 0.1797^2

$SIMULATION (12101987) ONLYSIM SUBPROBLEMS=500

$TABLE ID TIME AMT DOSE DV BW AGE MDV SIMUL
       NOHEADER NOPRINT FILE=sim1.tab
```

7.2. R code for covariates simulation

R code used for covariates simulation (step 1 of section 4.2.3).

```
#########################################################
########### Step 1 fo CTS: covariate simulations #########
#########################################################

rm(list=ls(all=TRUE))
load("BOYS")
load("GIRLS")

# setting the seed for results replication
set.seed(5)

# nsbj --> number of subjects (i.e. covariates) to be simulated in the two datasets
## Two datasets are needed because NONMEM 7.2 does not accept dataset with more than
## 1000 IDs

nsbj <- 1000

# d1 --> first dataset
d1 <- data.frame(ID=rep(1:nsbj))
# d2 --> second dataset
d2 <- data.frame(ID=rep(1:nsbj))
```

```
# TIME --> time column for NONMEM simulations (through concentration are sought)
d1$TIME <- 12
d2$TIME <- 12

# MGKG --> daily dose in mg/kg
MGKG <- 7

# dosing_interval --> 3.5 mgkg twice daily
dosing_interval <- MGKG/2

# DOSE --> nonmem dose column
d1$DOSE <- rep(dosing_interval,each=nsbj/length(dosing_interval))
d2$DOSE <- rep(dosing_interval,each=nsbj/length(dosing_interval))

# AMT --> amount nonmem column
d1$AMT <- NA
d2$AMT <- NA

# SS --> steady state NONMEM column (indicating that simulated conc are at steady state)
d1$SS <- NA
d2$SS <- NA

# II --> nonmem column indicating the time interval between two dosing
d1$II <- NA
d2$II <- NA

# CONC --> nonmem conentration column
d1$CONC <- NA
d2$CONC <- NA

# MDV --> nonmem missing dependet variable column
d1$MDV <- 0
d2$MDV <- 0

# EVID --> nonmem EVID coulmn
d1$EVID <- 0
d2$EVID <- 0

# SEX --> GENDER was not a covariate of the PK model, but it is needed because it will
#         determine age and weight values. Males and females have the same probability
#         to be included (as indicated by probs=0.5,0.5)
# SEX=0--> MALE
# SEX=1--> FEMALE
d1$SEX <- sample(c(0,1),size=nsbj,replace=T,prob=c(0.5,0.5))
d2$SEX <- sample(c(0,1),size=nsbj,replace=T,prob=c(0.5,0.5))


# BW --> patients' weight
d1$BW<-rep(0,nsbj)
d2$BW<-rep(0,nsbj)


# AGE --> simulated ages in the first dataset, assume mean values--> 2.5,3.5,4.5,....
# apart from age 10, for which wieght-age table was not available
d1$AGE <- floor(runif(nsbj,2,11)) + 0.5
d1$AGE[d1$AGE==10.5] <- 10.0
# AGE --> simulated ages in the second dataset
d2$AGE <- floor(runif(nsbj,2,11)) + 0.5
d2$AGE[d2$AGE==10.5] <- 10.0

# ADJ --> PK model binary covariate which is 1 if patients take more than 1 drug, 0
otherwise
d1$ADJ <- 1
d2$ADJ <- 1

## body weights are extracted from WHO statistics. BOYS and GIRLS datasets contain
## the mean and sd of the logarithms of body weights for different ages

## MALES
# d1
d1$BW[d1$AGE==2.5 & d1$SEX==0] <- exp(rnorm(length(d1$BW[d1$AGE==2.5 & d1$SEX==0]),
                                            BOYS$MEAN_Lbw[BOYS$AGE==2.5],
                                            BOYS$SD_Lbw[BOYS$AGE==2.5]))

d1$BW[d1$AGE==3.5 & d1$SEX==0] <- exp(rnorm(length(d1$BW[d1$AGE==3.5 & d1$SEX==0]),
                                            BOYS$MEAN_Lbw[BOYS$AGE==3.5],
                                            BOYS$SD_Lbw[BOYS$AGE==3.5]))

d1$BW[d1$AGE==4.5 & d1$SEX==0] <- exp(rnorm(length(d1$BW[d1$AGE==4.5 & d1$SEX==0]),
                                            BOYS$MEAN_Lbw[BOYS$AGE==4.5],
                                            BOYS$SD_Lbw[BOYS$AGE==4.5]))

d1$BW[d1$AGE==5.5 & d1$SEX==0] <- exp(rnorm(length(d1$BW[d1$AGE==5.5 & d1$SEX==0]),
                                            BOYS$MEAN_Lbw[BOYS$AGE==5.5],
                                            BOYS$SD_Lbw[BOYS$AGE==5.5]))

d1$BW[d1$AGE==6.5 & d1$SEX==0] <- exp(rnorm(length(d1$BW[d1$AGE==6.5 & d1$SEX==0]),
                                            BOYS$MEAN_Lbw[BOYS$AGE==6.5],
                                            BOYS$SD_Lbw[BOYS$AGE==6.5]))

d1$BW[d1$AGE==7.5 & d1$SEX==0] <- exp(rnorm(length(d1$BW[d1$AGE==7.5 & d1$SEX==0]),
                                            BOYS$MEAN_Lbw[BOYS$AGE==7.5],
                                            BOYS$SD_Lbw[BOYS$AGE==7.5]))

d1$BW[d1$AGE==8.5 & d1$SEX==0] <- exp(rnorm(length(d1$BW[d1$AGE==8.5 & d1$SEX==0]),
```

```
                                          BOYS$MEAN_Lbw[BOYS$AGE==8.5],
                                          BOYS$SD_Lbw[BOYS$AGE==8.5]))

d1$BW[d1$AGE==9.5 & d1$SEX==0] <- exp(rnorm(length(d1$BW[d1$AGE==9.5 & d1$SEX==0]),
                                          BOYS$MEAN_Lbw[BOYS$AGE==9.5],
                                          BOYS$SD_Lbw[BOYS$AGE==9.5]))

d1$BW[d1$AGE==10.0 & d1$SEX==0] <- exp(rnorm(length(d1$BW[d1$AGE==10.0 & d1$SEX==0]),
                                           BOYS$MEAN_Lbw[BOYS$AGE==10.0],
                                           BOYS$SD_Lbw[BOYS$AGE==10.0]))

# d2
d2$BW[d2$AGE==2.5 & d2$SEX==0] <- exp(rnorm(length(d2$BW[d2$AGE==2.5 & d2$SEX==0]),
                                          BOYS$MEAN_Lbw[BOYS$AGE==2.5],
                                          BOYS$SD_Lbw[BOYS$AGE==2.5]))

d2$BW[d2$AGE==3.5 & d2$SEX==0] <- exp(rnorm(length(d2$BW[d2$AGE==3.5 & d2$SEX==0]),
                                          BOYS$MEAN_Lbw[BOYS$AGE==3.5],
                                          BOYS$SD_Lbw[BOYS$AGE==3.5]))

d2$BW[d2$AGE==4.5 & d2$SEX==0] <- exp(rnorm(length(d2$BW[d2$AGE==4.5 & d2$SEX==0]),
                                          BOYS$MEAN_Lbw[BOYS$AGE==4.5],
                                          BOYS$SD_Lbw[BOYS$AGE==4.5]))

d2$BW[d2$AGE==5.5 & d2$SEX==0] <- exp(rnorm(length(d2$BW[d2$AGE==5.5 & d2$SEX==0]),
                                          BOYS$MEAN_Lbw[BOYS$AGE==5.5],
                                          BOYS$SD_Lbw[BOYS$AGE==5.5]))

d2$BW[d2$AGE==6.5 & d2$SEX==0] <- exp(rnorm(length(d2$BW[d2$AGE==6.5 & d2$SEX==0]),
                                          BOYS$MEAN_Lbw[BOYS$AGE==6.5],
                                          BOYS$SD_Lbw[BOYS$AGE==6.5]))

d2$BW[d2$AGE==7.5 & d2$SEX==0] <- exp(rnorm(length(d2$BW[d2$AGE==7.5 & d2$SEX==0]),
                                          BOYS$MEAN_Lbw[BOYS$AGE==7.5],
                                          BOYS$SD_Lbw[BOYS$AGE==7.5]))

d2$BW[d2$AGE==8.5 & d2$SEX==0] <- exp(rnorm(length(d2$BW[d2$AGE==8.5 & d2$SEX==0]),
                                          BOYS$MEAN_Lbw[BOYS$AGE==8.5],
                                          BOYS$SD_Lbw[BOYS$AGE==8.5]))

d2$BW[d2$AGE==9.5 & d2$SEX==0] <- exp(rnorm(length(d2$BW[d2$AGE==9.5 & d2$SEX==0]),
                                          BOYS$MEAN_Lbw[BOYS$AGE==9.5],
                                          BOYS$SD_Lbw[BOYS$AGE==9.5]))

d2$BW[d2$AGE==10.0 & d2$SEX==0] <- exp(rnorm(length(d2$BW[d2$AGE==10.0 & d2$SEX==0]),
                                           BOYS$MEAN_Lbw[BOYS$AGE==10.0],
                                           BOYS$SD_Lbw[BOYS$AGE==10.0]))


## FEMALES
# d1
d1$BW[d1$AGE==2.5 & d1$SEX==1] <- exp(rnorm(length(d1$BW[d1$AGE==2.5 & d1$SEX==1]),
                                          GIRLS$MEAN_Lbw[GIRLS$AGE==2.5],
                                          GIRLS$SD_Lbw[GIRLS$AGE==2.5]))

d1$BW[d1$AGE==3.5 & d1$SEX==1] <- exp(rnorm(length(d1$BW[d1$AGE==3.5 & d1$SEX==1]),
                                          GIRLS$MEAN_Lbw[GIRLS$AGE==3.5],
                                          GIRLS$SD_Lbw[GIRLS$AGE==3.5]))

d1$BW[d1$AGE==4.5 & d1$SEX==1] <- exp(rnorm(length(d1$BW[d1$AGE==4.5 & d1$SEX==1]),
                                          GIRLS$MEAN_Lbw[GIRLS$AGE==4.5],
                                          GIRLS$SD_Lbw[GIRLS$AGE==4.5]))

d1$BW[d1$AGE==5.5 & d1$SEX==1] <- exp(rnorm(length(d1$BW[d1$AGE==5.5 & d1$SEX==1]),
                                          GIRLS$MEAN_Lbw[GIRLS$AGE==5.5],
                                          GIRLS$SD_Lbw[GIRLS$AGE==5.5]))

d1$BW[d1$AGE==6.5 & d1$SEX==1] <- exp(rnorm(length(d1$BW[d1$AGE==6.5 & d1$SEX==1]),
                                          GIRLS$MEAN_Lbw[GIRLS$AGE==6.5],
                                          GIRLS$SD_Lbw[GIRLS$AGE==6.5]))

d1$BW[d1$AGE==7.5 & d1$SEX==1] <- exp(rnorm(length(d1$BW[d1$AGE==7.5 & d1$SEX==1]),
                                          GIRLS$MEAN_Lbw[GIRLS$AGE==7.5],
                                          GIRLS$SD_Lbw[GIRLS$AGE==7.5]))

d1$BW[d1$AGE==8.5 & d1$SEX==1] <- exp(rnorm(length(d1$BW[d1$AGE==8.5 & d1$SEX==1]),
                                          GIRLS$MEAN_Lbw[GIRLS$AGE==8.5],
                                          GIRLS$SD_Lbw[GIRLS$AGE==8.5]))

d1$BW[d1$AGE==9.5 & d1$SEX==1] <- exp(rnorm(length(d1$BW[d1$AGE==9.5 & d1$SEX==1]),
                                          GIRLS$MEAN_Lbw[GIRLS$AGE==9.5],
                                          GIRLS$SD_Lbw[GIRLS$AGE==9.5]))

d1$BW[d1$AGE==10.0 & d1$SEX==1] <- exp(rnorm(length(d1$BW[d1$AGE==10.0 & d1$SEX==1]),
                                           GIRLS$MEAN_Lbw[GIRLS$AGE==10.0],
                                           GIRLS$SD_Lbw[GIRLS$AGE==10.0]))


# d2
d2$BW[d2$AGE==2.5 & d2$SEX==1] <- exp(rnorm(length(d2$BW[d2$AGE==2.5 & d2$SEX==1]),
                                          GIRLS$MEAN_Lbw[GIRLS$AGE==2.5],
                                          GIRLS$SD_Lbw[GIRLS$AGE==2.5]))

d2$BW[d2$AGE==3.5 & d2$SEX==1] <- exp(rnorm(length(d2$BW[d2$AGE==3.5 & d2$SEX==1]),
                                          GIRLS$MEAN_Lbw[GIRLS$AGE==3.5],
```

```
                                          GIRLS$SD_Lbw[GIRLS$AGE==3.5]))

d2$BW[d2$AGE==4.5 & d2$SEX==1] <- exp(rnorm(length(d2$BW[d2$AGE==4.5 & d2$SEX==1]),
                                          GIRLS$MEAN_Lbw[GIRLS$AGE==4.5],
                                          GIRLS$SD_Lbw[GIRLS$AGE==4.5]))

d2$BW[d2$AGE==5.5 & d2$SEX==1] <- exp(rnorm(length(d2$BW[d2$AGE==5.5 & d2$SEX==1]),
                                          GIRLS$MEAN_Lbw[GIRLS$AGE==5.5],
                                          GIRLS$SD_Lbw[GIRLS$AGE==5.5]))

d2$BW[d2$AGE==6.5 & d2$SEX==1] <- exp(rnorm(length(d2$BW[d2$AGE==6.5 & d2$SEX==1]),
                                          GIRLS$MEAN_Lbw[GIRLS$AGE==6.5],
                                          GIRLS$SD_Lbw[GIRLS$AGE==6.5]))

d2$BW[d2$AGE==7.5 & d2$SEX==1] <- exp(rnorm(length(d2$BW[d2$AGE==7.5 & d2$SEX==1]),
                                          GIRLS$MEAN_Lbw[GIRLS$AGE==7.5],
                                          GIRLS$SD_Lbw[GIRLS$AGE==7.5]))

d2$BW[d2$AGE==8.5 & d2$SEX==1] <- exp(rnorm(length(d2$BW[d2$AGE==8.5 & d2$SEX==1]),
                                          GIRLS$MEAN_Lbw[GIRLS$AGE==8.5],
                                          GIRLS$SD_Lbw[GIRLS$AGE==8.5]))

d2$BW[d2$AGE==9.5 & d2$SEX==1] <- exp(rnorm(length(d2$BW[d2$AGE==9.5 & d2$SEX==1]),
                                          GIRLS$MEAN_Lbw[GIRLS$AGE==9.5],
                                          GIRLS$SD_Lbw[GIRLS$AGE==9.5]))

d2$BW[d2$AGE==10.0 & d2$SEX==1] <- exp(rnorm(length(d2$BW[d2$AGE==10.0 & d2$SEX==1]),
                                           GIRLS$MEAN_Lbw[GIRLS$AGE==10.0],
                                           GIRLS$SD_Lbw[GIRLS$AGE==10.0]))

# body weights rounding
d1$BW <- round(d1$BW, digits=2)
d2$BW <- round(d2$BW, digits=2)


# dos1, dos2 --> datasets with dosing records to be merged with d1 and d2 respectively
dos1 <- data.frame(ID=(1:nsbj))
dos2 <- data.frame(ID=(1:nsbj))

# dos$TIME --> dosing time
dos1$TIME <- 0
dos2$TIME <- 0

# dos$DOSE --> actual dose
dos1$DOSE <- rep(dosing_interval,each=nsbj/length(dosing_interval))
dos2$DOSE <- rep(dosing_interval,each=nsbj/length(dosing_interval))

# dos$AMT --> amount of dose adimnistered in mg
dos1$AMT <- dos1$DOSE*d1$BW
dos2$AMT <- dos2$DOSE*d2$BW

# dos$SS --> steady state concentrations
dos1$SS <- 1
dos2$SS <- 1

# dos$II--> time interval (in hours) between two doses
dos1$II <- 12
dos2$II <- 12

# dos$CONC --> DV NONMEM column
dos1$CONC <- NA
dos2$CONC <- NA

# dos$MDV --> equal 1 because it's a dosing record (thus no DV is present)
dos1$MDV <- 1
dos2$MDV <- 1

# dos$EVID --> equal 1 because it's a dosing record
dos1$EVID <- 1
dos2$EVID <- 1

# SEX
dos1$SEX <- d1$SEX
dos2$SEX <- d2$SEX

# body weight
dos1$BW <- d1$BW
dos2$BW <- d2$BW

# age
dos1$AGE <- d1$AGE
dos2$AGE <- d2$AGE

# ADJ --> PK model binary covariate which is 1 if patients take more than 1 drug, 0
otherwise
dos1$ADJ <- 1
dos2$ADJ <- 1

# datasets merging
dd1 <- rbind(dos1,d1)
dd1 <- dd1[order(dd1$ID,dd1$TIME),]

dd2 <- rbind(dos2,d2)
dd2 <- dd2[order(dd2$ID,dd2$TIME),]
```

```
# write simulated csv file
file_nm1 <- "PKsim_dataset1SS.csv"
file_nm2 <- "PKsim_dataset2SS.csv"
write.csv(dd1,file_nm1,na=".",quote=FALSE,row.names=FALSE)
write.csv(dd2,file_nm2,na=".",quote=FALSE,row.names=FALSE)
```

7.3. R code for study design simulation and statistical analysis

R code used for study designs simulation and statistical analysis (steps 3 and 4 of section 4.2.3).

7.3.1. Parallel design

```
##########################################
######### SIMULATIONS UNDER H1 ############
##########################################

rm(list=ls(all=T))

# setting the seed for results replication
set.seed(121087)

# delta <- minimum significance difference given by the PK-PD model
load("delta")

# PK data reading
d <- read.csv("PK.csv",sep=",",header=T)

# M --> number of indivduals simulated at step 1
M <- length(unique(d$ID))
# N_PK_sim --> number of PK simulations done at step 2
N_PK_sim <- length(unique(d$PKSIM))

## PK-PD parameters from Girgis et al
# "placebo effect"
beta0 <- 4.4830
# "TPM effect"
beta1 <- -0.0579
# "residual variability" in terms of standard deviation
sigma_epsilon <- 0.751664

# EPS --> variability that will be considered for PD simulations under H1
d$EPS <- rnorm(nrow(d),0,sigma_epsilon)

# PDsim --> number of CTS
PDsim <- 1000

# vectors used to randomise patients in the placebo or TPM arm
control_flag <- rep(c(T,F),M/2)
test_flag <- !control_flag

# alpha --> significance level for t-test
alpha <- 0.05
z_alpha <- qnorm((1-alpha),0,1)
# beta <- define the power of the study (80%)
beta <- 0.2
z_beta <- qnorm((1-beta),0,1)

# n --> sample size per arm
n <- ceiling(2*((z_alpha+z_beta)*sigma_epsilon/delta)^2)

## out --> output dataset
out <- data.frame(H1=rep(NA,PDsim), # result of the statistical test when simulating under
H1: 0--> accept h0, 1--> refuse H0
                  H0=NA,    # # result of the statistical test when simulating under H0
                  MEAN=NA,  # estimate of delta in the seizure percent reduction scale
                  MEAN_Y=NA,  # estimate of delta in the Y scale
                  SD=NA,    # standard deviation of delta in the Y scale
                  SE=NA,    # standard error of delta in the Y scale
                  N=n)   # sample size per arm (fixed a priori)

# ID_perm --> vector used to randomly extract covariates from dataset
ID_perm <- rep(0,M)

## control and test alternate the IDs
control_IDperm <- rep(0,M/2)
test_IDperm <- control_IDperm

# sim_perm --> vector used to randomly extract simulations of ID previously estracted from
dataset
sim_perm <- rep(0,N_PK_sim)
```

```
# tmp --> ancillary dataframe
tmp <- subset(d,PKSIM == 1)
# control_subjects --> subjects randomized to the control group
control_subjects <-  tmp[(1:n),]

# control --> endpoint in the control group
control <- rep(0,n)

# test_subjects --> subjects randomized to the treated group
test_subjects <-  tmp[(1:n),]

# test --> endpoint in the test group
test <- rep(0,n)


for (iter in (1:PDsim))
{
  ID_perm <- sample((1:M),size=M,replace=F)

  control_IDperm <- ID_perm[control_flag]
  test_IDperm <- ID_perm[test_flag]

  sim_perm <- sample((1:N_PK_sim),size=N_PK_sim,replace=F)

  tmp <- subset(d,PKSIM == sim_perm[1])

  control_subjects <-  tmp[control_IDperm[1:n],]

  control <- beta0 + beta1*0 + control_subjects$EPS

  test_subjects <-  tmp[test_IDperm[1:n],]

  test <- beta0 + beta1*test_subjects$CONC + test_subjects$EPS

  ## one sided p-value. Alternative hypothesis = less because the lower the reponse in the
TPM group the higher TPM efficacy
  out$H1[iter] <- t.test(test,control,var.equal=T,alternative="l")$p.value < 0.05
  ## estimate of treatments difference in the Y scale
  out$MEAN_Y[iter] <- mean(control) - mean(test)
  ## estimate of treatments difference in the perc-reduction scale
  out$MEAN[iter] <- exp(mean(control)) - exp(mean(test))
  ## sample standard deviation of delta in the Y scale
  out$SD[iter] <- sqrt(var(control)+var(test))
  ## standard error of estimate of delta in the Y scale
  out$SE[iter] <- out$SD[iter]/sqrt(n)

 }
##########################################
######### SIMULATIONS UNDER H0 ##############
##########################################

## PK-PD parameters from Girgis et al
# "placebo effect"
beta0 <- 4.4830
# "TPM effect"
# simulations under H0 --> beta1 is set to 0 in the TPM group
beta1 <- 0
# "residual variability" in terms of standard deviation
sigma_epsilon <- 0.751664

# EPS --> variability that will be considered for PD simulations under H0
d$EPS <- rnorm(nrow(d),0,sigma_epsilon)

for (iter in (1:PDsim))
{
  ID_perm <- sample((1:M),size=M,replace=F)

  control_IDperm <- ID_perm[control_flag]
  test_IDperm <- ID_perm[test_flag]

  sim_perm <- sample((1:N_PK_sim),size=N_PK_sim,replace=F)

  tmp <- subset(d,PKSIM == sim_perm[1])

  control_subjects <-  tmp[control_IDperm[1:n],]

  control <- beta0 + beta1*0 + control_subjects$EPS

  test_subjects <-  tmp[test_IDperm[1:n],]

  test <- beta0 + beta1*test_subjects$CONC + test_subjects$EPS

  ## one sided p-value under H0
  out$H0[iter] <- t.test(test,control,var.equal=T,alternative="l")$p.value < 0.05

}
```

7.3.2. Crossover design

```
##########################################
######### SIMULATIONS UNDER H1 #############
##########################################

rm(list=ls(all=T))
require(MASS) # for mvrnorm function

# setting the seed for results replication
set.seed(121087)

# delta <- minimum significance difference given by the PK-PD model
load("delta")

# PK data reading
d <- read.csv("PK.csv",sep=",",header=T)

# M --> number of indivduals simulated at step 1
M <- length(unique(d$ID))
# N_PK_sim --> number of PK simulations done at step 2
N_PK_sim <- length(unique(d$PKSIM))

## PK-PD parameters from Girgis et al
# "placebo effect"
beta0 <- 4.4830
# "TPM effect"
beta1 <- -0.0579
# "residual variability" in terms of standard deviation
sigma_epsilon <- 0.751664

# rho --> correlation coefficient
rho <- 0.5
# Sigma --> covariance matrix of the responses in the two periods
Sigma <- sigma_epsilon^2*matrix(c(1,rho,rho,1),nrow=2,ncol=2)

EPS <- mvrnorm(nrow(d),rep(0,2),Sigma)

# EPS1 --> variability that will be considered during the first period
d$EPS1 <- EPS[,1]
# EPS2 --> variability that will be considered during the second period
d$EPS2 <- EPS[,2]

# PDsim --> number of CTS
PDsim <- 1000

# flag indicating to which sequence patients are randomized:
# TP --> TPM and PCB
# PT --> PCB anb TPM
TP_flag <- rep(c(T,F),M/2)
PT_flag <- !TP_flag

# alpha --> significance level for t-test
alpha <- 0.05
z_alpha <- qnorm((1-alpha),0,1)
# beta <- define the power of the study (80%)
beta <- 0.2
z_beta <- qnorm((1-beta),0,1)

# n --> sample size per sequence
n <- ceiling(2*((z_alpha+z_beta)*sigma_epsilon/(2*delta))^2*(2*(1-rho)))

## out --> output dataset
out <- data.frame(H1=rep(NA,PDsim), # result of the statistical test when simulating under
H1: 0--> accept h0, 1--> refuse H0
                  H0=NA,   # # result of the statistical test when simulating under H0
                  MEAN=NA,  # estimate of delta in the seizure percent reduction scale
                  MEAN_Y=NA,  # estimate of delta in the Y scale
                  SD=NA,    # standard deviation of delta in the Y scale
                  SE=NA,   # standard error of delta in the Y scale
                  N=n)   # sample size per arm (fixed a priori)

# ID_perm --> vector used to randomly extract covariates from dataset
ID_perm <- rep(0,M)

## TP and PT alternate the IDs
TP_IDperm <- rep(0,M/2)
PT_IDperm <- rep(0,M/2)

# sim_perm --> vector used to randomly extract simulations of ID previously estracted from
dataset
sim_perm <- rep(0,N_PK_sim)

# tmp --> ancillary dataframe
tmp <- subset(d,PKSIM == 1)
# TP_subjects --> subjects randomized to TPM PCB sequence
TP_subjects  <-  tmp[(1:n),]

# TP_P --> response to PCB of patients randomized to sequence TPM PCB
TP_P <- rep(0,n)
# TP_T --> response to TPM of patients randomized to sequence TPM PCB
TP_T <- rep(0,n)
```

```
# TP --> whithin differences beween response during PCB and TPM
TP <- rep(0,n)

# PT_subjects --> subjects randomized to PCB TPM sequence
PT_subjects <-  tmp[(1:n),]

# PT_T --> response to TPM of patients randomized to sequence PCB TPM
PT_T <- rep(0,n)
# PT_P --> response to PCB of patients randomized to sequence PCB TPM
PT_P <- rep(0,n)
# PT --> whithin differences beween response during TPM and PCB
PT <- rep(0,n)

for (iter in (1:PDsim))
{
  ID_perm <- sample((1:M),size=M,replace=F)

  TP_IDperm <- ID_perm[TP_flag]
  PT_IDperm <- ID_perm[PT_flag]

  sim_perm <- sample((1:N_PK_sim),size=N_PK_sim,replace=F)

  tmp <- subset(d,PKSIM == sim_perm[1])

  # TPM-PCB sequence
  TP_subjects <-  tmp[TP_IDperm[1:n],]
  TP_P <- beta0 + beta1*0 + TP_subjects$EPS2
  TP_T <- beta0 + beta1*TP_subjects$CONC + TP_subjects$EPS1
  TP <- TP_P - TP_T

  # PCB-TPM sequence
  PT_subjects <-  tmp[PT_IDperm[1:n],]
  PT_T <- beta0 + beta1*PT_subjects$CONC + PT_subjects$EPS2
  PT_P <- beta0 + beta1*0 + PT_subjects$EPS1
  PT <- PT_T - PT_P

  ## one sided p-value.
  out$H1[iter] <- t.test(TP,PT,var.equal=T,alternative="g")$p.value < 0.05
  ## estimate of treatments difference in the Y scale
  out$MEAN_Y[iter] <- mean( c(mean(TP_P - TP_T) , mean(PT_P - PT_T) ))
  ## estimate of treatments difference in the perc-reduction scale
  out$MEAN[iter] <- exp(mean(cbind(TP_P, PT_P))) - exp(mean(cbind(TP_T, PT_T)))
  ## sample standard deviation of delta in the Y scale
  out$SD[iter] <- sqrt(0.25*(var(TP) + var(PT)))
  ## standard error of estimate of delta in the Y scale
  out$SE[iter] <- out$SD[iter]/sqrt(n)

}

##############################################
######### SIMULATIONS UNDER H0 ##############
##############################################

## PK-PD parameters from Girgis et al
# "placebo effect"
beta0 <- 4.4830
# "TPM effect"
# simulations under H0 --> beta1 is set to 0 in the TPM group
beta1 <- 0
# "residual variability" in terms of standard deviation
sigma_epsilon <- 0.751664

# EPS --> variability that will be considered for PD simulations under H0
EPS <- mvrnorm(nrow(d),rep(0,2),Sigma)

# EPS1 --> variability that will be considered during the first period
d$EPS1 <- EPS[,1]
# EPS2 --> variability that will be considered during the second period
d$EPS2 <- EPS[,2]

for (iter in (1:PDsim))
{
  ID_perm <- sample((1:M),size=M,replace=F)

  TP_IDperm <- ID_perm[TP_flag]
  PT_IDperm <- ID_perm[PT_flag]

  sim_perm <- sample((1:N_PK_sim),size=N_PK_sim,replace=F)

  tmp <- subset(d,PKSIM == sim_perm[1])

  # TPM-PCB sequence
  TP_subjects <-  tmp[TP_IDperm[1:n],]
  TP_P <- beta0 + beta1*0 + TP_subjects$EPS2
  TP_T <- beta0 + beta1*TP_subjects$CONC + TP_subjects$EPS1
  TP <- TP_P - TP_T

  # PCB-TPM sequence
  PT_subjects <-  tmp[PT_IDperm[1:n],]
  PT_T <- beta0 + beta1*PT_subjects$CONC + PT_subjects$EPS2
  PT_P <- beta0 + beta1*0 + PT_subjects$EPS1
  PT <- PT_T - PT_P

  ## one sided p-value.
  out$H0[iter] <- t.test(TP,PT,var.equal=T,alternative="g")$p.value < 0.05
```

```
}
```

7.3.3. Randomized withdrawal design

```
##########################################
######### SIMULATIONS UNDER H1 ############
##########################################

rm(list=ls(all=T))
require(MASS)

# delta <- minimum significance difference given by the PK-PD model
load("delta")
# theta --> responder rate given by the PK-PD model
load("theta")

# setting the seed for results replication
set.seed(121087)

# PK data reading
d <- read.csv("PK.csv",sep=",",header=T)

# M --> number of indivduals simulated at step 1
M <- length(unique(d$ID))
# N_PK_sim --> number of PK simulations done at step 2
N_PK_sim <- length(unique(d$PKSIM))

## PK-PD parameters from Girgis et al
# "placebo effect"
beta0 <- 4.4830
# "TPM effect"
beta1 <- -0.0579
# "residual variability" in terms of standard deviation
sigma_epsilon <- 0.751664

# rho --> correlation coefficent
rho <- 0.5
# Sigma --> covariance matrix of the variances of the responses in the two periods
Sigma <- sigma_epsilon^2*matrix(c(1,rho,rho,1),nrow=2,ncol=2)

EPS <- mvrnorm(nrow(d),rep(0,2),Sigma)

# EPS1 --> variability that will be considered during the first period
d$EPS1 <- EPS[,1]
# EPS2 --> variability that will be considered during the second period
d$EPS2 <- EPS[,2]

# PDsim --> number of CTS
PDsim <- 1000

# alpha --> significance level for t-test
alpha <- 0.05
z_alpha <- qnorm((1-alpha),0,1)
# beta <- define the power of the study (80%)
beta <- 0.2
z_beta <- qnorm((1-beta),0,1)

# n --> sample size per arm
n <- 2*((z_alpha+z_beta)*sigma_epsilon/delta)^2

# adding the number of non-reponders to the sample size
n <- ceiling(n/theta)

# open_label --> ALL subjects
open_label <- data.frame(ID=rep(NA,2*n),
                         CONC=NA,
                         PKSIM=NA,
                         EPS1=NA,
                         EPS2=NA,
                         RESPONSE=NA)

## out --> output dataset
out <- data.frame(H1=rep(NA,PDsim), # result of the statistical test when simulating under
H1: 0--> accept h0, 1--> refuse H0
                  H0=NA,   # # result of the statistical test when simulating under H0
                  MEAN=NA,  # estimate of delta in the seizure percent reduction scale
                  MEAN_Y=NA,  # estimate of delta in the Y scale
                  SD=NA,    # standard deviation of delta in the Y scale
                  SE=NA,   # standard error of delta in the Y scale
                  N=n,  # sample size per arm in the open label phase
                  N_DB=NA) # sample size per arm in the double blind phase

# sim_perm --> vector used to randomly extract simulations of ID previously estracted from
dataset
sim_perm <- rep(0,N_PK_sim)


for (iter in (1:PDsim))
{
  sim_perm <- sample((1:N_PK_sim),size=N_PK_sim,replace=F)
```

```
  ## OPEN LABEL PHASE

  open_label <-  transform(subset(d,PKSIM == sim_perm[1]),
                           RESPONSE = exp(beta0 + beta1*CONC + EPS1) < exp(beta0),
                           PKSIM=NULL)[1:(2*n),]

  # responders --> responders from the open label phase
  responders <- transform(subset(open_label,RESPONSE==T), RESPONSE=NULL)

  ## DOUBLE BLIND PHASE

  # randomization --> list with two datasets, the control and the experimental arm
  randomization <- split(responders,
                         sample(rep(1:2, nrow(responders)/2)))

  # test --> endpoint in the test group
  test <- with(randomization[[1]],
               beta0 + beta1*CONC + EPS2)

  # control --> endpoint in the control group
  control <- with(randomization[[2]],
                  beta0 + beta1*0 + EPS2)

  ## one sided p-value
  out$H1[iter] <- t.test(test,control,var.equal=T,alternative="l")$p.value < 0.05
  ## estimate of treatments difference in the Y scale
  out$MEAN_Y[iter] <- mean(control) - mean(test)
  ## estimate of treatments difference in the perc-reduction scale
  out$MEAN[iter] <- exp(mean(control)) - exp(mean(test))
  ## sample standard deviation of delta in the Y scale
  out$SD[iter] <- sqrt(var(control) + var(test))
  out$N_DB[iter] <- floor(nrow(responders)/2)
  ## standard error of estimate of delta in the Y scale
  out$SE[iter] <- out$SD[iter]/sqrt(out$N_DB[iter])

}
##############################################
######### SIMULATIONS UNDER H0 ##############
##############################################

## PK-PD parameters from Girgis et al
# "placebo effect"
beta0 <- 4.4830
# "TPM effect"
# simulations under H0 --> beta1 is set to 0 in the TPM group
beta1 <- 0
# "residual variability" in terms of standard deviation
sigma_epsilon <- 0.751664

EPS <- mvrnorm(nrow(d),rep(0,2),Sigma)

# EPS1 --> variability that will be considered during the first period
d$EPS1 <- EPS[,1]
# EPS2 --> variability that will be considered during the second period
d$EPS2 <- EPS[,2]

for (iter in (1:PDsim))
{
  sim_perm <- sample((1:N_PK_sim),size=N_PK_sim,replace=F)

  ## OPEN LABEL PHASE

  open_label <-  transform(subset(d,PKSIM == sim_perm[1]),
                           RESPONSE = exp(beta0 + beta1*CONC + EPS1) < exp(beta0),
                           PKSIM=NULL)[1:(2*n),]

  # responders --> responders from the open label phase
  responders <- transform(subset(open_label,RESPONSE==T), RESPONSE=NULL)

  ## DOUBLE BLIND PHASE

  # randomization --> list with two datasets, the control and the experimental arm
  randomization <- split(responders,
                         sample(rep(1:2, nrow(responders)/2)))

  # test --> endpoint in the test group
  test <- with(randomization[[1]],
               beta0 + beta1*CONC + EPS2)

  # control --> endpoint in the control group
  control <- with(randomization[[2]],
                  beta0 + beta1*0 + EPS2)

  ## one sided p-value
  out$H0[iter] <- t.test(test,control,var.equal=T,alternative="l")$p.value < 0.05

}
```

7.3.4. Sequential probability ratio test

```
##########################################
######### SIMULATIONS UNDER H1 ##############
##########################################

rm(list=ls(all=T))

# setting the seed for results replication
set.seed(121087)

# delta <- minimum significance difference given by the PK-PD model
load("delta")

# PK data reading
d <- read.csv("PK.csv",sep=",",header=T)

# M --> number of indivduals simulated at step 1
M <- length(unique(d$ID))
# N_PK_sim --> number of PK simulations done at step 2
N_PK_sim <- length(unique(d$PKSIM))

## PK-PD parameters from Girgis et al
# "placebo effect"
beta0 <- 4.4830
# "TPM effect"
beta1 <- -0.0579
# "residual variability" in terms of standard deviation
sigma_epsilon <- 0.751664

# EPS --> variability that will be considered for PD simulations under H1
d$EPS <- rnorm(nrow(d),0,sigma_epsilon)

# PDsim --> number of CTS
PDsim <- 1000

# vectors used to randomise patients in the placebo or TPM arm
control_flag <- rep(c(T,F),M/2)
test_flag <- !control_flag

# tolerated type I error
alpha <- 0.05
# tolerated type II error
beta <- 0.2

# delta_bar <- adjusted delta because beta != alpha
delta_bar <- delta*2*qnorm((1-alpha),0,1)/(qnorm((1-alpha),0,1)+qnorm((1-beta),0,1))

## out --> output dataset
out <- data.frame(H1=rep(NA,PDsim), # result of the statistical test when simulating under
H1: 0--> accept h0, 1--> refuse H0
                  H0=NA,   # # result of the statistical test when simulating under H0
                  MEAN=NA,  # estimate of delta in the seizure percent reduction scale
                  MEAN_Y=NA,  # estimate of delta in the Y scale
                  WCI=NA,    # width of confidence intervals
                  N=NA)   # sample size per arm (to be determined via CTS)

# G_half --> number of patients per group enrolled at each interim analysis
G_half <- 10

# I --> inspection interval
I <- 2*G_half/(4*sigma_epsilon^2)

# q --> absolute value of the intercept of the bounds
q <- 1/delta_bar*log((1-alpha)/alpha) - 0.583*sqrt(I)

# ID_perm --> vector used to randomly extract covariates from dataset
ID_perm <- rep(0,M/2)

## control and test alternate the IDs
control_IDperm <- rep(0,M/2)
test_IDperm <- control_IDperm

# sim_perm --> vector used to randomly extract simulations of ID previously estracted from
dataset
sim_perm <- rep(0,N_PK_sim)

# Z --> Z stastistic as defined in the paper
Z <- 0

# V --> V stastistic as defined in the paper
V <- 0

# theta_grid --> to be used to make inference
theta_grid <- seq(0,100,0.01)
# inference --> vector containing estimatye and confidence intervals of Y
inference <- rep(0,length(theta_grid))
# lambda --> treatment effect estimate under H0 (see pag 118)
lambda <- NA

for (iter in (1:PDsim))
{
```

```
  ID_perm <- sample((1:M),size=M,replace=F)

  control_IDperm <- ID_perm[control_flag]
  test_IDperm <- ID_perm[test_flag]

  sim_perm <- sample((1:N_PK_sim),size=N_PK_sim,replace=F)

  # n --> sample size per arm
  n <- G_half

  # label --> logical variable that takes false if a couclusion is drawn during the
sequential testing procedure
  label <- T

  while(label==T)
  {

    # control_subjects --> subjects randomized to the control group
    tmp <- subset(d,PKSIM == sim_perm[1])
    control_subjects <-  tmp[control_IDperm[1:n],]

    # control --> mean response in the control group
    control <- beta0 + beta1*0 + control_subjects$EPS

    # test_subjects --> subjects randomized to the treated group
    test_subjects <-  tmp[test_IDperm[1:n],]

    # test --> mean response in the test group
    test <- beta0 + beta1*test_subjects$CONC + test_subjects$EPS

    Z <- sum(control - test)/(2*sigma_epsilon^2)

    V <- n/(2*sigma_epsilon^2) - Z^2/n

    # upper --> boundary of rejection
    upper <-  q + delta_bar/2*V

    # lower --> boundary of acceptance
    lower <- - q + delta_bar/2*V

    # the test goes on until Z is in the continuation region
    label <- (Z > lower) & (Z < upper)

    # the sample size per arm increases by G at every step
    n <- n + G_half
  }

  # Decision:
  if (Z <= lower) # if Z falls below the lower boundry accept H0
    out$H1[iter] <- 0
  else            # if Z falls above the upper boundry accept H1
    out$H1[iter] <- 1

  inference <- seq_inference(alpha,delta_bar,theta_grid,V,0.95,"SPRT",out$H1[iter])

  lambda <- mean(c(test,control))

  out$MEAN[iter] <- exp(lambda + 0.5*inference[2]) - exp(lambda - 0.5*inference[2])
  out$MEAN_Y[iter] <- inference[2]
  # WCI --> 97.5% conf.int. - 2.5% conf.int.
  out$WCI[iter] <- inference[3] - inference[1]
  out$N[iter] <- n - G_half

 }

##############################################
######### SIMULATIONS UNDER H0 ##############
##############################################

## PK-PD parameters from Girgis et al
# "placebo effect"
beta0 <- 4.4830
# "TPM effect"
# simulations under H0 --> beta1 is set to 0 in the TPM group
beta1 <- 0
# "residual variability" in terms of standard deviation
sigma_epsilon <- 0.751664

# EPS --> variability that will be considered for PD simulations under H0
d$EPS <- rnorm(nrow(d),0,sigma_epsilon)

for (iter in (1:PDsim))
{
  ID_perm <- sample((1:M),size=M,replace=F)

  control_IDperm <- ID_perm[control_flag]
  test_IDperm <- ID_perm[test_flag]

  sim_perm <- sample((1:N_PK_sim),size=N_PK_sim,replace=F)

  # n --> sample size per arm
  n <- G_half

  # label --> logical variable that takes false if a couclusion is drawn during the
sequential testing procedure
```

```
  label <- T

  while(label==T)
  {

    # control_subjects --> subjects randomized to the control group
    tmp <- subset(d,PKSIM == sim_perm[1])
    control_subjects <-  tmp[control_IDperm[1:n],]

    # control --> mean response in the control group
    control <- beta0 + beta1*0 + control_subjects$EPS

    # test_subjects --> subjects randomized to the treated group
    test_subjects <-  tmp[test_IDperm[1:n],]

    # test --> mean response in the test group
    test <- beta0 + beta1*test_subjects$CONC + test_subjects$EPS

    Z <- sum(control - test)/(2*sigma_epsilon^2)

    V <- n/(2*sigma_epsilon^2) - Z^2/n

    # upper --> boundary of rejection
    upper <-  q + delta_bar/2*V

    # lower --> boundary of acceptance
    lower <- - q + delta_bar/2*V

    # the test goes on until Z is in the continuation region
    label <- (Z > lower) & (Z < upper)

    # the sample size per arm increases by G at every step
    n <- n + G_half
  }

  # Decision:
  if (Z <= lower) # if Z falls below the lower boundry accept H0
    out$H0[iter] <- 0
  else            # if Z falls above the upper boundry accept H1
    out$H0[iter] <- 1

  }
```

7.3.5. Triangular test

```
##############################################
######### SIMULATIONS UNDER H1 ##############
##############################################

rm(list=ls(all=T))

# setting the seed for results replication
set.seed(121087)

# delta <- minimum significance difference given by the PK-PD model
load("delta")

# PK data reading
d <- read.csv("PK.csv",sep=",",header=T)

# M --> number of indivduals simulated at step 1
M <- length(unique(d$ID))
# N_PK_sim --> number of PK simulations done at step 2
N_PK_sim <- length(unique(d$PKSIM))

## PK-PD parameters from Girgis et al
# "placebo effect"
beta0 <- 4.4830
# "TPM effect"
beta1 <- -0.0579
# "residual variability" in terms of standard deviation
sigma_epsilon <- 0.751664

# EPS --> variability that will be considered for PD simulations under H1
d$EPS <- rnorm(nrow(d),0,sigma_epsilon)

# PDsim --> number of CTS
PDsim <- 1000

# vectors used to randomise patients in the placebo or TPM arm
control_flag <- rep(c(T,F),M/2)
test_flag <- !control_flag

# tolerated type I error
alpha <- 0.05
# tolerated type II error
beta <- 0.2

# delta_bar <- adjusted delta because beta != alpha
delta_bar <- delta*2*qnorm((1-alpha),0,1)/(qnorm((1-alpha),0,1)+qnorm((1-beta),0,1))
```

```
## out --> output dataset
out <- data.frame(H1=rep(NA,PDsim), # result of the statistical test when simulating under
H1: 0--> accept h0, 1--> refuse H0
                  H0=NA,   # # result of the statistical test when simulating under H0
                  MEAN=NA,  # estimate of delta in the seizure percent reduction scale
                  MEAN_Y=NA,  # estimate of delta in the Y scale
                  WCI=NA,    # width of confidence intervals
                  N=NA)   # sample size per arm (to be determined via CTS)

# G_half --> number of patients per group enrolled at each interim analysis
G_half <- 10

# I --> inspection interval
I <- 2*G_half/(4*sigma_epsilon^2)

# q --> absolute value of the intercept of the bounds
q <- 2/delta_bar*log(1/(2*alpha)) - 0.583*sqrt(I)

# ID_perm --> vector used to randomly extract covariates from dataset
ID_perm <- rep(0,M/2)

## control and test alternate the IDs
control_IDperm <- rep(0,M/2)
test_IDperm <- control_IDperm

# sim_perm --> vector used to randomly extract simulations of ID previously estracted from
dataset
sim_perm <- rep(0,N_PK_sim)

# Z --> Z stastistic as defined in the paper
Z <- 0

# V --> V stastistic as defined in the paper
V <- 0

# theta_grid --> to be used to make inference
theta_grid <- seq(0,100,0.01)
# inference --> vector containing estimatye and confidence intervals of Y
inference <- rep(0,length(theta_grid))
# lambda --> treatment effect estimate under H0 (see pag 118)
lambda <- NA

for (iter in (1:PDsim))
{
  ID_perm <- sample((1:M),size=M,replace=F)

  control_IDperm <- ID_perm[control_flag]
  test_IDperm <- ID_perm[test_flag]

  sim_perm <- sample((1:N_PK_sim),size=N_PK_sim,replace=F)

  # n --> sample size per arm
  n <- G_half

  # label --> logical variable that takes false if a couclusion is drawn during the
sequential testing procedure
  label <- T

  while(label==T)
  {

    # control_subjects --> subjects randomized to the control group
    tmp <- subset(d,PKSIM == sim_perm[1])
    control_subjects <-  tmp[control_IDperm[1:n],]

    # control --> mean response in the control group
    control <- beta0 + beta1*0 + control_subjects$EPS

    # test_subjects --> subjects randomized to the treated group
    test_subjects <-  tmp[test_IDperm[1:n],]

    # test --> mean response in the test group
    test <- beta0 + beta1*test_subjects$CONC + test_subjects$EPS

    Z <- sum(control - test)/(2*sigma_epsilon^2)

    V <- n/(2*sigma_epsilon^2) - Z^2/n

    # upper --> boundary of rejection
    upper <-  q + 1/4*delta_bar*V

    # lower --> boundary of acceptance
    lower <- - q + 3/4*delta_bar*V

    # the test goes on until Z is in the continuation region
    label <- (Z > lower) & (Z < upper)

    # the sample size per arm increases by G at every step
    n <- n + G_half
  }

  # Decision:
  if (Z <= lower) # if Z falls below the lower boundary accept H0
    out$H1[iter] <- 0
  else            # if Z falls above the upper boundary accept H1
```

```
    out$H1[iter] <- 1

  inference <- seq_inference(alpha,delta_bar,theta_grid,V,0.95,"TT",out$H1[iter])

  lambda <- mean(c(test,control))

  out$MEAN[iter] <- exp(lambda + 0.5*inference[2]) - exp(lambda - 0.5*inference[2])
  out$MEAN_Y[iter] <- inference[2]
  # WCI --> 97.5% conf.int. - 2.5% conf.int.
  out$WCI[iter] <- inference[3] - inference[1]
  out$N[iter] <- n - G_half
}

##########################################
######### SIMULATIONS UNDER H0 ############
##########################################

## PK-PD parameters from Girgis et al
# "placebo effect"
beta0 <- 4.4830
# "TPM effect"
# simulations under H0 --> beta1 is set to 0 in the TPM group
beta1 <- 0
# "residual variability" in terms of standard deviation
sigma_epsilon <- 0.751664

# EPS --> variability that will be considered for PD simulations under H0
d$EPS <- rnorm(nrow(d),0,sigma_epsilon)

for (iter in (1:PDsim))
{
  ID_perm <- sample((1:M),size=M,replace=F)

  control_IDperm <- ID_perm[control_flag]
  test_IDperm <- ID_perm[test_flag]

  sim_perm <- sample((1:N_PK_sim),size=N_PK_sim,replace=F)

  # n --> sample size per arm
  n <- G_half

  # label --> logical variable that takes false if a couclusion is drawn during the
sequential testing procedure
  label <- T

  while(label==T)
  {

    # control_subjects --> subjects randomized to the control group
    tmp <- subset(d,PKSIM == sim_perm[1])
    control_subjects <-  tmp[control_IDperm[1:n],]

    # control --> mean response in the control group
    control <- beta0 + beta1*0 + control_subjects$EPS

    # test_subjects --> subjects randomized to the treated group
    test_subjects <-  tmp[test_IDperm[1:n],]

    # test --> mean response in the test group
    test <- beta0 + beta1*test_subjects$CONC + test_subjects$EPS

    Z <- sum(control - test)/(2*sigma_epsilon^2)

    V <- n/(2*sigma_epsilon^2) - Z^2/n

    # upper --> boundary of rejection
    upper <-  q + 1/4*delta_bar*V

    # lower --> boundary of acceptance
    lower <- - q + 3/4*delta_bar*V

    # the test goes on until Z is in the continuation region
    label <- (Z > lower) & (Z < upper)

    # the sample size per arm increases by G at every step
    n <- n + G_half
  }

  # Decision:
  if (Z <= lower) # if Z falls below the lower boundry accept H0
    out$H0[iter] <- 0
  else            # if Z falls above the upper boundry accept H1
    out$H0[iter] <- 1

}
```

7.3.6. Bayesian design

```
##########################################
######### SIMULATIONS UNDER H1 ############
##########################################
```

```
rm(list=ls(all=T))

# setting the seed for results replication
set.seed(121087)

# PK data reading
d <- read.csv("PK.csv",sep=",",header=T)

# M --> number of indivduals simulated at step 1
M <- length(unique(d$ID))
# N_PK_sim --> number of PK simulations done at step 2
N_PK_sim <- length(unique(d$PKSIM))

## PK-PD parameters from Girgis et al
# "placebo effect"
beta0 <- 4.4830
# beta1_p <- "TPM effect" in pediatrics
beta1_p <- -0.0579
# beta1_a <- "TPM effect" in adults
beta1_a <- -0.0627
# "residual variability" in terms of standard deviation
sigma_epsilon <- 0.751664

# EPS --> variability that will be considered for PD simulations under H1
d$EPS <- rnorm(nrow(d),0,sigma_epsilon)

# PDsim --> number of CTS
PDsim <- 1000

# vectors used to randomise patients in the placebo or TPM arm
control_flag <- rep(c(T,F),M/2)
test_flag <- !control_flag

# alpha --> significance level for t-test
alpha <- 0.05
z_alpha <- qnorm((1-alpha),0,1)
# beta <- define the power of the study (80%)
beta <- 0.2
z_beta <- qnorm((1-beta),0,1)

# delta_p_bar --> TPM vs PCB difference in paediatrics given by the PK-PD model
load("delta_p_bar")
delta_p_bar <- delta

# delta_hat_a --> ML estimate of delta based on adult trials (-0.0627=beta1 in adults, 8=
pop Cmin. From Girgis et al)
delta_hat_a <-  -beta1_a*8
# sa --> 2*sigma_epsilon
sa <- 2*sigma_epsilon
sp <- sa
# ma --> number of adults on which delta_hat_a has been estimated (from Girgis et al)
ma <- 663
# mp --> number of children to be enrolled in the upcoming trial
mp <- seq(10,500,2)
# ni -->  difference in improvement of TPM over placebo between children and adults
ni <- abs(delta_hat_a-delta_p_bar)/sqrt(2)
# omega --> as defined in the paper
omega <- sa^2/(sa^2/(ma) + 2*ni^2)
# x_bar --> to be used for power calulcation
x_bar <- (sp^2/(mp)) * (+1.64*sqrt(mp/sp^2+omega/sa^2) - omega/sa^2*delta_hat_a)

# z_beta_bayes--> to be used for power calculation
z_beta_bayes <- (sqrt(mp)/sp)*(delta_p_bar -  x_bar)
# n --> sample size per arm
n <- mp[z_beta_bayes>=z_beta][1]/2


## out --> output dataset
out <- data.frame(H1=rep(NA,PDsim), # result of the statistical test when simulating under
H1: 0--> accept h0, 1--> refuse H0
                  H0=NA,   # # result of the statistical test when simulating under H0
                  MEAN=NA,  # estimate of delta in the seizure percent reduction scale
                  MEAN_Y=NA,  # estimate of delta in the Y scale # cannot be obtained in
this bayesian framework
                  SD=NA,    # standard deviation of delta in the Y scale
                  SE=NA,   # standard error of delta in the Y scale # does not make sense in
bayes
                  N=n)   # sample size per arm (fixed a priori)

# ID_perm --> vector used to randomly extract covariates from dataset
ID_perm <- rep(0,M)

## control and test alternate the IDs
control_IDperm <- rep(0,M/2)
test_IDperm <- control_IDperm

# sim_perm --> vector used to randomly extract simulations of ID previously estracted from
dataset
sim_perm <- rep(0,N_PK_sim)

# tmp --> ancillary dataframe
tmp <- subset(d,PKSIM == 1)
# control_subjects --> subjects randomized to the control group
control_subjects <-  tmp[(1:n),]
```

```
# control --> endpoint in the control group
control <- rep(0,n)

# test_subjects --> subjects randomized to the treated group
test_subjects <-  tmp[(1:n),]

# test --> endpoint in the test group
test <- rep(0,n)

# delta_hat_p --> posterior estimate of treatment effect in children
delta_hat_p <- NA
# sp obtained from the simulations
sp <- NA
# z--> variable used to make inference
z <- NA

for (iter in (1:PDsim))
{
  ID_perm <- sample((1:M),size=M,replace=F)

  control_IDperm <- ID_perm[control_flag]
  test_IDperm <- ID_perm[test_flag]

  sim_perm <- sample((1:N_PK_sim),size=N_PK_sim,replace=F)

  tmp <- subset(d,PKSIM == sim_perm[1])

  control_subjects <-  tmp[control_IDperm[1:n],]

  control <- beta0 + beta1_p*0 + control_subjects$EPS

  test_subjects <-  tmp[test_IDperm[1:n],]

  test <- beta0 + beta1_p*test_subjects$CONC + test_subjects$EPS

  # BAYESIAN INFERENCE
  delta_hat_p <- mean(control)-mean(test)
  sp <- 2*sd(control-test)/sqrt(2)

  z <- -(delta_hat_p*(2*n)/sp^2+delta_hat_a*omega/sa^2)/
    sqrt((2*n)/sp^2+omega/sa^2)

  ## one sided p-value
  out$H1[iter] <- pnorm(z,0,1) < 0.05
  ## estimate of the mean of the posterior distribution of delta
  out$MEAN_Y[iter] <- (delta_hat_p*(2*n)/sp^2+delta_hat_a*omega/sa^2) /
((2*n)/sp^2+omega/sa^2)
  ## estimate of the standard deviation of the posterior distribution of delta
  out$SD[iter] <- ((2*n)/sp^2 + omega/sa^2)^(-0.5)

}

###########################################
######### SIMULATIONS UNDER H0 ##############
###########################################

## PK-PD parameters from Girgis et al
# "placebo effect"
beta0 <- 4.4830
# "TPM effect"
# simulations under H0 --> beta1 is set to 0 in the TPM group
beta1_p <- 0
# "residual variability" in terms of standard deviation
sigma_epsilon <- 0.751664

# EPS --> variability that will be considered for PD simulations under H0
d$EPS <- rnorm(nrow(d),0,sigma_epsilon)

for (iter in (1:PDsim))
{
  ID_perm <- sample((1:M),size=M,replace=F)

  control_IDperm <- ID_perm[control_flag]
  test_IDperm <- ID_perm[test_flag]

  sim_perm <- sample((1:N_PK_sim),size=N_PK_sim,replace=F)

  tmp <- subset(d,PKSIM == sim_perm[1])

  control_subjects <-  tmp[control_IDperm[1:n],]

  control <- beta0 + beta1_p*0 + control_subjects$EPS

  test_subjects <-  tmp[test_IDperm[1:n],]

  test <- beta0 + beta1_p*test_subjects$CONC + test_subjects$EPS

  # BAYESIAN INFERENCE
  delta_hat_p <- mean(control)-mean(test)
  sp <- 2*sd(control-test)/sqrt(2)

  z <- -(delta_hat_p*(2*n)/sp^2+delta_hat_a*omega/sa^2)/
    sqrt((2*n)/sp^2+omega/sa^2)
```

```
  ## one sided p-value
  out$H0[iter] <- pnorm(z,0,1) < 0.05

}
```

7.3.7. Bayesian sequential design with a non-hierarchical framework

```
##########################################
######### SIMULATIONS UNDER H1 ############
##########################################

rm(list=ls(all=T))

# setting the seed for results replication
set.seed(121087)

# PK data reading
d <- read.csv("PK.csv",sep=",",header=T)

# M --> number of indivduals simulated at step 1
M <- length(unique(d$ID))
# N_PK_sim --> number of PK simulations done at step 2
N_PK_sim <- length(unique(d$PKSIM))

## PK-PD parameters from Girgis et al
# "placebo effect"
beta0 <- 4.4830
# "TPM effect"
beta1 <- -0.0579
# "residual variability" in terms of standard deviation
sigma_epsilon <- 0.751664

# EPS --> variability that will be considered for PD simulations under H1
d$EPS <- rnorm(nrow(d),0,sigma_epsilon)

# PDsim --> number of CTS
PDsim <- 1000

# vectors used to randomise patients in the placebo or TPM arm
control_flag <- rep(c(T,F),M/2)
test_flag <- !control_flag

## out --> output dataset
out <- data.frame(H1=rep(NA,PDsim), # result of the statistical test when simulating under
H1: 0--> accept h0, 1--> refuse H0
                  H0=NA,   # # result of the statistical test when simulating under H0
                  MEAN=NA,  # estimate of delta in the seizure percent reduction scale
                  MEAN_Y=NA,  # estimate of delta in the Y scale
                  WCI=NA,    # width of confidence intervals
                  N=NA)   # sample size per arm (to be determined via CTS)

# G_half --> number of patients per group enrolled at each interim analysis
G_half <- 10

# PRIORS
# mu_pcb --> prior mean in the placebo group
mu_pcb <- 4.4469
# sd_pcb --> prior sd in the pcb group
sd_pcb <- 0.751664/sqrt(16.5)
# mu_tpm --> prior mean in the tpm group
mu_tpm <- mu_pcb - 0.0627*8
# sd_tpm --> prior sd in the tpm group
sd_tpm <- 0.751664/sqrt(16.5)

#### LIKELIHOOD
# m_pcb --> mean response rate in the placebo group
m_pcb <- NA
# V_pcb --> variance of the response in the placebo group
V_pcb <- NA
# m_tpm --> mean response rate in the tpm group
m_tpm <- NA
# V_tpm --> variance of the response in the tpm group
V_tpm <- NA

#### POSTERIORS
# mu_post_pcb --> posterior mean of the treatment effect in placebo group
mu_post_pcb <- NA
# sd_post_pcb --> posterior SD of the treatment effect in placebo group
sd_post_pcb <- NA
# mu_post_tpm --> posterior mean of the treatment effect in tpm group
mu_post_tpm <- NA
# sd_post_tpm --> posterior SD of the treatment effect in tpm group
sd_post_tpm <- NA

# delta_post_mean --> mean of the posterior distribution of the difference between
treatments
delta_post_mean <- NA
# delta_post_sd --> sd of the posterior distribution of the difference between treatments
```

```
delta_post_sd <- NA

# p_suc --> probability of a "statistically" successfull treatment
p_suc <- NA
# p_fut --> probability of a futile experimental treatment
p_fut <- NA

# DELTA --> decrease in percentage reduction of seizures of the exp arm vs control
DELTA <- 10
# Rc --> percent reduction in the control group
Rc <- exp(beta0) - 110
# d_min --> delta in the "log(+110)" scale

d_min <- log((Rc + 110) / (Rc - DELTA + 110))

# ID_perm --> vector used to randomly extract covariates from dataset
ID_perm <- rep(0,M/2)

## control and test alternate the IDs
control_IDperm <- rep(0,M/2)
test_IDperm <- control_IDperm

# sim_perm --> vector used to randomly extract simulations of ID previously estracted from
dataset
sim_perm <- rep(0,N_PK_sim)

Cs <- 0.99
Cf <- 0.5

for (iter in (1:PDsim))
{
  ID_perm <- sample((1:M),size=M,replace=F)

  control_IDperm <- ID_perm[control_flag]
  test_IDperm <- ID_perm[test_flag]

  sim_perm <- sample((1:N_PK_sim),size=N_PK_sim,replace=F)

  # n --> sample size per arm
  n <- G_half

  # label --> logical variable that takes false if a couclusion is drawn during the
sequential testing procedure
  label <- T

  while(label==T)
  {

    # control_subjects --> subjects randomized to the control group
    tmp <- subset(d,PKSIM == sim_perm[1])
    control_subjects <-  tmp[control_IDperm[1:n],]

    m_pcb <- mean(beta0 + beta1*0 + control_subjects$EPS)

    V_pcb <- var(beta0 + beta1*0 + control_subjects$EPS)

    # test_subjects --> subjects randomized to the treated group
    test_subjects <-  tmp[test_IDperm[1:n],]

    m_tpm <- mean(beta0 + beta1*test_subjects$CONC + test_subjects$EPS)

    V_tpm <- var(beta0 + beta1*test_subjects$CONC + test_subjects$EPS)

    mu_post_pcb <- (mu_pcb*V_pcb + n*sd_pcb^2*m_pcb) / (n*sd_pcb^2+V_pcb)

    sd_post_pcb <- sqrt(V_pcb*sd_pcb^2/(n*sd_pcb^2+V_pcb))

    mu_post_tpm <- (mu_tpm*V_tpm + n*sd_tpm^2*m_tpm) / (n*sd_tpm^2+V_tpm)

    sd_post_tpm <- sqrt(V_tpm*sd_tpm^2/(n*sd_tpm^2+V_tpm))

    delta_post_mean <- (V_pcb/n / (V_pcb/n + sd_pcb^2) * (mu_pcb - m_pcb) + m_pcb) -
      (V_tpm/n / (V_tpm/n + sd_tpm^2) * (mu_tpm - m_tpm) + m_tpm)


    delta_post_sd <- sqrt(V_pcb/n * sd_pcb^2 / (V_pcb/n + sd_pcb^2) +
                            V_tpm/n * sd_tpm^2 / (V_tpm/n + sd_tpm^2))


    p_suc <- 1 - pnorm(0,delta_post_mean,delta_post_sd)

    p_fut <- pnorm(d_min,delta_post_mean,delta_post_sd)

    label <- p_suc <= Cs  & p_fut <= Cf
    # the sample size per arm increases by G at every step
    n <- n + G_half
  }

  # Decision:
  if (p_fut > Cf) # accept H0
    out$H1[iter] <- 0
  else if (p_suc > Cs) # accept H1
    out$H1[iter] <- 1

  # MEAN = Rc - Rt
```

```
  out$MEAN[iter] <- (exp(mu_post_pcb) - 110) - (exp(mu_post_tpm) - 110)
  # MEAN_LOG = Yc-Yt
  out$MEAN_Y[iter] <- delta_post_mean
  # SD_LOG on the log
  out$WCI[iter] <- 2*qnorm(0.975)*delta_post_sd
  out$N[iter] <- n - G_half
}

##########################################
######### SIMULATIONS UNDER H0 ##########
##########################################

## PK-PD parameters from Girgis et al
# "placebo effect"
beta0 <- 4.4830
# "TPM effect"
# simulations under H0 --> beta1 is set to 0 in the TPM group
beta1 <- 0
# "residual variability" in terms of standard deviation
sigma_epsilon <- 0.751664

# EPS --> variability that will be considered for PD simulations under H0
d$EPS <- rnorm(nrow(d),0,sigma_epsilon)

for (iter in (1:PDsim))
{
  ID_perm <- sample((1:M),size=M,replace=F)

  control_IDperm <- ID_perm[control_flag]
  test_IDperm <- ID_perm[test_flag]

  sim_perm <- sample((1:N_PK_sim),size=N_PK_sim,replace=F)

  # n --> sample size per arm
  n <- G_half

  # label --> logical variable that takes false if a couclusion is drawn during the
sequential testing procedure
  label <- T

  while(label==T)
  {

    # control_subjects --> subjects randomized to the control group
    tmp <- subset(d,PKSIM == sim_perm[1])
    control_subjects <-  tmp[control_IDperm[1:n],]

    m_pcb <- mean(beta0 + beta1*0 + control_subjects$EPS)

    V_pcb <- var(beta0 + beta1*0 + control_subjects$EPS)

    # test_subjects --> subjects randomized to the treated group
    test_subjects <-  tmp[test_IDperm[1:n],]

    m_tpm <- mean(beta0 + beta1*0 + test_subjects$EPS)

    V_tpm <- var(beta0 + beta1*0 + test_subjects$EPS)

    mu_post_pcb <- (mu_pcb*V_pcb + n*sd_pcb^2*m_pcb) / (n*sd_pcb^2+V_pcb)

    sd_post_pcb <- sqrt(V_pcb*sd_pcb^2/(n*sd_pcb^2+V_pcb))

    mu_post_tpm <- (mu_tpm*V_tpm + n*sd_tpm^2*m_tpm) / (n*sd_tpm^2+V_tpm)

    sd_post_tpm <- sqrt(V_tpm*sd_tpm^2/(n*sd_tpm^2+V_tpm))

    delta_post_mean <- (V_pcb/n / (V_pcb/n + sd_pcb^2) * (mu_pcb - m_pcb) + m_pcb) -
      (V_tpm/n / (V_tpm/n + sd_tpm^2) * (mu_tpm - m_tpm) + m_tpm)


    delta_post_sd <- sqrt(V_pcb/n * sd_pcb^2 / (V_pcb/n + sd_pcb^2) +
                            V_tpm/n * sd_tpm^2 / (V_tpm/n + sd_tpm^2))


    p_suc <- 1 - pnorm(0,delta_post_mean,delta_post_sd)

    p_fut <- pnorm(d_min,delta_post_mean,delta_post_sd)

    label <- p_suc <= Cs  & p_fut <= Cf
    # the sample size per arm increases by G at every step
    n <- n + G_half
  }

  # Decision:
  if (p_fut > Cf) # accept H0
    out$H0[iter] <- 0
  else if (p_suc > Cs) # accept H1
    out$H0[iter] <- 1

}
```

7.3.8. Bayesian sequential design with a semi-hierarchical framework

```
##############################################
######### SIMULATIONS UNDER H1 ##############
##############################################

rm(list=ls(all=T))

# setting the seed for results replication
set.seed(121087)

# PK data reading
d <- read.csv("PK.csv",sep=",",header=T)

# M --> number of indivduals simulated at step 1
M <- length(unique(d$ID))
# N_PK_sim --> number of PK simulations done at step 2
N_PK_sim <- length(unique(d$PKSIM))

## PK-PD parameters from Girgis et al
# "placebo effect"
beta0 <- 4.4830
# beta1_p <- "TPM effect" in pediatrics
beta1_p <- -0.0579
# beta1_a <- "TPM effect" in adults
beta1_a <- -0.0627
# "residual variability" in terms of standard deviation
sigma_epsilon <- 0.751664

# EPS --> variability that will be considered for PD simulations under H1
d$EPS <- rnorm(nrow(d),0,sigma_epsilon)

# PDsim --> number of CTS
PDsim <- 1000

# vectors used to randomise patients in the placebo or TPM arm
control_flag <- rep(c(T,F),M/2)
test_flag <- !control_flag

## out --> output dataset
out <- data.frame(H1=rep(NA,PDsim), # result of the statistical test when simulating under
H1: 0--> accept h0, 1--> refuse H0
                  H0=NA,   # # result of the statistical test when simulating under H0
                  MEAN_Y=NA,  # estimate of delta in the Y scale
                  WCI=NA,    # width of confidence intervals
                  N=NA)   # sample size per arm (to be determined via CTS)

# G_half --> number of patients per group enrolled at each interim analysis
G_half <- 10

## PRIORS

# delta_p_bar --> TPM vs PCB difference in paediatrics given by the PK-PD model
load("delta_p_bar")
delta_p_bar <- delta

# delta_hat_a --> ML estimate of delta based on adult trials (-0.0627=beta1 in adults, 8=
pop Cmin. From Girgis et al)
delta_hat_a <-  -beta1_a*8
# sa --> 2*sigma_epsilon
sa <- 2*sigma_epsilon
sp <- sa
# ma --> number of adults on which delta_hat_a has been estimated (from Girgis et al)
ma <- 663
# ni -->  difference in improvement of TPM over placebo between children and adults
ni <- abs(delta_hat_a-delta_p_bar)/sqrt(2)
# omega --> as defined in the paper
omega <- sa^2/(sa^2/(ma) + 2*ni^2)


# delta_post_mean --> mean of the posterior distribution of the difference between
treatments
delta_post_mean <- NA
# delta_post_sd --> sd of the posterior distribution of the difference between treatments
delta_post_sd <- NA

# p_suc --> probability of a "statistically" successfull treatment
p_suc <- NA
# p_fut --> probability of a futile experimental treatment
p_fut <- NA

# DELTA --> decrease in percentage reduction of seizures of the exp arm vs control
DELTA <- 10
# Rc --> percent reduction in the control group
Rc <- exp(beta0) - 110
# d_min --> delta in the "log(+110)" scale

d_min <- log((Rc + 110) / (Rc - DELTA + 110))

# ID_perm --> vector used to randomly extract covariates from dataset
ID_perm <- rep(0,M/2)
```

```
## control and test alternate the IDs
control_IDperm <- rep(0,M/2)
test_IDperm <- control_IDperm

# sim_perm --> vector used to randomly extract simulations of ID previously estracted from
dataset
sim_perm <- rep(0,N_PK_sim)

# delta_hat_p --> posterior estimate of treatment effect in children
delta_hat_p <- NA
# sp obtained from the simulations
sp <- NA

Cs <- 0.99
Cf <- 0.5

for (iter in (1:PDsim))
{
  ID_perm <- sample((1:M),size=M,replace=F)

  control_IDperm <- ID_perm[control_flag]
  test_IDperm <- ID_perm[test_flag]

  sim_perm <- sample((1:N_PK_sim),size=N_PK_sim,replace=F)

  # n --> sample size per arm
  n <- G_half

  # label --> logical variable that takes false if a couclusion is drawn during the
sequential testing procedure
  label <- T

  while(label==T)
  {

    # control_subjects --> subjects randomized to the control group
    tmp <- subset(d,PKSIM == sim_perm[1])
    control_subjects <-  tmp[control_IDperm[1:n],]

    control <- beta0 + beta1_p*0 + control_subjects$EPS

    # test_subjects --> subjects randomized to the treated group
    test_subjects <-  tmp[test_IDperm[1:n],]

    test <- beta0 + beta1_p*test_subjects$CONC + test_subjects$EPS

    # BAYESIAN INFERENCE
    delta_hat_p <- mean(control)-mean(test)
    sp <- 2*sd(control-test)/sqrt(2)

    delta_post_mean <- (delta_hat_p*(2*n)/sp^2+delta_hat_a*omega/sa^2) /
((2*n)/sp^2+omega/sa^2)

    delta_post_sd <- ((2*n)/sp^2 + omega/sa^2)^(-0.5)

    p_suc <- 1 - pnorm(0,delta_post_mean,delta_post_sd)


    p_fut <- pnorm(d_min,delta_post_mean,delta_post_sd)

    label <- p_suc <= Cs  & p_fut <= Cf
    # the sample size per arm increases by G at every step
    n <- n + G_half
  }

  # Decision:
  if (p_fut > Cf) # accept H0
    out$H1[iter] <- 0
  else if (p_suc > Cs) # accept H1
    out$H1[iter] <- 1

  # MEAN_LOG = Yc-Yt
  out$MEAN_Y[iter] <- delta_post_mean
  # SD_LOG on the log
  out$WCI[iter] <- 2*qnorm(0.975)*delta_post_sd
  out$N[iter] <- n - G_half

}

##############################################
######### SIMULATIONS UNDER H0 ##############
##############################################

## PK-PD parameters from Girgis et al
# "placebo effect"
beta0 <- 4.4830
# "TPM effect"
# simulations under H0 --> beta1 is set to 0 in the TPM group
beta1_p <- 0
# "residual variability" in terms of standard deviation
sigma_epsilon <- 0.751664

# EPS --> variability that will be considered for PD simulations under H0
d$EPS <- rnorm(nrow(d),0,sigma_epsilon)
```

```
for (iter in (1:PDsim))
{
  ID_perm <- sample((1:M),size=M,replace=F)

  control_IDperm <- ID_perm[control_flag]
  test_IDperm <- ID_perm[test_flag]

  sim_perm <- sample((1:N_PK_sim),size=N_PK_sim,replace=F)

  # n --> sample size per arm
  n <- G_half

  # label --> logical variable that takes false if a couclusion is drawn during the
sequential testing procedure
  label <- T

  while(label==T)
  {

    # control_subjects --> subjects randomized to the control group
    tmp <- subset(d,PKSIM == sim_perm[1])
    control_subjects <-  tmp[control_IDperm[1:n],]

    control <- beta0 + beta1_p*0 + control_subjects$EPS

    # test_subjects --> subjects randomized to the treated group
    test_subjects <-  tmp[test_IDperm[1:n],]

    test <- beta0 + beta1_p*test_subjects$CONC + test_subjects$EPS

    # BAYESIAN INFERENCE
    delta_hat_p <- mean(control)-mean(test)
    sp <- 2*sd(control-test)/sqrt(2)

    delta_post_mean <- (delta_hat_p*(2*n)/sp^2+delta_hat_a*omega/sa^2) /
((2*n)/sp^2+omega/sa^2)

    delta_post_sd <- ((2*n)/sp^2 + omega/sa^2)^(-0.5)

    p_suc <- 1 - pnorm(0,delta_post_mean,delta_post_sd)


    # d_min --> delta in the "log(+110)" scale
    # Rc --> percent reduction in the control group
    Rc <- exp(mean(control)) - 110
    d_min <- log((Rc + 110) / (Rc - DELTA + 110))

    p_fut <- pnorm(d_min,delta_post_mean,delta_post_sd)

    label <- p_suc <= 0.99  & p_fut <= 0.5
    # the sample size per arm increases by G at every step
    n <- n + G_half
  }

  # Decision:
  if (p_fut > 0.5) # accept H0
    out$H0[iter] <- 0
  else if (p_suc > 0.99) # accept H1
    out$H0[iter] <- 1

 }
```

7.4. R function used for δ estimation in sequential designs

```
seq_inference <- function(alpha,thetaR,theta,V_obs,ci,test,result)
{
  ## INPUT:
  # alpha --> working significance level set to design the trial (nx1)
  # thetaR --> reference improvement set to design the trial (nx1)
  # theta --> indipendent variable
  # V_obs --> observed value of V at the final interim analysis
  # ci --> number indicating the confidence interval
  # test --> string variable indicating the test ("SPRT" or "TT")
  # result --> 1 means H0 refused (upper boundary crossed),
  #            0 means H0 accepted (lower boundary crossed)
  ## OUTPUT:
  # lower_ci --> (correct) lower confidence limit
  # delta_est --> (correct) estimate of treatment difference
  # upper_ci --> (correct) upper confidence limit
  # p_val --> p-value

Q <- function(alpha,thetaR,theta,V_obs,test)
  ## INPUT:
  # alpha --> working significance level set to design the trial (nx1)
  # thetaR --> reference improvement set to design the trial (nx1)
  # theta --> indipendent variable
```

```
  # V_obs --> observed value of V at the final interim analysis
  # test --> string variable indicating the test ("SPRT" or "TT")

  ## OUTPUT:
  # p --> Q(0,V;theta)
{
  # size --> length of Q
  size <- length(theta)
  p <- rep(0,size)

  # initialising parameters for computation time minimisation
  f1 <- 0
  b1 <- 0
  b2 <- 0
  f2 <- 0

  if (test=="SPRT")
  {
    # re-writing the parameters accdording to fromula 4.8.19 in Whithead
    psi <- log((1-alpha)/alpha)
    k <- theta/thetaR
    c <- V_obs*thetaR^2
    s <- (0:100)


    for(i in (1:size))
    {
      f1 <- ((k[i]-0.5)*c + psi)/sqrt(c)

      b1 <- ((k[i]-0.5)*c + (2*s+1)*psi)/sqrt(c)
      b2 <- (-(k[i]-0.5)*c + (2*s+1)*psi)/sqrt(c)

      # need to avoid division by zero for c values too low (very early stopping, see
Whithead)

      if (c>0.25*thetaR^2)
        f2 <- (-1)^s*exp(-2*s*(s+1)*psi^2/c) * ( (1-pnorm(b1))/dnorm(b1) + (1-
pnorm(b2))/dnorm(b2))
      else
        f2 <- (-1)^s*exp(-2*s*(s+1)*psi^2/c) * ( (1-pnorm(b1))/dnorm(b1))



      p[i] <- dnorm(f1)*sum(f2[!is.na(f2)])
    }
  }

  else if(test=="TT")
  {
    # re-writing the parameters accdording to fromula 4.9.3 in Whithead
    k <- theta/thetaR
    c <- V_obs*thetaR^2
    csi <- 2*log(1/(2*alpha))
    s <- (0:100)

    for (i in (1:size))
    {
      f1 <- ((k[i]-0.75)*c + csi) / sqrt(c)

      b1 <- ((k[i]-0.75)*c + (2*s+1)*csi) / sqrt(c)
      b2 <- (-(k[i]-0.75)*c + (2*s+1)*csi) / sqrt(c)

      f2 <- (-1)^s*exp(2*s*(s+1)*csi*(0.25 - csi/c)) *
        ( (1 - pnorm(b1))/dnorm(b1) + (1 - pnorm(b2))/dnorm(b2))

      p[i] <- dnorm(f1)*sum(f2[!is.na(f2)])
    }
  }

  else
    print("error, unknown variable test")

  return(p)
}

  prob <- (1-ci)/2

  if (result==1)
  {
    Q_p_val <- Q(alpha,thetaR,thetaR,V_obs,test)
    Q_ci <- Q(alpha,thetaR,(thetaR-theta),V_obs,test)
    # eq. 5.8.1 Whitehead
    ind_lower <-  which(abs(Q_ci-prob)==min(abs(Q_ci-prob)))
    # eq. 5.8.2 Whitehead
    ind_upper <-  which(abs(Q_ci-(1-prob))==min(abs(Q_ci-(1-prob))))
  }
  else
  {
    Q_p_val <- 1 - Q(alpha,thetaR,0,V_obs,test)
    Q_ci <- Q(alpha,thetaR,theta,V_obs,test)
    # eq. 5.8.2 Whitehead
    ind_upper <-  which(abs(Q_ci-prob)==min(abs(Q_ci-prob)))
    # eq. 5.8.1 Whitehead
    ind_lower <-  which(abs(Q_ci-(1-prob))==min(abs(Q_ci-(1-prob))))
  }
```

```
  ind_est <-  which(abs(Q_ci-0.5)==min(abs(Q_ci-0.5)))

  lower_ci <- theta[ind_lower]
  upper_ci <- theta[ind_upper]
  delta_est <- theta[ind_est]

  p_val <- Q_p_val

  return(c(lower_ci,delta_est,upper_ci,p_val))
}
```